

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200680036200.6

[43] 公开日 2008 年 10 月 29 日

[51] Int. Cl.
G10L 21/00 (2006.01)
G06F 17/30 (2006.01)
G06F 3/16 (2006.01)

[11] 公开号 CN 101297355A

[22] 申请日 2006.8.4

[21] 申请号 200680036200.6

[30] 优先权

[32] 2005. 8. 5 [33] US [31] 11/197,504

[86] 国际申请 PCT/US2006/030458 2006.8.4

[87] 国际公布 WO2007/019318 英 2007.2.15

[85] 进入国家阶段日期 2008.3.28

[71] 申请人 沃伊斯博克斯科技公司

地址 美国华盛顿

[72] 发明人 P·迪克里斯托 柯 敏

R·A·肯尼维克

L·E·阿姆斯特隆

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 李向英

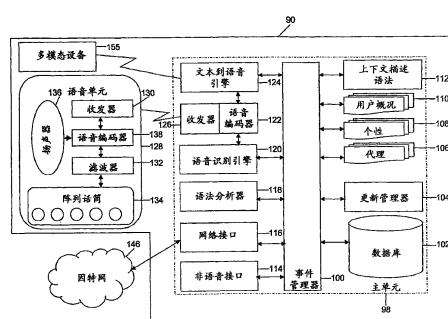
权利要求书 5 页 说明书 39 页 附图 8 页

[54] 发明名称

响应自然语言语音口头表达的系统和方法

[57] 摘要

公开了响应自然语言语音口头表达的系统和方法。提供的系统和方法用于接收自然语言问题和/或命令的语音和非语音通信，将所述语音和非语音通信转录为文本消息，并且执行这些问题和/或命令。本发明应用了上下文、先验信息、领域知识和用户专用的概况数据，以便实现一位或多用户跨越多个领域呈现问题或命令的自然环境。本系统和方法为每位用户创建、存储和使用了广泛的个人概况信息，因而改进了确定语音和非语音通信上下文的可靠性并为具体问题或命令呈现了所期待的结果。



1. 一种用于处理语音和非语音通信的系统，包括：
 终端设备，接收所述语音和所述非语音通信；
 转录模块，转录所述语音和所述非语音通信以创建基于语音的文本消息和基于非语音的文本消息；
 合并模块，将所述基于语音的文本消息和所述基于非语音的文本消息合并以产生查询；
 搜索模块，搜索所述查询中的文本组合；
 比较模块，将所述文本组合与上下文描述语法中的条目进行比较；
 多个领域代理，它们与所述上下文描述语法相关联；
 评分模块，基于来自所述比较模块的结果提供相关得分；
 领域代理选择器，基于来自所述评分模块的结果选择领域代理；
以及
 应答产生模块，与所述选择的领域代理通信以得到由所述选择的领域代理所汇集的内容，并且从所述内容产生应答，其中所述内容以基于来自所述评分模块的结果所选择的次序排列。

2. 根据权利要求 1 的系统，其中，所述应答产生模块产生聚集应答，包括由所述选择的领域代理所汇集的内容。
3. 根据权利要求 1 的系统，其中，所述终端设备包括(i)个人数字助理，(ii)蜂窝电话，(iii)便携式计算机，(iv)台式计算机，或者(i)至(iv)的任何组合。
4. 根据权利要求 1 的系统，其中，所述终端设备接收后续的语音和非语音通信，并且所述转录模块转录所述后续的语音和非语音通信以创建基于后续语音的文本消息和基于后续非语音的文本消息。
5. 根据权利要求 4 的系统，其中，合并模块将所述基于后续语音的文本消息和所述基于后续非语音的文本消息合并以产生后续的查询。

6. 根据权利要求 1 的系统，进一步包括便于所述应答格式编排的个性模块。

7. 根据权利要求 1 的系统，进一步包括上下文堆栈，它包括基于所述查询所选择的一个或多个上下文。

8. 根据权利要求 7 的系统，其中，所述评分模块确定所述一个或多个上下文时，基于对以下各项至少应用先验概率或模糊可能性：
(i)关键字匹配，(ii)用户概况，(iii)对话历史，或者(i)至(iii)的任何组合。

9. 根据权利要求 1 的系统，其中，所述领域代理至少其中之一创建请求，并且将其导向本地信息源和网络信息源至少其中之一。

10. 根据权利要求 1 的系统，其中，所述领域代理至少其中之一创建命令，并且将其导向远程或本地设备。

11. 根据权利要求 5 的系统，其中，所述领域代理至少其中之一评估来自多个数据源的多个查询。

12. 根据权利要求 5 的系统，其中，所述后续的查询与所述查询相同的上下文相关联。

13. 一种处理语音和非语音通信的方法，包括：

接收所述语音和非语音通信；

转录所述语音和非语音通信以创建基于语音的文本消息和基于非语音的文本消息；

将所述基于语音的文本消息和所述基于非语音的文本消息合并以产生查询；

搜索所述查询中的文本组合；

将所述文本组合与上下文描述语法中的条目进行比较；

访问多个领域代理，它们与所述上下文描述语法相关联；

基于所述文本组合与所述上下文描述语法中条目比较的结果产生相关得分；

基于来自所述相关得分的结果选择一个或多个领域代理；

得到由所述选择的领域代理所汇集的内容；以及

从所述内容产生应答，其中所述内容以基于来自所述相关得分的结果所选择的次序排列。

14. 根据权利要求 13 的方法，进一步包括产生聚集应答，包括由所述选择的领域代理所汇集的内容。

15. 根据权利要求 13 的方法，进一步包括：

接收后续的语音和非语音通信；

转录所述后续的语音和非语音通信以创建基于后续语音的文本消息和基于后续非语音的文本消息；以及

将所述基于后续语音的文本消息和所述基于后续非语音的文本消息合并以产生后续的查询。

16. 根据权利要求 13 的方法，进一步包括向用户传递所述应答的个性模块。

17. 根据权利要求 13 的方法，进一步包括产生上下文堆栈，它包括基于所述查询所选择的一个或多个上下文。

18. 根据权利要求 17 的方法，其中，产生所述一个或多个上下文时，基于对以下各项应用先验概率或模糊可能性：(i)关键字匹配，(ii)用户概况，(iii)对话历史，或者(i)至(iii)的任何组合。

19. 一种用于处理语音和非语音通信的多模态系统，包括：

终端设备，接收一种或多种类型的输入；

转录模块，将所述一种或多种类型的输入转录为一条或多条文本消息；

合并模块，将所述一条或多条文本消息合并以产生查询；

搜索模块，搜索所述查询中的文本组合；

比较模块，将所述文本组合与上下文描述语法中的条目进行比较；

多个领域代理，它们与所述上下文描述语法相关联；

评分模块，基于来自所述比较模块的结果提供相关得分；

领域代理选择器，基于来自所述评分模块的结果选择领域代理；以及

应答产生模块，与所述选择的领域代理通信以得到由所述选择的领域代理所汇集的内容，并且从所述内容产生应答，其中所述内容以基于来自所述评分模块的结果所选择的次序排列；

所述终端设备使用一种或多种类型的输出传递所述应答。

20. 根据权利要求 19 的多模态系统，其中，所述一种或多种类型的输入包括：(i)语音，(ii)文本，(iii)数字音频文件，或者(i)至(iii)的任何组合。

21. 根据权利要求 19 的多模态系统，其中，所述一种或多种类型的输出包括：(i)语音，(ii)文本，(iii)数字音频文件，或者(i)至(iii)的任何组合。

22. 根据权利要求 19 的多模态系统，其中，所述应答产生模块产生聚集应答，包括由所述选择的领域代理所汇集的内容，并且所述终端设备使用所述一种或多种类型的输出传递所述聚集应答。

23. 根据权利要求 19 的多模态系统，其中，所述终端设备包括(i)个人数字助理，(ii)蜂窝电话，(iii)便携式计算机，(iv)台式计算机，或者(i)至(iv)的任何组合。

24. 根据权利要求 19 的多模态系统，其中，所述终端设备接收一种或多种类型的后续输入，所述转录模块转录所述一种或多种类型的后续输入以创建一条或多条后续的文本消息，并且所述一种或多种类型的后续输入不同于所述一种或多种类型的输入。

25. 根据权利要求 24 的多模态系统，其中，接收所述一种或多种类型后续输入的所述终端设备不同于接收所述一种或多种类型输入的所述终端设备。

26. 根据权利要求 24 的多模态系统，其中，所述合并模块将所述一条或多条后续的文本消息合并以产生后续的查询。

27. 根据权利要求 19 的多模态系统，进一步包括便于所述应答格式编排的个性模块。

28. 根据权利要求 19 的多模态系统，进一步包括上下文堆栈，它包括基于所述查询所选择的一个或多个上下文。

29. 根据权利要求 28 的多模态系统，其中，所述评分模块确定所述一个或多个上下文时，基于对以下各项至少应用先验概率或模糊可能性：(i)关键字匹配，(ii)用户概况，(iii)对话历史，或者(i)至(iii)的任何组合。

响应自然语言语音口头表达的系统和方法

相关申请的交叉引用

本申请要求以 2005 年 8 月 5 日提交的美国专利申请序列号 11/197,504 作为优先权基础，其全部内容在此引用作为参考。

技术领域

本发明涉及通过语音接口和/或语音接口与非语音接口的组合检索信息或处理命令。更确切地说，本发明提供了完全集成的环境，允许用户通过语音接口和非语音接口提交自然语言的问题和命令。可以从广泛的学科中获得信息，以自然方式进行本地和网络查询以获得信息并表达结果，即使在所询问的问题或所收到的响应不完全、多义或主观的情况下。本发明可以进一步地允许用户或者在本地或者远程地控制若干设备和若干系统。

背景技术

机器以自然方式与人类交流的能力仍然是难题。关于人类互动的认知研究显示，口头交流比如某人询问问题或给出命令，典型情况下很大地依赖于上下文和目标人员的领域知识。相反，基于机器的查询（查询可以是问题、命令、请求和/或其他类型的通信）可能是高度结构化的，并且对人类用户可能并非天性自然。因此，口头交流以及从口头交流所取得查询的机器处理可能从根本上就不相容。然而允许人们进行基于自然语言语音的查询仍然是所希望的目标。

语音识别的准确度已经不断改进，现今成功地使用在范围广泛的应用中。自然语言处理已经应用于语音查询的语法分析。然而，现在的系统无法可靠地为用户提供完整环境，以便通过经处理后提供自然应答的自然语言查询提交口头和/或文本交流。为了创建完整的基于语

音的和/或基于非语音的自然语言查询和应答环境，仍然存在着许多重大障碍。

发明内容

根据本发明的某方面，本发明的一个目的是克服现有系统的这些和其他缺点。

根据本发明的一个方面，用户可以提供具有以自然语言格式口头表达和/或文本交流的系统，包括不完全的信息，比如不完整的想法、不完整的句子、不完整的短语、俚语术语、重复单词、单词变种、同义字或其他不完全的信息。为了使机器（如计算机设备）正确地响应以所述自然语言格式表达的问题和/或命令，可以对所述问题和/或命令进行语法分析和解释，表达为机器可处理的查询和算法。

在本发明的另一个方面中，提供的系统通过应用完整的基于语音的信息查询、检索、显示和命令环境或者基于语音和基于非语音的信息查询、检索、显示和命令环境的组合，克服现有系统的不足。这种环境最大限度地使用了上下文、先验信息、领域知识以及用户特定的概况数据，以完成使一位或多用户向多个领域提交查询和命令的自然环境。通过这种整合的方案，就可以创建完整的基于语音的自然语言查询和应答环境。作为替代，也可以创建基于语音和基于非语音的自然语言查询和应答环境的组合。此外，在所述过程的每一个步骤都可以对全部或部分故障以及适度恢复进行调节。通过在所述过程的几个阶段使用概率和模糊推理，实现了对部分故障的稳健性。对部分故障的这种稳健性提升了对问题和命令的自然应答感觉。

根据本发明的另一个方面，所述系统可以包括语音单元接口设备，它接收口头表达，包括来自用户的口头自然语言查询、命令和/或其他口头表达；以及计算机设备或系统，它从所述语音单元接收输入并处理所述输入（如响应所述查询而检索信息，采取与所述命令一致的行动并执行如本文详述的其他功能）。所述系统可以进一步产生自然语言应答和/或在没有来自用户的任何提示时产生自然语言消息，

比如报警消息。根据本发明的另一个实施例，可以提供非语音接口，以补充或替代所述语音单元接口。例如，所述非语音接口可以传达非语音信息，比如文本通信、图形或制表信息，或者其他非语音信息。

根据本发明的一个实施例，可以提供基础设施以便保存多模态互动期间的上下文信息，比如语音和/或非语音互动期间。根据本发明的一个示范实施例，通过在多模态设备或移动设备与所述系统之间提供通信信道，可以在多模态环境中保存上下文信息。所述通信信道允许所述系统接收多模态输入，比如基于文本的命令和问题以及/或者语音命令和问题。根据本发明的另一个实施例，所述多模态输入可以包括文本字符串，比如作为命令或问题所接收的关键字。根据本发明又一个实施例，所述系统可以使所述多模态设备与所述基于语音的单元之间的上下文同步。为了向对应设备发送应答，所述系统可以跟踪该来源并向对应的语音接口或非语音接口发送应答。

根据本发明的替代实施例，保存上下文信息时可以使用上下文管理器，它可以位于中心位置，以便从多个来源接收输入并向多个来源提供输出。根据一个实施例，与所述上下文管理器通信的设备可以通过注册模块进行注册，并且可以预订一种或多种事件。根据本发明的另一个实施例，所述上下文管理器可以接收例如 Context XML 形式的输入。通过上下文跟踪模块可以向其他已注册设备通报上下文的变化，以便在若干已注册模块之间使上下文能够同步。根据本发明的一个实施例，若干已注册模块可以添加到所述系统或从中去除。所述已注册模块可以包括专用于多模态设备的若干动态链接库（DLL）。

根据本发明又一个替代实施例，从以文本格式表示的命令或请求和/或呈现为口头表达的命令或请求中可以确定上下文信息，并且使用将所述口头表达转录为文本消息的多步式自动语音识别模块处理。所述命令或请求可以与上下文描述语法进行对比以识别出匹配。上下文描述语法中的任何有效语法都可以对所述命令或请求评分，并且可以将最佳的匹配发送到应答发生器模块。若干代理可以与对应的应答发生器模块相关联，并且可以为产生应答而检索所请求的信息。所述代

