(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102439659 A
(43)申请公布日 2012.05.02

(21)申请号 201080017511.4
(22)申请日 2010.02.22
(30)优先权数据
12/389,678 2009.02.20 US
(85)PCT申请进入国家阶段日
2011.10.19
(86)PCT申请的申请数据
PCT/US2010/024895 2010.02.22
(87)PCT申请的公布数据
WO2010/096752 EN 2010.08.26
(71)申请人 声钰科技
地址 美国华盛顿
(72)发明人 L·贝尔德文 克里斯·魏德
(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291
代理人 黄志华
(51)Int.Cl.
G10L 15/00(2006.01)

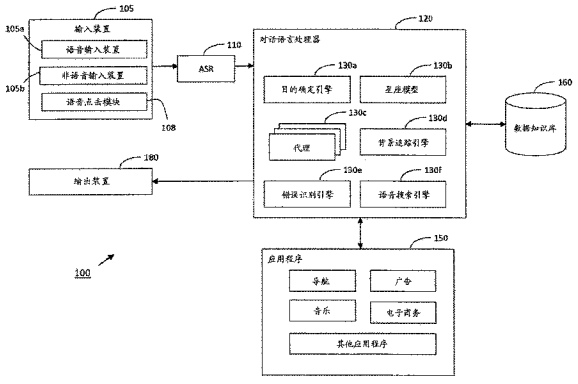
权利要求书 2 页 说明书 13 页 附图 3 页

(54)发明名称

在自然语言语音服务环境中处理多模式装置交互的系统和方法

(57)摘要

本发明可以提供在自然语言语音服务环境中处理多模式装置交互的系统和方法。具体地，可以在包括一个或多个电子装置的自然语言语音服务环境中接收一个或多个多模式装置交互。所述多模式装置交互可以包括与至少一个所述电子装置或与和所述电子装置有关的应用程序进行的非语音交互，且还可包括与所述非语音交互有关的自然语言语句。与所述非语音交互有关的背景和与所述自然语言语句有关的背景可以被提取并组合以确定所述多模式装置交互的目的，并且可以基于确定的所述多模式装置交互的目的来将请求路由到一个或多个所述电子装置。



CN 102439659 A

1. 一种用于在包括一个或多个电子装置的自然语言语音服务环境中处理一个或多个多模式装置交互的方法，所述方法包括：

检测至少一个多模式装置交互，其中所述多模式装置交互包括与所述电子装置中的至少一个或与和所述电子装置中的至少一个有关的应用程序进行的非语音交互，且其中所述多模式装置交互还包括与所述非语音交互有关的至少一个自然语言语句；

提取与所述多模式装置交互有关的背景信息，其中所提取的背景信息包括与所述非语音交互有关的背景，以及其中所提取的背景信息还包括与所述自然语言语句有关的背景；

组合与所述非语音交互有关的背景和与所述自然语言语句有关的背景；

基于与所述非语音交互和所述自然语言语句有关的组合的背景，确定所述多模式装置交互的目的；以及

基于所确定的所述多模式装置交互的目的，将至少一个请求路由到所述电子装置中的一个或多个。

2. 如权利要求 1 所述的方法，其中所述电子装置中的至少一个包括配置成接收所述自然语言语句的输入装置。

3. 如权利要求 2 所述的方法，所述方法还包括响应于检测到的所述非语音交互，用信号通知所述输入装置捕获自然语言语句。

4. 如权利要求 3 所述的方法，所述方法还包括：

在所述自然语言语音服务环境中建立一个或多个装置收听器，所述装置收听器被配置成检测所述非语音交互；以及

使与由所述装置收听器检测到的所述非语音交互有关的信息以及与由所述输入装置捕获的所述自然语言语句有关的信息对齐。

5. 如权利要求 1 所述的方法，所述方法还包括：

基于所确定的多模式装置交互的目的，产生至少一个交易提示；

接收与产生的所述交易提示有关的至少一个附加多模式装置交互；以及

响应于接收与所产生的交易提示有关的多模式装置交互，处理交易点进。

6. 如权利要求 5 所述的方法，其中产生的所述交易提示包括与所确定的所述多模式装置交互的目的有关的广告或推荐中的至少一个。

7. 如权利要求 1 所述的方法，其中所述非语音交互包括选择与所述电子装置中的一个或多个有关的部分、项目、数据或应用程序。

8. 如权利要求 1 所述的方法，其中所述非语音交互包括识别与所述电子装置中的一个或多个有关的注意点或关注点。

9. 如权利要求 1 所述的方法，其中所述非语音交互包括与所述电子装置中的一个或多个有关的一个或多个唯一且可区分的交互。

10. 一种用于在包括一个或多个电子装置的自然语言语音服务环境中处理一个或多个多模式装置交互的系统，其中所述系统包括一个或多个处理装置，所述一个或多个处理装置配置成：

检测至少一个多模式装置交互，其中所述多模式装置交互包括与所述电子装置中的至少一个或与和所述电子装置中的至少一个有关的应用程序进行的非语音交互，且其中所述多模式装置交互还包括与所述非语音交互有关的至少一个自然语言语句；

提取与所述多模式装置交互有关的背景信息，其中所提取的背景信息包括与所述非语音交互有关的背景，且其中所提取的背景信息还包括与所述自然语言语句有关的背景；

组合与所述非语音交互有关的背景和与所述自然语言语句有关的背景；

基于与所述非语音交互和所述自然语言语句有关的组合的背景，确定所述多模式装置交互的目的；以及

基于所确定的所述多模式装置交互的目的，将至少一个请求路由到所述电子装置中的一个或多个。

11. 如权利要求 10 所述的系统，其中所述电子装置中的至少一个包括配置成接收所述自然语言语句的输入装置。

12. 如权利要求 11 所述的系统，其中所述处理装置还配置成响应于正在被检测的所述非语音交互，用信号通知所述输入装置捕获自然语言语句。

13. 如权利要求 12 所述的系统，其中所述处理装置还配置成：

在所述自然语言语音服务环境中建立一个或多个装置收听器，所述装置收听器被配置成检测所述非语音交互；以及

使与由所述装置收听器检测到的所述非语音交互有关的信息以及与由所述输入装置捕获的自然语言语句有关的信息对齐。

14. 如权利要求 10 所述的系统，其中所述处理装置还配置成：

基于所确定的多模式装置交互的目的，产生至少一个交易提示；

接收与所产生的所述交易提示有关的至少一个附加多模式装置交互；以及

响应于接收与所产生的交易提示有关的多模式装置交互，处理交易点进。

15. 如权利要求 14 所述的系统，其中所产生的所述交易提示包括与所确定的所述多模式装置交互的目的有关的广告或推荐中的至少一个。

16. 如权利要求 10 所述的系统，其中所述非语音交互包括选择与所述电子装置中的一个或多个有关的部分、项目、数据或应用程序。

17. 如权利要求 10 所述的系统，其中所述非语音交互包括识别与所述电子装置中的一个或多个有关的注意点或关注点。

18. 如权利要求 10 所述的系统，其中所述非语音交互包括与所述电子装置中的一个或多个有关的一个或多个唯一且可区分的交互。

在自然语言语音服务环境中处理多模式装置交互的系统和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2009年2月20日提交的名称为“SYSTEMAND METHOD FOR PROCESSING MULTI-MODAL DEVICE INTERACTIONS IN A NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT”的第12/389,678号美国专利申请的权益，该美国专利申请的全部内容通过引用并入本文。

技术领域

[0003] 本发明涉及用于处理与一个或多个装置和/或应用程序进行的多模式交互的综合自然语言语音服务环境，其中所述多模式交互可以提供用于配合解释和另外处理伴随所述多模式交互的自然语言语句的额外背景。

