



(12)发明专利申请

(10)申请公布号 CN 108630203 A

(43)申请公布日 2018. 10. 09

(21)申请号 201810175617.5

G10L 15/18(2013.01)

(22)申请日 2018.03.02

G10L 15/19(2013.01)

(30)优先权数据

G10L 13/10(2013.01)

2017-040580 2017.03.03 JP

G06F 17/28(2006.01)

(71)申请人 国立大学法人京都大学

地址 日本京都府京都市

申请人 丰田自动车株式会社

(72)发明人 河原达也 高梨克也 中西亮辅

渡部生圣

(74)专利代理机构 北京集佳知识产权代理有限公司

公司 11227

代理人 杜诚 皇甫悦

(51)Int.Cl.

G10L 15/22(2006.01)

G10L 15/08(2006.01)

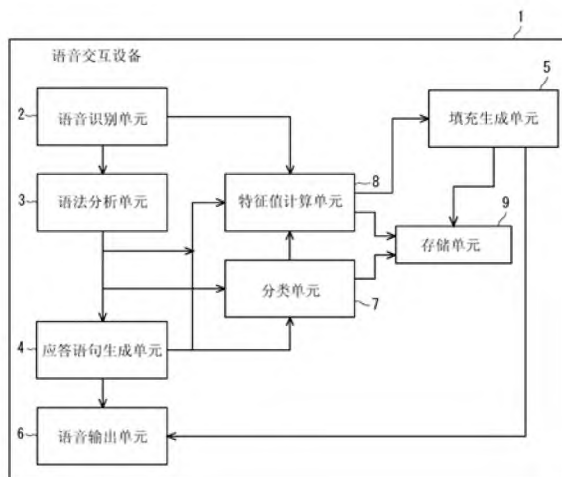
权利要求书2页 说明书8页 附图5页

(54)发明名称

语音交互设备及其处理方法和程序

(57)摘要

提供了语音交互设备及其处理方法和程序。该语音交互设备包括：用于识别用户的语音的语音识别装置；应答语句生成装置，用于基于所识别的语音来生成对用户的语音的应答语句；填充生成装置，用于生成要被插入到对话中的填充词；输出装置，用于输出所生成的应答语句和所生成的填充词；以及分类装置，用于将所生成的应答语句分类为指示预定义的话语类型的预定话语模式之一。当用户在第一应答语句之后发出语音之后输出装置输出填充词并输出第二应答语句时，分类装置将第一应答语句分类为话语模式之一，并且填充生成装置基于第一应答语句已被分类为的话语模式来生成填充词。



1. 一种语音交互设备,包括:
  - 用于识别用户的语音的语音识别装置;
  - 应答语句生成装置,用于基于由所述语音识别装置识别的所述用户的语音来生成对所述语音的应答语句;
  - 填充生成装置,用于生成要被插入到与所述用户的对话中的填充词;以及
  - 输出装置,用于输出由所述应答语句生成装置生成的所述应答语句和由所述填充生成装置生成的所述填充词,其中,
    - 所述语音交互设备还包括分类装置,所述分类装置用于将由所述应答语句生成装置生成的所述应答语句分类为指示预定义的话语类型的预定话语模式之一,并且
    - 当所述用户在第一应答语句之后发出语音之后所述输出装置输出所述填充词并输出第二应答语句时,
    - 所述分类装置将所述第一应答语句分类为话语模式之一,并且
    - 所述填充生成装置基于已由所述分类装置将所述第一应答语句分类为的话语模式来生成所述填充词。
2. 根据权利要求1所述的语音交互设备,其中,所述语音交互设备还包括:
  - 存储装置,用于存储表格信息,所述表格信息包括话语模式和关于与所述话语模式相关联的特征值的类型的信息;以及
  - 特征值计算装置,用于基于关于与已由所述分类装置将所述第一应答语句分类为的话语模式相关联的特征值的类型的信息来计算先前或后续话语的特征值,其中,
    - 所述填充生成装置基于由所述特征值计算装置计算的所述特征值来生成所述填充词。
3. 根据权利要求2所述的语音交互设备,其中,关于所述特征值的类型的所述信息包括以下各项中至少一者:所述先前话语的韵律信息、所述先前话语的语言信息、所述后续话语的语言信息和所述后续话语的韵律信息。
4. 根据权利要求2或3所述的语音交互设备,其中,
  - 所述存储装置存储与填充类型的各个特征值相关联的填充形式信息,每个所述填充类型包括至少一个填充词并且指示所述填充词的类型,并且
  - 所述填充生成装置基于已由所述分类装置将所述第一应答语句分类为的话语模式来缩小所述填充类型的数目、从缩小数目后的填充类型中选择与由所述特征值计算装置计算的所述特征值相关联的一种填充类型、并且通过选择所选择的填充类型中包括的填充词来生成所述填充词。
5. 一种用于语音交互设备的处理方法,所述语音交互设备包括:
  - 用于识别用户的语音的语音识别装置;
  - 应答语句生成装置,用于基于由所述语音识别装置识别的所述用户的语音来生成对所述语音的应答语句;
  - 填充生成装置,用于生成要被插入到与所述用户的对话中的填充词;以及
  - 输出装置,用于输出由所述应答语句生成装置生成的所述应答语句和由所述填充生成装置生成的所述填充词;所述处理方法包括:
  - 当所述用户在第一应答语句之后发出语音之后所述输出装置输出所述填充词并输出