理可以更新上下文堆栈以允许跟踪请求。

根据本发明的另一个方面，所述语音单元和/或多模态设备可以加入所述计算机设备或系统中，也可以是分离的构件。如果提供了分离的构件，所述语音单元和/或多模态设备就可以通过有线或无线连接与所述计算机设备相连。如果提供了无线连接，可以将基站单元内部地或外部地连接到所述计算机，以便与所述语音单元和/或多模态设备进行通信。

根据本发明的另一个方面，所述计算机设备或系统可以包括独立 PC 或网络 PC、个人数字助理（PDA）、蜂窝电话或者其他计算机设备或系统。为了方便起见，这些和其他计算机替代品都称为计算机。本发明的一个方面包括可以安装在所述计算机上的软件，所述软件可以包括以下模块的一种或多种：非语音信息接收模块；捕获用户口头表达的语音识别模块；分析所述口头表达的语法分析器；将所述文本转录为语音的文本到语音引擎模块；使所述计算机能够与一个或多个网络对接的网络接口；非语音接口模块和管理事件的事件管理器。优选情况下，所述事件管理器联系着上下文描述语法、能够创建、修改和访问用户概况的用户概况模块、能够创建和使用多种个性的个性模块、代理模块、更新管理器、提供用户与所述系统互动模式统计摘要的认知模块、一个或多个数据库以及其他部件。

根据本发明的另一个方面，领域特定的行为和信息可以组织在数据管理器中。数据管理器是若干自主的可执行程序，它们接收、处理并答复用户的问题、查询和命令。所述数据管理器提供了完整、方便和可再分配的若干程序包即功能模块，典型情况下用于特定领域的应用。数据管理器可以是可执行代码、脚本、信息链接以及其他形式的通信数据的完整程序包，提供了特定的功能程序包，通常在某特定领域中。换言之，数据管理器包括了将所述功能扩展到新领域的若干组件。此外，当添加了新行为或新信息变得可用时可以经过网络远程地更新数据管理器及其相关联的数据。数据管理器可以使用系统资源和其他，典型情况下是更专业的，数据管理器的服务。数据管理器可以

以许多方式分配和重新分配，包括在可拆卸存储介质上、通过网络传递或附在电子邮件和其他消息上。可以使用更新管理器将新的数据管理器添加到所述系统，或者更新现有的数据管理器。

根据本发明的另一个方面，可以提供许可证管理能力，它允许第三方将数据管理器以一次性或预订形式销售给一个或多个用户。另外，具有具体专长的用户可以创建数据管理器，通过添加新的行为和信息更新现有的数据管理器，以及使这些数据管理器成为其他用户的代理。

为了增强所述自然查询和应答的环境，所述系统可以编排结果的格式以增加用户的可理解性。编排结果的格式并呈现时可基于所述问题的上下文、所呈现应答的内容、与所述用户的互动历史、所述用户的偏爱和兴趣以及所述领域的性质。相反，很多用户可能认为苛刻的、高度格式化的或结构化的结果呈现是不自然的。

根据本发明的另一个实施例，所述系统可以仿真某些方面的人类“个性”。在某些情况下，所述应答表达和用于提供所述应答的术语可以随机化以避免出现苛刻地格式化的即呆板的应答。也期望使用其他仿真的个性化特征。例如，可以以表示同情的方式呈现可能使所述用户不安的应答。不仅如此，请求结果可能是长文本串、列表、表格或其他冗长的数据集。自然表现这种类型的信息带来了特定的挑战，因为简单读出长应答通常不是首选。相反，所述系统可以从所述应答中分析重要部分，并且最初可以仅提供报告。确定长应答的哪些部分要呈现时可以基于所述问题的上下文、所呈现应答的内容、与所述用户的互动历史、所述用户的偏爱和兴趣以及所述领域的性质。同时，所述系统可以给予所述用户互动控制权，以控制要呈现何种信息以及要呈现多少信息、一起停止所述应答或采取其他动作。

根据本发明的另一个方面，所述系统可以处理并响应若干问题、请求和/或命令。若干关键字和上下文可以用于判断所收到的口头表达和/或文本消息是否包括请求或命令。例如，口头表达可以包括问题、请求和/或命令的若干方面。例如，用户可以说出“录制我喜爱的 TV 节目”。对请求进行处理以确定所述用户喜爱 TV 节目的名称、频道

和时间。为了设置录象机捕获这个节目，必须执行一条命令。

对于包括若干问题和/或请求的口头表达，所述系统可以执行多个步骤，可能包括以下的一个或多个：

- 通过在多种多样真实环境中运行的语音识别组件，捕获所述用户的问题和/或请求；
- 对所述问题和/或请求进行语法分析和解释；
- 确定所述专长领域和上下文，调用适当的资源，包括若干代理；
- 将一个或多个请求表达为一个或多个本地和/或网络数据源，或者向本地或远程设备或所述系统自身发送适当的命令；
- 对显示进行格式编排、变量替换和变换以将所述请求修改为能够从所述可用源产生所期望结果的形式；
- 以异步方式执行所述多个请求或命令并适度地处理故障；
- 从可能以许多不同格式的任何一种返回的所述一个或多个结果中提取或分解所期望的信息；
- 评估并解释所述结果，包括处理错误以提供一个或多个被判定为“最佳”的结果，即使所述结果是歧义、不完整或冲突的；
- 进行格式编排、变量替换和变换以将所述结果修改为所述用户最容易理解的形式；以及
- 通过文本到语音引擎或多模态接口，以有用和所期望的方式向所述用户呈现所复合的结果。

可以利用专长领域的知识、所述问题或命令的上下文、领域特定的信息、所述用户的互动历史、用户偏爱、可用信息源或命令以及从所述源获得的应答，执行以上若干步骤。

概率或模糊集决策和匹配的方法可以应用于处理不一致、歧义、冲突和不完全的信息或应答。此外，异步查询可以用于对请求或命令提供快速而适度的失效，允许所述系统稳健、快速地返回结果，而且以对所述用户似乎自然的方式。

许多日常问题本质上是主观的，并且产生的答案事实上是大致看法或民意。这些问题往往其本性也很特别。根据本发明的另一个方面，所述系统可以使用自适应、概率和模糊集决策和匹配的方法，以便识别出所述问题的主观本性并估计出若干可能答案的范围，其中可以选择最准确地表示所述用户所期望的结果类型的一种和多种答案。

具体问题的上下文和所期待的结果可能在很大程度上取决于询问所述问题的个人。所以，所述系统可以创建、存储和使用用于每个用户的个人概况信息。所述概况中的信息可以随着所述用户使用所述系统而自动地添加和更新，也可以由所述用户或其他人手工地添加或更新。领域特定的代理可以收集、存储和使用特定的概况信息，因为最佳运行可能需要它。对于有规律地使用的报告、自动产生的警报以及用于结果格式编排和呈现的其他请求，用户可以创建命令。所述系统可以在解释问题、请求格式编排、向所述用户解释请求结果和呈现答案时使用概况数据。用户概况中的信息实例包括，所询问问题的历史、通话历史、格式编排和显示偏爱、专用字的拼法、关注术语、关注的特殊数据源、年龄、性别、教育、位置即地址、商家地点、商家类型、投资、业余爱好、体育兴趣、新闻兴趣以及其他概况数据。

根据本发明的一个方面，为了提供自然的问题和应答环境，所述系统可以尝试提供若干快速应答。可以在不获得附加信息的情况下提供所述快速应答。所述系统可以使用实时评分系统或其他技术判断用户问题或命令的代理组成、上下文和/或领域。基于这种判断，所述系统可以触发一个或多个代理，以响应所述用户的问题或命令。所述代理可以产生一个或多个请求并迅速地返回格式化的应答。因此，用户可以收到对一组问题的直接应答，每个问题都具有不同的应答或上下文。在某些情况下，包括所述请求结果的可用信息可能不足以回答所提出的问题。在这样的情况下，可以向所述用户询问一个或多个后续问题以解决歧义性。然后可以提出若干附加请求，再提供适当的应答。在这些情况下，所述系统可以使用上下文信息、用户概况信息和/或领域特定的信息，以便使为了所需的与所述用户的互动最少。

如果所述领域或上下文得分的置信度不够高，无法确保可靠的应答，所述系统可以请求所述用户验证所述问题或命令被正确理解。一般来说，可以用短语表达所述问题以指明所述问题的上下文，包括全部准则或参数。如果所述用户证实所述问题是正确的，所述系统就可以开始产生应答。否则，或者所述用户能够将原始问题重新用短语表达，也许添加了附加信息以消除歧义性，或者所述系统可以询问一个或更多的问题以尝试解决所述歧义性也可以采取其他行动。

根据本发明的一个方面，所述系统可以接受任何自然语言的问题或命令，结果可能会招致歧义的请求。为了帮助用户阐明简洁的问题和命令，所述系统可以支持声音查询语言。所述语言可以帮助用户连同所述参数或准则一起清楚地指定所述问题或命令的若干关键字或上下文。所述系统可以提供内建的训练能力以帮助所述用户学习如何阐明其问题和命令的最好方法。

为了使对用户问题和命令的应答似乎更自然，所述系统可以采用一种或多种动态可调用的个性和/或情感模型。个性和情感模型具有特定的特征，仿真真实人类的行为特征。这些特征的实例包括同情、愤怒和帮助以及相关联的情绪。所述个性还会使应答的若干方面随机化，正如真实人类的行为。这种行为包括所用术语和使信息显示次序的随机化。调用所述个性和/或情绪的特征时使用了概率或模糊集决策和匹配的方法，使用了若干准则，包括所述问题的上下文、所述用户互动的历史、用户偏爱、可用的信息源、从所述源得到的应答。

根据本发明的另一个方面，可以采用专用过程，以长文本串、表格、列表或其他长应答集的形式表达信息。以罗列方式简单地表达长信息集也许认为不自然，或者不符合大多数用户的心意。所述系统可以使用概率或模糊集匹配的方法，提取有关信息并优先呈现这些子集。此外所述系统可以提供的命令允许用户浏览所述列表，找到所述列表中的关键字或关键信息，或者完全停止处理所述列表。

除了处理用户问题和命令的能力之外，本发明可以为所述系统或为其他设备提供本地或远程的控制功能。用户可以本地或远程地启动

命令。可能有执行远程操作的几种技术，例如使用电话或其他声频连接。也可以使用其他远程命令技术。处理这些命令的方式可以与处理问题的方式几乎一致。一种差异是所述命令的结果通常是动作而不是应答。本发明支持的控制操作实例包括：找到并播放音乐、电影、游戏或其他娱乐媒体、控制家用电器或者其他电子或电子机械设备、安全系统控制、声频、文本和视频会议（包括 IP 语音会议）、操作集成的消息环境以及其他控制操作。

通过使用可选的模拟或数字电话接口或者数字网络接口，本发明能够提供对计算机或/和连接到所述计算机的辅助设备远程控制的能力。例如，如果所述计算机经过接口连接着一台或多台辅助设备，远程用户就能够使用这种能力远程地控制或编排所述辅助设备的操作。例如，这能够录制电视节目、检查家庭安全系统的状态、检查语音邮件消息、听取和响应电子邮件消息等等，不一而足。对于每位用户命令的口头表达，所述系统都可以执行许多步骤，可能包括：

- 通过在多种多样真实环境中运行的语音识别，捕获所述用户的命令；
- 对所述命令进行语法分析和解释；
- 确定所述命令的领域和上下文，调用适当的资源，包括若干代理；
- 为所述系统或外部设备制定设备专用的命令；
- 将命令路由到所述系统或外部设备，包括连接到数据网络的外部设备；
- 接收并处理命令的结果，包括错误；以及
- 可选地向用户提供应答，指示所述命令的成功或失败，并且可能包括状态信息。

所述系统可以和范围广泛的平台环境结合使用。这些环境可以包括：作为 PC、便携式或可佩戴计算机或其他计算机的外围设备、嵌入在 PC 或其他计算机中、在个人数字助理上比如带有无线网络功能的 PDA、在专用的设备中比如条形码扫描器或支付终端、在无线电话

上以及其他平台。如果本发明用于具有定位功能的移动或便携式设备，本发明就能够使用该位置数据作为用户问题上下文的一部分。用户可以在一台或多台设备上使用所述系统。在这种情况下，所述用户的历
史和概况信息可以在多台设备之间定期地或以其他方式同步。

根据本发明的另一个方面，所述系统可以部署在使用代理、数据、信息、用户概况和历史的公共基础的若干设备的网络中。然后每个用户都能够在网络上配备所需设备的任何位置处互动并接收所述相同的服务和应用程序。例如，使用本发明并连接到网络的多台设备可以放置在遍及家庭、商家地点或其他位置的不同位置处。在这样的情况下，所述系统可以使用由所述用户所涉及具体设备的位置，作为所询问问题上下文的一部分。

根据本发明的一个实施例，所述系统可以支持在不同时间访问所述系统的多个用户。根据本发明的另一个实施例，所述系统可以支持以交错或重叠方式在同一对话期间访问所述系统的多个用户。所述系统可以依据名字、声音或其他特征识别出所述多名用户，并且可以调用每个用户的正确概况。如果多个用户正在重叠或交错对话中访问所述系统，所述系统可以识别出所述多个用户并可以调用一个或多个对应概况。对于需要安全保护的应用，验证所述多个用户时可以使用声波纹匹配、密码或许可证短语匹配以及其他安全保护。

当多个用户处于交错对话中时，所述系统可以使用概率或模糊集决策方法适度地为每个用户解决冲突。这种过程可以仿真人类对待来自多个源的多个问题时会采用的方式。例如，所述系统可以在时间上首先回答简短的问题，而在时间上随后回答较长的问题。作为替代，在其他配置之中，所述系统可以依据收到问题的次序回答问题。

给定了对领域特定行为、用户特定行为和领域特定信息的需要，根据本发明的另一个方面，所述系统可以允许用户和内容提供商双方扩展所述系统的能力、向本地数据源添加数据以及向网络数据源添加引用。为了允许覆盖尽可能广的论题范围，所述系统可以允许第三方内容开发商开发、发行和销售专门的即领域特定的系统程序和信息。