背景技术

[0004] 近年来随着技术的进步，消费性电子装置已出现并几乎在很多人的日常生活中无处不在。为了满足移动电话、导航装置、嵌入式装置和其他这样的装置的功能性和移动性的增长引起的日益增长的需求，很多装置除了核心应用以外还提供大量特征和功能。然而，较大的功能性还带来了折衷，包括通常抑制用户完全利用他们的电子装置的所有性能的学习难度。例如，很多现有的电子装置包括复杂的人机界面，这些复杂的人机界面可能不是特别方便使用，这会抑制很多技术的大规模市场应用。而且，不方便的界面还经常导致难以找到或使用期望的特征（例如，因为菜单复杂或导航繁琐）。就这一点而言，很多用户趋向于不使用甚或不了解他们的装置的很多潜在性能。

[0005] 就这一点而论，电子装置的增加的功能往往趋向于浪费，市场研究表明，很多用户仅使用给定装置上可用的特征或应用的一部分。而且，在无线联网和宽带接入越来越普遍的社会中，消费者往往自然地希望他们的电子装置具有无缝移动性能。因此，由于消费者对更简单的与电子装置交互的机制的需求加强，因此妨碍快速且集约化交互的不方便的界面成为重要的议题。但是，在很大程度上仍未满足对以直观方式使用技术的机制的日益增长的需求。

[0006] 一种简化电子装置中的人机交互的方法包括使用语音识别软件，该语音识别软件有可能使用户利用原本不熟悉、不了解或难以使用的特征。例如，最近由Navteq公司（其提供比如自动导航和基于网页的应用的各种应用中使用的数据）进行的一项调查表明，语音识别在电子装置消费者最期望的特征中占首位。虽然如此，就用户而言，现有的语音用户界面在实际工作时仍需要大量学习。

[0007] 例如，很多现有的语音用户界面仅支持根据特定的命令与控制序列或语法制定的请求。而且，很多现有的语音用户界面因不准确的语音识别而导致用户沮丧或不满。类似地，通过强迫用户提供预先建立的命令或关键字来以系统可以理解的方式传递请求，现有的语音用户界面未能有效地使用户加入富有成效的、配合的对话中以解析请求并促进对话

朝着令人满意的目标进行（例如，当用户可能不确定具体需求、可用信息、装置性能等时）。就这一点而言，现有的语音用户界面往往有各种缺点，包括大大限制了使用户以配合方式和对话方式加入对话。

[0008] 此外，很多现有的语音用户界面达不到利用分布在不同领域、装置和应用程序中的信息以解析基于自然语言语音的输入。因此，现有的语音用户界面的缺陷在于局限于已经设计出的有限的一组应用程序或者局限于存在有它们的装置。尽管技术进步已使得用户通常利用若干装置来满足他们的各种需求，但现有的语音用户界面并不足以使用户摆脱装置的限制。例如，用户可能对与不同应用程序和装置关联的服务感兴趣，但现有的语音用户界面往往限制用户利用他们认为合适的应用程序和装置。而且，实际上，用户在任一给定时间通常仅能够携带有限数量的装置，而在各种情况下可能需要与用户目前使用的其他装置有关的内容或服务。

[0009] 因此，尽管用户往往具有不同的需求，其中在各种背景或环境中可能想要与不同的装置关联的内容或服务，但现有的语音技术往往达不到提供这样的综合环境：在该综合环境中，用户可以请求与几乎任何装置或网络关联的内容或服务。就这一点而言，现有的语音服务环境中对信息可用性和装置交互机制的限制往往妨碍用户以直观、自然且有效的方式体验技术。例如，当用户想要利用给定的电子装置执行给定的功能、但不一定了解如何着手执行该功能时，用户通常无法加入与该装置的多模式交互以仅发出自然语言的词来请求该功能。

[0010] 而且，利用不具有语音识别能力的电子装置，相对简单的功能通常可能执行起来繁琐。例如，为移动电话购买新的电话铃声往往是很简单的过程，但用户通常必须导航若干菜单并按下很多不同的按钮来完成该过程。就这一点而言，很明显，如果用户能够使用自然语言来开发隐藏或其他难以使用的功能，则与电子装置的交互会有效得多。现有的系统具有这些问题和其他问题。

发明内容

[0011] 根据本发明的一个方面，可以提供用于在自然语言语音服务环境中处理多模式装置交互的系统和方法。具体地，可以在包括一个或多个电子装置的自然语言语音服务环境中接收一个或多个多模式交互。所述多模式装置交互可以包括用户加入与所述电子装置中的一个或多个电子装置或与和所述装置有关的应用程序进行的非语音交互中，同时还提供与所述非语音交互有关的自然语言语句。例如，所述非语音装置交互可以包括用户选择特定的部分、项目、数据、注意点或关注点或者加入与所述电子装置或者与和所述电子装置关联的应用程序的一个或多个唯一且可区分的交互中。就这一点而言，可以从所述自然语言语句中提取出背景，且所述非语音装置交互可以提供用于所述自然语言语句的其他背景。接着可以使所述语句的背景和所述非语音装置交互的背景组合以确定所述多模式装置交互的目的，其中所述电子装置中的一个或多个电子装置可以基于所述多模式装置交互的目的处理请求。

[0012] 根据本发明的一个方面，所述电子装置中的至少一个电子装置可以包括配置成接收基于语音的输入的输入装置。在一个实现方式中，响应于检测与所述一个或多个电子装置或者应用程序的非语音交互，可以用信号通知所述基于语音的输入装置捕获所述自然语

言语句。而且，所述自然语言语音服务环境可以包括针对电子装置和相关应用程序建立的一个或多个收听器，其中所述收听器可以配置成检测与所述电子装置或应用程序的非语音交互。就这一点而言，与非语音交互有关的信息和与伴随的自然语言语句有关的信息可以被对齐以实现合作性处理所述语句和所述非语音装置交互。

[0013] 根据本发明的一个方面，可以基于所述多模式装置交互的目的产生至少一个交易提示。例如，可以接收附加多模式装置交互，其中所述附加多模式装置交互可以与针对第一多模式装置交互产生的交易提示有关。接着可以基于针对所述附加多模式装置交互确定的目的将至少一个请求路由到所述电子装置中的一个或多个电子装置，由此可以响应于接收到与所产生的交易提示有关的装置交互来处理交易点进。例如，所述交易提示可以包括基于最初的多模式装置交互的目的选择的广告或推荐，而附加多模式装置交互可以包括用户选择该广告或推荐。因此，选择该广告或者推荐可以被视为交易点进，这可以为具体的组织（例如，自然语言语音服务环境的提供商）产生收益。

[0014] 基于以下附图和详细描述，本发明的其他目的和优势将显而易见。

附图说明

[0015] 图 1 示出了根据本发明的各个方面的在自然语言语音服务环境中处理多模式装置交互的示例性系统的框图。

[0016] 图 2 示出了根据本发明的各个方面的用于在自然语言语音服务环境中使多模式装置同步的示例性方法的框图。

[0017] 图 3 示出了根据本发明的各个方面的用于在自然语言语音服务环境中处理多模式装置交互的示例性方法的流程图。

[0018] 图 4 示出了根据本发明的各个方面的用于在自然语言语音服务环境中处理多模式装置交互以产生一个或多个交易提示的示例性方法的流程图。

具体实施方式

[0019] 根据本发明的各个方面，图 1 示出了用于在自然语言语音服务环境中处理多模式装置交互的示例性系统 100 的框图。从本文要提供的进一步描述中将看出，图 1 中示出的系统 100 可以包括一个输入装置 105 或多个输入装置 105 的组合，输入装置 105 使用户能够以多模式方式与系统 100 交互。具体而言，系统 100 可以包括各种自然语言处理元件，所述的各种自然语言处理元件至少包括语音点击模块 108，其可以共同地处理用户与一个或多个输入装置 105 的多模式交互。例如，在一个实现方式中，输入装置 105 可以包括至少一个语音输入装置 105a（例如，话筒）和至少一个非语音输入装置 105b（例如鼠标、触摸屏显示器、滚轮选择器等）的任何适当组合。就这一点而言，输入装置 105 可以包括具有接收基于语音的输入和基于非语音的输入的机制的电子装置的任意适当组合（例如，连接到远程信息处理装置、个人导航装置、移动电话、VoIP 节点、个人计算机、媒体装置、嵌入式装置、服务器或其他电子装置中的一个或者多个的话筒）。就这一点而言，系统 100 可以使用户能够加入与一个或多个电子输入装置 105 或与和电子装置 105 有关的应用程序的多模式对话性交互中，其中系统 100 可以以适于路由任务或解析请求的自由形式和配合方式处理装置交互。