第二应答语句时，

将所述第一应答语句分类为指示预定义的话语类型的预定话语模式之一，并且基于所述第一应答语句已被分类为的话语模式来生成所述填充词。

6. 一种用于语音交互设备的程序，所述语音交互设备包括：

用于识别用户的语音的语音识别装置；

应答语句生成装置，用于基于由所述语音识别装置识别的所述用户的语音来生成对所述语音的应答语句；

填充生成装置，用于生成要被插入到与所述用户的对话中的填充词；以及

输出装置，用于输出由所述应答语句生成装置生成的所述应答语句和由所述填充生成装置生成的所述填充词，

所述程序适用于使计算机执行：

当所述用户在第一应答语句之后发出语音之后所述输出装置输出所述填充词并输出第二应答语句时，

将所述第一应答语句分类为指示预定义的话语类型的预定话语模式之一，并且基于所述第一应答语句已被分类为的话语模式来生成所述填充词。

## 语音交互设备及其处理方法和程序

### 技术领域

[0001] 本公开内容涉及一种与用户进行语音交互的语音交互设备及其处理方法和程序。

### 背景技术

[0002] 已知一种语音交互设备,其插入填充词(即用于填充对话中的沉默的词)以防止对话中的沉默被不自然地延长(参见日本未审查专利申请公开第2014-191030号)。

[0003] 然而,本发明人发现了以下问题。即,当对话中出现等待时间时,前述语音交互设备输出形式上的(即敷衍的)填充词作为用于填充沉默的词。因此,插入的填充词有可能不很适合对话的内容(例如含义),从而使对话变得不自然。

### 发明内容

[0004] 已经做出本公开内容以解决上述问题,并且本公开内容的目的是提供一种能够进行更自然的对话的语音交互设备及其处理方法和程序。

[0005] 为了实现上述目的,第一示例性方面是一种语音交互设备,包括:用于识别用户的语音的语音识别装置;应答语句生成装置,其用于基于由语音识别装置识别的用户的语音来生成对语音的应答语句;填充生成装置,其用于生成要被插入到与用户的对话中的填充词;以及输出装置,其用于输出由应答语句生成装置生成的应答语句和由填充生成装置生成的填充词,其中,该语音交互设备还包括分类装置,其用于将由应答语句生成装置生成的应答语句分类为指示预定义的话语类型的预定话语模式之一,并且当用户在第一应答语句之后发出语音之后输出装置输出填充词并输出第二应答语句时,该分类装置将第一应答语句分类为话语模式之一,并且填充生成装置基于已由分类装置将第一应答语句分类为的话语模式来生成填充词。

[0006] 根据该方面,通过根据与第一应答语句对应的话语模式生成填充词,可以生成适合于第一应答语句之后的用户语音的内容(例如含义)的最优填充词。通过输出该最优填充词,可以进行更自然的对话。

[0007] 在这方面,语音交互设备还可以包括:存储装置,其用于存储表格信息,表格信息包括话语模式和关于与话语模式相关联的特征值的类型的信息;以及特征值计算装置,用于基于关于与已由分类装置将第一应答语句分类为的话语模式相关联的特征值的类型的信息来计算先前或后续话语的特征值,其中,填充生成装置可以基于由特征值计算装置计算的特征值来生成填充词。

[0008] 以这样的方式,可以将第一应答语句分类为预定话语模式之一、计算与第一应答语句已被分类为的话语模式对应的最优先前或后续话语的特征值、并且通过使用所计算的特征值来生成用于对话的最优填充词。

[0009] 在这方面,关于特征值的类型的信息可以包括以下各项中至少一者:先前话语的韵律信息、先前话语的语言信息、后续话语的语言信息和后续话语的韵律信息。

[0010] 以这种方式,可以实现处理负荷的降低和适当填充词的生成二者,从而改善对话

的节奏和自然性。

[0011] 在这方面,存储装置可以存储与填充类型的各个特征值相关联的填充形式信息,每个填充类型包括至少一个填充词并指示该填充词的类型,并且填充生成装置可以基于已由分类装置将第一应答语句分类为的话语模式来缩小填充类型的数目、从缩窄后的填充类型中选择与由特征值计算装置计算的特征值相关联的一种填充类型,并且通过选择所选择的填充类型中包括的填充词来生成填充词。

