



(12) 发明专利申请

(10) 申请公布号 CN 103377028 A

(43) 申请公布日 2013. 10. 30

(21) 申请号 201310142182. 1

(22) 申请日 2013. 04. 22

(30) 优先权数据

13/452, 557 2012. 04. 20 US

(71) 申请人 纽安斯通讯公司

地址 美国马萨诸塞州

(72) 发明人 大卫·安德鲁·毛洛 亨利·鲍维尔

(74) 专利代理机构 北京银龙知识产权代理有限公司 11243

代理人 许静 郭凤麟

(51) Int. Cl.

G06F 3/16 (2006. 01)

G06F 3/0481 (2013. 01)

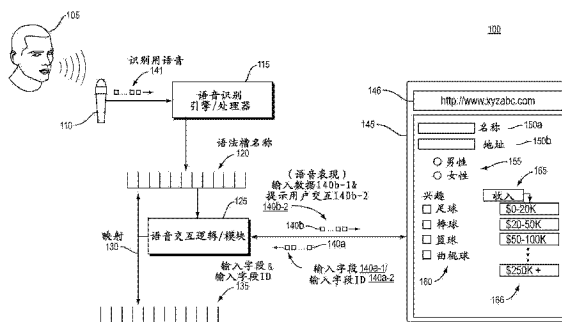
权利要求书5页 说明书7页 附图6页

(54) 发明名称

用于以语音启动人机界面的方法和系统

(57) 摘要

本发明涉及用于以语音启动人机界面的方法和系统。一般来说,人机界面被配置为接受来自用户的语音输入。但是,例如网络浏览器的界面必须被配置为能够接受来自用户的语音输入。例如移动浏览器的一些界面具有较少的配置适应性并且无法被配置为接受来自用户的语音输入。本发明实施例通过加载人机界面的内容并且将用于使语音能够与内容进行交互的逻辑添加到界面来语音启动人机界面。之后,实施例经由逻辑,为用户激活与内容的语音交互。从而,本发明实施例使语音能够与未被配置为适应于允许语音交互的界面进行交互,并且能够启动以无缝方式交互的语音交互。



1. 一种以语音启动人机界面的方法,所述方法包括:
加载所述人机界面的内容;
将用于启动与所述内容的语音交互的逻辑添加到所述界面;
将所述内容呈现给所述界面的用户;以及
经由所述逻辑,为所述用户激活与所述内容的语音交互。
2. 根据权利要求1所述的方法,其中,添加所述逻辑包括:
唯一地标识与所述内容相关的复数个输入字段和相应的输入字段标识(ID);
将所述输入字段标识映射为语法槽名称以产生语音到字段的映射,所述语法槽名称与预期由所述逻辑接收的语音的语音到文本合成相关联,并且输出与所述语法槽名称相应的所述语音的表现;以及
用于使所述语音的表现的流能够经由所述语音到字段的映射进入所述输入字段。
3. 根据权利要求1所述的方法,其中,添加所述逻辑包括:
解析与所述内容相关的代码,当由与所述界面相关的处理器执行所述代码时,所述代码使得所述界面将所述输入字段呈现给用户;以及
利用所述解析的结果,唯一地标识输入字段,其中,所述解析的结果包括所述输入字段和相应的输入字段标识。
4. 根据权利要求2所述的方法,其中,将所述输入字段标识映射为所述语法槽名称包括:从所述输入字段标识中确定关键词,并将确定的所述关键词与所述语法槽名称相关的关键词进行匹配。
5. 根据权利要求2所述的方法,其中,基于所述复数个输入字段的至少一个子集的用户选择,将所述输入字段标识映射为所述语法槽名称;以及其中,所述映射还包括:将所述相应的输入字段标识与响应于在所述用户选择之后接收的语音相应的语法槽名称进行关联。
6. 根据权利要求2所述的方法,其中,将所述输入字段标识映射为所述语法槽名称还包括:
提示用户对特定输入字段的声音输入;以及
将所述特定输入字段的相应的字段标识与在所述用户选择之后接收的语音相应的语法槽名称进行关联。
7. 根据权利要求1所述的方法,还包括:使用在语音中提供的值的表现以向所述用户提示所述语音,所述值的表现应用于所述复数个输入字段的特定输入字段。
8. 根据权利要求7所述的方法,其中,向所述用户提示语音包括执行以下操作的至少之一:加亮所述特定输入字段、改变经由所述界面呈现的文本显示的状态、和呈现对与特定输入字段相关的语音的音频指示。
9. 根据权利要求7所述的方法,其中,向所述用户提示语音包括:向所述用户提示与所述复数个输入字段的多个输入字段相关的语音。
10. 根据权利要求2所述的方法,还包括:使语音能够与所述输入字段进行交互,所述输入字段包括:显示在屏幕视图上的交互元素,所述交互元素包括以下输入字段结构的至少之一:单选按钮、文本字段、按钮和下拉菜单。
11. 一种以语音启动人机界面的系统,所述系统包括:

加载模块,用于加载所述人机界面的内容;

应用模块,用于将逻辑应用到所述内容,所述逻辑被用于使语音能够与所述内容进行交互;

呈现模块,用于将所述内容呈现给所述界面的用户;以及

语音交互模块,用于经由所述逻辑,为所述用户激活与所述内容的语音交互。

12. 根据权利要求 11 所述的系统,其中,所述逻辑包括:

标识模块,用于唯一地标识与所述内容相关的复数个输入字段和相应的输入字段标识(ID);

映射模块,用于将所述输入字段标识映射为语法槽名称以产生语音到字段的映射,所述语法槽名称与预期由所述逻辑接收的语音的语音到文本合成相关联,并且输出与所述语法槽名称相应的所述语音的表现;以及

转录模块,用于经由所述语音到字段的映射使所述语音的表现的流进入所述输入字段。

13. 根据权利要求 11 所述的系统,还包括:

解析模块,用于解析与所述内容相关的代码,当由与所述界面相关的处理器执行所述代码时,所述代码使得所述界面将所述输入字段呈现给用户;以及

标识模块,用于利用来自所述解析模块的结果以唯一地标识输入字段,其中,所述解析模块的结果包括所述输入字段和相应的输入字段标识。

14. 根据权利要求 12 所述的系统,其中,所述映射模块还用于:从所述输入字段标识中确定关键词并将确定的所述关键词与所述语法槽名称相关的关键词进行匹配。

15. 根据权利要求 12 所述的系统,其中,所述映射模块还用于:基于所述复数个输入字段的至少一个子集的用户选择,将所述输入字段标识映射为所述语法槽名称,以及将所述相应的输入字段标识与响应于在所述用户选择之后接收的语音的语法槽名称进行关联。

16. 根据权利要求 12 所述的系统,还包括:提示模块,用于:

提示用户对特定输入字段的声音输入;以及

将所述特定输入字段的相应的字段标识与在所述用户选择之后接收的语音相应的语法槽名称进行关联。

17. 根据权利要求 11 所述的系统,还包括:提示模块,用于使用在语音中提供的值的表现来向所述用户提示语音,所述值的表现应用于所述复数个输入字段的特定输入字段。

18. 根据权利要求 17 所述的系统,其中所述提示模块还用于:通过执行以下操作的至少之一来向所述用户提示语音:加亮所述特定输入字段、改变经由所述界面呈现的文本显示的状态、和呈现对与特定输入字段相关的语音的音频指示。