[0020] 如上所述，在一个实现方式中，该系统可以包括能够支持自由形式语句和 / 或其他形式的装置交互的各种自然语言处理元件，所述各种自然语言处理元件可以将用户从与制定命令、查询或其他请求的方式有关的约束中解放出来。就这一点而言，用户可以利用对语音输入装置 105a 讲话或与非语音输入装置 105b 交互中的任一方式来与输入装置 105 交互以请求系统 100 中可用的内容或服务。例如，用户可以通过将自然语言语句提供给语音输入装置 105a 来请求在系统 100 中可用的任何内容或服务。在一个实现方式中，接着可以利用 2008 年 7 月 8 日授权的名称为“Systems and Methods for Responding to Natural Language Speech Utterance”的第 7,398,209 号美国专利和 2003 年 6 月 15 日提交的名称为“Mobile Systems and Methods for Responding to Natural Language Speech Utterance”的美国专利申请 10/618,633 中描述的技术处理该语句，所述的美国专利和美国专利申请公开的全部内容通过引用并入本文。此外，用户可以与一个或多个非语音输入装置 105b 交互以提供与该语句和 / 或请求的内容或服务有关的进一步语境或其他信息。

[0021] 在一个实现方式中，系统 100 可以连接到包括额外多模式装置的各种其他系统，所述其他系统具有和图 1 中所示的自然语言处理性能相似的自然语言处理性能。因此，系统 100 可以为多装置环境提供一界面，在该界面中，用户可以请求通过该环境中的各个额外装置可得到的内容或服务。例如，在一个实现方式中，系统 100 可以包括星座模型 130b，该星座模型 130b 提供与通过该环境中的其他系统和装置可以得到的内容、服务、应用程序、目的确定性能和其他特征有关的知识。例如，在一个实现方式中，系统 100 可以与该环境中的装置、应用程序或其他系统交互以合作性地解析请求，如 2008 年 5 月 27 日提交的名称为“System and Method for an integrated, Multi-Modal, Multi-Device Natural Language Voice Services Environment”的共同待决的美国专利申请 12/127,343 中所述，该美国专利申请公开的全部内容通过引用并入本文。例如，该多装置环境可以在各个系统和装置中共享信息以提供解析请求的合作性环境，其中，所共享的信息可以涉及比如装置性能、背景、先前的交互、领域知识、短期性知识、长期性知识和认知模型等方面。

[0022] 如上所述，除了别的以外，图 1 中示出的系统 100 可以包括一个或多个电子输入装置 105，所述一个或多个电子输入装置 105 共同提供用于接收来自用户的一个或多个多模式装置交互的界面（或界面组合），其中装置交互至少包括用户口语语句。尽管图 1 中示出的实现方式包括分立的语音输入装置 105a 和非语音输入装置 105b，但是显然，在一个或多个实现方式中，语音输入装置 105a 和非语音输入装置 105b 可以是相同装置或不同装置的元件。例如，输入装置 105 可以包括连接到移动电话的话筒（即，语音输入装置 105a），且还可以包括连接到该移动电话的一个或多个按钮、可选显示器、滚轮选择器或其他元件（即，非语音输入装置 105b）。在另一示例中，输入装置 105 可以包括连接到远程信息处理装置的话筒组合（即，语音输入装置 105a）且还可以包括连接到媒体播放器的按钮、触摸屏显示器、轨迹滚轮或其他非语音输入装置 105b，该媒体播放器可通信地连接到该远程信息处理装置、然而与该远程信息处理装置分立。因此，输入装置 105 可以包括可通信地连接的电子装置的任意适当组合，该任意适当组合包括用于接收自然语言语句输入的至少一个输入装置和用于接收多模式非语音输入的至少一个输入装置。

[0023] 在一个实现方式中，可通信地连接到一个或多个输入装置 105 的语音点击模块 108 可以实现对语音输入装置 105a 和一个或多个非语音输入装置 105b 接收到的多模式装

置交互进行配合处理。例如，语音点击模块 108 可以为系统 100 提供能够用以鉴于通过非语音输入装置 105b 接收到的一个或多个非语音装置交互处理通过语音输入装置 105a 接收到的自然语言语句的信息。因此，语音点击模块 108 使用户能够与各种输入装置 105 以直观且自由形式的方式交互，由此，当试图发起行动、检索信息或请求系统 100 中可用的内容或服务时，用户可以将各种类型的信息提供给系统 100。

[0024] 语音输入装置 105a 可以包括具有用于接收自然语言语句或其他形式的口语输入的性能的任何适当的装置或装置的组合。例如，在一个实现方式中，语音输入装置 105a 可以包括定向话筒、话筒阵列或能够创建编码语音的其他装置。在一个实现方式中，语音输入装置 105a 可以配置成最大化编码语音的保真度。例如，语音输入装置 105a 可以配置成最大化沿着用户方向的增益、消除回音和零点噪声源、执行可变速率采样、滤去环境噪声或背景对话、或者使用其他技术来最大化编码语音的保真度。就这一点而言，语音输入装置 105a 可以以容忍噪声或可能干扰系统 100 准确解释自然语言语句的其他因素的方式创建编码语音。

[0025] 非语音输入装置 105b 可以包括具有支持非语音装置交互的性能的任何适当装置或装置的组合。例如，在一个实现方式中，非语音输入装置 105b 可以包括手写笔和触摸屏或写字板界面组合、黑莓®滚轮选择器、iPod®点击式转盘、鼠标、键盘、按钮或支持可区分的非语音装置交互的任何其他装置。因此，用户可以利用非语音输入装置 105b 进行数据选择或识别待与通过语音输入装置 105a 提供的相关自然语言语句连同处理的注意点（或关注点）。例如，用户可以将手写笔指向触摸屏显示器的特定部分、利用鼠标突出文本、点击按钮、与一应用程序交互、或加入用于选择数据或识别注意点的任何适当的装置交互中（即，语音激活或“语音点击”所选择的数据和 / 或识别的注意点）。

[0026] 而且，除了可用于进行数据选择、识别注意点、或激活与一个或多个语句有关的待解释的数据，用户还可以使用非语音输入装置 105b 来加入系统 100 中的具有意义的专用的装置交互中。例如，专用的装置交互（其可以被称为“点击”或者“语音点击”）可以包括持续给定时间段的点击、连续保持给定时间段的点击、按预定顺序进行的点击、或输入装置 105 和 / 或语音点击模块 108 可以识别、检测或以其他方式区分的任何其他交互或交互序列。

[0027] 在一个实现方式中，专用的装置交互可以与和系统 100 中可用的应用程序或服务有关的一个或多个动作、查询、命令、任务或其他请求关联。在一个实现方式中，专用的装置交互还可以包括与部署在多装置环境中的各个装置中的任一装置有关的一个或多个动作、查询、命令、任务或其他请求，如以上提及的 2008 年 5 月 27 日提交的名称为“System and Method for an Integrated, Multi-Modal, Multi-Device Natural Language Voice Services Environment”的共同待决的美国专利申请 12/127,343 中所述。例如，在显示于触摸屏显示器上的具体部分或项目上用手写笔点击的不同顺序可以被定义为用于在移动电话上发起电话呼叫、在导航装置上计算路径、为媒体播放器购买歌曲或其他类型的请求的专用装置交互或语音点击。