通过创建新的数据管理器、改写现有数据管理器、向代理或数据库添加新数据以及添加或修改到信息源的链接，可以创建内容。这种信息的发行对所述用户兴趣和使用历史敏感，以及对他们是否愿意支付敏感。

根据本发明的另一个方面，所述系统可以允许用户传播他们的知识而不编程。所述系统包括的机制可以允许用户在他们具体专长领域中张贴和发行数据管理器和信息，以便改进系统性能。此外，用户可以按他们自己的偏爱扩展所述系统并进行配置、向他们的概况添加信息以定义新问题或查询、扩展和修改现有的问题和查询、添加新数据源、更新数据源、设置优先选择以及规定结果的显示参数。

由于本发明可以在许多环境中运行，包括具有背景噪声、点噪声源和人们持续会话的环境，所以可以对话音输入进行滤波。本发明可以使用例如或者一维或者二维的阵列话筒接收人类话音。所述阵列话筒可以是固定的，也可以采用动态射束形成技术。可以调整所述阵列模式以使所述用户方向上的增益最大并使点噪声源无效。然后可以利用模拟或数字滤波器处理所述话筒收到的话音，以便优化带宽、消去回声并陷波消除窄带噪声源。滤波之后，所述系统可以使用可变速率采样以使编码后话音的保真度最高，同时使所需带宽最小。在所述编码后话音经过无线网络或链接传输的情况下，这种过程尤其有用。

本发明的某些实例应用包括但是不限于以下各项：

- 白页和黄页查找以找到商家和个人的电子邮件地址、电话号码、街道地址以及其他信息；
- 每位用户的个人地址簿、日历本和待办事项本；
- 通过话音和其他通信控制功能进行自动电话拨号、读取和发送电子邮件和页面；
- 地图、定位和定向应用；
- 电影或其他娱乐预告，评论信息以及购票；
- 电视、广播或其他家庭娱乐时间表，评论信息和来自本地或远程用户的设备控制；

- 本地区或其他地点的天气信息；
- 股票和其他投资信息，包括价格、公司报告、概况、公司信息、商务新闻报道、公司报告、分析、价格警报、新闻警报、资产组合报告、资产组合计划等；
- 航班或其他定期运输信息和售票；
- 旅店预定，租车和其他旅行服务；
- 本地、国内和国际新闻信息，包括按主题或地点的受关注标题、故事摘要、故事全文、故事的音频和视频检索和播放；
- 体育成绩、新闻报道、时间表、警报、统计表、背景和历史信息等；
- 交互预订多媒体信息频道的能力，包括体育、新闻、商务、不同类型的音乐和娱乐、使用用户特定偏爱提取并显示信息；
- 对所使用或发表的信息或内容进行版权管理；
- 占星术、每日笑话和喜剧、纵横拼字谜的检索和显示以及有关的娱乐或消费；
- 食谱、进餐计划、营养信息和计划、购货单和与家庭组织有关的其他活动；
- 作为拍卖和在线购物的接口，并且所述系统可以管理付款或电子钱夹；
- 管理网络通信和会议，包括电信、电子邮件、即时消息、IP语音通信和会议、局域和广域视频音频会议、页面和警报传递话音；
- 对播放列表的定位、选择、管理以及对来自本地或网络源互动娱乐的播放控制，包括：点播视频、数字音频比如 MP3 格式材料、互动游戏、网络广播和视频广播；
- 对家庭、商家和其他用户团体的组织和日程表管理，包括：会议、约会和事件的管理；以及
- 使用本地和网络资料的互动教育节目，课程资料等级基于用户概况设置，并且包括互动多媒体课程、宗教指导、计算器、

字典和拼写、语言训练、外语翻译、百科全书以及其他参考资料。

应当认识到，对于本发明若干特点的以上陈述并非试图囊括或限制由参考本全部公开材料所认识到的其适当范围以及其合理地显而易见的若干变种和扩展。

附图说明

将要连同附图参考本发明的若干优选和替代实施例介绍本发明，其中：

图 1 是根据本发明一个实施例的总图解视图；

图 2 是一幅示意框图，显示了根据本发明一个实施例的代理架构；

图 3 是根据本发明一个实施例的高级别过程，用于接收基于自然语言语音的查询和/或命令并产生应答；

图 4A 是根据本发明一个实施例的过程，用于接收基于自然语言语音的查询和/或命令并产生应答；

图 4B 是根据本发明一个实施例的过程，用于接收基于自然语言语音的命令，以便在本地或远程地控制设备功能；

图 5 是根据本发明一个实施例的过程，用于正确地解释用户的口头表达；

图 6 是根据本发明一个实施例的过程，用于确定要调用的恰当领域代理以及将要提交到该代理的恰当格式化的查询和/或命令；

图 7 展示了根据本发明一个实施例，集中地保持上下文的图。

具体实施方式

对于在人与机器之间的自然人请求-应答互动的障碍包括：自然语言询问可能在其定义上不完整。不仅如此，为了恰当地解释若干问题和/或命令，可能需要先前问题的上下文、领域知识和/或用户的兴趣和偏爱历史。因此，自然语言问题无法容易地转录为机器可处理的形式。混合了这个难题时，自然语言问题可能会模糊或主观。在这些

情况下，可能难以形成机器可处理的查询并返回自然语言应答。

为了使机器（如计算机设备）正确地响应以自然语言形式所提交的问题和命令，在已经对自然形式的问题或命令进行了语法分析和解释之后，可以表达为机器可处理的查询和算法。若干算法描述这些机器应当如何集合数据以响应这些问题或命令。取决于这些问题或命令的性质，也许没有一组简单的查询和算法将返回适当的应答。为了实现完整应答可能需要启动几种查询和算法，甚至可能需要将这些查询和算法环环相扣即连接起来。此外，任何单一可用源都不会包含产生完整应答所需要的完全结果集。因此，为了访问有的位于本地有的位于远程的多个数据源，可以产生多个查询和/或算法，也许分为几个部分。有的数据源、查询和/或算法可能不返回有用的结果，或者根本不返回任何结果。

返回的有用结果往往嵌入在其他信息中，并且可能需要从所述其他信息中提取。例如，可能需要从文本字符串、表格、列表、页面、视频流或其他信息中的大量其他信息里“分解出”几个关键字或数字。与此同时，为了处理该应答可以去除包括图形或图片的无关信息。在任何情况下都必须评估和组合多个结果以形成最好的可能应答，即使在某些查询没有返回有用的结果或者完全无法产生结果的情况下。在判定所述问题有歧义或所述结果本质上主观的情况下，确定在应答中要呈现的结果是复杂的过程。最后，为了保持自然互动，需要向用户迅速地返回应答。管理和评估复杂而不确定的查询，同时保持实时性能是重大的挑战。

以下详细说明涉及附图，介绍了本发明的示范实施例。其他实施例也是可能的并且可以对示范实施例做出若干修改而不脱离本发明的实质、功能和范围。所以，以下详细说明不意味着限制本发明。

本发明提供了完全基于语音的信息查询、检索、处理和显示的环境，或者基于语音和基于非语音结合的信息查询、检索、处理和显示的环境。另外，本发明可用于控制系统自身和/或外部设备。这种集成环境最大限度地使用了上下文、先验信息以及领域和用户专用的概况

数据，以便实现使一个和多个用户提交多个领域中请求或命令的自然环境。通过这种集成方案，就可以创建完全基于语音的或者基于语音和基于非语音结合的自然语言查询、算法和应答环境。图 1 显示了根据本发明实施例的系统 90 的一幅示范示意视图。

系统 90 可以包括主单元 98、语音单元 128 和多模态设备 155。作为替代，系统 90 也可以由主单元 98、语音单元 128 和多模态设备 155 都是完全分开的系统组成。事件管理器 100 可以仲裁主单元 98 的其他组件之间的互动。事件管理器 100 提供的多线程环境允许系统 90 以高效方式对来自多个用户会话多条命令或问题操作而没有冲突，保持着实时响应的能力。

代理 106 可以包括语法、准则处理器和算法的集合，在响应一组请求和/或命令时受到访问。代理 106 可以进一步包含若干软件包，适合于系统 90 的一般行为和领域特定行为的都有。代理 106 可以使用非易失性存储器存储数据、参数、历史信息以及在系统数据库 102 和其他本地源中提供的本地存储的内容。可以提供一个或多个用户概况 110，它包括用户专用数据、参数以及会话和历史信息，用于确定代理 106 的行为。在数据判断系统中可以提供一个或多个个性模块 108，它包括代理 106 所用的若干个性特征。更新管理器 104 管理着代理 106 的自动和手工加载和更新以及它们的相关联数据，它们通过网络接口 116 来自因特网 146 或其他网络。

根据本发明的一个实施例，系统 90 基于语音的接口可以包括一个或多个语音单元 128。语音单元 128 可以包括一个或多个话筒，例如阵列话筒 134，接收用户口头表达。话筒 134 收到的语音可以由滤波器 132 处理并传递到语音编码器 138 进行编码和压缩。在一个实施例中，收发器模块 130 将编码后语音传送到主单元 98。从主单元 98 收到的编码后语音由收发器模块 130 检测，然后由语音编码器 138 解码和解压缩，并且由扬声器 136 宣告。

根据本发明的一个实施例，系统 90 基于非语音的接口可以包括一台或多台多模态设备 155，它们可以包括移动设备、独立或网络计

算机、个人数字助理（PDA）、便携式计算机设备或其他多模态设备。

语音单元 128、多模态设备 155 和主单元 98 可以通过通信链路进行通信。通信链路可以包括有线或无线链接。根据一个实施例，通信链路可以包括 RF 链接。语音单元上的收发器 130 可以与主单元 98 上的收发器 126 经过通信链路双向地传递编码后语音数据。根据另一个实施例，RF 链接可以使用任何标准的局域无线数据协议，包括 IEEE 802.11、蓝牙或其他标准。作为替代，也可以使用符合任何合适标准的红外线数据链接，比如 IrDA 或其他红外线标准。在替代实施例中，导线连接着语音单元 128 和主单元 98，不再需要一台语音编码器 138。也可以使用其他有线或无线的模拟或数字传输技术。

根据本发明的一个实施例，主单元 98 上收发器 126 收到的编码后语音可以传递到语音编码器 122 进行解码和解压缩。解码后语音可以由语音识别引擎 120 使用上下文描述语法模块 112 与其他信息进行处理。识别后的任何信息都可以由语法分析器 118 进行处理，它使用知识代理所提供的数据将信息转录为完整的算法和问题。使用语法堆栈可以从多个代理 106 中选出若干知识代理，这些知识代理提供的信息用于产生对于问题或命令的响应。这些知识代理然后可以处理这些命令或问题，创建若干查询，提交到本地数据库 102 或通过网络接口 116 经过因特网 146 或外部其他网络提交到外部数据源。典型情况下，算法将导致系统 90 自身采取行动（即暂停或停止），或者经过到因特网的网络接口或其他数据接口，对远程设备或数据源采取行动（即下载数据或程序或者控制远程设备）。

根据本发明的一个实施例，知识代理可以返回问题的结果作为对用户的应答。创建所述应答时可以使用信息查询结果、系统的个性 108、用户偏爱、用户概况 110 中的其他数据和/或其他信息。代理 106 可以使用语音单元 128 呈现这些结果。代理 106 可以创建应答串，它将被发送到文本到语音引擎 124。文本到语音引擎 124 产生若干口头表达，可以由语音编码器 122 进行编码和压缩。一旦编码后，收发器 126 就将这些口头表达从主单元 98 传递到语音单元 128 上的收发器

130. 这些口头表达然后由语音编码器 138 进行解码和解压缩并由扬声器 136 输出。作为替代，代理 106 也可以使用多模态设备 155 呈现这些结果。

非语音接口 114 可以是多模态设备 155 的一部分也可以与其分开，并且可用于代替或补充所述语音接口。例如，非语音接口 114 可用于以用户更容易领会的方式呈现非语音（如图形或表格）信息并与之互动。根据本发明的一个实施例，可以提供多模态支持以便在语音互动和经过非语音接口 114 的互动期间都保存上下文。在一个示范实施例中，在多模态设备 155 与主用户接口系统 90 之间可以开放通信通道，以便允许多模态设备 155 输入文本命令和问题。多模态设备 155 可以接收和/或提供一种或多种类型的通信，包括语音、文本、数字音频文件以及其他通信。根据本发明的另一个实施例，多模态设备 155 可以发送命令或问题所用的文本或关键字的字符串。主接口系统 90 可以使多模态设备 155 与语音单元 128 之间的上下文同步。为了向对应的设备发送应答，主用户接口系统 90 可以跟踪输入来自何方，以便应答可以发送到 TTS 或多模态设备 155。

根据图 7 展示的本发明替代实施例，可以集中地保持着上下文管理器 702 以允许来自多台移动设备 36 的输入以及向它们的输出。每台移动设备都可以与上下文管理器 702 通信以便通过注册模块 712 进行注册，其中注册可以表明移动设备 36 可以预订的事件。上下文管理器 702 可以以例如 Context XML 的形式接收输入。通过上下文跟踪模块 714 可以通知其他已注册移动设备 36 上下文的变化，从而在若干已注册移动设备 36 之间使上下文同步。根据本发明的一个实施例，可以添加或去除已注册的移动设备 36。已注册移动设备 36 可以是专用于该移动设备 36 的动态链接库（DLL）。

根据本发明的另一个替代实施例，从表示为文本消息的命令或请求以及/或者表示为口头表达的命令或请求中可以确定上下文信息，并且使用多步式自动语音识别模块进行处理，它将口头表达改编为文本消息。多步式自动语音识别模块可以使用听写语法或大词汇语法与其

他资源，将口头表达改编为文本消息。在无听写语法可用的平台上，多步式自动语音识别可以使用虚拟听写语法，对于词汇表之外的词它使用饵词。饵词包括实用词、无意义词、孤立音节、孤立独特音和其他饵词。