[0012] 通过如上所述预先缩小填充类型的数目,可以更进一步降低处理负荷。

[0013] 为了实现上述目的,另一示例性方面可以是一种用于语音交互设备的处理方法,该语音交互设备包括:用于识别用户的语音的语音识别装置;应答语句生成装置,其用于基于由语音识别装置识别的用户的语音来生成对语音的应答语句;填充生成装置,其用于生成要被插入到与用户的对话中的填充词;以及输出装置,其用于输出由应答语句生成装置生成的应答语句和由填充生成装置生成的填充词,该处理方法包括:当用户在第一应答语句之后发出语音之后输出装置输出填充词并输出第二应答语句时,将第一应答语句分类为指示预定义的话语类型的预定话语模式之一,并且基于第一应答语句已被分类为的话语模式来生成填充词。

[0014] 为了实现上述目的,另一示例性方面可以是一种用于语音交互设备的程序,该语音交互设备包括:用于识别用户的语音的语音识别装置;应答语句生成装置,其用于基于由语音识别装置识别的用户的语音来生成对语音的应答语句;填充生成装置,其用于生成要被插入到与用户的对话中的填充词;以及输出装置,其用于输出由应答语句生成装置生成的应答语句和由填充生成装置生成的填充词,该程序适用于使计算机执行:当用户在第一应答语句之后发出语音之后输出装置输出填充词并输出第二应答语句时,将第一应答语句分类为指示预定义的话语类型的预定话语模式之一,并且基于第一应答语句已被分类为的话语模式来生成填充词。

[0015] 根据本公开内容,可以提供一种能够进行更自然的对话的语音交互设备及其处理方法和程序。

[0016] 从下文中给出的详细描述和附图中将更充分地理解本公开内容的以上和其他目的、特征和优点,附图仅以图示的方式给出,因此不被认为是对本公开内容的限制。

#### 附图说明

[0017] 图1是示出根据本公开内容的实施例的语音交互设备的示意性系统配置的框图;

[0018] 图2示出了话语模式的示例;

[0019] 图3示出了由设备生成的话语和由用户发出的话语的示例;

[0020] 图4示出了前述由设备生成的话语与由用户发出的话语之间的对应关系的示例;

[0021] 图5示出了将话语模式和特征值的类型相关联的表格信息的示例;

[0022] 图6示出了填充类型的示例;以及

[0023] 图7是示出由根据本公开内容的实施例的语音交互设备执行的处理方法的流程的流程图。

#### 具体实施方式

[0024] 以下参照附图来说明根据本公开内容的实施例。

[0025] 图1是示出根据本公开内容的实施例的语音交互设备的示意性系统配置的框图。根据本实施例的语音交互设备1安装在例如机器人、PC(个人计算机)或移动终端(诸如智能电话或平板计算机)中并与用户进行对话。

[0026] 语音交互设备1包括例如识别用户的语音的语音识别单元2、分析语音的语句结构的语法分析单元3、生成对用户的语音的应答语句的应答语句生成单元4、生成填充词的填充生成单元5、输出语音的语音输出单元6、对话语进行分类的分类单元7、计算特征值的特征值计算单元8以及存储数据的存储单元9。

[0027] 注意,语音交互设备1例如由主要使用微型计算机的硬件形成,所述微型计算机包括执行算术处理等的CPU(中央处理单元)、由ROM(只读存储器)和RAM(随机存取存储器)组成并且存储由CPU执行的算术程序等的存储器、从外部接收和输出信号的接口单元(I/F)等。CPU、存储器和接口单元通过数据总线等相互连接。

[0028] 语音识别单元2是语音识别装置的具体示例。语音识别单元2例如基于关于通过麦克风输入的用户语音的信息执行语音识别处理,将用户语音的信息转换成文本,从而将其识别为字符串信息。

[0029] 例如,语音识别单元2将通过麦克风输入的用户语音的信息转换为数字信号。语音识别单元2通过从上述数字化信息中检测话语片段并通过参考统计语言模型等对检测到的话语片段中的语音信息进行模式匹配来进行语音识别。

[0030] 注意,统计语言模型是例如用于计算语言表达的出现概率(例如词的出现分布和在特定词之后出现的词的分布)的概率模型,并且该模型通过在语素的基础上学习连接概率来获得。统计语言模型被预先存储在存储单元9等中。语音识别单元2针对用户的语音信息中的每个语素生成已添加成分信息的语素信息,即,通过向每个语素添加成分类型(例如名词、形容词、动词、副词等)而获得的信息。语音识别单元2将所识别的用户的语音信息输出到语法分析单元3。

[0031] 存储单元9是存储装置的具体示例。存储单元9由例如存储器等构成。

[0032] 语法分析单元3分析由语音识别单元2识别的语音信息的语句结构。例如,语法分析单元3利用常规语素分析器对表示经口语识别的用户语音信息的字符串信息执行语素分析等并且对字符串信息执行语义分析。语法分析单元3将对字符串信息的分析结果(包括语素信息、修饰信息、通过识别获得的文本等)输出到应答语句生成单元4。

[0033] 应答语句生成单元4是应答语句生成装置的具体示例。应答语句生成单元4基于经语法分析单元3分析的语音信息的语句结构来生成对用户的语音信息的应答语句。

[0034] 例如,应答语句生成单元4基于从语法分析单元3输出的对字符串信息的分析结果来生成对用户的语音信息的应答语句。更具体地,应答语句生成单元4从字符串信息“我玩围棋”中提取一个谓术语“玩围棋”。语法分析单元3通过将所提取的语素串“玩围棋”用作特征向量并且使用SVM(支持向量机)模型来分析该字符串信息,并且确定该字符串信息属于两个预定义类别中的哪一类。