19. 根据权利要求 17 所述的系统,其中,所述提示模块还用于:向所述用户提示语音包括向所述用户提示与所述复数个输入字段的多个输入字段相关的语音。

20. 根据权利要求 12 所述的系统,还包括:语音交互模块,用于使语音能够与所述输入字段进行交互,所述输入字段包括显示在屏幕视图上的交互元素,所述交互元素包括以下输入字段结构的至少之一:单选按钮、文本字段、按钮和下拉菜单。

21. 一种计算机可读介质,在其中具有计算机可读程序代码,以用于呈现对人机界面的语音启动,所述计算机可读程序代码包括指令,当由处理器执行时,所述指令使得所述处理

器执行：

加载所述人机界面的内容；

将逻辑应用到所述内容，所述逻辑用于使语音能够与所述内容进行交互；

将所述内容呈现给所述界面的用户；以及

经由所述逻辑，为所述用户激活与所述内容的语音交互。

22. 一种以语音启动人机界面的方法，所述方法包括：

唯一地标识与人机界面相关的复数个输入字段和相应的输入字段标识(ID)；

将所述输入字段标识映射为语法槽名称以产生语音到字段的映射，所述语法槽名称与预期由语音到文本合成器接收的语音的语音到文本合成相关联，所述语音到文本合成器用于输出与所述语法槽名称相应的语音的表现；以及

经由所述语音到字段的映射使所述语音的表现的流能够进入所述输入字段。

23. 根据权利要求 22 所述的方法，其中，所述输入字段包括：显示在屏幕视图上的交互元素，所述交互元素包括以下输入字段结构的至少之一：单选按钮、文本字段、按钮和下拉菜单。

24. 根据权利要求 22 所述的方法，其中，所述人机界面包括以下至少之一：网页、交互式软件程序、电脑化输入表格、图形用户界面、视觉显示或音频接口。

25. 根据权利要求 22 所述的方法，其中，唯一地标识所述输入字段和相应的输入字段标识包括：对所述人机界面检查人类互动元素，所述人类互动元素被标识为与所述人类互动元素相关的输入字段和相应的标识。

26. 根据权利要求 22 所述的方法，其中，唯一地标识输入字段包括：解析代码，当执行所述代码时使得所述人机界面向用户呈现所述输入字段，其中，所述解析的结果包括所述输入字段和相应的输入字段标识。

27. 根据权利要求 22 所述的方法，其中，将所述输入字段标识映射为所述语法槽名称包括：从所述输入字段标识中确定关键词，并将确定的所述关键词与所述语法槽名称相关的关键词进行匹配。

28. 根据权利要求 22 所述的方法，其中，将所述输入字段标识映射为所述语法槽名称是基于所述复数个输入字段的至少一个子集的用户选择，并且还包括将所述相应的输入字段标识与响应于在所述用户选择之后接收的语音的语法槽名称进行关联。

29. 根据权利要求 22 所述的方法，其中，将所述输入字段标识映射到所述语法槽名称包括：

提示用户对特定输入字段的声音输入；以及

将所述特定输入字段的相应的字段标识与在所述用户选择之后接收的语音相应的语法槽名称进行关联。

30. 根据权利要求 22 所述的方法，还包括：使用在语音中提供的值的表现来向所述用户提示语音，所述值的表现应用于所述复数个输入字段的特定输入字段。

31. 根据权利要求 30 所述的方法，其中，向用户提示语音包括：执行以下操作的至少之一：加亮所述特定输入字段、改变文本显示的状态、和呈现音频指示。

32. 根据权利要求 30 所述的方法，其中，向所述用户提示语音包括：向所述用户提示与所述复数个输入字段的多个输入字段相关的语音。

33. 根据权利要求 22 所述的方法,还包括:

基于与所述输入字段相关的输入字段名称或输入字段值,自动地或动态地构建语音识别词汇表;

将所述语音识别词汇表的值与输入字段标识进行关联;以及

其中,将所述输入字段标识映射为语法槽名称包括:经由所述输入字段标识将所述语音识别词汇表的值与所述语法槽名称进行映射。

34. 根据权利要求 33 所述的方法,其中,构建所述语音识别词汇表包括:

对所述人机界面的检查预期由所述语音到文本合成器接收的语音;

将预期被接收的所述语音的表现存储在数据存储器中。

35. 根据权利要求 22 所述的方法,还包括:

接受语音;

将所述语音发送到所述语音到文本合成器以获取所述语音的表现;以及

经由所述语音到字段的映射将所述表现插入输入字段。

36. 一种用于以语音启动人机界面的系统,所述系统包括:

标识模块,用于唯一地标识与人机界面相关的复数个输入字段和相应的输入字段标识(ID);

映射模块,用于将所述输入字段标识映射为语法槽名称以产生语音到字段的映射,所述语法槽名称与预期由语音到文本合成器接收的语音的语音到文本合成相关联,所述语音到文本合成器被用于输出与所述语法槽名称相应的语音的表现;以及

转录模块,用于经由所述语音到字段的映射使所述语音的表现的流进入所述输入字段。

37. 根据权利要求 36 所述的系统,其中,输入字段包括:显示在屏幕视图上的交互元素,所述交互元素包括以下输入字段结构的至少之一:单选按钮、文本字段、按钮和下拉菜单。

38. 根据权利要求 36 所述的系统,其中,所述人机界面包括以下至少之一:网页、交互式软件程序、电脑化输入表格、图形用户界面、视觉显示或音频接口。

39. 根据权利要求 36 所述的系统,其中,用于唯一地标识输入字段和相应的输入字段标识的所述标识模块还用于对人机界面检查人类互动元素,所述人类互动元素被标识为与所述人类互动元素相关的输入字段和的相应的标识。

40. 根据权利要求 36 所述的系统,其中,所述标识模块还用于解析代码,当执行所述代码时,使得所述人机界面向用户呈现所述输入字段,其中,所述解析的结果包括所述输入字段和相应的输入字段标识。

41. 根据权利要求 36 所述的系统,其中,所述映射模块还用于:从所述输入字段标识中确定关键词,并将确定的所述关键词与所述语法槽名称相关的关键词进行匹配。

42. 根据权利要求 36 所述的系统,其中,所述映射模块用于:基于所述复数个输入字段的至少一个子集的用户选择,将所述输入字段标识映射为复数语法插槽名称,并且所述映射模块还用于将复数相应的输入字段标识与响应于在所述用户选择之后接收的语音的语法槽名称进行关联。

43. 根据权利要求 36 所述的系统,其中,所述映射模块还用于:

提示用户对特定输入字段的语音输入 ; 以及

将所述特定输入字段的相应的字段标识与在所述用户选择之后接收的语音相应的语法槽名称进行关联。

44. 根据权利要求 36 所述的系统, 还包括提示模块, 用于使用在语音中提供的值的表现来向所述用户提示语音, 所述值的表现应用于所述复数个输入字段的特定输入字段。

45. 根据权利要求 44 所述的系统, 其中, 所述提示模块通过执行以下操作的至少之一来向所述用户提示语音 : 加亮所述特定输入字段、改变文本显示的状态、和呈现音频指示。