[0028] 因此，连接到输入装置 105 的语音点击模块 108 可以持续地监测用户与非语音输入装置 105b 的交互以检测至少一个非语音装置交互的发生，非语音装置交互在此可以被称为“语音点击”。因此，检测到的语音点击可以提供处理多模式装置交互的进一步背景，该

多模式装置交互可以包括至少一个语音点击和一个或多个自然语言语句，它们每一个都可以提供任务说明的背景。因此，语音点击通常可以用信号通知系统 100 当前语句或其他基于语音的输入要和与一个或者多个装置 105 的当前交互一起处理。例如，在一个实现方式中，当前装置交互可以包括与一个或多个装置 105 关联的用户选择、突出或识别具体的关注点、对象或者其他项目。就这一点而言，当前装置交互可以提供用于加强辨别、解释和理解伴随的语句的背景，而且，当前语句可以提供用以增强由所伴随的装置交互提供的背景的信息。

[0029] 在一个实现方式中，语音点击模块 108 可以基于非语音输入装置 105b 的具体特性确定待检测的各种语音点击交互（例如，语音点击交互可以包括非语音输入装置 105b 支持的可区分的交互）。例如，多触摸显示器通常包括触摸屏显示器装置，该触摸屏显示器装置被配置成支持与显示在该触摸屏显示器装置中的信息交互的各种可区分的手势（例如，用户可以利用特定的手势或者其他交互技术放大、缩小、旋转、或以其他方式控制显示在多触摸屏上的图形信息）。因此，在一个示例中，非语音输入装置 105b 可以包括多触摸显示器，在该情况下，语音点击模块 108 可以被配置成在用户加入由非语音多触摸显示器 105b 支持的一个或多个可区分的手势时检测语音点击的发生。

[0030] 在一个实现方式中，用户可以自定义或修改待由语音点击模块 108 检测的语音点击交互。具体地，由语音点击模块 108 检测的特定装置交互可以被删除或修改，或可以添加新的装置交互。就这一点而言，由语音点击模块 108 检测的语音点击装置交互可以包括非语音输入装置 105b 和 / 或语音点击模块 108 可以区分的任何适当的交互或交互的组合。

[0031] 当语音点击模块 108 检测到用户加入语音点击装置交互中时，语音点击模块 108 可以提取出与语音点击装置交互有关的背景信息以用于语音激活。具体地，语音点击模块 108 可以识别与用户选择的部分、项目、注意点、关注点或者其他数据有关的信息，或者以其他方式识别与用户加入的具体的装置交互或装置交互序列有关的信息。因此，语音点击模块 108 提取出所识别的与检测到的语音点击有关的信息，该信息可以用作与一个或者多个先前的、同时发生的或随后的自然语言语句有关的背景信息。

[0032] 因此，响应于语音点击模块 108 检测到语音点击（例如，选择图标、一段文本、地图显示器上的特定坐标或其他信息），语音点击模块 108 可以用信号通知系统 100 利用自然语言语句语音输入（其可通过语音输入装置 105a 接收）作为用于确定待执行的动作、查询、命令、任务或其他请求的进一步背景以服务于检测到的语音点击。就这一点而言，系统 100 中的各种自然语言处理元件可以使用语音点击和伴随的自然语言语句的组合背景来确定语音点击装置交互的目的并适当地将一个或多个动作、查询、命令、任务或其他请求路由到部署在多装置环境中的各个装置中的任何装置。

[0033] 例如，在一个实现方式中，多装置环境可以包括语音启用导航装置。因此，示例性语音点击装置交互可以包括用户用手写笔触碰与语音启用导航装置关联的触摸屏显示器 105b 上显示的特定交叉点，同时还将比如“这周围有什么餐馆？”的语句提供到话筒 105a。在该示例中，语音点击模块 108 可以提取出与语音点击的交叉点有关的信息，该信息可以用作处理伴随的语句的背景（即，与用户的当前位置或一些其他含义相对比，所选择的交叉点可以为解释“这周围”提供背景）。而且，如上所述，语音输入可以用作确定任务说明的附加背景。因此，可以利用系统 100 的各个自然语言处理元件进一步处理所述语句以用于

识别和对话解释，这将在下文更详细地描述。

[0034] 在一个实现方式中，自动语音识别器 (ASR)110 可以产生通过语音输入装置 105a 接收到的语句的一个或多个初步解释。例如，ASR 110 可以利用一个或多个动态适应识别语法识别语句的音节、单词、短语或其他声学特征。在一个实现方式中，动态识别语法可以用来利用基于一个或多个声学模型的语音听写识别一连串音位（例如，如 2005 年 8 月 5 日提交的名称为“Systems and Methods for Responding to Natural Language Speech Utterance”的共同待决的美国专利申请 11/197,504 中所述，该美国专利申请公开的全部内容通过引用并入本文）。

[0035] 在一个实现方式中，ASR 110 可以配置成执行多遍语音识别，其中第一语音识别引擎可以产生语句的初级转录（例如，利用大列表听写语法），且随后可以向一个或多个第二语音识别引擎请求一个或者多个次级转录（例如，利用具有未登录词的假词的虚拟听写语法）。在一个实现方式中，第一语音识别引擎可以基于初级转录的可信度请求次级转录。

[0036] ASR 110 中使用的识别语法可以包括用于识别语句的各种词汇表、词典、音节、单词、短语或其他信息。在一个实现方式中，识别语法中包括的信息可以被动态地优化以提高准确识别给定语句的可能性（例如，在不正确地解释一单词或短语之后，可以将该不正确解释从语法中删除以降低重复该不正确解释的可能性）。另外，各种形式的知识可用来在动态的基础上持续优化识别语法中包括的信息。例如，系统 100 可以具有如下知识，包括环境知识（例如，点对点关系、该环境中的各种装置的性能等）、历史知识（例如，频繁的请求、先前背景等）或与当前对话性谈话或交互有关的短期共享知识，以及其他类型的知识。

[0037] 在一个实现方式中，识别语法中的信息可以根据背景或特定应用领域而进一步优化。具体地，相似的语句可以根据该语句所涉及的背景而被不同地解释，所述背景包括导航、音乐、电影、天气、购物、新闻、语言、时间或地理相邻性或者其他背景或领域。例如，包括词“traffic”的语句可以根据该背景与导航（即，路况）、音乐（即，1960 年的摇滚乐队）、还是电影（即，由 Steven Soderbergh 导演的影片）有关而面临不同的解释。因此，ASR 110 可以使用各种技术来产生自然语言语句的初步解释，比如以上提及的共同待决的美国专利申请和 / 或 2006 年 8 月 31 日提交的名称为“Dynamic Speech Sharpening”的共同待决的美国专利申请 11/513,269 中所述，该美国专利申请 11/513,269 公开的全部内容通过引用并入本文。

[0038] 就这一点而言，ASR 110 可以将语音点击中包括的自然语言语句的一个或多个初步解释提供给对话语言处理器 120。对话语言处理器 120 可以包括各种自然语言处理元件，所述各种自然语言处理元件共同配置成模拟人与人对话或交互。例如，对话语言处理器 120 可以包括目的确定引擎 130a、星座模型 130b、一个或多个领域代理 130c、背景追踪引擎 130d、错误识别引擎 130e 以及语音搜索引擎 130f 等。而且，对话语言处理器 120 可以连接到一个或多个数据知识库 160 和与各种背景或领域有关联的一个或多个应用程序 150。

[0039] 因此，系统 100 可以使用与对话语言处理器 120 有关联的各种自然语言处理元件以使用户加入合作性对话中并基于用户发起语音点击的目的解析语音点击装置交互。更具体地，目的确定引擎 130a 可以基于系统 100 的性能以及多装置环境中的任何其他装置的性能建立给定多模式装置交互的含义。例如，参照用户语音点击具体交叉点以确定“这周围有什么餐馆”的以上示例，对话语言处理器 120 可以确定语音点击的对话目的（例如，“什