[0035] 当作为字符串信息的分析结果确定该字符串信息属于两个预定义类别之一时,应答语句生成单元4生成应答语句“那听起来不错”。另一方面,当作为字符串信息的分析结果确定该字符串信息属于两个预定义类别中的另一类时,应答语句生成单元4生成应答

语句“这似乎很麻烦”。注意,上述用于生成应答语句的方法仅仅是示例,并且该方法不限于上述示例。也就是说,可以使用用于生成应答语句的任意方法(例如增加了预定义的类别的数量的方法)。此外,预定义的类别可以是用于诸如“积极的”和“消极的”之类的情绪的类别,并且应答语句生成单元4可以做出关于情绪的确定的。应答语句生成单元4将生成的应答语句输出到语音输出单元6。

[0036] 填充生成单元5是填充生成装置的具体示例。填充生成单元5生成被插入到用户与语音交互设备1之间的对话中的填充词。填充词是用于填充对话中的沉默的词,例如“I mean (我的意思是)”、“Let’s see (让我们来看看)”以及“Yah (呀)”。填充生成单元5将所生成的填充词输出到语音输出单元6。

[0037] 语音输出单元6是输出装置的具体示例。语音输出单元6将由应答语句生成单元4生成的应答语句和由填充生成单元5生成的填充词输出为语音。例如,语音输出单元6通过使用扬声器等向用户输出与应答语句和填充词对应的语音。

[0038] 例如,语音输出单元6基于填充词和应答语句来合成与由填充生成单元5生成的填充词和由应答语句生成单元4生成的应答语句对应的语音,并且输出经合成的语音。语音输出单元6可以通过选择预先设置(即存储)在存储单元9等中的语音文件之一来输出与填充词和应答语句对应的语音。

[0039] 顺便提及,相关技术中的语音交互设备在对话中出现等待时间时输出形式上的(即敷衍的)填充词作为用于填充沉默的词。因此,这些填充词可能不很适合对话的内容(例如含义)从而使对话变得不自然。

[0040] 与此相反,根据本实施例的语音交互设备1包括分类单元7,其将由应答语句生成单元4生成的应答语句分类为指示预定义的话语类型的预定话语模式之一。在用户于第一应答语句后发出语音之后,语音输出单元6输出填充词并输出第二应答语句。在这种情况下,分类单元7将第一应答语句分类为话语模式之一。填充生成单元5基于由分类单元7将该第一应答语句分类为的话语模式来生成填充词。

[0041] 后续的用户语音和第二应答语句的内容(例如含义)根据第一应答语句的内容(例如含义)而改变。因此,通过基于与第一应答语句对应的话语模式生成填充词,可以生成适合于用户语音和第二应答语句的内容(例如含义)的最优填充词。通过输出该最优填充词,可以进行更自然的对话。例如,在小型谈话的情况下,与插入形式上的(即敷衍的)填充词的对话不同,通过输出最优填充词可以以自然的方式填充对话中的沉默,就像在与真人的对话中一样。

[0042] 分类单元7是分类装置的具体示例。分类单元7将由语音输出单元6输出的应答语句(以下称为设备话语)和由用户发出的语音(以下称为用户话语)分类为多个话语模式之一。话语模式指示预定义的话语类型。

[0043] 例如,根据用户或由语音交互设备1进行的对话行为将话语模式分类为多个话语模式,并将经分类的话语模式设置(即存储)在存储单元9等中。对话行为指示用户话语与设备话语之间的关系。

[0044] 例如,如图2所示,话语模式被分类为:(1)用于询问事实的询问模式,“是/否询问模式”,(2)用于从组中进行选择的询问模式,“Wh(何时、哪里、什么或为什么)询问模式”,(3)包括回答、同意、接受等的模式,“回答模式”,(4)包括自我批露、请求等的模式,“批露模

式”，(5)包括问候、介绍等的模式，“其他模式”，等等。注意，以上所示的话语模式仅仅是示例，并且话语模式不限于以上所示的示例。

[0045] 图3示出了设备话语和用户话语的示例。在图3中，设备话语和用户话语按时间顺序排列。时间T-3至T-1是过去(即在当前时间T之前)的设备话语和用户话语的时间。此外，时间T+1至T+3是在当前时间T之后的设备话语和用户话语的时间。

[0046] 例如，如图3所示，分类单元7可以基于规则或者通过使用分类模型，基于从时间T-3至T+3的用户话语和设备话语，对这些用户话语和这些设备话语进行分类。更具体地，分类单元7将时间T处的设备话语“你有什么爱好吗？”分类为针对用户的“Wh询问模式”。此外，分类单元7将后续在时间T+1处的用户话语“我玩围棋”分类为“回答模式”。此后，分类单元7通过使用状态序列估计模型(条件随机场)等将后续在时间T+3处的用户话语“那听起来不错”分类为“回答模式”。