46. 根据权利要求 44 所述的系统, 其中, 所述提示模块还用于向所述用户提示与所述复数个输入字段的多个输入字段相关的语音。

47. 根据权利要求 36 所述的系统, 还包括 :

词典模块, 用于基于与所述输入字段相关的输入字段名称或输入字段值, 自动地或动态地构建语音识别词汇表 ;

关联模块, 用于将所述语音识别词汇表的值与输入字段标识进行关联 ; 以及

其中, 所述映射模块还用于将所述输入字段标识映射为语法槽名称, 这包括 : 经由所述输入字段标识将所述语音识别词汇表的值与所述语法槽名称进行映射。

48. 根据权利要求 47 所述的系统, 其中, 所述词典模块还用于 :

对所述人机界面检查预期由所述语音到文本合成器接收的语音 ;

将预期被接收的所述语音的表现存储在数据存储器中。

49. 根据权利要求 36 所述的系统, 还包括 :

接收模块, 用于接受语音并将所述语音发送到所述语音到文本合成器以获得所述语音的表现 ; 以及

所述转录模块还用于经由所述输入字段标识到所述语法槽名称的映射, 将所述表现插入输入字段。

50. 一种计算机可读介质, 在其中具有计算机可读程序代码, 以用于呈现对人机界面的语音启动, 所述计算机可读程序代码包括指令, 当由处理器执行时, 所述指令使得所述处理器执行的 :

唯一标识与人机界面相关的复数个输入字段和相应的输入字段标识 (ID) ;

将所述输入字段标识映射为语法槽名称以产生语音到字段的映射, 所述语法槽名称与预期由语音到文本合成器接收的语音的语音到文本合成相关联, 所述语音到文本合成器被用于输出与所述语法槽名称相应的语音的表现 ; 以及

经由所述语音到字段的映射使所述语音的表现的流能够进入所述输入字段。

用于以语音启动人机界面的方法和系统

技术领域

[0001] 本发明涉及用于以语音启动人机界面的方法和系统。

背景技术

[0002] 一般来说,人机界面(human-to-machine interfaces)允许用户经由输入设备(例如,鼠标、键盘、触摸屏和控制面板 / 按钮)的触摸操作与界面进行交互。另外,一些人机界面被配置为接受来自用户的语音输入。例如,桌面网络浏览器是用于允许用户与互联网上的内容进行交互的人机界面的一个例子。为了使传统的桌面网络浏览器接受语音输入,浏览器必须在某些方面进行改进,例如,被配置为支持使浏览器能够接受来自用户的语音输入的插件(即,增强机制)。

发明内容

[0003] 本发明实施例涉及语音启动人机界面。本实施例加载人机界面的内容。例如,人机界面可以是网站的网页,本实施例可以将例如交互元素、文本元素和图形元素(例如,人机界面的内容)加载在显示网站网页的网络浏览器上。之后,本实施例将用于使语音能够与内容进行交互的逻辑添加到界面上。例如,逻辑可以被附加到界面。之后,本实施例将内容呈现给界面的用户并经由逻辑,为用户激活与内容的语音交互。

[0004] 另外,本实施例可以唯一标识与界面的内容相关的复数个输入字段和相应的输入字段标识(ID)。进一步地,本实施例可以将输入字段 ID 映射为语法槽名称,以产生语音到字段的映射。语法槽名称可以与预期由逻辑接收的语音的语音到文本合成相关。之后,本实施例可以输出与语法槽名称相应的语音的表现,以及经由语音到字段的映射使语音的表现能够进入输入字段。

[0005] 本实施例还可以解析与内容相关的代码,当由与界面相关的处理器执行代码时,该代码使得界面将输入字段呈现给用户。之后,本实施例可以利用解析的结果,唯一地标识输入字段,其中,解析的结果包括输入字段和相应的输入字段 ID。

[0006] 进一步地,本实施例通过从输入字段 ID 中确定关键词并将确定的关键词与语法槽名称相关的关键词进行匹配来将输入字段 ID 映射为语法槽名称。本实施例还可以基于复数个输入字段的至少一个子集的用户选择来将输入字段 ID 映射为语法槽名称。本实施例还可以通过响应于在用户选择之后接收到的语音,将相应的输入字段 ID 与语法槽名称进行关联来映射语法槽名称。

[0007] 本实施例还可以提示用户对特定输入字段的语音输入,以及之后将特定输入字段的相应的字段 ID 与在用户选择之后接收的语音相应的语法槽名称进行关联。本实施例还可以使用在语音中提供的应用于复数个输入字段的特定输入字段的值来向用户提示语音。

[0008] 此外,本实施例可以通过执行以下操作的至少之一来向用户提示语音:加亮特定输入字段、改变经由界面呈现的文本显示的状态、和呈现对与特定输入字段相关的语音的音频指示。进一步地,本实施例可以通过向用户提示与复数个输入字段的多个输入字段相

关的语音来向用户提示语音。

[0009] 本实施例还可以使语音能够与输入字段进行交互,其中输入字段包括显示在屏幕视图上的交互元素。输入元素包括以下输入字段结构中的至少之一:单选按钮、文本字段、虚拟按钮和下拉菜单。

[0010] 人机界面可以包括以下至少之一:网页、交互式软件程序、电脑化输入表格、图形用户界面、视觉显示或音频接口。

[0011] 进一步地,本实施例可以通过对人机界面检查人类互动元素来唯一标识输入字段和相应的输入字段 ID,其中,人类互动元素被标识为与人类互动元素相关的输入字段和的相应的 ID。

[0012] 本实施例可以基于与输入字段相关的输入字段名称或输入字段值,自动地或动态地构建语音识别词汇表,并且将语音识别词汇表的值与输入字段 ID 进行关联。进一步地,本实施例可以通过经由输入字段 ID,将语音识别词汇表的值与语法槽名称进行映射来将输入字段 ID 映射为语法槽名称。

[0013] 此外,本实施例通过对人机界面检查预期由语音到文本合成器接收的语音来构建语音识别词汇表。之后,本实施例将预期被接收的语音的表现存储在数据存储器中。

[0014] 本实施例还可以接受语音作为输入,将语音发送到语音到文本合成器以获取语音的表现,以及经由语音到字段的映射将表现插入输入字段。

[0015] 需要理解的是,为了辅助用户交易的通信,本发明实施例可以以方法、系统或计算机可读介质(其具有包含在其上的程度代码)的形式来实现。为了易于阅读,在说明的各部分使用了术语“系统”,其可以表示本发明的实施例和实现的一些或所有形式,例如,系统、方法、装置、设备、计算机可读介质、网络节点和网络。

附图说明

[0016] 从以下对于如附图所示的本发明实施例更具体的描述,前述内容将会显而易见,在附图中,贯穿不同视图,相似的参考字符表示相同的部分。附图不必按比例绘出,而是重点示出本发明的实施例。

[0017] 图 1 是示出根据本发明实施例,用于语音启动人机界面的语音交互模块的通信图;

[0018] 图 2 是示出根据本发明实施例,用于语音启动人机界面的语音交互模块的网络图;

[0019] 图 3A 是根据本发明实施例的语音交互模块的框图;

[0020] 图 3B 是根据本发明实施例的语音交互服务模块的框图;