根据本发明的一个实施例，在文本消息中可以搜索具体字符、字符组、词、词组和其他文本组合。所述文本组合可以与每个代理 106 相关联的上下文描述语法中的条目进行对比。如果识别出了上下文描述语法中有效语法与该命令和/或请求之间的匹配，那么该匹配可以评分。根据所确定的分数可以对代理 106 定级。根据从代理 106 中收到一个或多个应答而产生聚集应答时，根据代理 106 的等级可以确定来自各个代理的应答的次序。聚集应答可以由应答发生器模块产生。代理 106 可以更新上下文堆栈，它包括命令上下文的有序列表，以便能够跟踪请求。

根据本发明的另一个实施例，如果在文本消息与有效语法之间未发现匹配，或者仅仅发现了部分匹配，那么知识强化的语音识别系统可以用于扩大语义搜索。知识强化的语音识别系统可以用于判断请求的意图和/或改正错误的识别。知识强化的语音识别可以访问上下文堆栈中存储的一组预期上下文以确定最合适上下文。知识强化的语音识别可以使用上下文专用的匹配程序，它们能够识别出上下文，比如时间、位置、数字、日期、种类（如音乐、电影、电视、地址等）和其他上下文。进行匹配过程时可以通过比较字符、字符组、词、词组和其他文本组合。作为替代，或者对基于文本匹配的补充，进行匹配过程时可以使用语音学匹配与其他技术。任何匹配的结果都可用于产生命令和/或请求，被传达到代理 106 做另外的处理。根据本发明的一个实施例，非语音接口 114 可以显示系统、状态和历史信息，其方式比通过语音接口可能的方式更简洁。为了创建或扩展代理 106 的能力，可以访问非语音接口 114。这些操作可以包括把代理编排为脚本、向代理或该代理所用的数据库 102 添加数据、添加到信息源的链接以及其他操作。

根据本发明的另一个实施例，系统 90 可以包括不同类型的代理 106。例如，一般的和领域专用的行为和信息可以组织到领域代理中。相反，系统代理可以提供默认功能和基本服务。领域代理为每个应用领域提供了完整、方便和可重新分配的软件包或模块。换言之，领域代理包括了在现有或新领域中扩展或修改系统 90 功能所需要的数据。此外，在添加了新行为或新信息变为可用时，领域代理及其相关联的数据能够经过网络远程地更新。领域代理可以使用其他的、典型情况下更专业化的数据管理器和系统代理 150 的服务。代理以许多方式分配和重新分配，包括在可拆卸存储介质上、通过网络传递或附在电子邮件和其他消息上。本发明可以提供许可证管理能力，它允许第三方将数据管理器以一次性或预订形式销售给一个或多个用户。另外，具有具体专长的用户可以创建数据管理器，通过添加新的行为和信息更新现有的数据管理器，以及使这些数据管理器成为其他用户的代理。图 2 显示了根据本发明实施例的代理架构的框图。

代理 106 可以从事件管理器 100 中接收若干事件以及向其返回事件。系统代理 150 和领域代理 156 都可以从语法分析器 118 中接收问题和命令。根据问题和命令中的关键字以及问题和命令的结构，语法分析器可以调用所选定的代理。代理使用非易失性存储器存储数据、参数、历史信息以及在系统数据库 102 中提供的本地内容。

根据本发明的一个实施例，当系统启动或自举时，代理管理器 154 可以加载和初始化系统代理 150 和一个或多个领域代理 156。代理管理器 154 包括代理 106 的知识，并且将代理 106 映射到代理库 158。关闭时，代理管理器可以卸载代理 106。代理管理器 154 还对领域代理 156 和数据库 102 中的内容执行许可证管理功能。

系统代理 150 管理着若干准则处理器 152，它们处理用于确定问题和命令上下文的特定参数或数值（准则）。根据本发明的一个实施例，准则处理器 152 包括若干语法分析例程，专用于识别语音的特定部分，比如时间、位置、电影名称以及语音的其他部分。准则处理器 152 可以从若干短语中识别出匹配短语并从其提取出若干语义属性。

系统代理 150 和领域代理 156 都可以使用准则处理器 152。多种领域代理 156 都可以使用系统代理 150 的服务，以及其他典型情况下更专用的领域代理 156 的服务。系统代理 150 和领域代理 156 可以使用代理库 158 的服务，它含有公用功能的实用程序。根据本发明的一个实施例，代理库可以是执行一个或多个代理的动态链接库。代理库可以包括若干实用程序，用于文本和字符串处理、网络通信、数据库查找和管理、模糊和概率评估、文本到语音的格式转录以及其他实用程序。

领域代理 156 可以是数据驱动的、脚本化的或以编译代码创建的。类属代理的基础可以用作数据驱动或脚本化代理起点。以编译代码创建的代理典型情况下建立在动态可链接即可加载的库中。代理开发商能够根据需要向代理库 158 添加新功能。在以下段落中将讨论代理分配和更新的细节以及代理创建或修改。

根据本发明的另一个实施例，提供的若干性能使用更新管理器 104 通过无线或有线网络 136，包括拨号上网，分配和更新系统代理 150、领域代理 156、代理库 158 组件、数据库 102 以及上下文描述语法 112。网络接口 116 可以提供到一个或多个网络的连接。更新管理器 104 还可以管理核心系统更新的下载和安装。代理管理器 154 可以对领域代理和数据库执行许可证管理功能。更新管理器 104 和代理管理器 154 可以对全部的代理和数据库内容执行这些功能，包括对全部用户都可用的代理和内容，或者对所选定用户可用的代理和/或内容。定期添加或更新的代理和数据库组件的实例包括：

- 新领域的代理；
- 代理的附加领域知识；
- 领域的新关键字，可以包括新近知名的政治家、运动员、专业演员的名字、新电影或歌曲的名称等；
- 对所覆盖领域的优选信息源组的链接，包括对娱乐、新闻、体育、天气等的链接；
- 根据例如税法变化、公司合并、政治边界变化而对领域信息的更新；

- 对内容的更新，包括字典、百科全书和年鉴；以及
- 其他内容和数据库组件。

当用户需要或选择新领域代理 156 或数据库组件 102 时，更新管理器 104 可以通过网络接口 116 连接到网络 146 上它们的资源，下载并安装代理或数据。为了节省系统资源并遵守所有的许可证条件，更新管理器 104 可以卸载不再使用的代理。在本发明的一个实施例中，更新管理器 104 可以定期查询受许可代理和数据库组件的一个或多个源，以便当代理可执行程序、脚本或数据变为可用时定位并下载若干更新。作为替代，当已注册即受许可代理的代理更新变为可用时，代理源可以启动将其下载到更新管理器。

代理管理器 154 可以提供许可证管理客户，它们能够执行几乎任何许可证条款和条件。根据提交的命令选择具体代理 106 和/或数据库组件 102 时，代理管理器 154 验证使用的代理或数据分量在所允许的条款和条件之内，并且假若如此，调用该代理或允许访问该数据分量。通过代理管理器 154 能够实现的许可证管理模式包括即时购买、预约更新、一次或有限时间使用。代理管理器 154 还可以管理共享的代理和数据分量的使用（比如从由若干领域专家组所维护网站下载的代理和数据分量）。

如果问题或命令与系统上当前加载的代理不匹配，代理管理器 154 可以通过网络接口 116 搜索网络 146 以发现合适的代理源。一旦定位后，就能够在更新管理器 104 的控制下、在许可证协定的条款和条件之内加载该代理，正如由代理管理器所实施的。

通过改变代理数据和脚本能够向任何领域代理 156 添加新的命令、关键字、信息或信息源。这些配置能力可以允许用户和内容开发商扩展和修改现有领域代理 156 的行为，或者从类属代理创建新的领域代理 156，而不需要创建新的编译代码。因此，修改领域代理 156 的范围可以从甚至最不经意的用户进行的较小的数据驱动更新，比如指定词的拼法，到使用脚本语言的复杂行为的开发，正如典型情况下会由领域专家所完成的。通过语音接口命令或使用非语音接口 114，

用户可以创建和管理对领域代理 156 的修改。领域代理 156 的用户特定修改连同用户概况 110 一起存储并由领域代理 156 在运行时访问。

配置数据驱动代理 156 所用的数据以易于高效评估和帮助开发商组织的方式构造。这些数据不仅由代理 156 使用，而且由语音识别引擎 120、文本到语音引擎 124 和语法分析器 118 使用。某些主要种类数据的实例包括：

1. 内容数据包包括问题或命令。每个命令或问题或者命令或问题组都包括创建一个或多个请求所用的若干上下文。领域代理 156 将上下文描述语法表达传递给语法分析器 118 以进行上下文或问题的评估。典型情况下为每个命令或问题都提供了初始的即默认的上下文。命令或问题可以包括管理和评估上下文堆栈所用的语法。
2. 页面列表或指针，指向其他本地或网络内容源。对于每个页面或内容源都有指向该页面或源的指针（如 URL、URI 或其他指针）。每个页面都有特定的分解信息，用于提取所关注的数据。分解信息包括匹配模式、HTML 或其他格式的语法分析信息。
3. 应答列表，给定了上下文、用户概况和检索信息后，确定领域代理 156 对具体命令或问题的应答。应答可以包括诊断的错误消息或请求更多信息，用从已知信息还不能解决该问题或命令时。应答可以基于或取决于变量的阈值或者概率或模糊权重。
4. 代入列表，包含变量替换和变换，往往由代理 150 在查询和结果的格式编排中应用。例如，股票领域的专用代理 156 将使用公司交易符号、公司名称和共用缩写的代入列表。可以对命令和问题执行代入和变换以创建精确的查询，这些查询能够对一个或多个信息源或者对结果进行，以便为用户创建更有意义的输出。代入列表包括的信息也用于最理想地处理结构化信息，比如 HTTP 格式网页的语

法分析和评估。

5. 应答所用的个性。以加权方式结合多种特质构造个性。对每个代理的领域范围都可以规定权重以创建一种或多种特定个性。个性特质的实例包括讽刺、幽默、愤怒和同情以及其他特质。

6. 信息源、替换、变换、变量或准则的公共和用户专用参数。公共参数表是代理数据包 156 的一部分。用户专用参数包含在用户概况 110 中。

根据用户个性或用户概况 110 的数值解释命令和问题、表达查询、创建应答并呈现结果。个人概况包括的信息可以专用于个人、他们的兴趣、他们专门使用的术语、他们与系统互动的历史以及所关注领域。个人概况数据可以由代理 106、语音识别引擎 120、文本到语音引擎 124 以及语法分析器 118 使用。偏爱可以包括特殊的（修改过的）命令、过去的行为即历史、问题、信息源、格式、报告和警告。用户概况数据可以由该用户手工输入和/或由系统 90 根据用户行为学得。用户概况值可以包括：

- 拼写偏爱；
- 用户、家人和朋友的出生日期；
- 收入水平
- 性别
- 职业
- 位置信息，比如家庭地址、邻域和商家地址；
- 轿车车型
- 电信和其他服务的提供商和服务；
- 财务和投资信息；
- 别名（即某人的昵称）；
- 专用拼法
- 关键字
- 变换和替换变量；

- 所关注领域；以及
- 其他值

最终用户可以使用数据驱动代理 156 的扩展和修改程序以及用户概况 110 中存储的数值，创建专用报告、查询包、警告和输出格式。单一报警或报告也可以配置为使用多个数据源和其他变量值(即时间、位置)以确定何时应当发送警报。例如，通过每隔 15 分钟对股票价格采样可以产生警报，并且如果该价格下跌至某个值以下便发送警报。为了创建报告，用户首先指定一组命令或问题。下一步，用户为该报告创建或选择格式。最后用户可以为该报告命名。报告可以具有若干可变的参数。例如，用户可以创建公司股票报告，并且通过陈述其名称和公司名称执行该报告，它给出了用户选择的信息并且以适合于这个公司的指定格式。在另一个实例中，用户可以创建“上午”报告，它以所期望的次序和格式呈现了从不同信息源（新闻、体育、交通、天气）选出的多媒体信息。仅仅使用语音命令和应答、通过非语音接口 114 使用命令和应答或者两者的组合，都能够创建警报和报告。为了创建报告、警报或其他专门化的行为，用户执行了许多步骤，包括：

- 指定命令以运行报告或报警；
- 指定问题或若干问题，包括查询所使用的若干关键字；
- 设置运行报告的准则，比如立即或者当满足了某具体条件时；
- 定义优选的信息源；
- 定义由信息源、数值等进行结果评估的次序偏爱；
- 指定报告或报警的呈现介质，比如电子邮件、文本的语音引擎、寻呼机消息，或者文本和图形显示；以及
- 指定报告的优选格式，比如要呈现的信息、要呈现信息的次序、优选的缩写或其他变量代人。

滤波和噪声消除在促进系统 90 的多种功能中可能很重要。向语音识别引擎 120 输入高信噪比信号，进行用户语音的识别和语法分析。为了提供可接受的结果，可以使用一组声学模型、阵列话筒 134、滤波器 132 或其他组件。如果不能实现高信噪比，可以使用噪声识别算

法，可以选择适当的声学模型，例如，在类似于所识别噪声条件下已经训练的模型。根据本发明的实施例，话筒阵列、滤波器和语音编码器 138 都从主单元 98 物理地分离到语音单元 128 中，并且使用无线链接相连。由于无线连接带宽非常珍贵，所以语音编码器动态地适应所捕获语音的数字化速率和压缩。

语音单元 128 可以使用话筒阵列 134，以便使提供的定向信号捕获和噪声消除优于用单话筒所能够达到的程度。话筒阵列可以是一维的（线性阵列）也可以是二维的（圆、正方形、三角形或其他适宜形状）。所述阵列的方向特性可以固定，也可以通过使用模拟或数字相移电路使其自适应。有源阵列的模式被调向一位或多用户谈话的方向。与此同时，可以向该模式添加零点，对点或有限区域噪声源进行陷波。使用阵列话筒也有助于减少从文本到语音引擎 124 通过扬声器 136 的输出与用户语音检测之间的串音。

可以用模拟或数字（即 IP 语音）语音接口补充话筒。这种接口允许远程用户连接系统并与系统互动，方式与他们本身就在现场的可能方式相同。

根据本发明的实施例，语音单元 128 可以使用阵列话筒 134 与语音编码器 138 之间的模拟或数字滤波器 132。设置滤波器的通带以使向语音识别引擎 120 的输入的信噪比最优。在某些实施例中，滤波器是自适应的，使用与陷波组合的频带整形技术滤除窄带噪声。在本发明的一个实施例中，系统 90 可以使用滤波器中的自适应回声消除技术。回声消除技术有助于防止文本到语音引擎的输出与用户语音检测之间的串音以及压制环境导致回声。为了比较背景噪声与从用户语音收到的信号进行了若干计算，并用于使自适应滤波器的频带整形参数最优。本领域技术人员应当熟悉构建适宜的数字或模拟滤波器所用的多种技术。