[0047] 分类单元7可以将先前的用户话语和在其之后的设备话语进行组合，并对经组合的话语进行分类。例如，当话语模式是“是/否询问模式”时，在其之后的话语模式必然被确定为“回答模式”。因此，当先前的用户话语是“是/否询问模式”时，分类单元7将在其之后的设备话语分类为“回答模式”。类似地，当先前的用户话语是“回答模式”时，分类单元7将在其之后的设备话语分类为“回答模式”。

[0048] 分类单元7可以基于规则根据在用户话语之前的设备话语对该后续的用户话语进行分类。例如，如图4所示，当在时间T处的设备话语是“是/否询问模式”或“Wh询问模式”时，分类单元7基于规则将在时间T+1处的用户话语分类为“回答模式”。类似地，当时间T处的设备话语是“回答模式”时，分类单元7基于规则将时间T+1处的用户话语分类为“批露模式”。

[0049] 分类单元7可以基于包括成分及其类型、谓项及其类型、正确分类在内的数据预先学习分类模型，并且通过使用学习结果，对时间T+1处的用户话语进行分类。分类单元7可以基于从时间T-3至T+1的话语类型对时间T+2处的设备话语进行分类。

[0050] 特征值计算单元8是特征值计算装置的具体示例。特征值计算单元8基于关于与已由分类单元7将第一应答语句(设备话语)分类为的话语模式相关联的特征值的类型的信息来计算先前或后续话语的特征值。

[0051] 注意，在对话中插入的最优填充形式根据该对话的场景而变化。例如，在用户在时间T处的第一应答语句之后在时间T+1处发出语音后、在时间T+2处输出填充词并且在时间T+3处输出第二应答语句的对话场景中，在第一应答语句之后插入的填充的最优形式根据第一应答语句的内容(例如含义)而变化。此外，如何识别最适合用于确定最优填充形式的(一个或多个)特征也是变化的。

[0052] 例如，在用于询问事实的“是/否询问模式”之后插入填充的情况下，先前话语的韵律信息(例如话语的长度)显著影响该填充的形式。因此，为了进行更自然的对话，尤其有效的是：识别先前话语的韵律信息的特征、选择适合于所识别的特征的填充形式并且根据所选择的填充形式生成填充词。

[0053] 因此，特征值计算单元8基于关于与已由分类单元7将第一实施例分类为的话语模式相关联的特征值的类型的信息来计算先前或后续话语的特征值。填充生成单元5根据由特征值计算单元8计算出的特征值来生成最优填充词。

[0054] 以这种方式，可以将第一应答语句分类为预定话语模式之一、计算与第一应答语



句已被分类为的话语模式(在下文中简称为“分类的话语模式”)对应的最优先前或后续话语的特征值,并且通过使用所计算的特征值来生成针对该对话最合适的填充词。

[0055] 如图5所示,例如,关于特征值的类型的信息与话语模式相关联。图5示出了将话语模式与特征值的类型相关联的表格信息的示例。该表格信息被预先设置(即存储)在存储单元9等中。注意,图5所示的表格信息仅仅是示例,表格信息不限于该示例。

[0056] 在图5中,例如,“是/否询问模式”与“先前话语的韵律信息”相关联。“Wh询问模式”与“先前话语的韵律信息”和“先前话语的语言信息”相关联。此外,“回答模式”与“后续话语的语言信息”相关联。

[0057] 注意,“先前话语的韵律信息”例如包括先前话语中的:在口语话语末尾的预定时段(大约100毫秒)中的F0(基频)、最大值、最小值、平均值和功率的一阶回归系数(倾斜度)、话语的长度、停顿的长度、话语速度等。

[0058] “先前话语的语言信息”包括例如先前话语中的:话语末尾的成分、话语末尾的从句边界标签、词的数目、从句的数目等。根据紧接在从句边界之后的间隙的大小将从从句边界标签分为三个等级,即诸如“我认为”的绝对边界、诸如“虽然”的强边界以及诸如“如果”的弱边界。“后续话语的语言信息”包括后续话语中的:话语开头的成分、词的数目、从句的数目等。

[0059] 特征值计算单元8基于话语模式和存储在存储单元9中的表格信息来计算与已由分类单元7将第一应答语句分类为的话语模式(在下文中简称为“由分类单元7分类的话语模式”)对应的特征值(特征向量)。特征值计算单元8基于从语音识别单元2提供的用户的语音信息、从语法分析单元3提供的字符串信息、从应答语句生成单元4提供的应答语句(文本数据)等来计算先前话语的韵律信息、先前话语的语言信息以及后续话语的语言信息的特征向量。

[0060] 例如,特征值计算单元8基于存储在存储单元9中的表格信息来计算与由分类单元7分类的话语模式“回答模式”对应的“后续话语的语言信息”(话语开头的成分、词的数目、从句的数目等)的特征向量。

