



(12)发明专利申请

(10)申请公布号 CN 112347233 A

(43)申请公布日 2021.02.09

(21)申请号 202010373462.3

(22)申请日 2020.05.06

(30)优先权数据

10-2019-0095495 2019.08.06 KR

(71)申请人 现代自动车株式会社

地址 韩国首尔

申请人 起亚自动车株式会社

(72)发明人 李廷旻 朴永敏 金宣我

(74)专利代理机构 北京康信知识产权代理有限  
责任公司 11240

代理人 徐丽华

(51)Int.Cl.

G06F 16/332(2019.01)

G10L 15/22(2006.01)

B60R 16/02(2006.01)

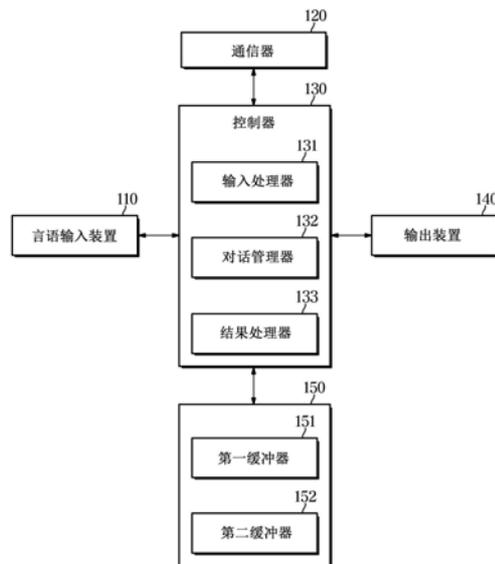
权利要求书3页 说明书12页 附图5页

(54)发明名称

对话处理设备、包括对话处理设备的车辆和  
对话处理方法

(57)摘要

本发明公开了对话处理设备、包括对话处理设备的车辆和对话处理方法。对话处理设备包括：配置为接收用户的言语信号的言语输入装置；以及配置为在其中存储接收到的言语信号的第一缓冲器；输出装置；以及控制器。控制器被配置为：基于存储的言语信号来检测话语结束时间点；基于是否从话语结束时间点之前的言语信号所对应的第一言语识别结果中识别出用户的意图，来生成话语结束时间点之后的言语信号所对应的第二言语识别结果；以及控制输出装置输出与基于第一言语识别结果或第二言语识别结果中的至少一个确定的用户的意图相对应的响应。



1. 一种对话处理设备,包括:

言语输入装置,被配置为接收用户的言语信号;

第一缓冲器,被配置将所接收的所述言语信号存储在所述第一缓冲器中;

输出装置;以及

控制器,被配置为基于所存储的所述言语信号来检测话语结束时间点,基于是否从与  
所述话语结束时间点之前的言语信号对应的第一言语识别结果中识别出所述用户的意图,  
生成与所述话语结束时间点之后的言语信号对应的第二言语识别结果,以及所述控制器被  
配置为控制所述输出装置输出与基于所述第一言语识别结果或所述第二言语识别结果中  
的至少一者确定的所述用户的意图对应的响应。

2. 根据权利要求1所述的对话处理设备,其中,当不能从所述第一言语识别结果中识别  
出所述用户的意图时,生成与所述话语结束时间点之后的所述言语信号对应的所述第二言  
语识别结果。

3. 根据权利要求2所述的对话处理设备,还包括第二缓冲器,

其中,当不能从所述第一言语识别结果识别出所述用户的所述意图时,所述控制器将  
所述第一言语识别结果存储在所述第二缓冲器中。

4. 根据权利要求2所述的对话处理设备,其中,当不能从所述第一言语识别结果中识别  
出所述用户的所述意图时,所述控制器基于言语识别的次数生成所述第二言语识别结果。

5. 根据权利要求4所述的对话处理设备,其中,当所述言语识别的次数小于预定参考值  
时,所述控制器基于在所述话语结束时间点之后的所述言语信号生成所述第二言语识别结  
果。

6. 根据权利要求4所述的对话处理设备,其中,当所述言语识别的次数大于或等于预定  
参考值时,所述控制器删除所述第一缓冲器中存储的数据,并生成与不能识别所述用户的  
意图的情况对应的响应。

7. 根据权利要求1所述的对话处理设备,其中,当输出与所述用户的言语信号对应的响  
应时,所述控制器将言语识别的次数设置为初始值。

8. 根据权利要求1所述的对话处理设备,其中,当生成所述第二言语识别结果时,所述  
控制器基于所述第一言语识别结果或所述第二言语识别结果中的至少一者,确定用于确定  
所述用户的意图的意图候选组,并将从所确定的所述意图候选组中选出的一个意图候选  
确定为所述用户的意图。

9. 根据权利要求8所述的对话处理设备,其中,所述控制器确定所述意图候选组的准确  
性,并且将所述意图候选组中具有最高准确性的意图候选确定为所述用户的意图。

10. 根据权利要求1所述的对话处理设备,其中,当检测到所述话语结束时间点时,所述  
控制器被配置为删除所述第一缓冲器中存储的数据,并将在所述话语结束时间点之后输入  
的言语信号存储在所述第一缓冲器中。

11. 根据权利要求1所述的对话处理设备,其中,当能够从所述第一言语识别结果中识  
别出所述用户的意图时,所述控制器删除所述第一缓冲器中存储的数据。

12. 一种车辆,包括:

言语输入装置,被配置为接收用户的言语信号;

第一缓冲器,被配置将所接收的所述言语信号存储在所述第一缓冲器中;

输出装置;以及

控制器,被配置为基于所存储的所述言语信号来检测话语结束时间点,基于是否从与  
所述话语结束时间点之前的言语信号对应的第一言语识别结果中识别出所述用户的意图,  
生成与所述话语结束时间点之后的言语信号对应的第二言语识别结果,以及所述控制器被  
配置为控制所述输出装置输出与基于所述第一言语识别结果或所述第二言语识别结果中  
的至少一者确定的所述用户的意图对应的响应。

13.根据权利要求12所述的车辆,其中,当不能从所述第一言语识别结果中识别出所述  
用户的意图时,所述控制器生成与所述话语结束时间点之后的所述言语信号对应的所述第  
二言语识别结果。

14.根据权利要求12所述的车辆,其中,当生成所述第二言语识别结果时,所述控制器  
基于所述第一言语识别结果或所述第二言语识别结果中的至少一者,确定用于确定所述用  
户的意图的意图候选组,并将从所确定的所述意图候选组中选出的一个意图候选确定为所  
述用户的意图。

15.一种对话处理方法,包括:

接收用户的言语信号;

将所接收的所述言语信号存储在所述第一缓冲器中;

基于所存储的所述言语信号检测话语结束时间点;

基于是否从与所述话语结束时间点之前的言语信号对应的第一言语识别结果中识别  
出所述用户的意图,生成与所述话语结束时间点之后的言语信号对应的第二言语识别结  
果;以及

输出与基于所述第一言语识别结果或所述第二言语识别结果中的至少一者确定的所  
述用户的意图对应的响应。

16.根据权利要求15所述的对话处理方法,其中,生成与所述话语结束时间点之后的言  
语信号对应的第二言语识别结果包括:

当不能从所述第一言语识别结果中识别出所述用户的意图时,生成与所述话语结束时  
间点之后的所述言语信号对应的所述第二言语识别结果。

17.根据权利要求16所述的对话处理方法,其中,生成与所述话语结束时间点之后的言  
语信号对应的第二言语识别结果包括:

当不能从所述第一言语识别结果识别出所述用户的所述意图时,将所述第一言语识别  
结果存储在第二缓冲器中。

18.根据权利要求16所述的对话处理方法,其中,生成与所述话语结束时间点之后的言  
语信号对应的第二言语识别结果包括:

当不能从所述第一言语识别结果中识别出所述用户的所述意图时,基于言语识别的次  
数生成所述第二言语识别结果。

19.根据权利要求18所述的对话处理方法,其中,生成与所述话语结束时间点之后的言  
语信号对应的第二言语识别结果包括:

当所述言语识别的次数小于预定参考值时,基于在所述话语结束时间点之后的所述言  
语信号生成所述第二言语识别结果。

20.根据权利要求15所述的对话处理方法,还包括:当生成所述第二言语识别结果时,

基于所述第一言语识别结果或所述第二言语识别结果中的至少一者,确定用于确定所述用户的意图的意图候选组,并将从所确定的所述意图候选组中选出的一个意图候选确定为所述用户的意图。

## 对话处理设备、包括对话处理设备的车辆和对话处理方法

### 技术领域

[0001] 本公开涉及用于通过与用户的对话来识别用户的意图并提供用户期望的信息或服务的对话处理设备,包括该对话处理设备的车辆及其对话处理方法。

### 背景技术

[0002] 对话处理设备是用于与用户进行对话的装置。对话处理设备运行以识别用户的言语,通过识别言语的结果来识别用户的意图,并输出用于向用户提供期望的信息或服务的响应。

[0003] 当识别用户的言语时,需要确定用户实际说出话语的部分,并且为此,难以检测用户完成话语的点。这种检测称为结束点检测 (EPD)。

[0004] 传统的对话处理设备在检测到大于或等于预定时间的无声部分时识别出用户已经结束话语。传统的对话处理设备基于直到话语结束为止获得的言语数据生成言语识别结果。在这种情况下,当用户在没有意图结束话语的情况下在预定时间内没有讲话,在生成言语识别结果时不考虑用户输入的下一言语。因此,输出了与用户的意图不匹配的响应。

### 发明内容

[0005] 因此,本公开的目的是提供一种能够接收用户的言语信号并输出与该用户的言语信号相对应的响应的对话处理设备,包括该对话处理设备的车辆及其对话处理方法。

[0006] 本公开的其他方面部分地在随后的描述中阐述,并且部分地从描述中应当是显而易见的,或者可以通过本公开的实践而获知。

[0007] 因此,本公开的一方面在于提供对话处理设备。该对话处理设备包括:配置为接收用户的言语信号的言语输入装置,配置为在其中存储接收到的言语信号的第一缓冲器,输出装置,以及控制器。控制器被配置为:基于存储的言语信号来检测话语结束时间点;基于是否从话语结束时间点之前的言语信号所对应的第一言语识别结果中识别出用户的意图,来生成话语结束时间点之后的言语信号所对应的第二言语识别结果;以及控制输出装置输出与基于第一言语识别结果或第二言语识别结果中的至少一个确定的用户的意图相对应的响应。

[0008] 当从第一言语识别结果中不能识别出用户的意图时,可以生成与话语结束时间点之后的言语信号相对应的第二言语识别结果。

[0009] 对话处理设备可以进一步包括第二缓冲器。当根据第一言语识别结果无法识别出用户的意图时,控制器可以将第一言语识别结果存储在第二缓冲器中。

[0010] 当从第一言语识别结果中不能识别出用户的意图时,控制器可以基于言语识别的次数来生成第二言语识别结果。

[0011] 当言语识别的次数小于预定参考值时,控制器可以基于在话语结束时间点之后的言语信号生成第二言语识别结果。

[0012] 当言语识别的次数大于或等于预定参考值时,控制器可以删除存储在第一缓冲器

中的数据,并生成与无法识别用户的意图的情况相对应的响应。

[0013] 当输出与用户的言语信号相对应的响应时,控制器可以将言语识别的次数设置为初始值。

[0014] 当生成第二言语识别结果时,控制器可以基于第一言语识别结果或第二言语识别结果中的至少一个确定用于确定用户的意图的意图候选组。控制器可以进一步确定从确定的意图候选组中选择一个为用户的意图。

[0015] 控制器可以确定意图候选组的准确性,并且将意图候选组中具有最高准确性的意图候选确定为用户的意图。

[0016] 当检测到话语结束时间点时,控制器可被配置为删除存储在第一缓冲器中的数据,并将在话语结束时间点之后输入的言语信号存储在第二缓冲器中。

[0017] 当从第一言语识别结果中可以识别出用户的意图时,控制器可以删除存储在第二缓冲器中的数据。

[0018] 本公开的另一方面在于提供一种车辆,包括:配置为接收用户的言语信号的言语输入装置;配置为在其中存储接收到的言语信号的第二缓冲器;输出装置;以及控制器。该控制器被配置为:基于存储的言语信号来检测话语结束时间点;基于是否从话语结束时间点之前的言语信号所对应的第一言语识别结果中识别出用户的意图,来生成话语结束时间点之后的言语信号所对应的第二言语识别结果;以及控制输出装置输出与基于第一言语识别结果或第二言语识别结果中的至少一个确定的用户的意图相对应的响应。

[0019] 当从第一言语识别结果中不能识别出用户的意图时,控制器可以生成与话语结束时间点之后的言语信号相对应的第二言语识别结果。

[0020] 当生成第二言语识别结果时,控制器可以基于第一言语识别结果或第二言语识别结果中的至少一个,确定用于确定用户的意图的意图候选组。控制器可以进一步确定从确定的意图候选组中选择一个为用户的意图。

[0021] 本公开的另一方面在于提供一种对话处理方法。该方法包括:接收用户的言语信号;将接收的言语信号存储在第二缓冲器中;基于存储的言语信号检测话语结束时间点;基于是否从话语结束时间点之前的言语信号所对应的第一言语识别结果中识别出用户的意图,来生成话语结束时间点之后的言语信号所对应的第二言语识别结果;以及输出与基于第一言语识别结果或第二言语识别结果中的至少一个确定的用户的意图相对应的响应。

[0022] 生成与话语结束时间点之后的言语信号相对应的第二言语识别结果可以包括:当从第一言语识别结果中不能识别出用户的意图时,生成与生成话语结束时间点之后的言语信号相对应的第二言语识别结果。

[0023] 生成与话语结束时间点之后的言语信号相对应的第二言语识别结果可以包括:当从第一言语识别结果中不能识别出用户的意图时,将第一言语识别结果存储在第二缓冲器中。

[0024] 生成与话语结束时间点之后的言语信号相对应的第二言语识别结果可包括:当从第一言语识别结果中不能识别出用户的意图时,基于言语识别的次数来生成第二言语识别结果。

[0025] 生成与话语结束时间点之后的言语信号相对应的第二言语识别结果可以包括:当言语识别的次数小于预定参考值时,基于在话语结束时间点之后的言语信号生成第二言语

识别结果。

[0026] 对话处理方法还可以包括：当生成第二言语识别结果时，基于第一言语识别结果或第二言语识别结果中的至少一个，确定用于确定用户的意图的意图候选组，并确定从确定的意图候选组中选择一个为用户的意图。

### 附图说明

[0027] 通过以下结合附图对实施例的描述，本公开的这些和/或其他方面将变得显而易见并更容易理解：

[0028] 图1是示出根据实施例的对话处理设备的控制框图。

[0029] 图2是用于描述根据实施例的对话处理设备的运行的图。

[0030] 图3是用于描述根据实施例的对话处理设备的运行的图。

[0031] 图4是示出根据实施例的对话处理方法的流程图。

[0032] 图5A和图5B是示出根据另一实施例的对话处理方法的流程图。

### 具体实施方式

[0033] 在整个说明书中，相同的数字表示相同的元件。并未描述本公开的实施例的所有元件。省略了对本领域公知的或在实施例中彼此重叠的描述。在整个说明书中使用的术语，例如“~部分”，“~模块”，“~成员”，“~块”等，可以用软件和/或硬件来实现，并且多个“~部分”，“~模块”，“~成员”或“~块”可以在单个元件中实现，或者单个“~部分”，“~模块”，“~成员”或“~块”可以包括多个元件。

[0034] 应当进一步理解，术语“连接”或其派生词同时指直接和间接连接。间接连接包括通过无线通信网络的连接。

[0035] 应当进一步理解，术语“包括”和/或“包含”在本说明书中使用，指定存在所述特征、整数、步骤、运行、元件和/或部件，但不排除一个或多个其他特征、整数、步骤、运行、元件、部件和/或他们的组的存在或添加，除非上下文另外明确指出。