由阵列话筒 134 收到并通过滤波器 132 传递的语音可以发送到语音数字化器或编码器 138。为了使带宽要求最优化，语音编码器 138 可以使用自适应有损音频压缩，以便经过无线链接向语音识别引擎

120 传输编码后语音。所述有损编码可以优化以仅仅保存最佳识别所需要的语音信号分量。此外，可以使用的有损压缩算法可以设计为预防信号流中的均匀瞬时间隙，它能够导致语音识别引擎中的严重错误。在编码器中可以缓冲数字化的语音，并且该编码器可以适应输出数据速率，以便最优地使用可用的带宽。在编码器和语音识别引擎之间使用有限频带的无线链接时，使用自适应语音编码器尤其有利。

在替代实施例中，阵列话筒可以由一组物理上分布的话筒或用户佩戴的耳麦取代。分布的话筒可以放置在房间的不同部分或在不同房间中。分布的话筒能够创建三维阵列以改进信噪比。耳麦可以使用无线或有线连接。

虽然本发明力图能够接受几乎所有自然语言的问题或命令，但是歧义性仍是难题。为了帮助用户表达简洁的问题和命令，系统 90 可以支持语音查询语言。为了允许歧义性最小的多种查询而构造了所述语言。因此，语音查询语言帮助用户连同参数和准则一起清楚地指定问题或命令的关键字或上下文。所述语言提供的语法清楚地规定了确定上下文所用的关键字和一组准则或参数。用户以语音查询语言提出问题或陈述命令时几乎总保证能够收到应答。

语音查询语言对上下文堆栈的内容敏感，其中上下文定义了在会话期间可能是活动的或不活动的一组问题。根据一个实施例，每个代理都可以将一个上下文指定为根上下文，它定义了该代理实施的基本算法。因此，使用简化语法能够提出后续问题，因为从上下文堆栈能够继承关键字和准则。例如，如果问题的准则保持不变，用户可以简单地提出另一个关键字。

系统 90 可以提供若干内建的训练功能，以便帮助用户学习表达其问题和命令的最佳方法。互动训练允许用户可听地或可视地了解机器对其查询的解释，并且提供如何更好构建查询的建议。使用互动训练用户可以快速地变得适应于语音查询语言，同时学到如何使对话中每个步骤所需要的信息量最优化。

语音编码器 122 的输出可以提供给语音识别引擎 120。语音识别

引擎 120 使用上下文描述语法 112 中的信息识别出词和短语，并且将这些结果传递到语法分析器 118 进行解释。语音识别引擎 120 可以通过每个口头表达的语音和名字确定用户的身份。识别出的词和短语可以在全部进一步处理中以这个身份标记。因此，当多名用户处于交迭会话中时，由语音识别引擎 120 向每个口头表达添加的标记将允许系统 90 的其他组件将这个口头表达结合到正确的用户和对话。用户识别功能可以进一步用作若干应用的安全措施，比如需要安全措施的拍卖或在线购物。每名用户的语音特征可以包含在用户概况 110 中。

根据本发明的一个实施例，当用户初次与系统 90 说话时，他们便开始了与其的会话。这能够由讲出普通单词（“计算机”）或陈述出具体名字（“Fred”）完成，通常可以将其结合到系统的个性 108。一旦用户开始了对话，它就能够被语音识别引擎 120 使用用户语音的若干独特特征识别出。会话结束或中断会话时，用户可以说出撤退的词（“再见”）。

根据本发明的另一个实施例，系统 90 采用的语音识别引擎 124 可以使用来自上下文描述语法 112、用户概况 110、代理 106 以及其他组件的数据，改进词识别准确度。与此同时，这些词在上下文描述语法中的模糊集可能性即先验概率可以动态更新，以便使每个会话阶段正确识别的概率最大。所述概率或可能性可以根据许多准则动态更新，包括应用领域、问题或命令、上下文、用户概况和偏爱、用户对话历史、识别器字典和短语表、词拼写以及其他准则。

对于不常见词或词汇表中新词，可以给用户拼写该词的选择。通过讲出名称或字母，或者使用音标可以完成拼写。音标可以是默认的或用户选择的。

作为替代，如果语音识别引擎 120 根本未识别出或者未正确认别出用户提交的词，那么可以请求用户拼写该词。语音识别引擎 120 基于评分过程的置信度确定这种条件。该词可以在字典中查找，并且有该词的发音或者添加到字典、代理 106 中或者添加到用户的概况 110 中。然后该词的发音可以与领域、问题、上下文和用户相关联。经过

这种过程，语音识别引擎随时学习并改进了准确度。为了帮助用户拼写词，可以使用个性化的音标。每位用户都能够用词修改标准的音标，他们可以更容易地记住它。

一旦语音识别引擎 120 已经识别出了若干词和短语，就会向语法分析器 118 传递标记和用户身份。语法分析器 118 检查用于问题或命令、上下文和准则的标记。语法分析器 118 通过对关键字匹配、用户概况 110 和对话历史应用先验概率或模糊可能性，确定口头表达的上下文。问题或命令的上下文决定领域，进而决定要调用的领域代理 156，如果有的话。例如，带有关键字“温度”的问题暗示着该问题的上下文值天气。语法分析器从系统代理 150 或已经活动的领域代理 156 中动态地接收关键字和相关联的先验概率或模糊可能性更新。根据这些概率即可能性，对这些可能的上下文评分，最高分的一个或几个用于进一步处理。

对于用户的问题和/或命令，语法分析器 118 使用评分系统确定最可能的上下文或领域。该得分通过对许多因子的加权确定，所述因子包括用户概况 110、领域代理的数据内容以及先前上下文。基于这种评分，系统 90 调用正确的代理。如果所述分数的置信度不足以确保可靠的应答，系统 90 可以要求用户验证是否正确地理解了问题和/或命令。

一般来说，由系统 90 询问的问题可以用短语表达以指明问题的上下文，包括全部准则或参数。例如，问题的形式可以是“我是否理解了你想要某物”。如果用户确认该问题是正确的，系统继续产生应答。否则，用户可以将原始问题重新构词，也许添加另外的信息以消除歧义性，系统也可以询问一个或多个问题以尝试解决歧义性。

一旦问题或命令的上下文已经确定，语法分析器 118 就可以调用正确的代理 156、150。为了以代理所用的规则语法表达问题或命令，语法分析器 118 对若干准则或参数可以确定所需的和可选的数值。这些准则可能已经由用户显式地提供了，也可能需要进行推理。语法分析器 118 可以利用系统代理 150 提供的准则处理器 152。准则处理器

152 可以提供上下文敏感的过程，从用户的问题或命令中提取若干准则或参数。某些准则由执行所述代理中的若干算法而确定，而其他准则可以由对可能数值的若干表应用模糊推理概率而确定。先验概率或模糊可能性以及相关联数值从许多数据源收到，包括对话历史、用户概况 110 和代理。根据用户的应答，先验概率或模糊可能性随着系统学习所期望的行为而更新。对于天气上下文，准则实例包括位置、日期和时间。其他准则可以包括命令准则（即是/不是、接通/关闭、暂停、停止）以及拼写。特殊准则处理器可以来自系统代理，用于处理列表、表格、插嘴命令、文本长串和系统命令。

准则处理器 152 对提取的准则进行迭代或递归操作以消除歧义性。这种处理有助于减少用户问题或命令中的歧义性。例如，假若用户在其口头表达中有地名（或者其他专有名词），语法分析器 118 就能够使用领域代理 156 的服务在数据库 102 的表中查找地名，也能够尝试确定哪个词是来自口头表达句法的专有名词。在另一个实例中，用户询问“航班一百二十又如何？” 。语法分析器和领域代理连同上下文使用数据库和网络信息中的航班信息在以下诸项中确定看似最可信的解释：航班 100 还有航班 20、航班 100 和航班 22、航班 122 等。

一旦确定了上下文和准则，语法分析器 118 就可以以代理 150、156 进行处理所用的标准格式或分层数据结构形成所述问题或命令。语法分析器 118 可以为上下文的语法填写全部需要的和某些可选的标记。这些标记往往必须转录为代理可接受的数值和形式。语法分析器从代理、对话历史或用户概况 110 获得所需要的变换。由语法分析器对标记执行的变换或替换的实例包括：

- 用股票符号代替公司名称或缩写；
- 用数值代替某个词或若干词；
- 将邮政编码添加到地址；以及
- 将地方或其他名称改变为公用的标准缩写。

只要语法分析器 118 已经将命令或问题置于所需的标准格式中，代理 150、156 就可以接收它。基于所述上下文，语法分析器 118 调用

正确的代理处理问题或命令。

命令可以导向系统 90 或外部实体。系统命令一般导向系统代理 150。用于外部实体的命令一般由领域代理 156 处理，它包括命令上下文和外部实体的行为。

具体问题一般导向领域代理 156 之一。基于问题或上下文以及参数或准则，领域代理创建对一个或多个本地或外部信息源的一个或多个查询。问题在性质上可以是客观的也可以是主观的。客观问题的结果常常可以利用对一个或多个本地或网络信息源进行结构化查询而得到。即使对客观问题，系统 90 也可能需要应用概率或模糊集分析，以便应对信息冲突或信息不完整的情况。回答主观问题的信息一般利用对本地或网络数据源进行一个或多个特别查询而得到，继之以对一个或多个结果进行概率或模糊集评估以确定最佳答案。

一旦领域代理 156 已经表达了一个或多个查询，就可以将其发送给本地和/或网络信息源。针对信息源以不同速度响应或根本未能响应的事实，以异步方式执行查询。完全相同的查询被发送到不同的信息源以确保至少一个信息源及时地以有用结果响应。此外，如果以及时方式收到了多个结果，系统可以对它们评分以确定哪个数据最可靠或最适宜。所接纳的数据源实例包括：HTTP 数据源、具有多种格式包括 XML 元数据的数据源、娱乐音频、视频和游戏文件包括 MP3、使用查询语言和结构化响应比如 SQL 的数据库，以及其他数据源。

本地信息源可以存储在一个或多个系统数据库 102 中，也可以在任何本地数据存储器上，比如播放机中的一组 CD 或 DVD 或者其他本地数据存储器。网络信息源可以连接到因特网 136 或其他网络，并且可以通过网络接口 116 中称为可插入源的一系列插卡或适配器进行访问。网络接口 116 中的可插入源能够执行若干协议并解释所关注数据源的数据格式。可插入源可以提供信息、将每个数据源的数据和过程分解到领域代理 156。如果要使用新类型的数据源，可以向网络接口 116 添加新的插卡或适配器。

领域代理 156 在一个或多个查询的结果到达时评估它们。领域代

理 156 根据已经收到的结果、上下文、准则、对话历史、用户概况 110 和领域专用信息，使用概率或模糊评分技术的对这些结果的相关性评分。部分对话历史被保存在上下文堆栈中。评分时每个上下文权重基于一个上下文与另一个上下文的相关性，以及这些上下文的已存在时间。其他评分变量可以通过上下文堆栈相关联。上下文也可以是唯一的，所以先前的上下文在评分中没有权重。

根据连续的评分过程，领域代理 156 可以判断是否可以提取单一的最佳答案。对于大多数问题，所期望的结果具有要表达答案时必须找到的一组标记。只要为这些标记的每一个都找到了的数值，向用户呈现的结果就准备好了。例如，对于天气问题，所述标记可以包括日期、星期几、预报高温、预备低温、降水的可能性、预期云量、预期降水类型以及其他标记。以这种方式处理的结果包括错误消息。对于主观问题，在由所收到结果进行匹配而提取的答案中，通过确定最可能答案或若干答案而做出这种决定。如果从查询结果不能推理出令人满意的答案，代理可以实行以下步骤之一：

1. 询问用户更多信息，典型情况下通过语音接口，并且基于得到的结果表达新的查询。如果在表达应答时出现了不能解决的歧义性则应用这种方式。
2. 基于从最初查询组收到的结果表达新的查询，典型情况下，在收到的应答不包含全部所需信息时应用这种方式。从已经得到的结果（即在 HTML 文档中的链接）或从其他信息源中可以推理出要查询的信息源。使用这种方式时一组或多组查询和应答能够环环相扣而不需要用户动作。
3. 等待另外的查询返回结果。

在任何情况下，领域代理 156 都可以继续进行请求并评估结果，直到构造出令人满意的应答。这样做时，代理可以起动几个重叠查询通道即询问线程，典型情况下由事件管理器 100 仲裁。这种技术与使用从多个数据源异步查询结合，提供了与用户自然互动所需要的实时

应答性能。

领域代理 156 可以在收到每个查询应答时对其实施条件分解操作。条件分解动作依赖于上下文、准则、用户概况 110 和领域代理编码和数据。对于要提取的每个标记，都可以使用系统代理 150 的若干服务创建分解准则 152。所述分解准则使用格式专用的分解方法，包括：表格、列表、文本和其他分解方法。一种或多种分解准则可以应用于页面或结果组。一旦收到了另外的结果，领域代理 156 就可以创建新的分解准则以应用于已经获取的结果。条件分解过程删除了无关的信息，比如图形，不必对其做进一步处理或存储，改进了系统性能。

领域代理 156 可以强烈地影响如何呈现结果。例如，一旦领域代理 156 对问题已经创建了令人满意的应答，代理 156 就可以对该应答进行格式编排以便呈现。典型情况下，领域代理 156 将该应答格式编排为文本到语音引擎 124 所用的置标格式。领域代理 156 还可以使用可用的格式模版并基于上下文、准则和用户概况 110 对结果表达进行格式编排。代理 156 可以执行变量替换和变换以产生用户最容易理解和最自然的应答。此外，代理 156 可以改变标记的表达次序和所用的提取技术以创建对用户更自然的应答。领域代理 156 也可以选择要使用的表达个性 108。