[0061] 如图5所示,关于多种类型的特征值的信息可以与一种话语模式相关联。例如,先前话语的韵律信息和先前话语的语言信息可以与话语模式“Wh询问模式”相关联。在这种情况下,特征值计算单元8基于存储在存储单元9中的表格信息来计算与由分类单元7分类的话语模式“Wh询问模式”对应的“先前话语的韵律信息”的特征向量和与该话语模式对应的“先前话语的语言信息”的特征向量。然后,特征值计算单元8计算这些特征向量的统一特征向量。

[0062] 注意,基于后续话语的韵律信息执行的处理涉及后续话语及其语法分析。因此,其处理负荷大并且提前时间增加(约5秒)。类似地,基于先前话语的语言信息执行的处理涉及其语法分析。因此,其处理负荷大并且提前时间增加(约4秒)。由于如此大的处理负荷,对话的节奏和自然性可能劣化。因此,如果可能的话,仅基于先前话语的韵律信息执行的处理是优选的。然而,存在以下情况:为了生成适当的填充词,使用语言信息是优选的。

[0063] 因此,在该实施例,在表格信息中将“韵律信息”和“语言信息”与话语模式相关联以便实现处理负荷的降低和适当的填充词的生成二者。以这种方式,可以实现处理负荷的降低和适当的填充词的生成二者,从而改善对话的节奏和自然性。

[0064] 如上所述,可以将话语的特征值大致分类为语言信息中的语言特征(例如成分串)和韵律信息中的声学特征(诸如音高)。如上所述,语言特征的使用施加了大量的处理负荷并增加了输出结果所需的时间。因此,在该实施例中主要使用声学特征,但是会根据需要使用语言特征。以这种方式,可以有效地识别话语的特征并由此生成适当的填充词,同时降低处理负荷。

[0065] 填充生成单元5基于由特征值计算单元8计算的特征值来生成填充词。例如,填充生成单元5通过使用关于预定义的填充形式的信息来生成填充词。

[0066] 填充形式信息例如被预先设置(即存储)在存储单元9等中。例如,如图6所示,在填充形式信息中,(一个或多个)特征值与填充词的每种类型(在下文中称为填充类型)相关联。填充类型包括固有类型、应答词类型、指示词类型、状语类型、意识类型、其他类型、无等。

[0067] 一个或多个填充词与每种填充类型相关联。例如,填充词“Let's see(让我们来看看)”、“Well(好吧)”等与固有类型相关联。通过使用具有像诸如随机森林的决策树的组合的结构的学习机器,将特征值优选地与相应的填充类型相关联。注意,可以使用任意的学习机器。例如,可以使用诸如逻辑回归的学习机器。

[0068] 填充生成单元5基于所计算的特征值和存储在存储单元9中的填充形式信息来选择与由特征值计算单元8计算的特征值相关联的填充类型。然后,填充生成单元5根据所选择的填充类型选择一个填充词来生成填充词。以这种方式,可以通过使用指示对话场景的特征的特征值来选择适合于该对话场景的填充形式,并且基于所选择的填充形式来生成最优填充词。

[0069] 例如,填充生成单元5根据所选择的填充类型随机选择一个填充词。填充生成单元5可以根据预定的优先顺序根据所选择的填充类型选择一个填充词。可替代地,可以使用任意其他选择方法。

[0070] 尽管填充生成单元5基于由特征值计算单元8计算的特征值和存储在存储单元9中的填充形式信息来从七种类型的填充词中选择与该特征值相关联的一种填充类型,但是选择方法不限于该示例。填充生成单元5可以基于由分类单元7分类的话语模式来缩小填充类型的数目、从缩窄后的填充类型中选择与特征值相关联的一种填充类型、并且基于所选择的填充类型生成填充词。通过如上所述预先缩小填充类型的数目,可以更进一步降低处理负荷。

[0071] 例如,当话语模式是“回答”时,自然可以将适合于该话语模式的可能的填充类型预先缩窄为四种填充类型(即固有类型、状语类型、其他类型以及无)。因此,填充生成单元5基于由分类单元7分类的话语模式“回答”而将七种填充类型缩窄为四种填充类型(即固有类型、状语类型、其他以及无),从缩窄后的填充类型中选择与特征值相关联的一种填充类型,并且基于所选择的填充类型生成填充词。例如,将缩窄后的填充类型与话语模式相关联的信息被设置(即存储)在存储单元9等中。填充生成单元5可以基于在存储单元9中存储的该信息来执行上述对填充类型的缩窄。

[0072] 当填充生成单元5基于由特征值计算单元8计算的特征值和存储在存储单元9中的填充形式信息而选择一种填充类型“无”时,填充生成单元5不生成填充词。在这种情况下,不输出填充词。这意味着不输出填充词适合于该对话场景。

[0073] 填充生成单元5将如上所述生成的填充词输出到语音输出单元6。例如,语音输出单元6输出由填充生成单元5在时间T+2处生成的填充词并输出由应答语句生成单元4在时间T+3处生成的应答语句。