[0021] 图 4 是用于语音启动人机界面的方法实例的流程图;以及

[0022] 图 5 是用于语音启动人机界面的方法的另一实例的流程图。

具体实施方式

[0023] 以下描述本发明的示例实施例。

[0024] 一般来说,人机界面区域可以被配置为接受来自用户的语音输入。但是,例如浏览器的传统界面,必须被配置为支持使浏览器能够接受来自用户的语音输入的Flash®或者插

件。未被配置为支持Flash®或者插件的界面(例如,浏览器),如移动浏览器,不能接受来自用户的语音输入。本发明实施例,通过将人机界面的内容加载到介面,并且将被配置为能与内容进行语音交互的逻辑添加到界面来语音启动人机界面。之后,实施例经由逻辑为用户激活语音与内容的交互。从而,本发明实施例能够实现与未被配置为适合允许语音交互的界面进行语音交互。

[0025] 需要注意的是,声音输入还可以包括语音输入。声音输入和语音输入都来自于话语(例如,来自用户的讲话)。话语(utterance)可以是单个单词、整条短语、句子或者甚至是几个句子。需要进一步注意的是,术语“声音输入”和“语音输入”在本文中是可以被互换使用的,不管在何种情况下,都意指最宽的意义。

[0026] 图1是示出根据本发明实施例,用于语音启动人机界面145的语音交互模块125的通信图。

[0027] 在图中,经由人机界面145(例如,网络浏览器),用户105可以与网站146进行交互。网站可以包括几个网页,其中网站的至少一个网页子集请求用户105与网页进行交互。用户105可以经由人机界面使用典型的外围设备与网页进行交互,所述外围设备例如是键盘、鼠标或触摸屏。如上所述,本发明实施例能够经由界面145实现与内容(例如,网站的网页)的语音交互。

[0028] 例如,本发明实施例利用语音交互逻辑/模块125来经由界面145实现与内容的语音交互。语音交互模块125经由语音识别引擎/处理器115接收输入的语音,该语音与经由麦克风110从用户105输入到语法槽名称120的语音相匹配。例如,用户105对着麦克风110讲话,麦克风可操作地耦合至转换模块或声音采样模块,其将语音转换为数据包141并将该数据包(例如,用于识别的语音)提供至语音识别处理器115。之后,处理器115对接收到的语音/话语与语法槽名称进行匹配,以将音频语音转换为等效文本。

[0029] 同时,语音交互模块125将语法槽名称120(被语音识别处理器115用来将语音转换为文本)映射为经由人机界面145呈现给用户的内容的交互字段的字段标识符。利用映射130,语音交互模块125能够获得转换后的语音(例如,语音的文本或任意其他表现)并且将转换后的语音应用于经由人机界面145呈现给用户的内容的字段。换句话说,经由语音到字段的映射130,语音交互模块125能够实现语音表现(例如文本)到输入字段的流。

[0030] 为了将语法槽名称120映射为呈现给用户的内容的字段,语音交互模块125首先唯一地标识与经由人机界面145呈现给用户105的内容相关的输入字段和相应的输入字段标识符(ID)135。例如,为了唯一地标识输入字段,语音交互模块125可以解析与内容相关的代码。当由与界面145相关的处理器执行时,该代码会使得界面向用户105呈现内容和相关的输入字段。因此,通过标识对输入字段唯一的元数据,语音交互模块125可以解析代码并标识输入字段。例如,大多数输入字段需要来自用户105的输入/交互,并且大多数输入字段可以包括向界面145发信号以使用户105能够与字段135交互的元数据/旗标。

[0031] 一旦语音交互模块125标识了输入字段和相应的输入字段ID135,语音交互模块125就能够将语法槽名称映射为输入字段。例如,语音交互模块125可以通过对从语法槽名称120提取的关键词和从输入字段和相应的字段ID135提取的关键词进行匹配来完成映射。

[0032] 另外,语音交互模块125可以通知用户105内容是由语音启动的并提示用户105

进行语音输入。例如,该通知可以经由在界面 145 上的视觉提示呈现给用户,或者是经由与界面 145 相关的音频输出设备的音频输出。经由包括对用户交互的提示 140b-2 的通信消息 140b,语音交互模块 125 可以使界面 145 能够向用户 105 呈现提示。另外,利用包括输入数据 140b-1 (例如,语音的表现)的通信消息 140-b,语音交互模块 125 使语音的表现(例如,文本)的流能够经由映射 130 被输入到输入字段 135 中。输入数据 140b-1 可以包括使界面 145 能够将输入数据 140b-1 输入至适当的输入字段 135 的标识(例如,映射信息)。

[0033] 输入字段 135 可以是经由界面 145 (例如,网络浏览器)显示在屏幕上的交互元素,该界面 145 包括以下输入字段结构中的至少之一:文本字段 150a-b、单选按钮 155、复选框 160 和包括选项 166 的下拉菜单 165。虽然描述了构建的几个输入字段,但是应该了解的是,根据本发明实施例可以实现本领域公知的或者将会公知的很多其他输入字段结构。

[0034] 图 2 是在本发明的实现的实现的环境中可以使用的通信网络 200 的高层框图。通信网络 200 包括经由人机界面(例如,网络浏览器(未显示))与多个节点互相连接的通信链接的集合,以形成节点的互联网,这些节点例如是通信单元 205a-b、接入点 270a-b、中间节点 280a-n、语音交互模块 225 和网络服务器 290。这些互联网节点通过根据预先定义的网络协议组交换数据包来进行通信,这些网络协议例如是传输控制协议 / 因特网互联协议(TCP/IP)和会话初始化协议(SIP)。本文中使用的网络协议是正式的规则组,其定义在通信网络中的节点之间如何交换数据。

[0035] 如下文更详细的描述,语音交互模块 225 可以被用于改善、定制或其他改进用户对网络服务器 290 的体验。在更详细的描述语音交互模块 225 之前,对通信网络 200 进行描述。需要理解的是,语音交互模块 225 可以被使用在其他网络拓扑结构或其他应用中,例如,单个处理器机器。

[0036] 中间节点 280a-n 是典型的传统中间节点,例如被配置为在通信网络 285(例如,互联网协议电话(VoIP)网络)中操作的路由器。接入点 270a-b 包括使通信单元 205a-b 能够将信息(例如,数据)经由中间节点 280a-n 通过接入点 270a、270b 传输至网络服务器 290 的逻辑。为此,接入点 270a-b 包括被配置为经由无线链路 211a、211b 发射和接收携带信息的信号(例如,无线电频率(RF)信号)的电路。可以与本发明一起使用的接入点 270a-b 的实例包括某些符合电气与电子工程师协会(IEEE) 802. 11 的接入点和支持通信流量(例如,数据流量)的传输的某些蜂窝电话无线系统。在本发明实施例中使用目前已知的或以后将开发的接入点的其他形式是可以预期的。