领域代理 156 可以选择表达模版、确定标记的表达次序和使用概率或模糊集决策方法确定变量替换和变换。形成表达所用的模版可以来自领域代理自身也可以来自用户概况 110。用户概况 110 完全能够指定表达格式，也能够用于选择、然后修改现有的表达格式。表达模版的选择和格式编排还可以依赖于表达个性模块 108。与此同时，使用从上下文、准则、领域代理自身和用户概况 110 导出的概率或模糊可能性动态地确定用于应答的个性特征。

领域代理 156 可以先对标记实施许多变换，再呈现给用户。这些变量替换和变换从许多信息源导出，包括由代理携带的领域信息、上下文、标记值、准则、要使用的个性模块 108 和用户概况 110。

变量替换和变换的实例包括：

- 用单词代替数字；
- 用名称代替首字母缩写词或符号（即交易符号）；
- 使用从信息源（即 HTML 标签）导出的格式编排信息；
- 响应的性质，包括文本、长文本、列表、表格；
- 可能的丢失信息或错误；
- 度量单位（即英制的或公制的）；以及
- 来自用户概况 110 或表达个性 108 的优选术语。

根据本发明的另一个实施例，系统 90 可以提供特殊目的的表达能力，用于长文本字符串、表格、列表和其他大结果集。领域代理 156 可以对这样的结果使用特殊的格式编排模版。系统代理 150 可以提供特殊的准则处理器 152，用于大结果集的表达和用户命令。由领域代理 156 对大结果集所用的表达模板典型情况下包括若干方法，用于归纳若干结果然后允许用户更详细地查询结果。例如，最初仅仅呈现了简短的摘要，比如标题行或关键数字。然后用户能够进一步查询结果集。准则处理器 152 为用户提供了浏览大结果集的能力。准则处理器 152 为大结果集提供的命令包括：停止、暂停、跳过、反绕、开始以及向前移动。

某些信息，格式为比如视频、图片和图形，以显示格式可以表现得最好。在这些情况下领域代理 156 可以应用合适的表达模版，并且通过图形用户界面 114 呈现该信息。为了显示表达和控制，系统代理 150 提供的特殊准则处理器 152 用于表达和用户命令。

图 3 展示了根据本发明一个实施例的高级别过程 300，用于接收基于自然语言语音的查询和/或命令并产生应答。过程 300 包括若干单独的步骤，其中的几步可以重复，以容纳部分故障。当在 302 收到用户的口头表达并处理时，过程 300 便开始。用户的口头表达可以是查询和/或命令并可以是基于自然语言语音的形式。收到了该自然语言查询和/或命令后，可以处理该查询和/或命令，使得该数据为机器更容易理解的格式，比如数字格式。只要该口头表达已经被转录为可理解

的格式，就可以处理该口头表达，使得在 304 至少能够确定该口头表达的大致意义。例如利用存储的字典、短语、用户概况、领域代理数据以及其他数据可以完成这个步骤。根据口头表达的大致意义，可以产生一个或多个正确地格式编排的查询和/或命令。

用户创建的查询和/或命令可能要求为了产生所期望的应答或动作可以需要多条请求和/或命令。此外，取决于其中已经提交了查询和/或命令的上下文（如用户是谁、查询和/或命令指向的主题是什么、查询和/或命令指向什么时间或什么地点，以及定义查询和/或命令时所使的其他参数），查询和/或命令可能需要特定格式。例如，领域代理 156 的每一个都可以要求以特定方式对请求和/或命令进行格式编排。因此在 306，可以产生一条或多条适当格式编排的查询和/或命令。操作 304 和 306 可以重复以便得到用户口头表达的正确解释，并且产生所期望的应答和/或动作。一旦产生了适当格式编排的查询和/或命令，就可以在 308 采取适合的动作。这可能要求将适当格式编排的查询和/或命令发送到能够满足所述查询和/或命令要求的特定领域代理、信息源、设备或者其他适当目的地。一旦适当的动作已经执行，就可以在 310 把这个事件记录到例如用户概况、数据库和/或一个或多个代理中。这样的数据对于未来的用户询问和命令可能有用。在动作已经执行之后，在 312 如果需要可以产生应答并转发到用户和/或第三方。

在为了检索数据而查询的情况下，应答会包含所请求的信息。在命令的情况下，应答可以是已经执行了特定动作的确认。应答的形式可以是自然语言格式。应答还可以格式编排为反映应答的具体个性或声调，以便使应答更加“人性化”。应答可以作为音频消息和/或用户界面上显示的可视消息传递给用户和/或第三方。

图 4A 是根据本发明另一个实施例的过程 400，使用系统 90 接收基于自然语言语音的查询和/或命令并产生应答。在 402 通过在多种真实世界环境中运行的语音识别捕获了用户口头表达（即用户查询和/或命令）时，过程 400 便开始。一旦捕获了口头表达，就在 404 对其

进行语法分析和解释以确定在口头表达中包含的查询和/或命令。下一步，在 406 调用适宜的资源包括例如若干代理评审该查询和/或命令以确定所需要的专长领域和查询的上下文。用户提交的原始查询和/或命令可能需要产生多条查询和/或命令。

例如，假设用户对检索她股票资产组合的数值有兴趣。用户会说出“请获得我的股票资产组合数值”。系统 90 可以连同存储的数据比如用户概况评审这个请求，并且确定关键字比如“获得数值”和“我的股票资产组合”。然后系统 90 可以产生查询以确定用户资产组合中的股票、股票数量和用于当前价格信息的信息源。这些查询然后可以发送到一个或多个代理领域，比如可以访问其中包含用户概况的数据库的领域和访问股票价格数据源的领域，以确定这些问题答案。

为了得到所期望的数据，可以向这些领域代理或数据源发送查询。因此，在操作 408，可以表达一个或多个查询并发送到一个或多个本地和/或网络数据源，并且/或者将适宜的命令发送到本地或远程设备或系统自身。这些查询然后被发送到所指定的代理。这些代理然后又产生它们自己的查询和/或命令，发送到例如本地或远程信息源以检索所需数据。在 410 根据目标源的需求可以表达代理所产生的查询和/或命令，并且执行变量替换和变换，将查询的形式修改为最可能从这些可用信息源中产生所期望的结果。一旦正确地表达了这些查询，在 412 就可以以异步的方式执行它们并适度地处理故障。作为执行查询的结果，这些结果可以由领域代理和/或数据源返回。然后在 414 系统 90 可以从一个或多个结果中提取或分解所期望的信息，这些结果可能以许多不同格式的任何一种返回。也就是说，用户寻求的结果可以是汇总即从例如几个信息源所得到的信息进一步处理的结果。

下一步，在 416 可以评估和解释这些结果，包括处理错误，并且将它们汇集和结合为被判定“最佳的”的单一最佳结果，即使这些结果是歧义的、不完整的或矛盾的。一旦确定了最佳结果，就进行所有需要的格式编排。在操作 418，可以使用变量替换和变换修改结果。最后，在操作 420，可以通过文本到语音引擎 124 以有用和预期的方

式向用户呈现该复合的结果。在执行过程 400 的同时可以考虑到可用的所需专长领域、在其中呈现问题或命令的上下文、可用的领域专用信息、用户互动历史、用户偏爱、信息源或命令以及从这些信息源得到的应答。

在过程 400 的每个阶段，概率或模糊集决策与匹配的方法都可以应用于处理不一致、歧义、冲突和不完整的信息或响应。另外，使用可能导致某些查询或命令出现快速和适度故障的异步查询，可以准许系统 90 稳健地快速返回结果并且方式对用户看起来自然。

图 4B 展示了根据本发明另一个实施例的过程 450，用于接收基于自然语言语音的命令，以便在本地或远程地控制系统 90 或其他设备的功能。在 452 通过在多种真实世界环境中运行的语音识别捕获了用户口头表达（即用户查询和/或命令）时，过程 450 便开始。一旦捕获了口头表达，就在操作 454 对其进行语法分析和解释以确定在口头表达中包含的命令。

下一步，在操作 456 调用包括若干代理的适宜的资源评审该命令以确定领域和上下文。用户提交的原始命令将往往需要产生多条查询和/或命令。例如，假设用户对录制他喜爱的电视节目有兴趣。用户会说出“请录制我喜爱的 TV 节目”。系统 90 可以连同存储的数据比如用户概况评审这个请求，并且确定关键字比如“录制”和“我喜爱的 TV 节目”。然后系统 90 可以产生查询以确定用户喜爱 TV 节目的名称、频道和时间。这些查询然后可以发送到代理领域，它可以访问包含用户概况的数据库并确定这些问题的答案。然后可以向录像机发送命令，命令该录像机录制所选定的 TV 节目。因此，在操作 458，可以表达一个或多个查询并将其发送到一个或多个本地和/或网络数据源。将适宜的命令发送到本地或远程设备或系统自身。下一步，在操作 460，可以将产生的命令路由到适合的系统和/或外部设备。只要已经执行了这些命令，在 462 就可以接收包括错误的结果并处理。在操作 464 可以有选择地向用户呈现执行命令的结果。

图 4A 描述的某些操作可以使用能够重复的多个步骤执行。例如

在操作 404 中，为了判断用户的口头表达的意义，对口头表达进行了语法分析和解释。系统 90 可以根据例如上下文描述语法模块 112、用户概况 110、代理 106 和数据库 104 中存储的数据，对口头表达进行初始解释。为了产生领域或上下文的得分，对解释结果可以应用实时的评分系统或其他技术。如果所述领域或上下文得分的置信度不足以确保可靠的应答，系统 90 可以请求用户验证是否正确地理解了问题或命令。一般来说，问题可以用短语表达以指明问题的上下文，包括全部准则或参数。如果用户确认该问题是正确的，系统继续产生应答。否则，或者用户可以将原始问题重新构词，也许添加另外的信息以消除歧义性，或者系统可以询问一个或多个问题以尝试解决歧义性或可以采取其他行动。

图 5 展示了根据本发明一个实施例的过程 500，用于正确地解释用户的口头表达。过程 500 一般地表示了过程 400 的操作 402 和 404。一开始在操作 502 将用户的口头表达输入到系统 90 中。然后使用例如多种存储的数据比如用户概况、代理数据、字典和短语以及其他相关数据对输入进行解释。然后在 506 使用例如先前介绍的加权评分系统对解释评分。一旦获得了得分，在 508 就可以分析解释的置信度以判断它是否令人满意。换言之，对解释给定的得分是否超过了某个数值做出判断。如果判定置信度不令人满意，那么在 510 和 512 就可以向用户提交请求，请求用户验证该解释。如果用户对该解释不满意，在 514 可以要求他/她重新用短语表示该口头表达和/或提供另外的信息。一旦用户提供了重新用短语表达的口头表达和/或另外的信息，过程 500 就返回到操作 502 的起点。相反，如果用户对该解释满意，那么用于接收基于自然语言语音的查询和/或命令并产生应答的过程 400 就可以在操作 516 继续。一旦已经确定了该口头表达的意义，就可以为适当的领域代理恰当地进行代理以及查询和/或命令格式编排。

图 6 展示的过程 600 用于确定要调用的适当领域代理以及将要提交到代理 106 的查询和/或命令的适当格式，正如在图 4A 的步骤 406 至步骤 408 中一般描述过的。为了以代理 106 所用的规则语法表达问

题或命令，对于问题和/或命令的准则或参数的必需和可选值可以做出判断。用户可能已经显式地提供了这些准则，也可能需要进行推理。因此在操作 602，对于例如由图 4A 的操作 404 和图 4B 的操作 454 所产生的恰当解释的查询和/或命令进行语法分析。

然后在 604 分析已恰当解释的查询和/或命令的内容以确定已恰当解释的查询和/或命令的准则和/或参数。为了从已恰当解释的查询和/或命令中提取准则或参数，可以使用上下文敏感的过程。某些准则通过执行代理中的算法确定，而其他的却可以通过对可能值的若干表应用概率或模糊推理确定。先验概率或模糊可能性以及相关联的数值从许多数据源收到，包括对话历史、用户概况 110 和代理。

基于用户的响应，随着系统学习所期望的行为而更新先验概率或模糊可能性。对于天气上下文，准则实例包括位置、日期和时间。其他准则可以包括命令准则（即，是/否、接通/关闭、暂停、停止）以及拼写。确定准则的过程可以是迭代的或递归的以便消除用户问题或命令中的歧义性。例如，假若用户在其口头表达中有地名（或者其他专有名词），就可以在数据库 102 的若干表中查找地名，或者对从口头表达句法确定哪个单词是专有名词进行尝试。在另一个实例中，用户询问“航班一百二十又如何？”在这种情况下，数据库和网络信息中的航班信息连同上下文可以用于在以下诸项中确定看似最可信的解释：航班 100 还有航班 20、航班 100 和航班 22、航班 122 等。只要已经建立了用于查询和/或命令的参数和准则，在 606 就可以选择适当的代理 106。

在选定了适当的代理 106 后，在操作 608 可以确定将提交给代理 106 的恰当地表达的查询和/或命令。查询和/或命令可以是标准的格式也可以是代理 106 进行处理所用的分层数据结构。为了向代理 106 提交恰当地表达的查询和/或命令，上下文语法的全部必需的和某些可选的标记都可以填入。这些标记往往必须转录为代理可接受的数值和形式。可以从代理、对话历史或用户概况 110 中获得所需要的变换。以上提供了可以进行的变换或替换的实例。一旦产生了恰当地表达的查

询和/或命令，在 610 该过程（如过程 400）就可以继续。

虽然以上已经介绍了本发明的若干特定实施例，但是应当认识到，可以以不同于介绍的方式实践本发明。例如，本发明可以采取计算机程序的形式，包含机器可读指令的一个或多个序列，描述以上公开的方法，也可以采取数据存储介质（如半导体存储器、磁盘或光盘）的形式，使这样的计算机程序存储其中。