[0074] 图7是示出由根据该实施例的语音交互设备执行的处理方法的流程的流程图。

[0075] 分类单元7基于在存储单元9中存储的表格信息将由应答语句生成单元4生成的第一应答语句分类为该表格信息中定义的话语模式之一(步骤S101)。

[0076] 特征值计算单元8基于在存储单元9中存储的表格信息和该话语模式来计算与由分类单元7分类的话语模式对应的特征值的类型的特征向量(步骤S102)。

[0077] 填充生成单元5基于由特征值计算单元8计算的特征向量和填充形式信息从在存储单元9中存储的填充形式信息中定义的多种填充类型中选择一种填充类型。然后,填充生成单元5通过根据所选择的填充类型选择一个填充词来生成填充词(步骤S103)。

[0078] 语音输出单元6输出由填充生成单元5生成的填充词(步骤S104)。

[0079] 如上所述,根据该实施例的语音交互设备1包括分类单元7,其将由应答语句生成单元4生成的应答语句分类为指示预定义的话语类型的预定话语模式之一。在用户于第一应答语句之后发出语音之后,语音输出单元6输出填充词并输出第二应答语句。在这种情况下,分类单元7将第一应答语句分类为话语模式之一。填充生成单元5基于由分类单元7分类的话语模式来生成填充词。

[0080] 通过根据与第一应答语句对应的话语模式生成填充词,可以生成适合于第一应答语句之后的用户语音的内容(例如含义)的最优填充词。通过输出该最优填充词可以进行更自然的对话。

[0081] 注意,本公开内容不限于上述实施例,并且可以在不脱离本公开内容的精神和范围的情况下做出各种修改。

[0082] 尽管在上述实施例中使用输出语音的语音输出单元6作为输出装置,但是输出装置不限于该示例。例如,可以使用输出(即显示)文本的显示单元作为输出装置。

[0083] 在本公开内容中,例如,可以通过使CPU执行计算机程序来实现图7所示的处理。

[0084] 可以利用任何类型的非暂态计算机可读介质来存储该程序并将其提供给计算机。非暂态计算机可读介质包括任何类型的有形存储介质。非暂态计算机可读介质的示例包括磁存储介质(诸如软盘、磁带、硬盘驱动器等)、光磁存储介质(例如磁光盘)、CD-ROM(光盘只读存储器)、CD-R(可记录光盘)、CD-R/W(可重写光盘)和半导体存储器(诸如掩模ROM、PROM(可编程ROM)、EPROM(可擦除PROM)、闪存ROM、RAM(随机存取存储器)等)。

[0085] 可以利用任何类型的暂态计算机可读介质将该程序提供给计算机。暂态计算机可读介质的示例包括电信号、光信号和电磁波。暂态计算机可读介质可以经由有线通信线路(例如电线和光纤)或无线通信线路将该程序提供给计算机。

[0086] 根据如此描述的公开内容,将明显的是,本公开内容的实施例可以以许多方式变化。这样的变化不被认为是脱离本公开内容的精神和范围,并且对于本领域技术人员而言明显的是,所有这样的修改旨在被包括在所附权利要求书的范围内。