[0036] 尽管术语“第一”、“第二”、“A”、“B”等可用于描述各种部件，但这些术语并不限制相应的部件，而是仅用于将一个部件与另一部件区分开的目的。

[0037] 如本文中所使用的，单数形式“一”、“一个”和“该”也意图包括复数形式，除非上下文另外明确指出。

[0038] 用于方法步骤的参考数字仅是为了便于解释，而不是限制步骤的顺序。因此，除非上下文另外明确指出，否则可以书面方式执行书面命令。

[0039] 在下文中，下面参照附图描述本公开的运行原理和实施例。

[0040] 图1是示出根据实施例的对话处理设备的控制框图。

[0041] 参考图1，根据实施例的对话处理设备100包括言语输入装置110、通信器120、控制器130、输出装置140和存储器150。

[0042] 言语输入装置110可以接收用户的言语形式的命令。换句话说，言语输入装置110可以接收用户的言语信号。为此，言语输入装置110可以包括麦克风，该麦克风接收声音并将声音转换成电信号。

[0043] 存储器150可以存储由对话处理设备100直接或间接使用的各种类型的数据，以输

出与用户的言语相对应的响应。

[0044] 此外,存储器150可以包括第一缓冲器151和第二缓冲器152。第一缓冲器151可以存储输入的言语信号,并且第二缓冲器152可以存储言语识别的结果。

[0045] 可以使用非易失性存储装置,例如高速缓存,只读存储器(ROM),可编程ROM(PROM),可擦除可编程ROM(EPROM),电可擦除可编程ROM(EEPROM)以及闪存,易失性存储装置(例如随机存取存储器(RAM))或其他存储介质(例如硬盘驱动器(HDD),CD-ROM)等中的至少一个来实现存储器150,但是存储器150的实现方式不限于此。存储器150可以是被实现为与处理器(下面结合控制器130对其进行描述)分离的芯片的存储器,或者被实现为与处理器集成的单个芯片。

[0046] 控制器130包括:被配置为识别输入言语信号以产生言语识别结果的输入处理器131;以及被配置为基于言语识别结果识别用户的意图并确定与用户的意图相对应的动作的对话管理器132;以及被配置为生成用于执行所确定的动作的对话响应的结果处理器133。

[0047] 输入处理器131可以识别用户的输入言语信号,并且可以将用户的言语信号转换为文本型话语。输入处理器131可以将自然语言理解算法应用于话语文本以识别用户的意图。

[0048] 话语文本或用户意图中的至少一项可以由输入处理器131输出作为言语识别结果。输入处理器131可以将言语识别结果发送到对话管理器132。

[0049] 为此,输入处理器131可以包括言语识别模块,并且可以使用执行用于处理输入言语的操作的处理器(未示出)来实现。

[0050] 可以在输入言语时执行输入处理器131的言语处理操作,或者可以在输入言语并且满足特定条件时执行输入处理器131的言语处理操作。

[0051] 详细地,当输入预定的呼叫词时或当从用户接收到言语识别开始命令时,输入处理器131可以执行上述言语处理操作。

[0052] 此外,输入处理器131可以生成与针对特定部分输入的言语信号相对应的言语识别结果。详细地,输入处理器131可以生成与从话语开始时间点到话语结束时间点的言语信号相对应的言语识别结果。

[0053] 为此,输入处理器131可以基于输入言语信号来检测话语开始时间点或话语结束时间点。输入处理器131可以确定输入预定呼叫词或从用户接收到言语识别开始命令的时间点作为话语开始时间点。在这种情况下,用户可以通过说出预定的呼叫词或通过单独的按钮来输入言语识别开始命令,并且输入处理器131可以从话语开始时间点识别用户的言语。

[0054] 另外,当在输入言语信号中存在大于或等于预定时间的无声部分时,输入处理器131可以识别出用户的话语终止。在这种情况下,输入处理器131可以将从无声部分的开始时间起经过预定时间的时间点确定为话语结束时间点。

[0055] 另一方面,当用户在短时间内停止话语而不打算终止话语时,在短时间内停止话语期间的静音部分会导致用户的话语被误认为已终止。因此,在产生言语识别结果时,可能不考虑用户在话语停止后的言语信号。在这种情况下,当难以仅基于在用户停止话语之前的言语信号来识别用户的意图时,就不能准确地识别用户的意图。因此,很难输出适合用户

情况的对话响应。

[0056] 为此,输入处理器131可以基于是否可识别与话语结束时间点之前的言语信号对应的用户意图来确定是否识别在话语结束时间点之后的言语信号。其细节在下面描述。

[0057] 另一方面,除了上述操作之外,输入处理器131可以执行以下操作:去除输入言语信号的噪声并且识别与输入言语信号相对应的用户。

[0058] 对话管理器132可以基于从输入处理器131接收的言语识别结果来确定用户的意图,并确定与用户的意图相对应的动作。

[0059] 结果处理器133可根据对话管理器132的输出结果提供特定服务或输出系统话语以继续对话。结果处理器133可产生对话响应和执行所接收动作所需的命令并可以输出命令。对话响应可以被输出为文本,图像或音频。当输出命令时,可以执行与输出命令相对应的服务,例如车辆控制和外部内容提供。

[0060] 控制单元130可以包括存储器(未示出),该存储器用于存储关于用于控制对话处理设备100的部件的运行的算法或代表该算法的程序的数据。控制单元130还可以包括处理器(未示出),该处理器使用存储在存储器中的数据来执行上述操作。在这种情况下,存储器和处理器可以被实现为分开的芯片。可替代地,存储器和处理器可以被实现为一个芯片。

[0061] 可替代地,输入处理器131、对话管理器132和结果处理器133可以被集成到一个处理器中或者可以被实现为分开的处理器。

[0062] 输出装置140可以视觉或听觉地输出由控制器130生成的响应。为此,输出装置140可以包括显示器(未示出)或扬声器(未示出)。显示器(未示出)和扬声器(未示出)不仅可以输出对用户的话语的响应,对用户的询问或要提供给用户的信息,而且在另一端以视觉或听觉的方式输出对用户意图的确认并响应人的话语。

[0063] 通信器120可以与诸如服务器的外部装置通信。为此,通信器120可以包括使得能够与外部装置进行通信的一个或多个部件,例如,以下中的至少一项:短距离通信模块,诸如蓝牙模块、红外通信模块、以及射频识别(RFID)通信模块;有线通信模块,诸如控制器局域网(CAN)通信模块和局域网(LAN)模块;以及无线通信模块,诸如Wi-Fi模块和无线宽带模块。

[0064] 当通信器120包括无线通信模块时,无线通信模块可以包括无线通信接口,该无线通信接口包括用于发送信号的天线和发射器。另外,无线通信模块可以进一步包括信号转换模块,该信号转换模块用于在对话处理设备100的控制下将从对话处理设备100通过无线通信接口输出的数字控制信号转换为模拟类型的无线信号。

[0065] 无线通信模块可以包括无线通信接口,该无线通信接口包括用于接收信号的天线和接收器。另外,无线通信模块可以进一步包括信号转换模块,用于将通过无线通信接口接收的模拟类型的无线信号解调为数字控制信号。

[0066] 可以添加或省略至少一个部件以对应于图1所示的对话处理设备100的部件的性能。此外,部件的相互位置可以改变以对应于系统的性能或结构。

[0067] 图1所示的每个部件可以指软件部件和/或硬件部件,例如现场可编程门阵列(FPGA)和专用集成电路(ASIC)。

[0068] 图1所示的对话处理设备100的全部或一些部件可以被包括在车辆中,可以识别包括车辆的驾驶员和乘客的用户的言语并且可以提供适当的响应。

[0069] 图2是用于描述根据实施例的对话处理设备的运行的图。

[0070] 参考图2,根据实施例,对话处理设备100可以接收包括话语“圣塔菲……让我知道燃料效率”的言语信号Sin。在这种情况下,假设用户在“圣达菲”之后短时间停止话语,然后说出“让我知道燃油效率”。