以上说明旨在展示而非限制。因此，对于本领域的技术人员显而易见，对所介绍的本发明可以做出若干修改而不脱离以下阐述的权利要求书的范围。

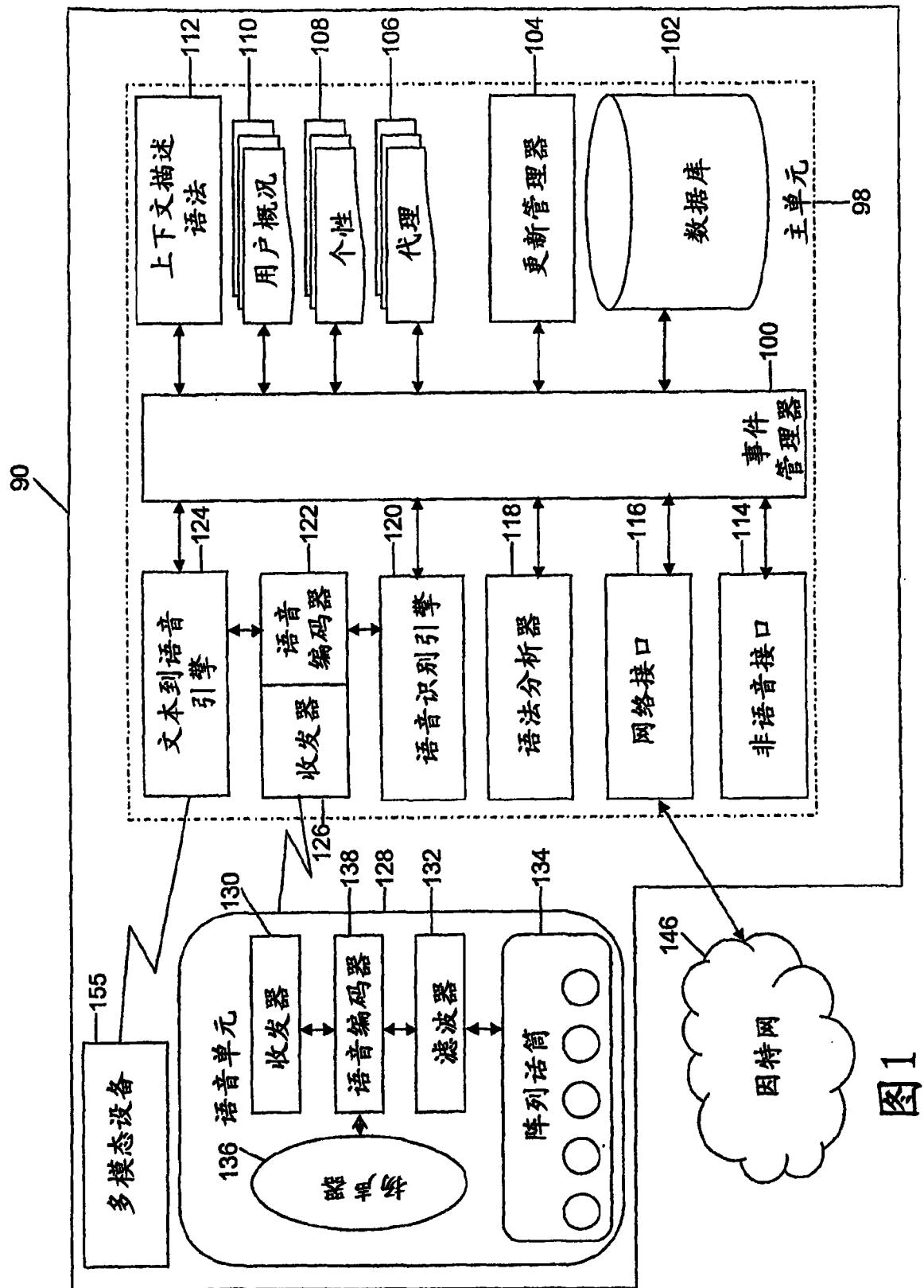
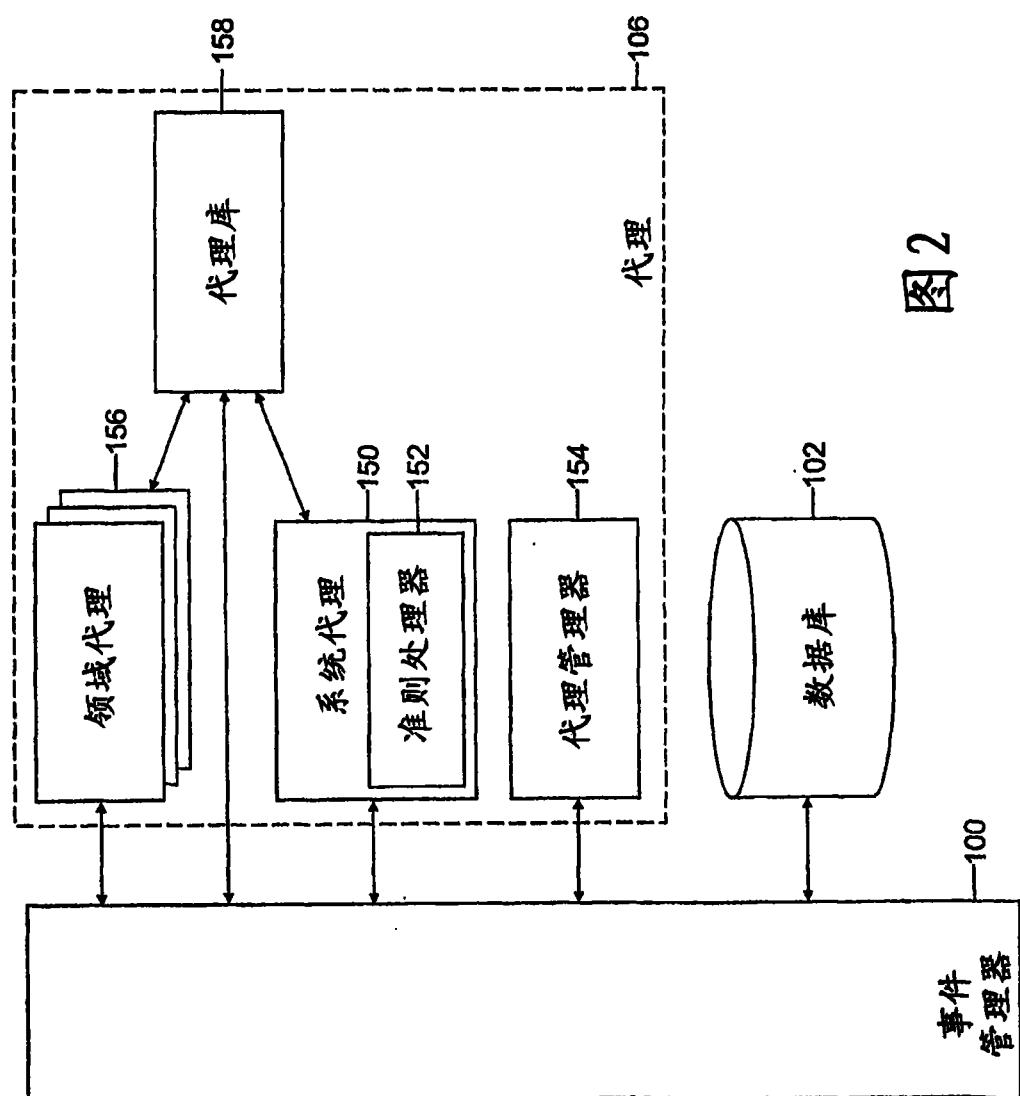