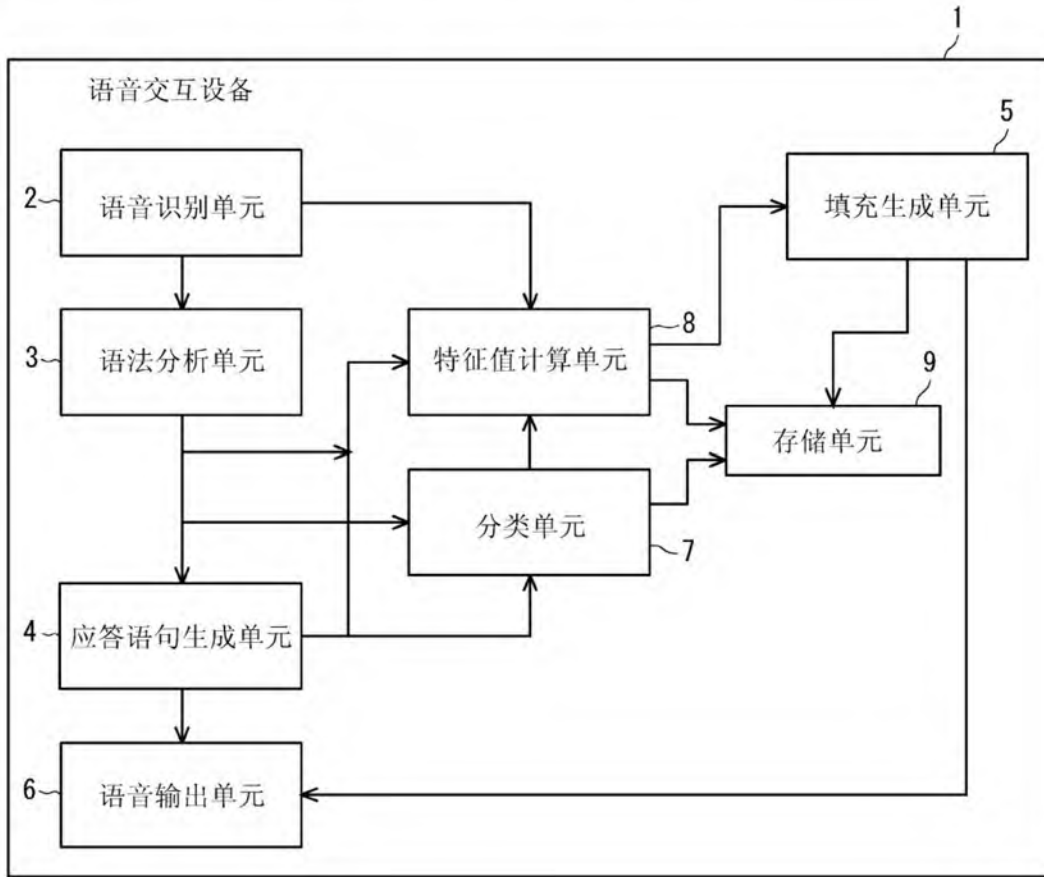


图1

话语模式	内容
(1) 是/否询问	用于询问事实的询问
(2) Wh询问	用于从组中进行选择的询问
(3) 回答	回答、同意、接受等
(4) 批露	自我批露、请求等
(5) 其他	问候、介绍等

图2

时间	设备话语	用户话语	话语模式
T-3	你好。		(5) 其他
T-2		你好。	(5) 其他
T-1	我喜欢烹饪。		(4) 批露
T	你有什么爱好吗?		(2) Wh询问
T+1		我玩围棋。	(3) 回答
T+2	哦。		(填充)
T+3	那听起来不错。		(3) 回答

图3

设备话语	后续的用户话语
(1) 是/否询问	(3) 回答
(2) Wh询问	(3) 回答
(3) 回答	(4) 批露
(4) 批露	(3) 回答
(5) 其他	(1) 是/否询问/ (2) Wh询问

图4

话语模式	特征值的类型
(1) 是/否询问	先前话语的韵律信息
(2) Wh询问	先前话语的韵律信息 先前话语的语言信息
(3) 回答	后续话语的语言信息
(4) 批露	先前话语的韵律信息
(5) 其他	先前话语的韵律信息 先前话语的语言信息

图5

填充物类型	填充词
固有类型	Let's see (让我们来看看)、Well (好吧)
应答词类型	Yah (呀)
指示词类型	I mean (我的意思是)、Well (好吧)
状语类型	Gee (哇)、Oh (哦)
意识类型	Ah (啊)、Uh (嗯)
其他类型	How should I say (我该怎么说)
无	

图6

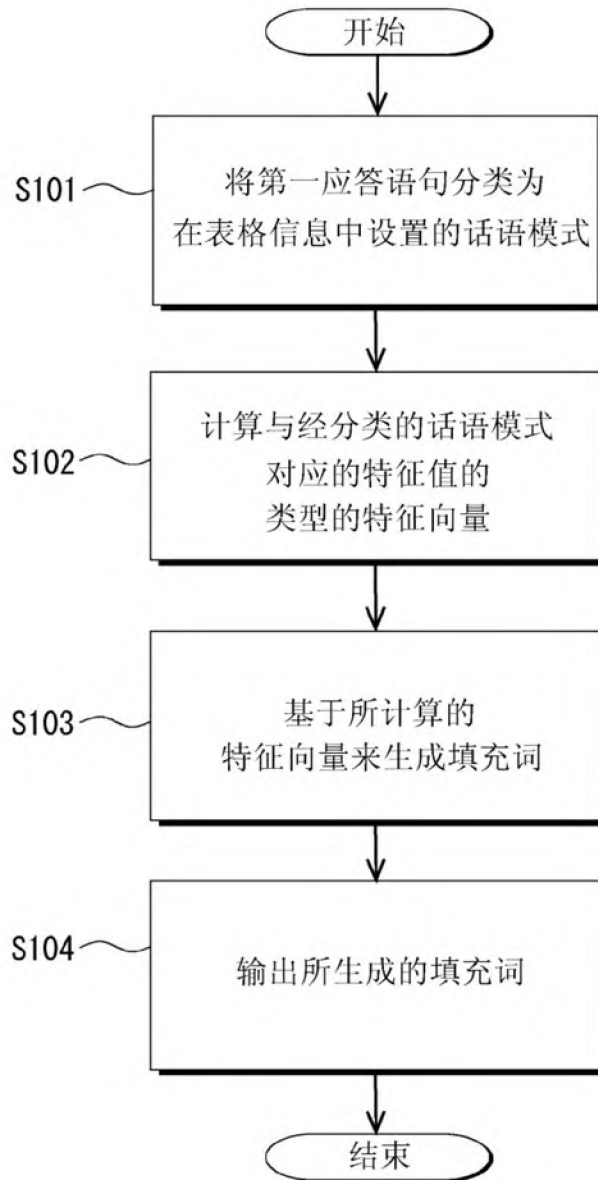


图7