[0071] 输入处理器131可以确定用户说出预定的呼叫词或输入言语识别开始命令的时间点t1作为话语开始时间点。输入处理器131还可将在话语开始时间点之后输入的言语信号存储在缓冲器151中。

[0072] 输入处理器131可以基于存储在缓冲器151中的言语信号来检测话语结束时间点。

[0073] 详细地,当存储的言语信号中存在从用户停止输入言语的时间点t2到在时间点t2之后已经经过预定时间的时间点t3对应的无声部分TS1时,输入处理器131可以将无声部分TS1结束的时间点确定为话语结束时间点。

[0074] 当检测到话语结束时间点时,输入处理器131可以识别在话语结束时间点之前的部分(在下文中,称为第一部分A)中输入的言语信号。输入处理器131可以生成话语文本X[i]:“圣达菲”作为与言语信号相对应的言语识别结果。

[0075] 此时,输入处理器131可以通过删除存储在缓冲器151中的第一部分A的言语信号来初始化缓冲器151。此外,输入处理器131可以将在第一部分A的话语结束时间点之后输入的言语信号存储在缓冲器151中。

[0076] 另外,输入处理器131可以根据是否可根据对应于在第一部分A中输入的言语信号的言语识别结果识别用户的意图,生成与在话语结束时间点之后的言语信号相对应的言语识别结果。

[0077] 详细地,当仅使用在第一部分A中输入的言语信号不能识别用户的意图时,输入处理器131可以生成与在话语结束时间点之后的部分(以下称为第二部分B)中输入的言语信号相对应的言语识别结果。

[0078] 在这种情况下,输入处理器131可以将第一部分A的言语识别结果存储在缓冲器152中。

[0079] 此后,当通过另一个无声部分TS2的存在检测到第二部分B的话语结束时间点时,输入处理器131可以识别存储在缓冲器151中的第二部分B中输入的言语信号。输入处理器131可以生成话语文本X[i+1]:“让我知道燃料效率”作为与言语信号相对应的言语识别结果。

[0080] 此时,输入处理器131可以通过删除存储在缓冲器151中的第二部分B的言语信号来初始化缓冲器151。在这种情况下,输入处理器131可以将在第二部分B的话语结束时间点之后输入的言语信号存储在缓冲器151中。

[0081] 另外,输入处理器131可以根据是否可基于第一部分A的言语识别结果或第二部分B的言语识别结果中的至少一个识别用户的意图,生成与在话语结束时间点之后的言语信号相对应的言语识别结果。

[0082] 在这种情况下,当仅使用与第二部分B相对应的言语识别结果的话语文本X[i+1]:“让我知道燃油效率”不能识别用户的意图时,输入处理器131可以将作为第一部分A的言语识别结果的话语文本X[i]:“圣达菲”与作为第二部分B的言语识别结果的话语文本X[i+1]:

“让我知道燃油经济性”组合。输入处理器131可以基于组合的言语识别结果进一步识别用户的意图。

[0083] 当即使结合第一部分A的言语识别结果和第二部分B的言语识别结果也无法识别用户的意图时,输入处理器131可以生成与在第二部分B的话语结束时间点之后的部分中输入的言语信号相对应的言语识别结果。此后,可以重复上述后续操作。

[0084] 另一方面,当基于至少一个部分的言语识别结果可识别用户的意图时,输入处理器131可将确定的用户的意图发送给对话管理器132。对话管理器132可以确定与用户意图相对应的动作。结果处理器133可以生成用于执行从对话管理器132接收的动作的对话响应。可以通过输出装置140输出结果处理器133的响应。

[0085] 另外,当基于至少一个部分的言语识别结果可识别用户的意图时,输入处理器131可以通过将言语输入装置110切换为关闭状态来阻止输入言语信号。另外,输入处理器131可以通过删除存储在第一缓冲器151或第二缓冲器152中的至少一个中的数据来初始化第一缓冲器151或第二缓冲器152中的至少一个。

[0086] 输入处理器131可以根据用户的意图是否可识别对话结束时间点之后的言语信号执行处理操作。输入处理器131不仅可以使使用话语结束时间点之后的言语识别结果,而且还可以将话语结束时间点之前的言语识别结果用作控制识别用户意图的操作的基础。因此,准确地识别出用户的意图,输出适合于用户的响应,并且增加了用户的便利性。

[0087] 图3是用于描述根据另一实施例的对话处理设备的操作的图。

[0088] 参考图3,根据另一个实施例,对话处理设备100可以接收包括话语“圣塔菲……不,索纳塔,让我知道燃料效率”的言语信号Sin'。在这种情况下,假设用户在“圣达菲”之后短时间停止话语,然后说出“不,索纳塔,让我知道燃油效率”。

[0089] 如上面参考图2所述,输入处理器131可以确定用户说出预定的呼叫词或输入言语识别开始命令的时间点t1'作为话语开始时间点。输入处理器131可将话语开始时间点之后输入的言语信号存储在缓冲器151中。

[0090] 输入处理器131可以基于存储在缓冲器151中的言语信号来检测话语结束时间点。具体地,当存储的言语信号中存在从用户停止输入言语的时间点t2'到在时间点t2'之后已经经过预定时间的时间点t3'的无声部分TS3时,输入处理器131可以将无声部分TS3结束的时间点确定为话语结束时间点。

[0091] 当检测到话语结束时间点时,输入处理器131可以识别在话语结束时间点之前的部分(在下文中,称为第一部分C)中输入的言语信号。输入处理器131可以生成话语文本X[i']:“圣达菲”作为与言语信号相对应的言语识别结果。

[0092] 此时,输入处理器131可以通过删除存储在缓冲器151中的第一部分C的言语信号来初始化缓冲器151。此外,输入处理器131可以将第一部分C的话语结束时间点之后输入的言语信号存储在缓冲器151中。

[0093] 另外,输入处理器131可以,基于是否可根据对应于在第一部分C中输入的言语信号的言语识别结果识别用户的意图,生成与在话语结束时间点之后的言语信号相对应的言语识别结果。

[0094] 详细地,当仅使用在第一部分C中输入的言语信号不能识别用户的意图时,输入处理器131可以生成与在话语结束时间点之后的部分(以下称为第二部分D)中输入的言语信

号相对应的言语识别结果。

[0095] 在这种情况下,输入处理器131可以将第一部分C的言语识别结果存储在第二缓冲器152中。

[0096] 此后,当通过另一个无声部分TS4的存在检测到第二部分D的话语结束时间点时,输入处理器131可以识别存储在第二缓冲器151中的第二部分D中输入的言语信号。输入处理器131可以生成话语文本X[i'+1]:“不,索纳塔,让我知道燃料效率”作为与言语信号相对应的言语识别结果。

[0097] 此时,输入处理器131可以通过删除存储在第二缓冲器151中的第二部分D的言语信号来初始化第二缓冲器151。另外,输入处理器131可以将第二部分D的话语结束时间点之后输入的言语信号存储在第二缓冲器151中。

[0098] 另外,输入处理器131可以,根据是否可基于第一部分C的言语识别结果或第二部分D的言语识别结果中的至少一个识别用户的意图,生成与在话语结束时间点之后的言语信号相对应的言语识别结果。

[0099] 在这种情况下,当仅使用作为与第二部分D相对应的言语识别结果的话语文本X[i'+1]:“不,索纳塔,让我知道燃油效率”能够识别用户的意图时,输入处理器131可以将用户的意图发送给对话管理器132。当用户的意图经由对话管理器132被发送给结果处理器133时,结果处理器133生成对话响应。

[0100] 可替换地,输入处理器131可以基于作为第二时段D的言语识别结果的话语文本X[i'+1]:“不,索纳塔,让我知道燃油效率”,或作为第一时段C的言语识别结果的话语文本X[i']:“圣达菲”中的至少一个确定用于确定用户意图的意图候选组。输入处理器131可以将意图候选组中选择一个确定为用户的意图。