图 1



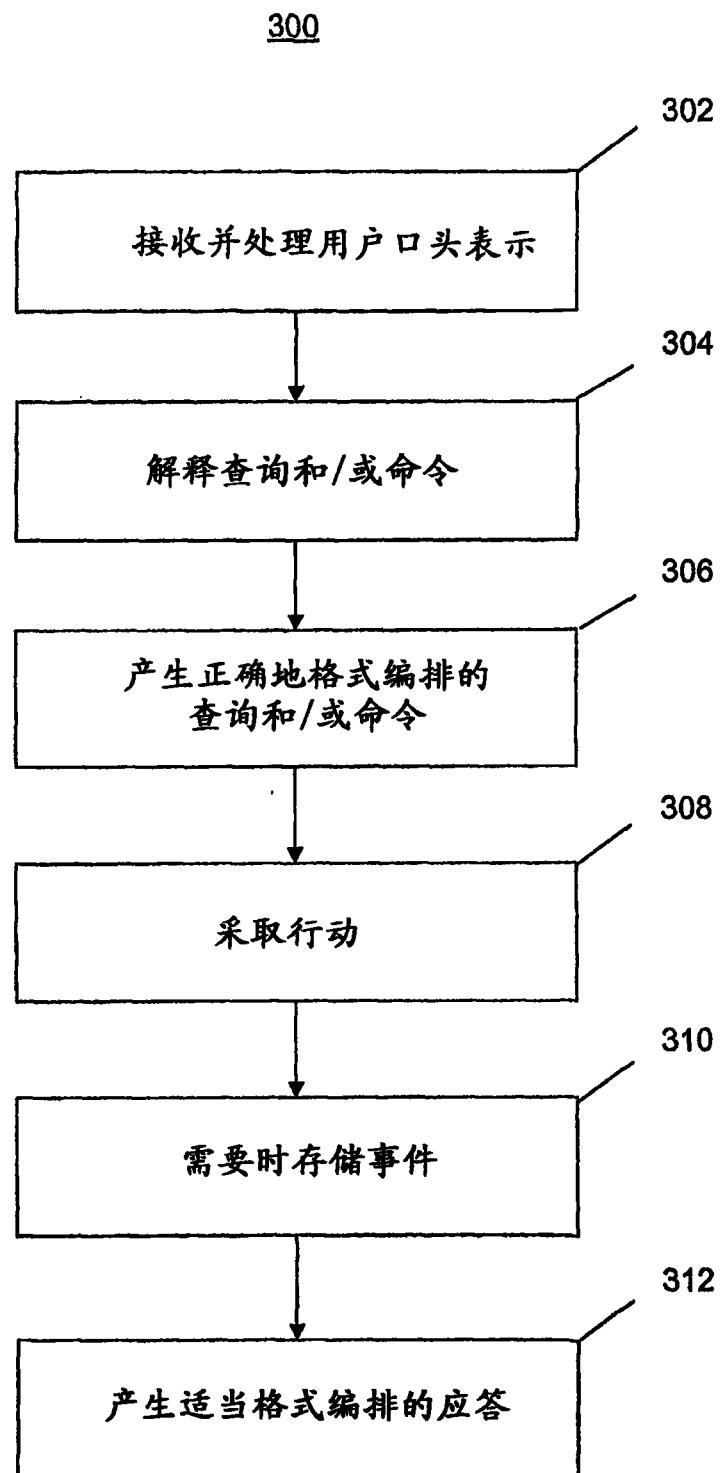


图 3

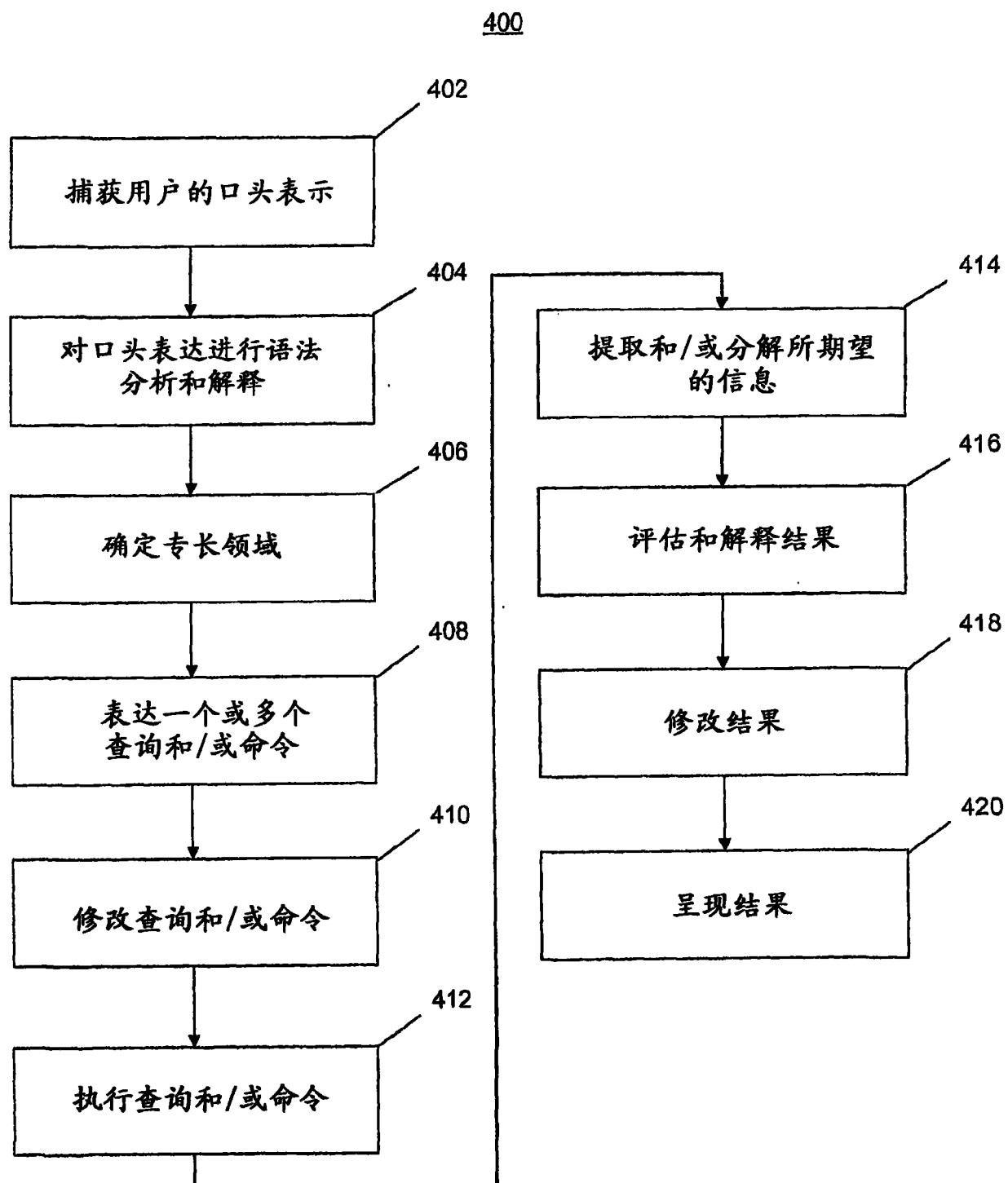


图 4A

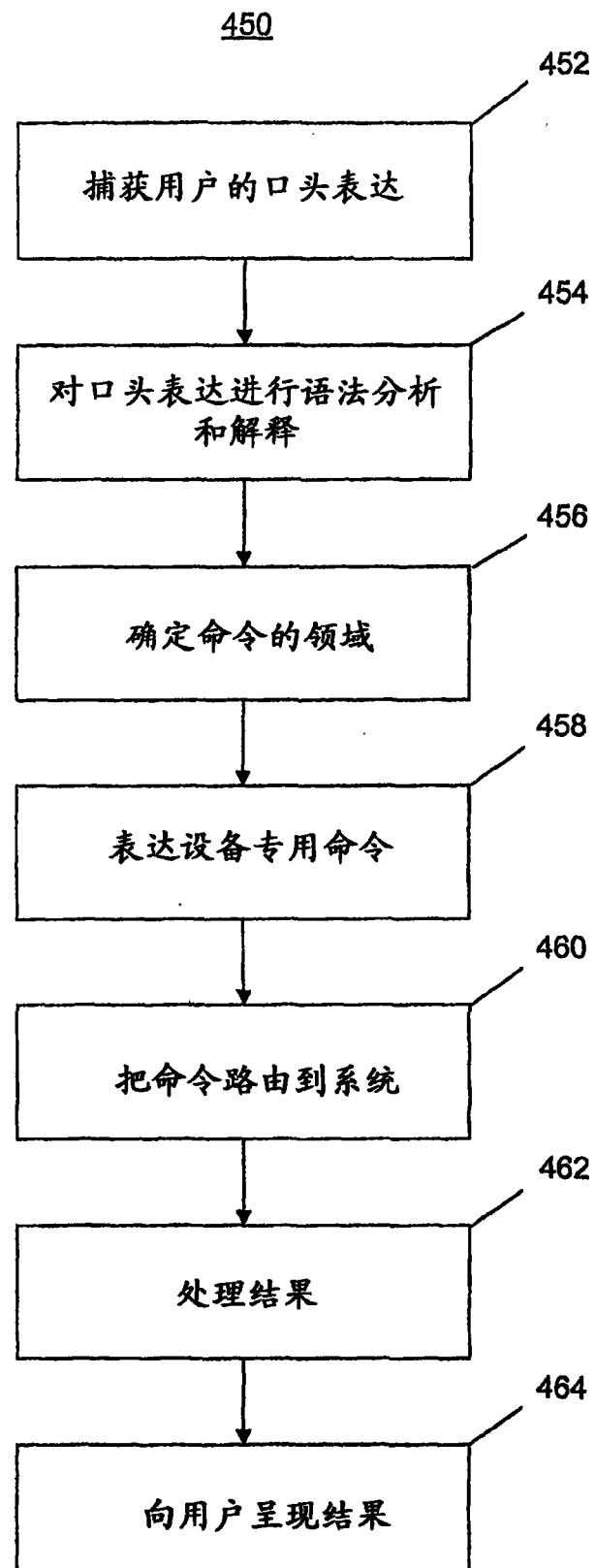


图 4B

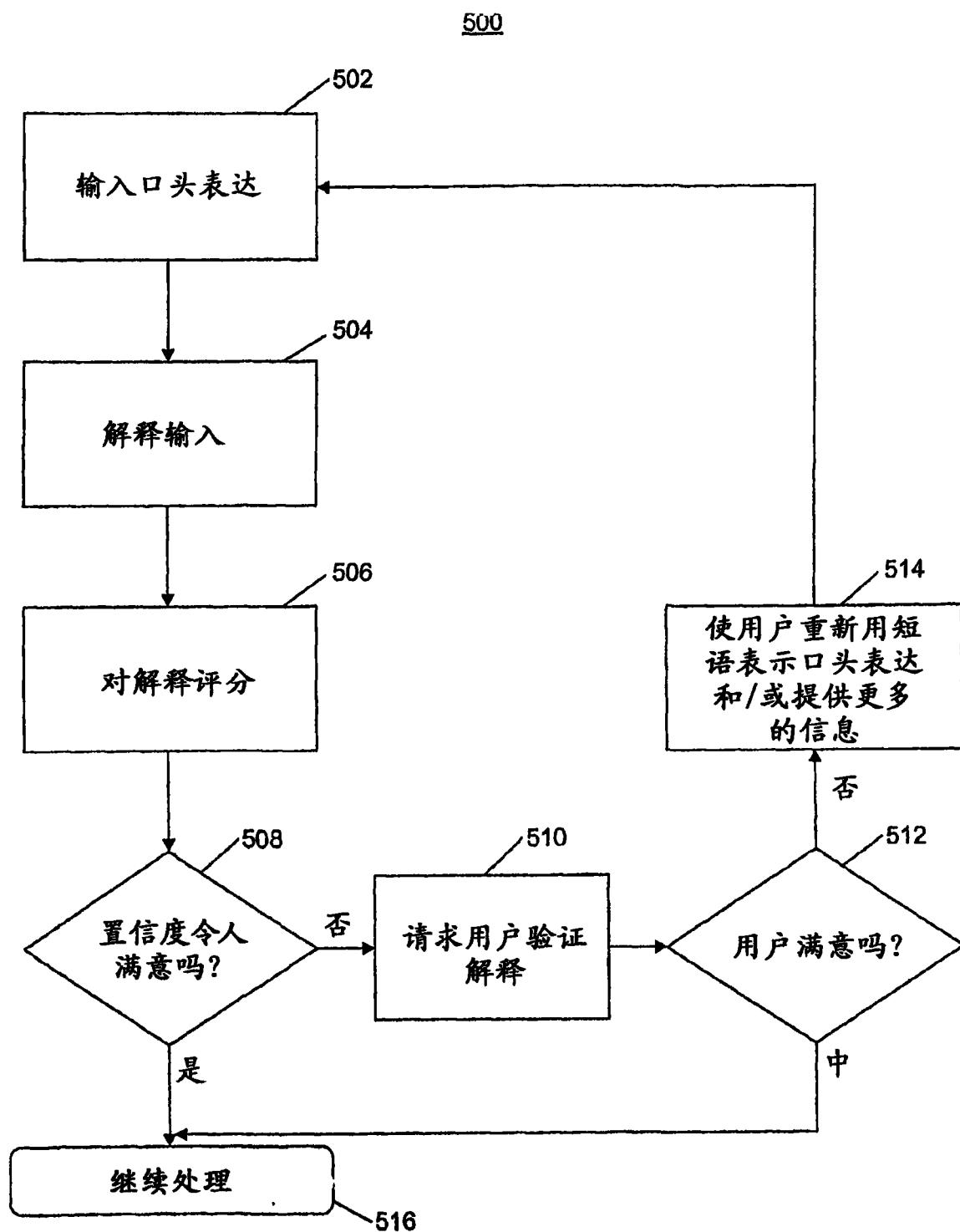


图 5

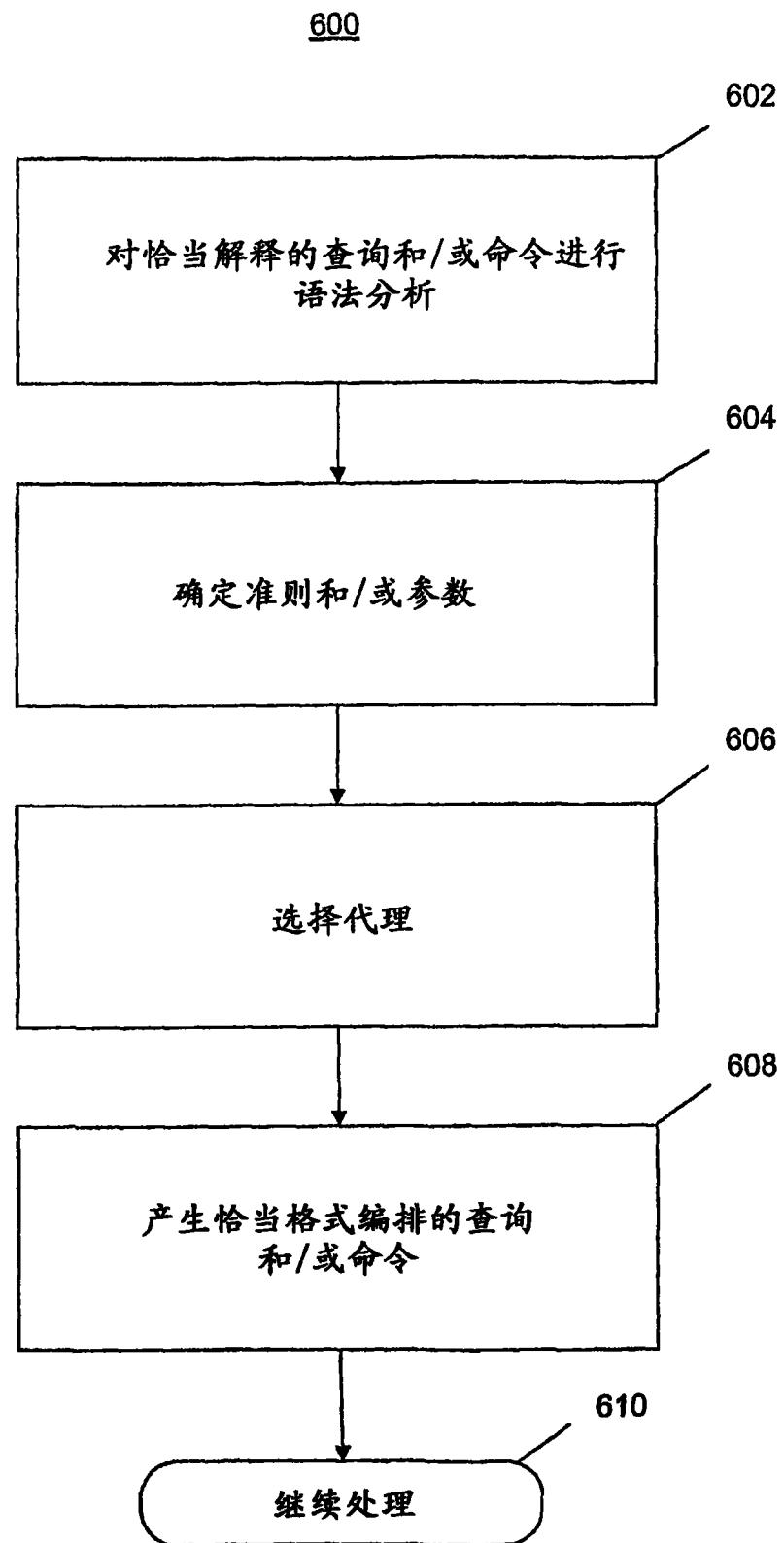


图 6

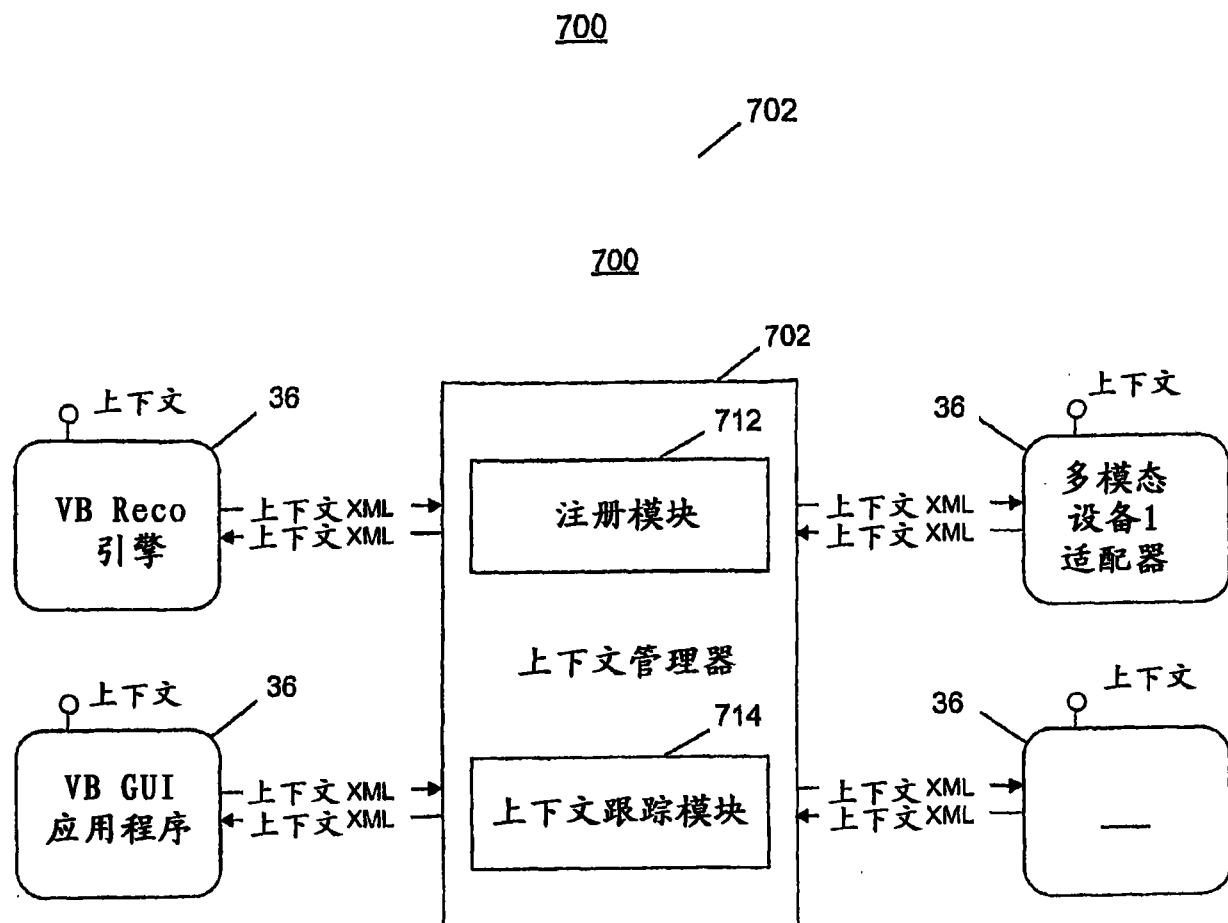


图 7