么”可以表示所述语句与请求数据检索的查询有关）。此外，对话语言处理器 120 可以调用背景追踪引擎 130d 以确定该语音点击的背景。例如，为了确定语音点击背景，背景追踪引擎 130d 可以将与识别的注意点有关的背景（即，选择的交叉点）和与所述语句有关的背景（即餐馆）组合起来。

[0040] 因此，语音点击的组合背景（其包括装置交互和伴随的语句）可以为路由特定查询提供充足信息。例如，该查询可以包括与餐馆和识别出的交叉点有关的各种参数或准则。接着对话语言处理器 120 可以选择可以向其路由该查询以进行处理的具体装置、应用程序或其他元件。例如，在一个实现方式中，对话语言处理器 120 可以评估星座模型 130b，星座模型 130b 包括多装置环境中的每一装置的性能的模型。在一个实现方式中，星座模型 130b 可以包括该环境中的每一装置可用的处理知识和存储资源以及每一装置的领域代理、背景、性能、内容、服务和其他信息的性质和范围等。

[0041] 就这一点而言，利用星座模型 130b 和 / 或其他信息，对话语言处理器 120 可以确定哪一装置或哪些装置的组合具有可以被调用以处理给定的语音点击装置交互的适当性能。例如，再次参照以上给出的示例，对话语言处理器 120 可以确定语音点击的背景涉及与导航装置的交互且因此路由该查询以利用导航应用程序 150 进行处理。查询结果可以随后被处理（例如，基于用户的知识比如对素食餐馆的偏好权衡结果）并通过输出装置 180 返给用户。

[0042] 根据本发明的各个方面，图 2 示出了用于在自然语言语音服务环境中同步不多模式装置的示例性方法的框图。如上所述，多模式装置交互（或“语音点击”）通常可以发生在以下时候 :用户加入与一个或多个多模式装置的一个或多个交互中同时提供和与多模式装置的交互有关的一个或多个自然语言语句。在一个实现方式中，和与多模式装置的交互有关的背景信息可以与和自然语言语句有关的背景信息组合以确定语音点击的目的（例如，以发起特定的动作、查询、命令、任务或其他请求）。

[0043] 在一个实现方式中，各种自然语言处理元件可以配置成持续收听或以其他方式监测多模式装置以确定语音点击何时发生。就这一点而言，图 2 中示出的方法可以用来调整或配置负责持续收听或监测多模式装置的元件。例如，在一个实现方式中，自然语言语音服务环境可以包括多个提供不同性能或服务的多模式装置，且用户可以加入一个或多个语音点击中以请求与各个装置中的任一装置有关的服务或任一给定装置交互的性能。

[0044] 为了能够持续收听多模式装置交互或语音点击，该环境中的多个装置中的每一装置可以配置成接收与语音点击有关的信息。因此，在一个实现方式中，操作 210 可以包括为该环境中的多个装置中的每一装置建立装置收听器。另外，可以响应于一个或多个新装置添加到该环境中而执行操作 210。操作 210 中建立的装置收听器可以包括配置成在一个或多个处理装置或其他硬件元件上执行的指令、固件或其他程序的任何适当组合。对于该环境中的每一装置，相关的装置收听器可以与该装置进行通信以确定与该装置有关的性能、特征、支持的领域或其他信息。在一个实现方式中，装置收听器可以配置成利用针对辅助计算机装置设计的通用即插即用协议与该装置进行通信。然而，显然可以使用与多模式装置进行通信的任何适当机制。

[0045] 当已经为该环境中的每一装置建立装置收听器时（或者当已经为添加到该环境中的新装置建立装置收听器时），可以在操作 220 中同步各个装置收听器。具体地，所述各

个装置中的每一装置可能具有不同的内部时钟或其他计时机制，其中操作 220 可以包括根据装置各自的内部时钟或计时机制来同步各个装置收听器。在一个实现方式中，同步装置收听器可以包括各个装置收听器中的每一个装置收听器公布与相关装置的内部时钟或计时有关的信息。

[0046] 因此，当随后发生针对一个或多个装置的一个或多个多模式交互或语音点击时，在操作 230 中，相关装置收听器可以检测与语音点击有关的信息。例如，在一个实现方式中，在操作 210 中建立的各个装置收听器可以与上文描述和图 1 中示出的语音点击模块有关。因此，操作 230 可以包括一个或多个装置收听器或语音点击模块检测用户与一个或者多个装置交互的发生（例如，选择与该装置有关的数据、识别与该装置有关的注意点或关注点、或者以其他方式加入与该装置的一个或多个交互或交互序列中）。而且，操作 240 于是可以包括捕获来自用户的与操作 230 中检测到的装置交互有关的语句。

[0047] 例如，浏览显示装置上呈现的网页的用户可能在该网页上看到产品名称并且想要得到关于购买该产品的更多信息。该用户可以从该网页中选择出包括该产品名称的文本（例如，使用鼠标或键盘突出文本），接着发起语音点击以询问“这可以在 Amazon.com 上买到吗？”。在该示例中，操作 230 可以包括与该显示装置关联的装置收听器检测对与该产品名称关联的文本的选择，而操作 240 可以包括捕获询问是否可在 Amazon.com 上买到该产品的语句。

[0048] 如上所述，接收来自用户的输入的每一装置都可以具有内部时钟或计时机制。因此，在操作 250 中，每一装置可以从本地角度来确定何时接收到该输入并通知语音点击模块接收到该输入。具体而言，除了与一个或多个其他装置的一个或多个其他交互之外，给定的语音点击可以至少还包括通过语音输入装置接收到的自然语言语句。该语句可以在装置交互之前、与其同时或之后接收到，由此操作 250 包括确定装置交互的时间以与所述相关语句关联。具体而言，利用参照操作 220 描述的被同步的装置收听器信号，操作 260 可以包括使装置交互的信号和该语句的信号对齐。在使装置交互信号和语句信号匹配时，可以产生包括对齐的语音和非语音成分的语音点击输入。接着，语音点击输入可以经受进一步的自然语言处理，如下文详细描述。

[0049] 根据本发明的各个方面，图 3 示出了用于在自然语言语音服务环境中处理多模式装置交互的示例性方法的流程图。如上所述，多模式装置交互（或“语音点击”）通常可以在以下时候发生：用户与一个或多个多模式装置交互，同时还提供与所述装置交互有关的一个或多个自然语言语句。就这一点而言，在一个实现方式中，图 3 中示出的方法可以在以下时候执行：一个或多个自然语言处理元件持续收听或以其他方式监测一个或多个多模式装置以确定一个或多个语音点击何时发生。

[0050] 在一个实现方式中，一个或多个装置交互可以被限定为发起语音点击。例如，任一给定的电子装置通常可以支持各种不同的交互，所述各种不同的交互可以引起执行给定的动作、命令、查询或其他请求。因此，在一个实现方式中，给定的电子装置可以唯一识别或者使用以产生可唯一识别信号的装置交互的任何适当组合可以被定义为语音点击，其中该语音点击可以提供这样的信号：该信号指示自然语言语句要和与相关装置交互关联的背景一起被处理。例如，具有四通导航键或五通导航键的装置可以支持特定不同的交互，其中以特定方式按导航键可以引起执行特定任务或其他动作，比如控制地图显示或计算路径。在另

一示例中，具有滚轮选择器的BlackBerry®装置可以支持多种交互，比如在具体的注意点或关注点上滚动光标、按压滚轮以选择特定数据或给定的应用程序、或者各种其他交互。各种其他装置交互可以用来指示自然语言语句何时与和所述装置交互有关的背景一起处理，但不限于此，其中在任一给定的实现方式中，具体的装置交互可以变化。例如，相关装置交互还可以包括下列中的一个或多个：利用定向仪器或绘图仪器在触摸感应显示屏上用动作示意（例如，绘制耳状波形曲线），比如长触摸或者双击的独特交互方法，和/或如果系统在以上所述的持续收听模式下工作，则预定义的背景命令字可以表示当前装置背景要和跟在该背景命令字之后的一部分基于语音的输入一起处理（例如，命令字为“可以(OK)”、“请”、“计算机”或其他适当的字，其中用户可以在地图上选择特定的点并说“请放大”，或者当显示电子邮件时说“可以读取”，等）。