[0101] 详细地,输入处理器131可以确定关于意图候选组的准确性,并且可以将意图候选组中具有最高准确性的意图候选确定为用户的意图。在这种情况下,可以将意图候选组的准确性计算为概率值。输入处理器131可以将具有最高概率值的意图候选确定为用户的意图。

[0102] 例如,输入处理器131可以基于作为第二部分D的言语识别结果的话语文本X[i'+1]:“不,索纳塔,让我知道燃油效率”确定第一意图候选。另外,输入处理器131可以基于组合第一部分C和第二部分D:“圣达菲,不,索纳塔,让我知道燃油效率”的言语识别结果的结果值来确定第二意图候选。输入处理器131可以将第一意图候选和第二意图候选中具有最高准确性的意图候选确定为用户的意图。

[0103] 此后,输入处理器131可以将确定的用户意图发送给对话管理器132。当用户的意图通过对话管理器132发送给结果处理器133时,结果处理器133可以生成对话响应。

[0104] 另外,当确定用户的意图时,输入处理器131可以通过将言语输入装置110切换为关闭状态来阻止输入言语信号。另外,输入处理器131可以通过删除存储在第二缓冲器151或第二缓冲器152中的至少一个中的数据来初始化第二缓冲器151或第二缓冲器152中的该至少一个。

[0105] 输入处理器131可以通过组合基于话语结束时间点划分的至少一个言语识别部分的言语识别结果来确定用户的意图候选组。输入处理器131可以进一步基于意图候选组的准确性来确定用户的最终意图。因此,准确地识别出用户的意图,输出适合于用户的响应,

并且增加了用户的便利性。

[0106] 图4是示出根据实施例的对话处理方法的流程图。

[0107] 参考图4,对话处理设备100可以识别是否识别出呼叫命令(401)。在这种情况下,可以将呼叫命令设置为预定的呼叫词,并且用户可以通过说出预定的呼叫词或输入言语识别开始命令(例如通过操纵按钮)来发出呼叫命令。

[0108] 当识别出呼叫命令时(操作401中为“是”),对话处理设备100可以将输入言语信号存储在\*\*第一缓冲器151中(402)。在这种情况下,对话处理设备100可以将实时输入的言语信号存储在\*\*第一缓冲器151中。

[0109] 对话处理设备100可以基于存储在\*\*第一缓冲器中的言语信号来生成言语识别结果(403)。

[0110] 对话处理设备100可以识别是否检测到话语结束时间点(404)。详细地,当存储的言语信号中存在从用户停止输入言语的时间点到已经经过预定时间的\*\*时间点对应的无声部分时,对话处理设备100可以将无声部分结束的时间点确定为话语结束时间点。

[0111] 当检测到话语结束时间点时(操作404中为“是”),对话处理设备100可以通过删除存储在\*\*第一缓冲器151中的言语信号(在第n言语识别部分中输入的言语信号)来初始化\*\*第一缓冲器151(405)。对话处理设备100可以将\*\*在话语结束时间点之后输入的言语信号(在第n+1言语识别部分中输入的言语信号)存储在\*\*第一缓冲器151中(406)。

[0112] 此后,对话处理设备100可以识别是否使用基于在话语结束时间点之前的言语信号生成的言语识别结果可识别识别用户的意图(407)。在这种情况下,言语识别结果可以表示在操作403中生成的言语识别结果,即,与在第n个言语识别部分中输入的言语信号相对应的言语识别结果。

[0113] 当无法识别用户的意图时(操作407中为“否”),对话处理设备100可以将所生成的言语识别结果存储在\*\*第二缓冲器152中(410)。换句话说,对话处理设备100可以将\*\*在操作403中生成的言语识别结果(与在第n个言语识别部分中输入的言语信号相对应的言语识别结果)存储在\*\*第二缓冲器152中。此后,对话处理设备100可以基于存储在\*\*第一缓冲器151中的言语信号(在第n+1言语识别部分中输入的言语信号)来生成言语识别结果(403)。在这种情况下,存储在\*\*第一缓冲器151中的言语信号可以表示在话语结束时间点之后输入的言语信号(即,在第n+1言语识别部分中输入的言语信号)。

[0114] 此后,对话处理设备100可以识别是否检测到第n+1言语识别部分的话语结束时间点(404)。当检测到话语结束时间点时(在操作404中为“是”),对话处理设备100可以如上所述执行操作405和406。此后,对话处理设备100可以检查是否可基于存储在\*\*第二缓冲器152中的第n+1个言语识别部分的言语识别结果或第n个言语识别部分的言语识别结果中的至少一个来用户的意图识别(407)。此后,可以重复上述后续处理。

[0115] 作为另一示例,当用户的意图是可识别的时(在操作407中为“是”),对话处理设备100可以阻止输入言语信号并初始化\*\*第一缓冲器151(408)。详细地,对话处理设备100可以通过将言语输入装置110切换为关闭状态来阻止输入言语信号,并且可以通过删除存储在\*\*第一缓冲器151中的数据来初始化\*\*第一缓冲器151。在这种情况下,对话处理设备100可以通过删除存储在\*\*第二缓冲器152中的数据来初始化\*\*第二缓冲器152。

[0116] 对话处理设备100可以生成并输出与用户的意图相对应的响应(409)。

[0117] 输入处理器131可以根据用户的意图是否可识别对话语结束时间点之后的言语信号执行处理操作。输入处理器131不仅可以使使用话语结束时间点之后的言语识别结果,而且还可以将话语结束时间点之前的言语识别结果用作控制识别用户意图的操作的基础。因此,准确地识别出用户的意图,输出适合于用户的响应,并且增加了用户的便利性。

[0118] 参考图4,操作404在操作403之后执行,但是操作403和操作404可以同时执行,操作403也可以在操作404之后执行。但是,操作403在操作404之后执行表示当检测到话语结束时间点时(操作404中为“是”),执行操作403基于存储在第一缓冲器151中的言语信号来生成言语识别结果。

[0119] 图5A和5B是示出根据另一实施例的对话处理方法的流程图。

[0120] 参考图5A和图5B,根据实施例的对话处理设备100可以识别是否识别出呼叫命令(501)。当识别出呼叫命令时(操作501中为“是”),对话处理设备100可以将输入言语信号存储在第二缓冲器152中(502)。在这种情况下,对话处理设备100可以将实时输入的言语信号存储在第二缓冲器152中。

[0121] 对话处理设备100可以基于存储在第二缓冲器中的言语信号来生成言语识别结果(503)。

[0122] 对话处理设备100可以识别是否检测到话语结束时间点(504)。

[0123] 当检测到话语结束时间点时(操作504中为“是”),对话处理设备100可以通过删除存储在第二缓冲器152中的言语信号(在第n言语识别部分中输入的言语信号)来初始化第二缓冲器152(505)。对话处理设备100可以将话语结束时间点之后输入的言语信号(在第n+1言语识别部分中输入的言语信号)存储在第二缓冲器152中(506)。

[0124] 此后,对话处理设备100可以检查是否可使用基于在话语结束时间点之前的言语信号生成的言语识别结果来识别用户的意图(507)。在这种情况下,言语识别结果可以表示在操作503中生成的言语识别结果,即,与在第n个言语识别部分中输入的言语信号相对应的言语识别结果。

[0125] 当不能识别用户的意图时(操作507中为“否”),对话处理设备100可以识别在言语识别结果中是否存在结果值(510)。

[0126] 在这种情况下,当言语识别结果中存在话语文本时,对话处理设备100可以确认结果值存在。换句话说,当在言语识别结果中没有生成话语文本时,例如,由于用户不说出话语,对话处理设备100可以确定在言语识别结果中不存在结果值。

[0127] 当结果值存在于言语识别结果中时(操作510中为“是”),对话处理设备100可以将所生成的言语识别结果存储在第二缓冲器152中(511)。换句话说,对话处理设备100可以将操作503中生成的言语识别结果(与在第n个言语识别部分中输入的言语信号相对应的言语识别结果)存储在第二缓冲器152中。此后,对话处理设备100可以基于存储在第二缓冲器152中的言语信号(在第n+1言语识别部分中输入的言语信号)来生成言语识别结果(503)。在这种情况下,存储在第二缓冲器152中的言语信号可以表示在话语结束时间点之后输入的言语信号(即,在第n+1言语识别部分中输入的言语信号)。