[0037] 通信单元 205a-b 可以是传统的通信单元,例如,膝上型计算机、台式计算机、无线发射 / 接收单元(WTRU) (例如,无线电话和个人数字助理(PDA))、互联网协议(IP)电话等,他们使可听见的和 / 或可视的通信能够被转换为经由无线链路 211a、211b 被传输至接入点 270a-b 的信号。接入点 270a-b 将通信单元 205a-b 和网络服务器 290 与网络 100 连接,并且使信号能够在通信单元 205a-b、网络服务器 290 和网络 285 之间传输。具体地,接入点 270a-b 将经由无线链路 211a、211b 接收的来自通信单元 205a-b 和网络服务器 290 的信号转换为在网络 200 上传输的数据包(未显示),并同样将接收到的来自网络的包转换为被传输至通信单元 205a-b 和人机界面 115 的信号。典型地,在通信单元 205a-b 和服务器 290 之间传输信息(例如,数据、声音或视频)。需要注意的是,本发明实施例可以适用于与能够经由通信网络进行通信的固定设备以及移动设备使用。这些固定设备可以包括接线到网络

的电话单元、个人计算机等。

[0038] 如所示的,语音交互模块 225 可以被放置在通信单元 205a-b 和网络服务器 290 之间的通信网络 200 的中间点上。可选地,语音交互模块 225 可以与通信单元 205a-b 或网络服务器 290 逻辑耦合或物理耦合。

[0039] 图 3A 是可以与本发明实施例关联使用的语音交互模块的框图。语音交互模块 315 包括经由存储器总线 322 耦合至处理器 325 的存储器 310,和经由输入 / 输出(I/O) 总线 334 耦合至处理器 325 的存储设备 335 和网络接口 330。需要注意的是,语音交互模块 315 可以包括其他设备,例如键盘、显示单元等。网络接口 330 连接语音交互模块 315 和网络 285 (参见图 2),并且使数据(例如,数据包)能够在网络服务器 290 (参见图 2)和其他节点之间传输,所述其他节点例如是,在网络 285 中的通信单元 205a-b 和网络服务器 290。为此,网络接口 330 可以包括传统的电路,该电路包含信号、电和机械特性、以及与网络 200 的物理介质和在该介质上运行的协议接口的互换电路。存储设备 335 是传统的存储设备(例如,磁盘),尤其是,能够存储人机界面的状态、状态的决策树结构、目标交易、状态的可视表现和目标交易的可视表现。

[0040] 存储器 310 是计算机可读介质的实例,可选地,实现为采用 RAM 设备的 RAM,例如 DRAM 设备和 / 或闪存设备。存储器 310 包括由处理器 325 使用的各种软件和数据结构,例如,实现本发明各方面的软件和数据结构。具体地,存储器 310 可以存储被配置作为操作系统 320 或经由语音交互服务模块 315 提供语音交互服务的软件。操作系统 320 可以被用于通过调用操作来在功能上组织语音交互服务模块 325,所述操作支持在语音交互模块 325 执行的软件处理和服务,如语音交互服务。如下文描述的,语音交互服务模块 315 可以包括非瞬态的计算机可执行指令,通过确定用户的目标交易、构建和呈现目标交易的可视表现、并让用户能够通过可视表现与人机界面进行交互来辅助用户交易的通信。

[0041] 存储设备 335 可以包括信息数据库 345,该信息数据库可以以这样的数据结构的形式来实现:其被配置为保持用以辅助用户交易通信的各种信息,例如,人机界面的状态、状态的决策树结构、目标交易、状态的可视表现和目标交易的可视表现。

[0042] 图 3B 是语音交互服务模块 315 的框图,该语音交互服务模块 315 包括标识模块 350、映射模块 355、转录模块 360、提示模块 365、辞典 370、关联模块 375、呈现 / 加载单元 380、语音交互模块 385、应用模块 390 和接收模块 395。

[0043] 呈现 / 加载单元 380 加载人机界面的内容(例如,通过网络浏览器显示的网页)并将内容呈现给用户。例如,可以通过通信设备 205a-b (参见图 2)的可视显示将内容显示给用户。应用模块 390 将逻辑应用到内容,并且启动语音交互模块 315 来启动经由逻辑的与内容的语音交互。进一步地,语音交互模块 385 为用户激活经由逻辑的与内容的语音交互。

[0044] 为了启动语音交互,标识 / 解析模块 350 唯一标识与内容相关的内容的多个输入字段和相应的输入字段 ID。标识 / 解析模块可以通过解析与用于标识输入字段的内容相关的代码来标识输入字段和相应的输入字段 ID。另外,标识模块 350 将与所标识的输入字段相关的信息发送至映射模块 355。

[0045] 映射模块 355 将所标识的输入字段和输入字段 ID 映射为用于将语音输入转换为文本的语法槽名称。一旦完成映射并启动语音交互,转录模块 360 使语音表现(例如,文本)的流能够通过语音到字段的映射被输入到输入字段。

[0046] 映射模块 355 还可以通过确定输入字段和相应的输入字段 ID 的关键词,将标识的输入字段和相应的输入字段 ID 映射为语法槽名称(其被用于将语音转换为文本),并将上述关键词和与语法槽名称相关的关键词进行匹配。另外,基于输入字段的用户选择,通过将用户选择的输入字段相应的输入字段 ID 和响应于在输入字段的用户选择之后接收到的语音的语法槽名称进行关联,映射模块 355 可以动态地将字段 ID 映射为语法槽名称。换句话说,映射模块 355 预测由用户在输入字段的用户选择之后接收到的语音与输入字段相关。

[0047] 进一步地,提示模块 365 可以提示用户对特定输入字段的语音输入,并且映射模块 355 从提示模块 365 接收与提示输入字段的输入字段 ID 相关的信息,其将相应的字段 ID 与在用户选择之后接收到的语音进行关联。以此,与在用户选择之后接收到的语音匹配的语法槽名称被映射为提示输入字段的输入字段 ID。

[0048] 通过执行至少一个以下示例操作,提示模块 365 可以向用户提示与特定输入字段相关的语音输入:加亮特定输入字段、改变经由界面呈现的文本显示的状态、或者呈现对与特定输入字段相关的语音的音频指示。也可以通过本领域中公知的或将会公知的很多其他方法来实现。进一步地,提示模块 365 可以向用户提示与所呈现内容的多个输入字段相关的语音。

[0049] 基于与输入字段相关的输入字段名称或输入字段值,词典模块 370 可以自动地或动态地建构语音识别词汇表(例如,语法槽名称)。例如,词典模块可以从输入字段中提取关键词并利用提取的关键词创建语法槽名称。之后,关联模块 375 可以将语音识别词汇表的值与输入字段 ID 进行关联。

[0050] 在执行任意上述操作之前,接收模块 395 接收语音并将语音传输至语音到文本合成器,以获得接收到的语音的表现(例如,文本)。之后,将语音的表现与经由映射模块 350 映射为输入字段的语法槽名称进行匹配。之后,通过输入字段 ID 到语法槽名称的映射,转录模块 360 将表现插入输入字段。

[0051] 图 4 是用于以语音启动人机界面的示例方法 400 的流程图。方法开始于 405。在 410,该方法将人机界面的内容加载到通信设备。在 415,该方法将逻辑增加到内容,以使语音与界面的内容能进行交互。此外,在 420,该方法通过可视显示将内容呈现给用户。在 425,该方法使用逻辑为用户激活语音与内容的交互。在 430,方法 400 结束。