[0051] 就这一点而言，操作310可以包括在自然语言语音服务环境中处理多模式装置交互以检测表示发起语音点击的一个或者多个装置交互的发生。具体地，操作310中检测到的装置交互可以包括引起电子装置产生唯一的、可识别的或其他可区分的信号的任何适当交互，该可区分的信号涉及用户选择数据、识别注意点或关注点、调用应用程序或任务、或者根据装置的特定性能以另一方式和该装置交互。

[0052] 除了装置响应于用户交互而产生的特定信号外，操作310中检测到的交互可以表示发起语音点击，由此先前的、同时发生的或随后的自然语言语音输入可提供用于解释操作310中检测到的装置交互的进一步背景。例如，自然语言处理系统通常可以配置成在具体的装置交互发生（例如，按按钮以打开话筒）时接受语音输入。因此，在图3中示出的方法中，表示进入的语音输入的装置交互还可以包括与电子装置的任何适当的交互或交互组合，包括与用户选择数据、识别注意点或关注点、调用应用程序或任务、或根据装置的特定性能以另一方式与该装置进行交互有关的交互。

[0053] 就这一点而言，当操作310中已检测到语音点击装置交互时，可以在操作320中产生语音点击信号以表示自然语言语音输入应当与操作320中检测到的交互关联。随后，操作330可以包括捕获要和操作310中检测到的交互关联的用户语句。在一个实现方式中，操作310中检测到的交互可以表示随后将提供语音输入，但是显然，在一个或多个实现方式中，操作330中捕获的语句可以在操作310中检测到的交互之前或与其同时提供（例如，用户可以提供比如“在iTunes®上查找此艺术家”的语句并随后在媒体播放器上语音点击该艺术家的名字，或者该用户可以在语音点击该艺术家的名字的同时提供此语句，或者该用户可以语音点击该艺术家的名字且接着提供该语句）。

[0054] 当已接收到与语音点击装置交互有关的信息和相关自然语言语句时，操作340可以包括提取并组合装置交互的背景信息和相关语句的背景信息。具体而言，从语音点击装置交互中提取出的背景信息可以包括与用户选择的部分、项目、注意点、关注点或数据、或者用户加入的具体的装置交互或装置交互序列有关的信息。提取出的装置交互的背景接着可以与针对在操作330中捕获的自然语言语句提取出的背景组合，其中在操作350中，组合的背景信息可以用来确定语音点击的目的。

[0055] 例如，在示例性语音点击装置交互中，用户可以选择性地将来自媒体播放器的音乐合集拷贝到备份存储装置。当在媒体播放器上浏览音乐时，用户可能遇到具体的歌曲并语音点击该歌曲，同时说“拷贝此整个唱片集”（例如，在突出该歌曲的同时通过长时间按该

媒体播放器上的具体的按键）。在该示例中，操作 310 可以包括检测长时间的按钮按下的交互，该长时间的按钮按下引起操作 320 中产生语音点击信号。接着，可以在操作 330 中捕获语句“拷贝此整个唱片集”，并且与该语音点击装置交互有关的背景信息和所述语句的背景信息可以在操作 340 中组合。具体地，装置交互的背景可以包括与所选择的歌曲有关的信息等（例如，该背景还可以包括与该歌曲有关的元数据中包括的信息，比如音乐文件的 ID3 标记）。此外，所述语句的背景可以包括识别拷贝操作和包括所选歌曲的唱片集的信息。

[0056] 就这一点而言，与和多模式装置的语音点击交互有关的背景信息可以和与自然语言语句有关的背景信息组合，由此操作 350 可以确定语音点击交互的目的。例如，参照以上示例，操作 350 中确定的目的可以包括将包括来自媒体播放器的突出显示的歌曲的唱片集拷贝到备份存储装置上的目的。因此，响应于操作 350 中确定语音点击的目的，在操作 360 中可以适当地路由一个或多个请求。在本文讨论的示例中，操作 360 可以包括将一个或多个请求路由到该媒体播放器，以识别与包括该突出显示的歌曲的唱片集有关的所有数据，以及将一个或多个请求路由到能够管理将所识别的数据从媒体播放器拷贝到备份存储装置的装置的任何适当组合（例如，与媒体播放器和存储装置连接的个人计算机）。

[0057] 根据本发明的各个方面，图 4 示出了用于在自然语言语音服务环境中处理多模式装置交互以产生交易提示或“点进”的示例性方法的流程图。具体而言，图 4 中示出的方法可以用于结合响应于检测到的一个或多个语音点击装置交互而执行的一个或多个动作来产生交易提示或点进。

[0058] 例如，操作 410 可以包括检测从用户接收到的一个或多个语音点击装置交互，其中所述语音点击装置交互可以包括与一个或多个相关自然语言语句结合的一个或多个装置交互的任一适当组合。接着可以在操作 420 中确定用户加入语音点击装置交互的目的，且随后的操作 430 可以包括基于确定的目的将一个或多个请求路由到一个或多个处理装置以解析语音点击交互。在一个实现方式中，可以与以上参照图 2 和图 3 所述的方式相似的方式执行操作 410、420 和 430，由此用于装置交互的信号可以与用于一个或多个自然语言语句的信号对齐，且可以从所述信号中提取出背景信息以确定语音点击装置交互的目的。

[0059] 除了基于用户目的来路由一个或多个请求外，图 4 中示出的方法还可包括产生一个或多个交易提示，所述交易提示可以导致一个或多个点进。例如，点进通常可以指用户点击或选择电子广告以访问与刊登广告的人有关的一个或多个服务的示例。在很多电子系统中，点进或点进率可以提供用于测量用户与电子广告交互的机制，该机制可以提供各种测量，刊登广告的人可以使用所述各种测量以确定要付给将广告提供给用户的组织的金额。

[0060] 就这一点而言，图 4 中示出的方法可以产生包括广告或推荐的交易提示，由此用户的与特定的装置交互组合的基于语音的输入可以提供用于产生交易提示的进一步的注意点。以此方式，提供给用户的广告或推荐可以与用户可能与其交互的特定信息更相关。而且，利用自然语言认知模型与用户的偏好有关的共享知识可以为对特定用户量身订制的目标交易提示提供进一步的背景，因此这可以更有可能导致能够产生向语音服务提供商支付费用的点进。

[0061] 因此，除了基于用户加入语音点击装置交互的目的而路由一个或多个请求之外，操作 440 还可包括基于确定的目的产生一个或多个交易提示。具体而言，基于装置交互和相关自然语言语句的组合背景，可以以“更接近”用户的方式处理交易提示，其中本地语音

背景和非语音背景可用作执行目标广告的任一适当系统中的状态数据。例如，参照以上给出的用户选择导航装置上显示的交叉点并同时说“找到这周围的餐馆”的示例，操作 440 中产生的交易提示可以包括交叉点附近的餐馆的一个或多个广告或推荐，所述广告和推荐可以基于用户的短期偏好和长期偏好（例如，偏好的餐馆类型、偏好的价格范围等）的知识而面向用户。

[0062] 接着，所述交易提示可以呈现给用户（例如，作为地图显示上的可选点）。接着可以在操作 450 中监测用户的随后的多模式装置交互以确定一个或多个其他多模式装置交互是否发生或何时发生。如果未发生其他交互，则可以确定用户未按照交易提示行动，则该过程可以结束。然而，如果发生附加的多模式交互，则可以在操作 480 中处理该多模式输入以确定输入的目的并相应地路由一个或多个请求。此外，操作 460 可以包括确定所述多模式输入是否与操作 440 中产生的交易提示有关。例如，用户可以通过提供语句、非语音装置交互或请求与所述交易提示有关的进一步动作或信息的语音点击装置交互来选择做广告的或推荐的餐馆中的一个餐馆。在这样的情况下，操作 470 还可以包括处理与操作 440 中产生的交易提示有关的交易点进，其中所述交易点进可以用来为具体组织（例如，语音服务的提供商或与所述交易提示或交易点进有关的另一组织）确定支付费用或产生收益。