[0128] 此后,对话处理设备100可以识别是否检测到第n+1言语识别部分的话语结束时间点(504)。当检测到话语结束时间点时(在操作504中为“是”),对话处理设备100可以执行如上所述操作505和506。此后,对话处理设备100可以检查是否可基于存储在第二缓冲器152

中的第n+1个言语识别部分的言语识别结果或第n个言语识别部分的言语识别结果中的至少一个来识别用户的意图(507)。此后,可以重复上述后续过程。

[0129] 在另一示例中,当在言语识别结果中不存在结果值时(操作510中为“否”),对话处理设备100可以确定言语识别的次数是否大于或等于参考值(512)。在这种情况下,言语识别的次数可以代表生成言语识别结果的次数。另外,用于言语识别的次数的参考值可以表示考虑到存储器150的存储容量而获得的言语识别的最大次数。

[0130] 当言语识别的次数小于参考值时(操作512中的“否”),对话处理设备100可以基于存储在第一缓冲器151中的言语信号来生成言语识别结果(503)。在这种情况下,存储在第一缓冲器151中的言语信号可以表示在话语结束时间点之后输入的言语信号(即,在第n+1言语识别部分中输入的言语信号)。此后,可以重复上述后续过程。

[0131] 在另一个示例中,当言语识别的次数等于或大于参考值时(操作512中为“是”),或者当用户的意图可识别时(操作507中为“是”),对话处理设备100可以阻止言语信号的输入并初始化第一缓冲器151(508)。详细地,对话处理设备100可以通过将言语输入装置110切换为关闭状态来阻止输入言语信号,并且可以通过删除存储在第一缓冲器151中的数据来初始化第一缓冲器151。在这种情况下,对话处理设备100可以通过删除存储在第二缓冲器152中的数据来初始化第二缓冲器152。

[0132] 对话处理设备100可以生成并输出与用户的意图相对应的响应(509)。在这种情况下,对话处理设备100可以将言语识别的次数设置为初始值。当在言语识别的次数被设置为初始值之后识别出呼叫命令时,对话处理设备100可以生成与第一言语识别部分相对应的言语识别结果。

[0133] 输入处理器131可以根据用户的意图是否可识别对话语结束时间点之后的言语信号执行处理操作。输入处理器131不仅可以使使用话语结束时间点之后的言语识别结果,而且还可以将话语结束时间点之前的言语识别结果用作控制识别用户意图的操作的基础。因此,准确地识别出用户的意图,输出适合于用户的响应,并且增加了用户的便利性。

[0134] 此外,由于根据言语识别的次数来生成针对用户输入的言语的言语识别结果,因此可以考虑存储容量来进行有效的言语识别。

[0135] 参考图5A,操作504在操作503之后执行,但是操作503和操作504可以同时执行,操作503也可以在操作504之后执行。但是,操作503在操作之后执行504表示当检测到话语结束时间点时(操作504中为“是”),执行操作503基于存储在第一缓冲器151中的言语信号来生成言语识别结果。

[0136] 所公开的实施例可以以存储可由计算机执行的指令的记录介质的形式体现。指令可以以程序代码的形式存储,并且当由处理器执行时,可以生成程序模块以执行所公开的实施例的操作。记录介质可以体现为计算机可读记录介质。

[0137] 计算机可读记录介质包括其中存储了可以由计算机解码的指令的各种记录介质,例如,只读存储器(ROM)、随机存取存储器(RAM),磁带磁盘、闪存、光学数据存储装置等。

[0138] 从以上显而易见,对话处理设备、包括该对话处理设备的车辆和其对话处理方法可以提高识别用户言语和处理对话的准确性。通过准确识别用户的意图,可以提高用户的便利性。