[0052] 图 5 是用于以语音启动人机界面的另一示例方法 500 的流程图。方法开始于 505。在 510,该方法唯一标识与人机界面相关的多个输入字段和相应的输入字段标识(ID)。在 515,方法 500 将输入字段 ID 映射为语法槽名称以产生语音到字段的映射,在该映射中,语法槽名称与语音的语音到文本合成相关联。在 520,方法随后经由语音到字段的映射,使语音的表现的流能够进入输入字段。在 525,方法 500 结束。

[0053] 需要理解的是,本文中呈现的框图和流程图仅仅是实施例的示例表现。框图和流程图可以包括更多或更少的框或状态,具有更多或更少的联系等。可以采用本领域中已知的其他功能,并且被预期为在本文呈现的示例实施例的范围内。

[0054] 本发明的多个实施例或多个方面可以在硬件、固件或软件中实现。如果在软件中实现,该软件以能够执行本发明实施例的任意软件语言来实现。软件可以被存储在任意非瞬态的计算机可读介质上,例如, RAM、ROM、CD-ROM 等。软件包括能够由通用或专用处理器来加载和执行的指令,所述处理器能够支持本发明实施例。

[0055] 当结合本发明示例实施例来具体显示和描述本发明的同时,本领域技术人员将理解在不背离附加的权利要求所覆盖的本发明的范围的情况下,可以对其作出各种形式上和细节上的改变。