[0063] 本发明的实现方式可以在硬件、固件、软件或它们的各种组合中进行。本发明还可作为存储在机器可读介质上的指令实现，所述指令可以由一个或多个处理器读取并执行。机器可读介质可以包括用来存储或发送机器（例如计算装置）可读形式的信息的各种机构。例如，机器可读存储介质可以包括只读存储器、随机存取存储器、磁盘存储媒介、光学存储媒介、闪存装置或其他存储媒介，机器可读传输媒介可以包括多种形式的传播信号，比如载波、红外线信号、数字信号或其他传输媒介。而且，可以在以上公开内容中就本发明的具体示例方面和实现方式以及执行某些动作这些方面来描述固件、软件、程序或指令。然而，显然，这些描述仅是为了方便，且这些动作实际上由执行所述固件、软件、程序或指令的计算装置、处理器、控制器或其他装置产生。

[0064] 尽管本文提供的描述主要集中在用于在自然语言语音服务环境中处理多模式装置交互的技术，但是显然，各种其他自然语言处理性能可以用于结合、附加于或代替与本文讨论的具体方面和实现方式关联描述的自然语言处理性能。例如，除了以上提及的共同待决的美国专利申请描述的技术外，本文描述的系统和方法还可利用 2005 年 8 月 5 日提交的名称为“Systems and Methods for Responding to Natural Language Speech Utterance”的共同待决的美国专利申请 11/197,504、2005 年 8 月 10 日提交的名称为“System and Method of Supporting Adaptive Misrecognition in Conversational Speech”的美国专利申请 11/200,164、2005 年 8 月 29 日提交的名称为“Mobile Systems and Methods of Supporting Natural Language Human-Machine Interactions”的美国专利申请 11/212,693、2006 年 10 月 16 日提交的名称为“System and Method for a Cooperative Conversational Voice User Interface”的美国专利申请 11/580,926、2007 年 2 月 6 日提交的名称为“System and Method for Selecting and Presenting Advertisements based on Natural Language Processing of Voice-Based Input”的美国专利申请 11/671,526 以及 2007 年 12 月 11 日提交的名称为“System and Method for Providing a Natural Language Voice User Interface in an Integrated Voice Navigation Services

Environment”的美国专利申请 11/954,064 中描述的自然语言处理性能，所述美国专利申请公开的全部内容通过引用并入本文。

[0065] 因此，本发明的方面和实现方式可以在本文中描述为包括具体的特征、结构或性质，但将明显的是，每一方面或实现方式可以或者可以不一定包括具体的特征、结构或性质。此外，当具体的特征、结构或者性质已结合一给定的方面或实现方式予以描述时，应当理解，无论是否明确描述，这样的特征、结构或性质也可以包括在其他的方面或实现方式中。因此，可以对以上描述进行各种改变或修改，而不脱离本发明的精神或范围，因此，本说明书和附图应当仅看作示例性的，本发明的范围仅由所附权利要求确定。

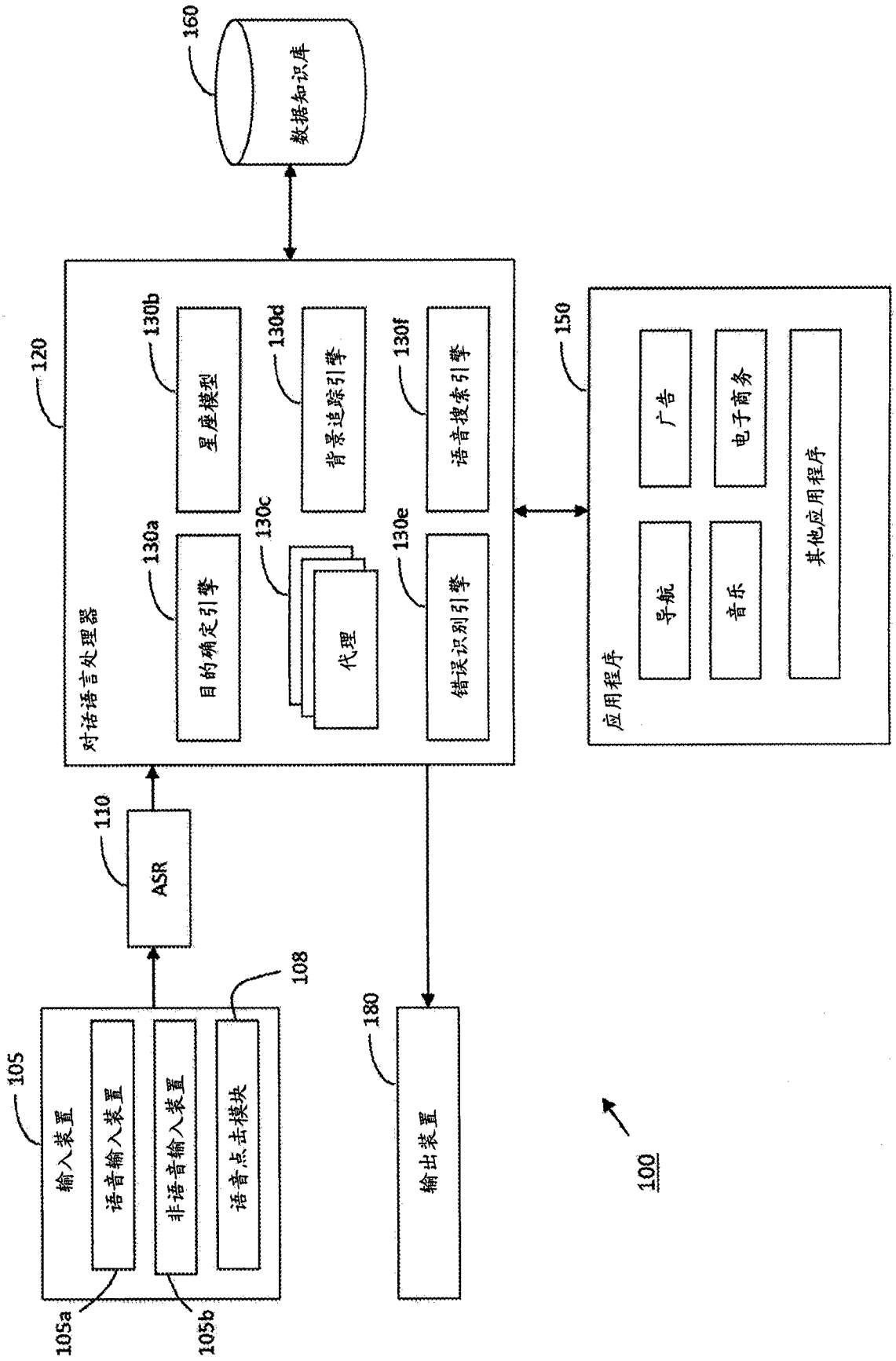

160
数据知识库
120
对话语言处理器
130b
星座模型
130d
背景追踪引擎
130f
语音搜索引擎
130a
目的确定引擎
130c
代理
130e
错误识别引擎
150
应用程序
广告
电子商务
导航
音乐
其他应用程序
110
ASR
105
输入装置
语音输入装置
非语音输入装置
语音点击模块
108
105a
105b
180
输出装置
100