[0139] 尽管已经出于说明性目的描述了本公开的实施例,但是本领域普通技术人员应当

理解,在不脱离本公开的范围和精神的情况下,可以进行各种修改、增加和替换。因此,没有出于限制目的描述本公开的实施例。

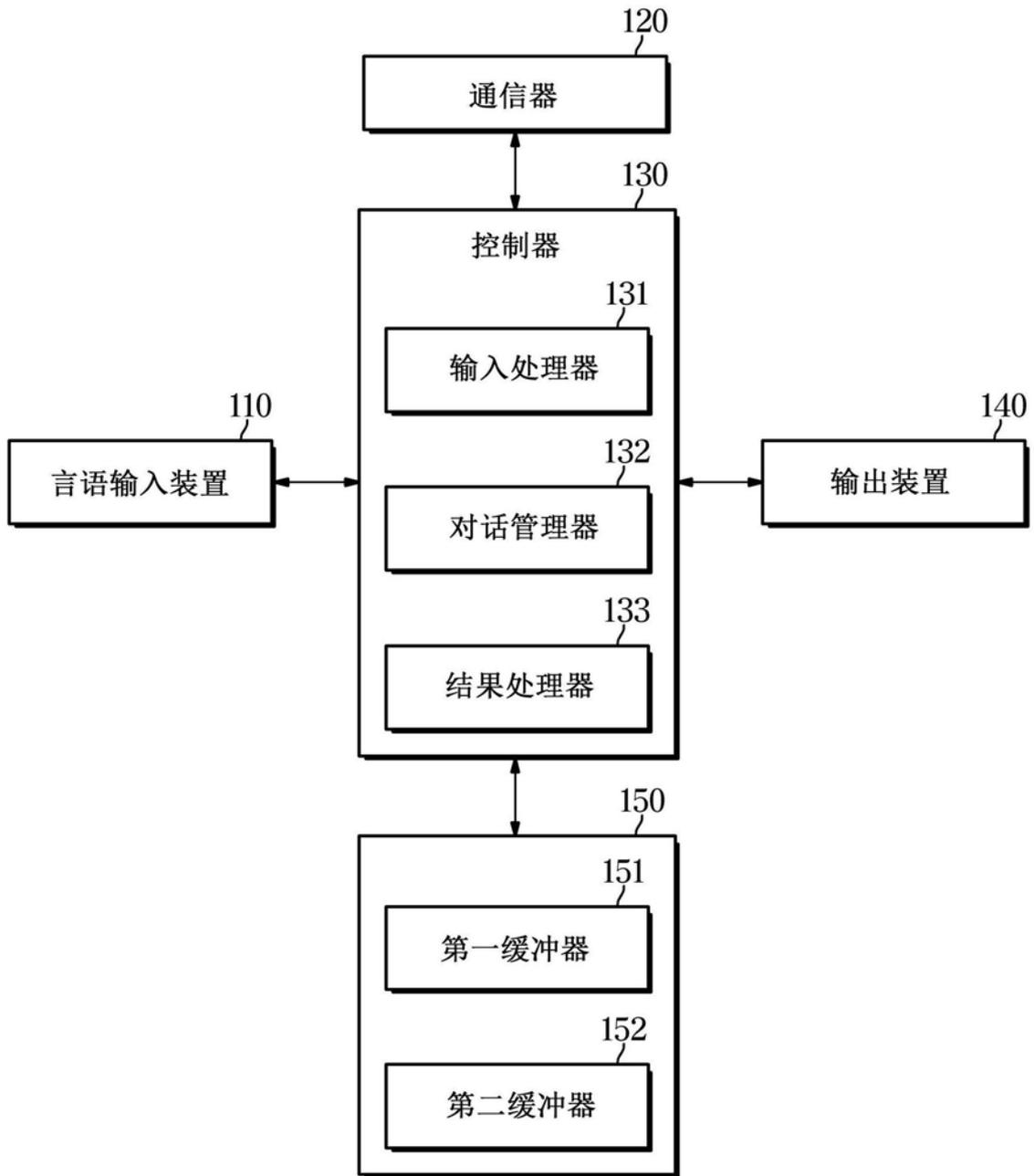


图1

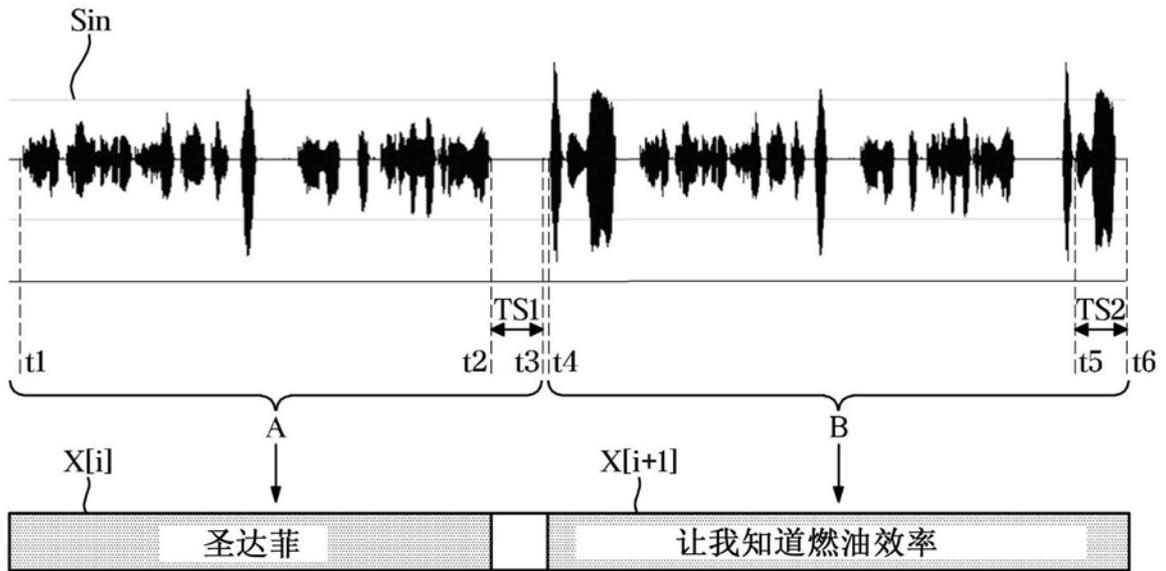