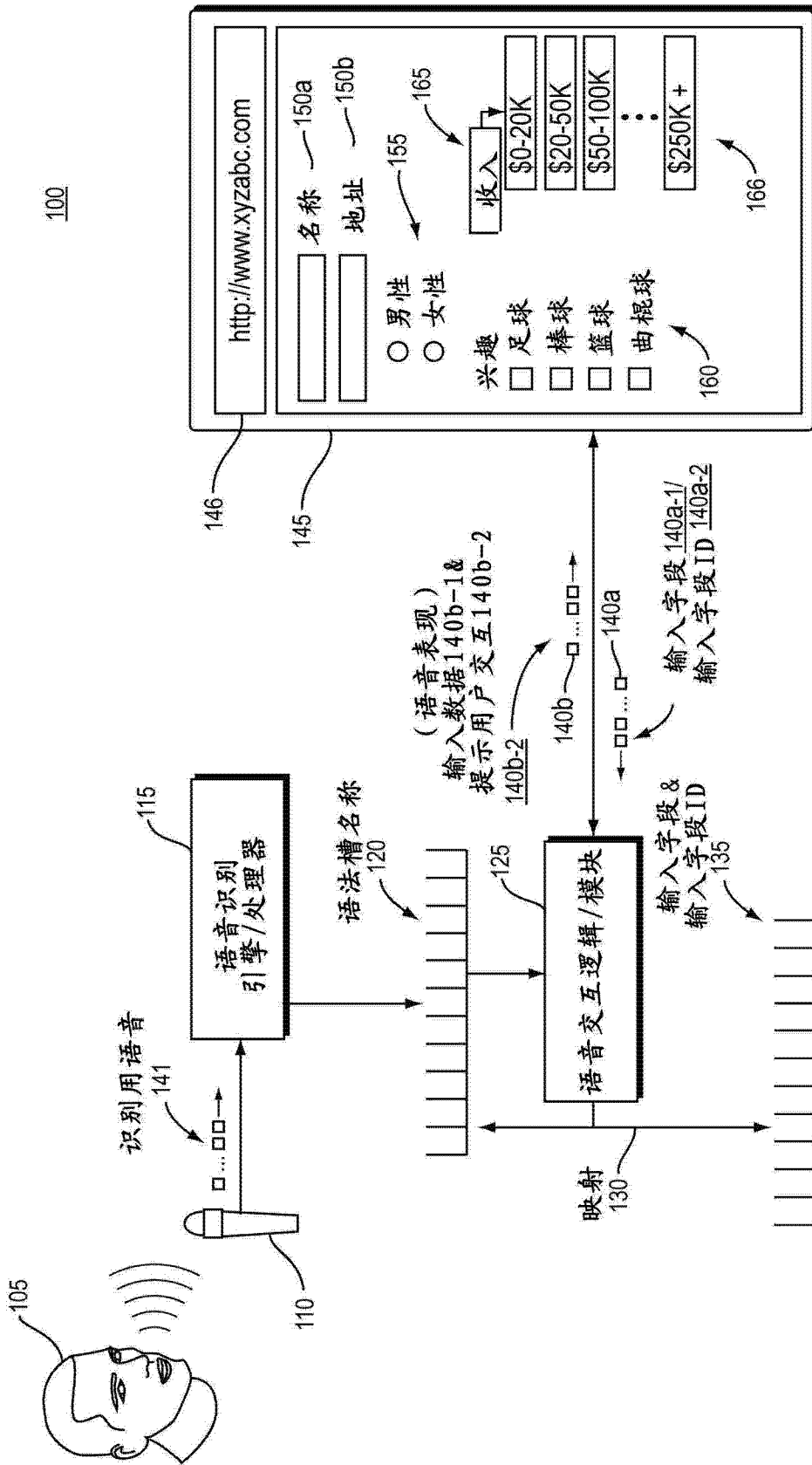


图 1

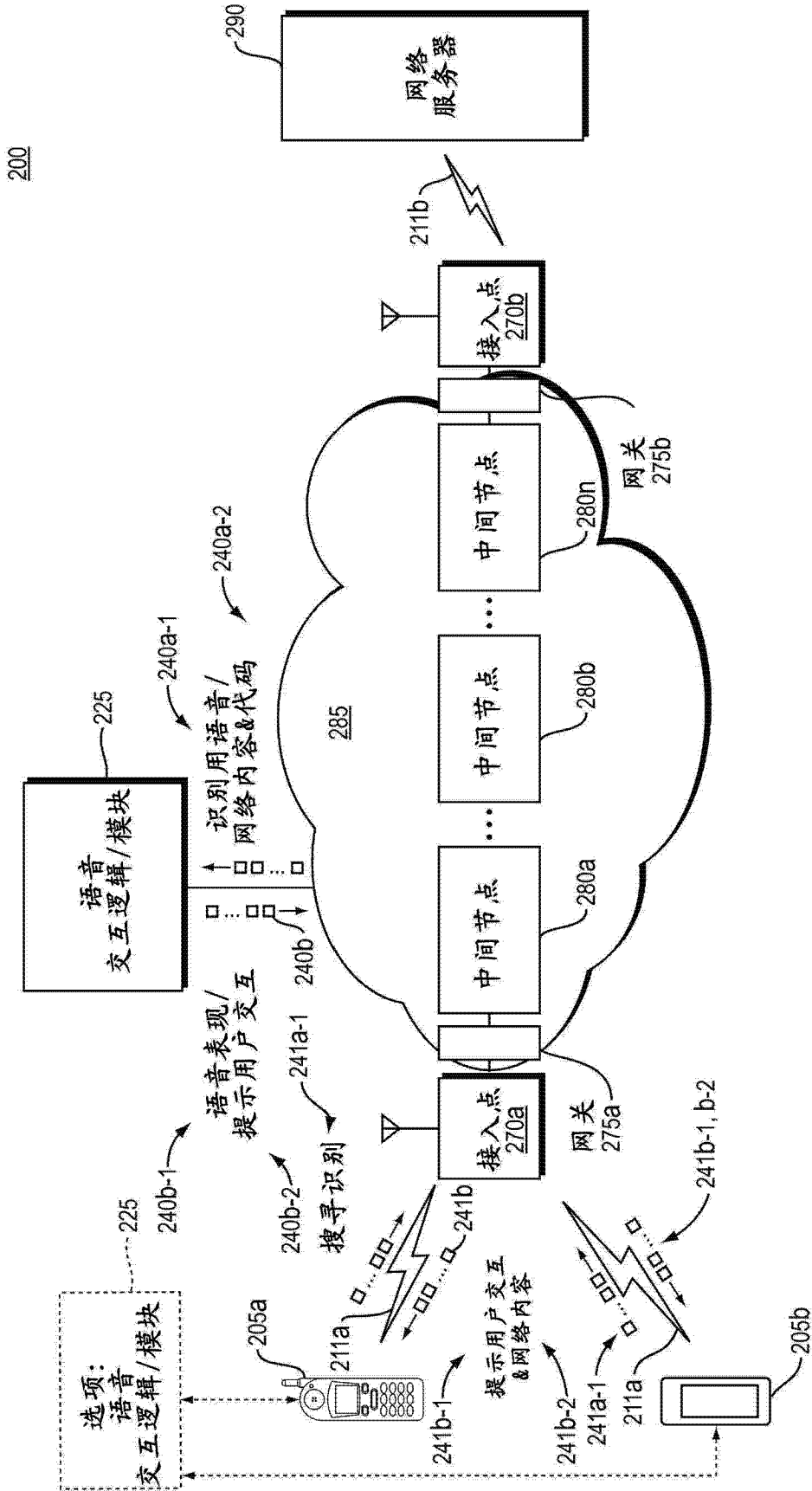


图 2

语音识别引擎 315/处理器 & 语音交互逻辑/模块 325

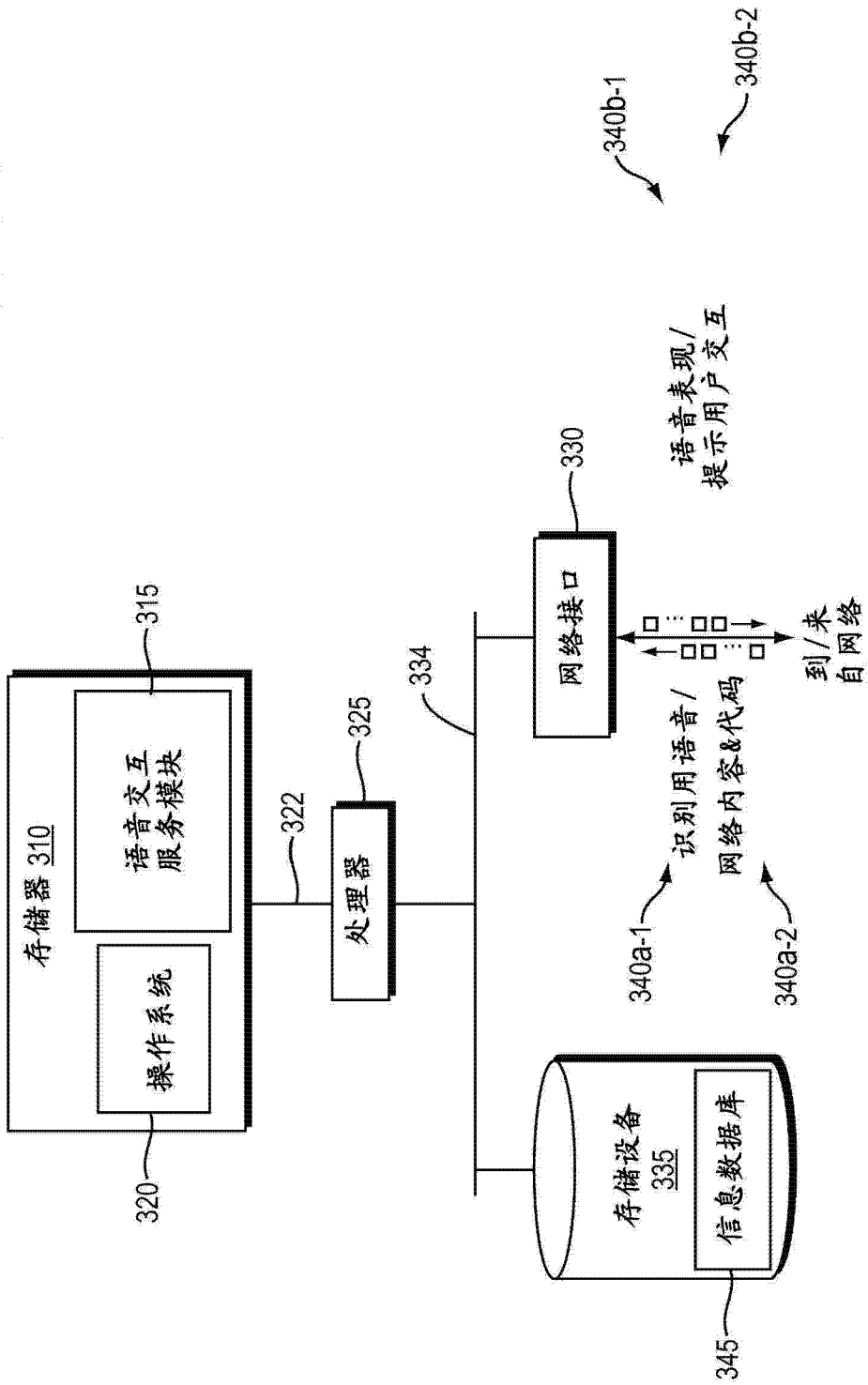