图 1

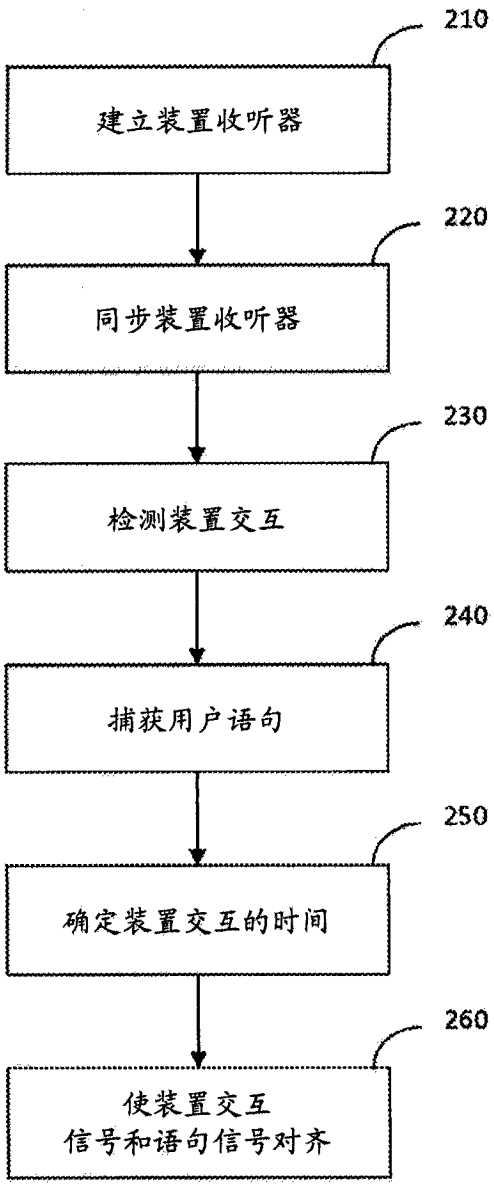

210
建立装置收听器
220
同步装置收听器
230
检测装置交互
240
捕获用户语句
250
确定装置交互的时间
260
使装置交互
信号和语句信号对齐

图 2

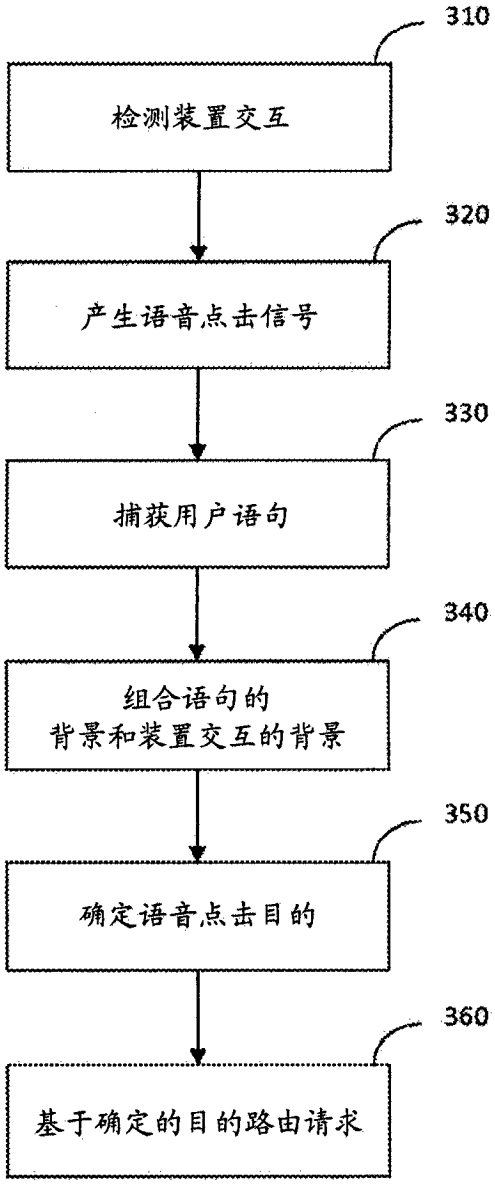

310
检测装置交互
320
产生语音点击信号
330
捕获用户语句
340
组合语句的
背景和装置交互的背景
350
确定语音点击目的
360
基于确定的目的路由请求

图 3

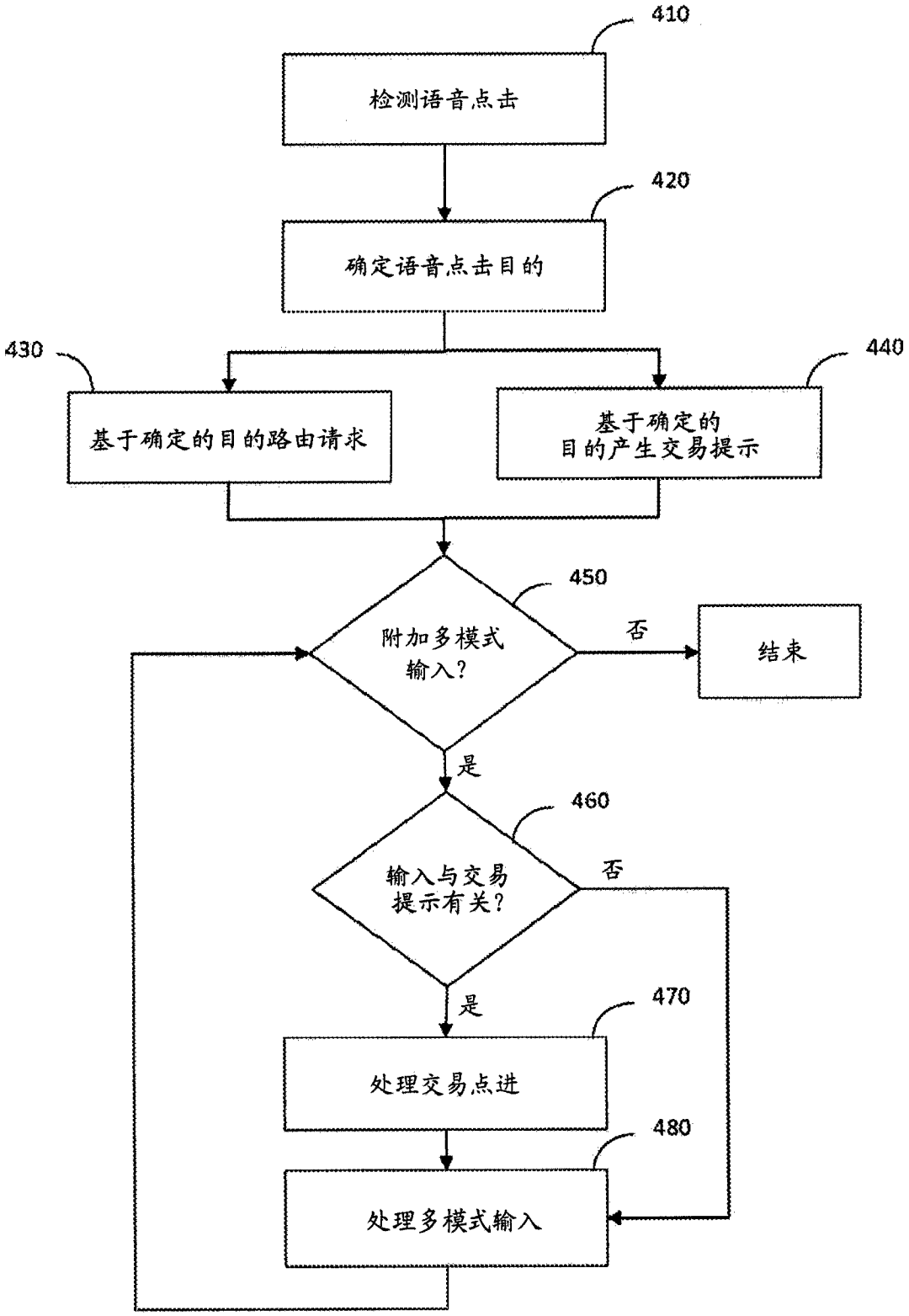

410
检测语音点击
420
确定语音点击目的
430
基于确定的目的路由请求
440
基于确定的
目的产生交易提示
450
附加多模式
输入?
否
结束
是
460
输入与交易
提示有关?
否
是
470
处理交易点进
480
处理多模式输入

图 4