图2

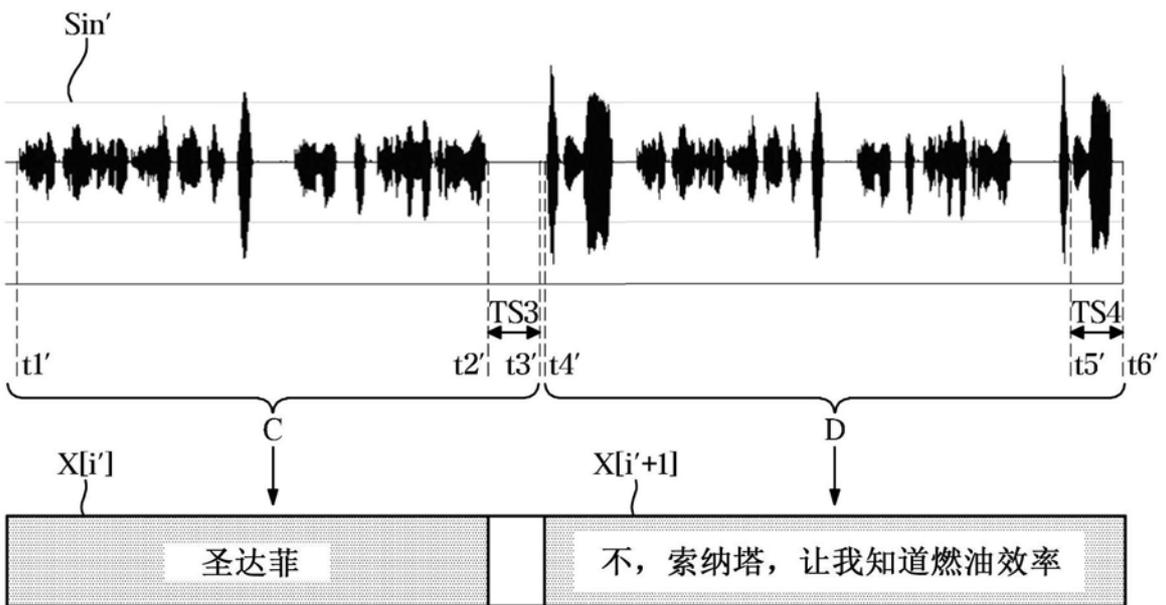


图3

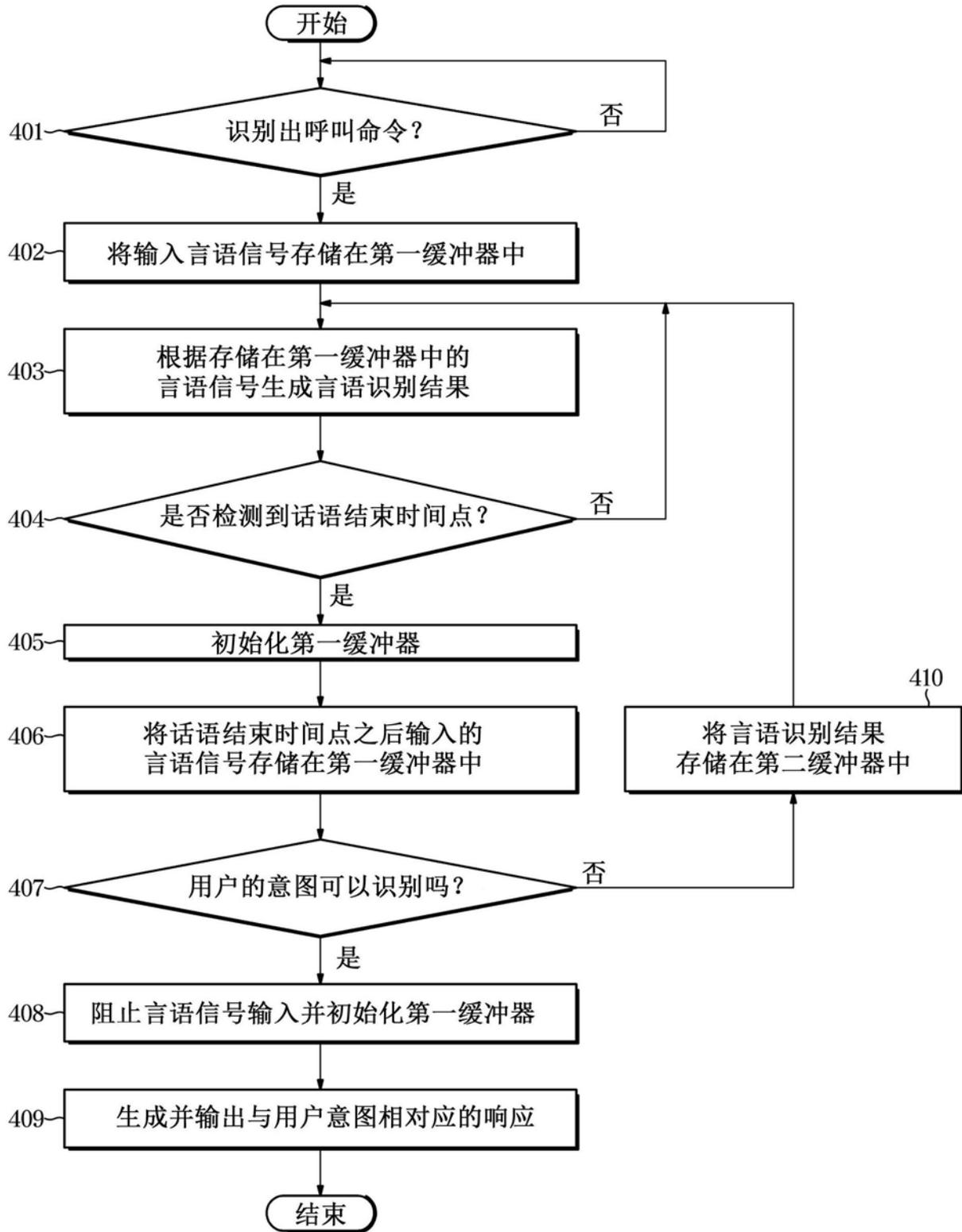


图4

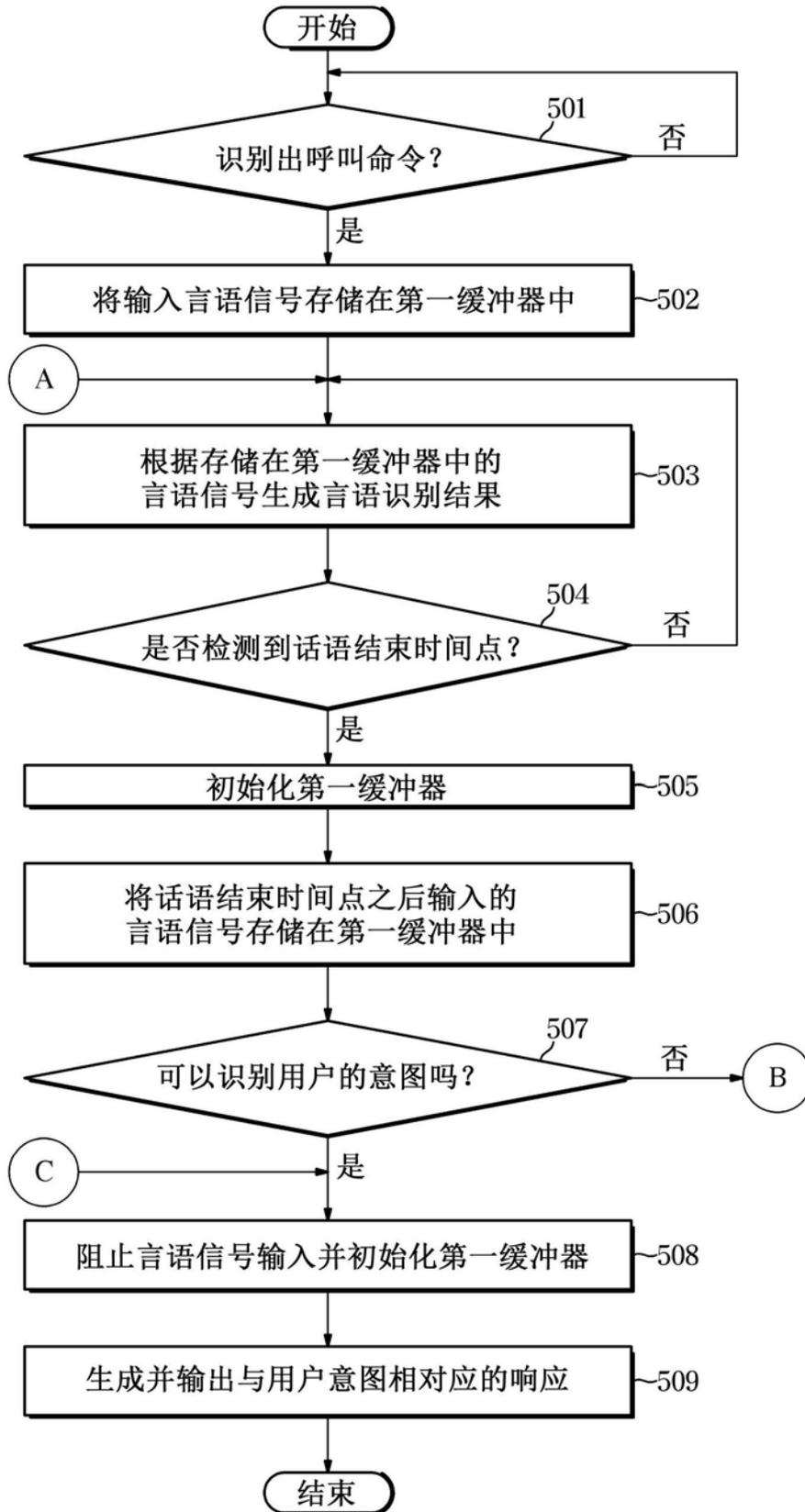


图5A

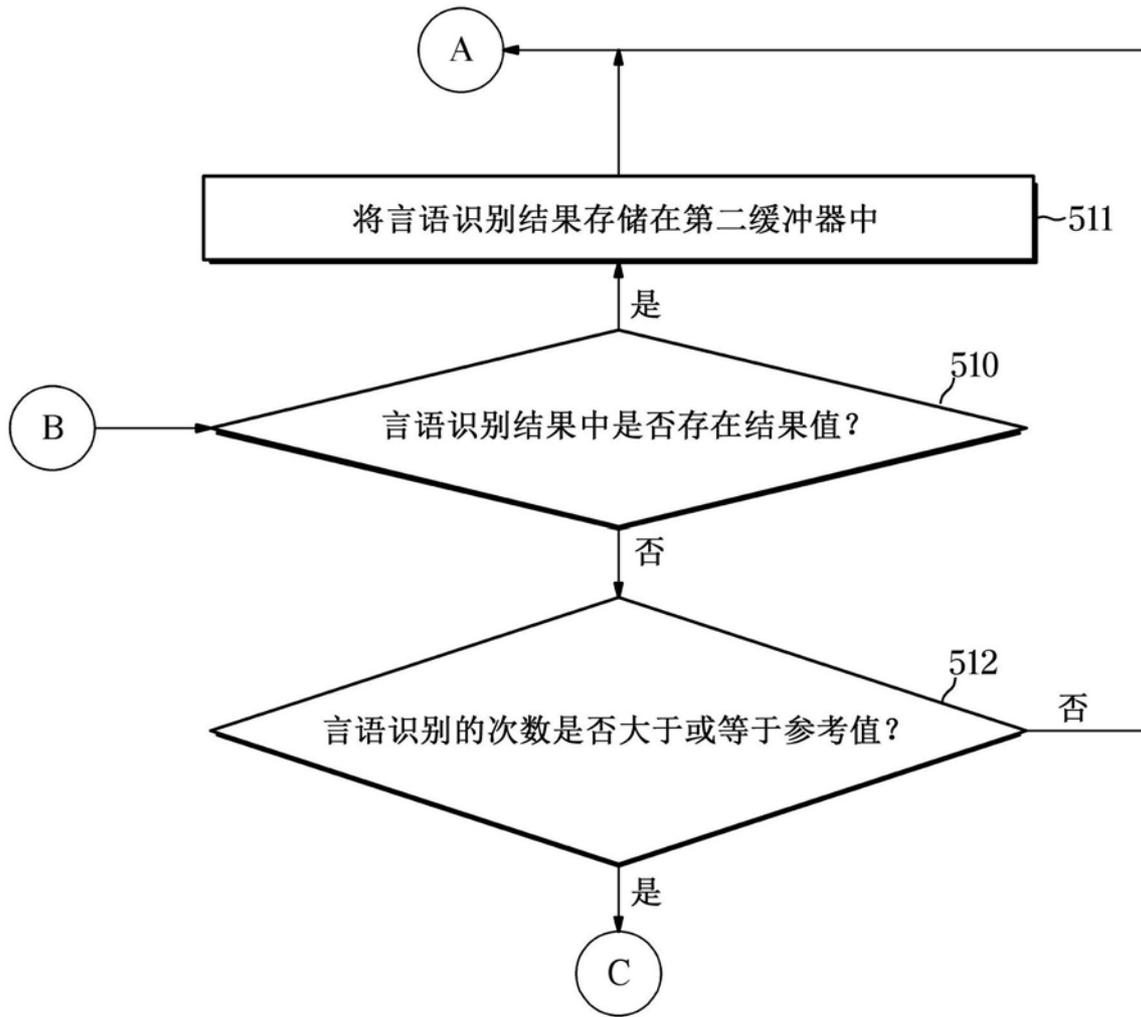


图5B