图 3A

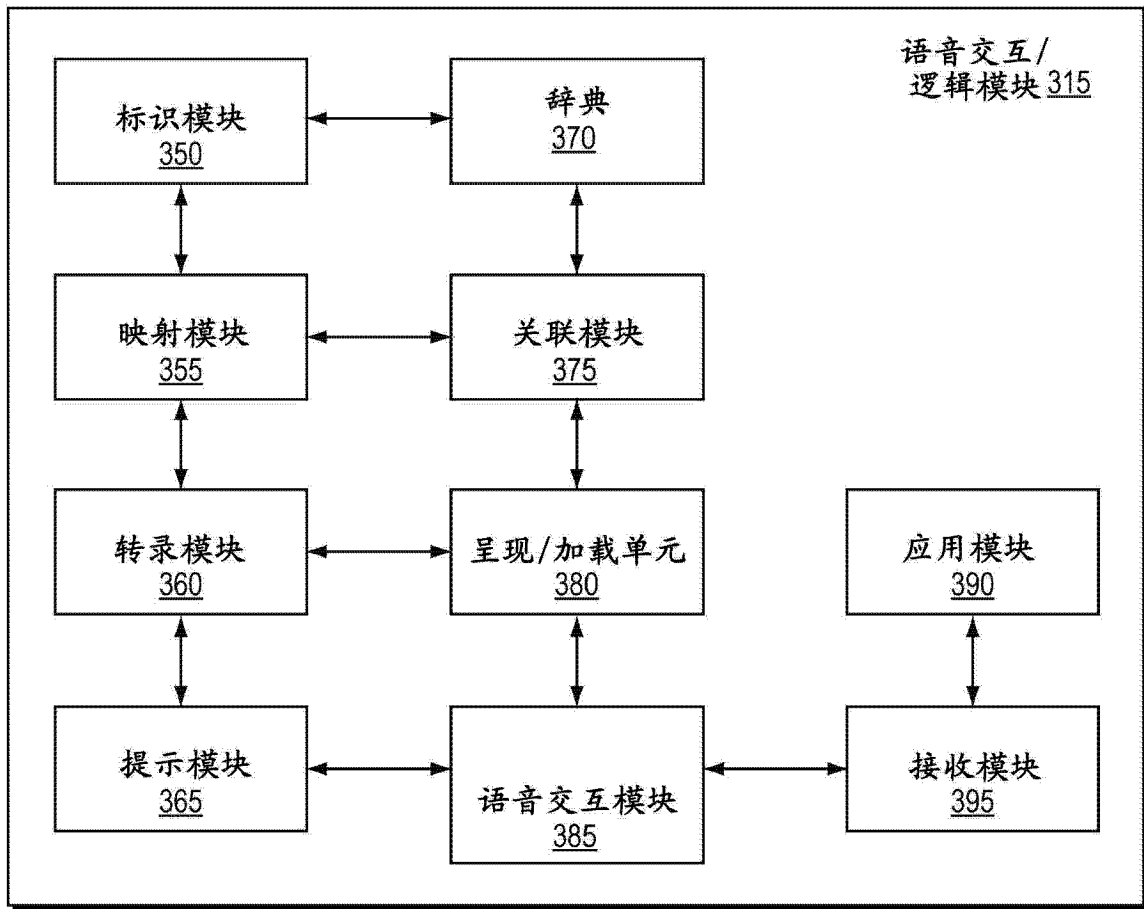


图 3B

400

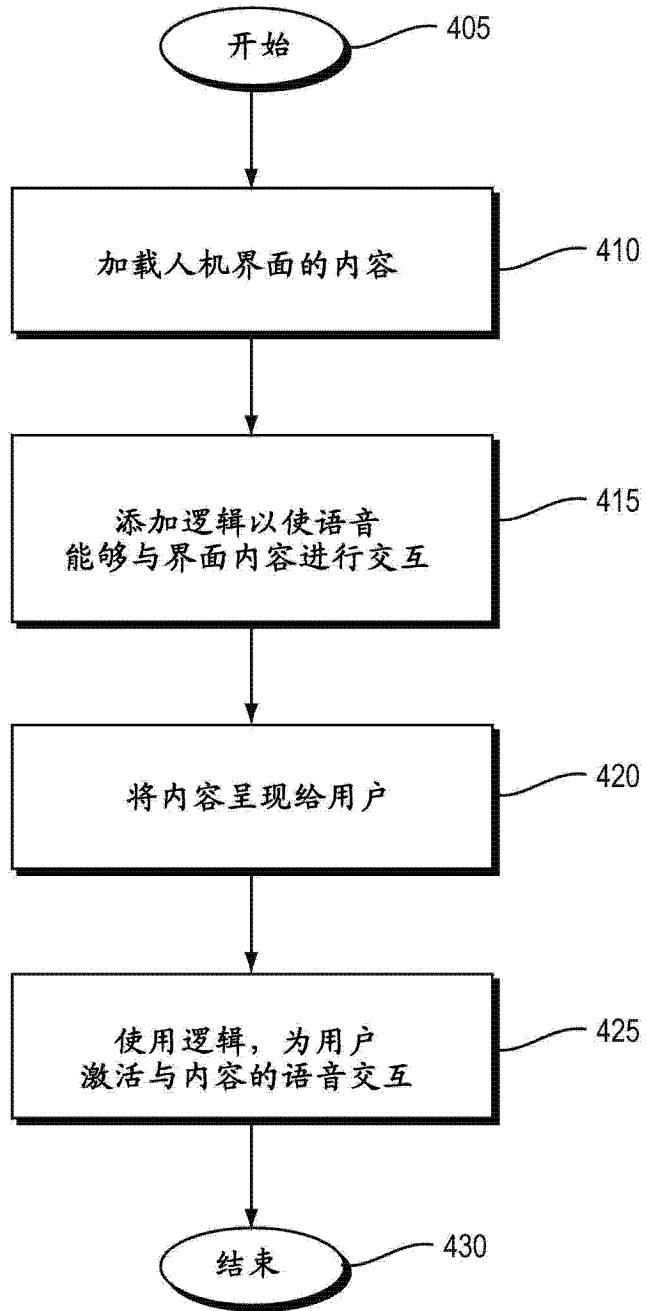


图 4

500

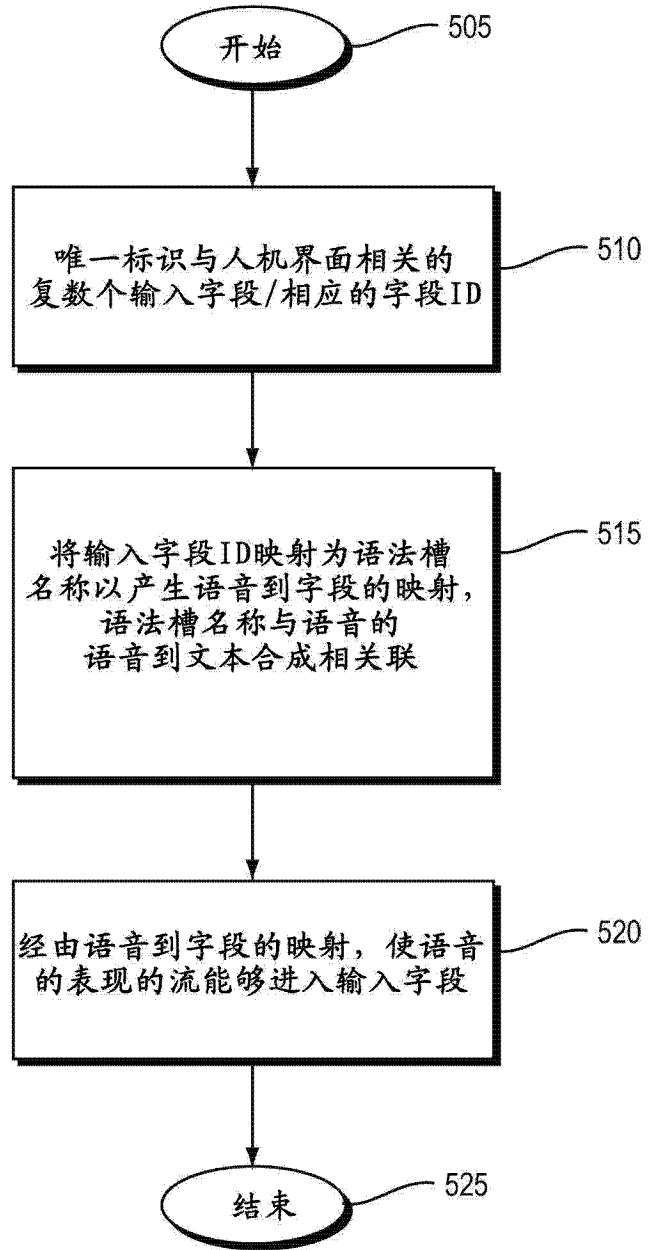


图 5