

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G10L 15/18 (2006.01)

G10L 15/04 (2006.01)

G10L 21/00 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200680019481.4

[43] 公开日 2008年7月30日

[11] 公开号 CN 101233559A

[22] 申请日 2006.6.27

[21] 申请号 200680019481.4

[30] 优先权

[32] 2005.6.27 [33] US [31] 11/167,414

[86] 国际申请 PCT/US2006/024848 2006.6.27

[87] 国际公布 WO2007/002606 英 2007.1.4

[85] 进入国家阶段日期 2007.11.30

[71] 申请人 微软公司

地址 美国华盛顿州

[72] 发明人 E·J·霍维茨

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 陈斌

权利要求书 2 页 说明书 17 页 附图 18 页

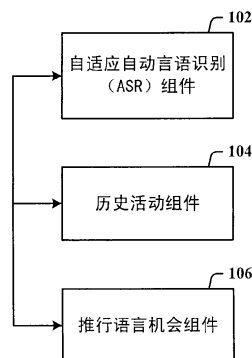
[54] 发明名称

用于不同语言的说话者之间的增强的交互和理解
的上下文敏感通信和翻译方法

[57] 摘要

一种与不同语言的用户交互以提高言语翻译的架构。集中所识别的思想，并以对用户与讲另一种语言的其他人的通信提供简化推断的设备对混合发起的用户交互消除歧义。基于当前注意力焦点或可能的注意力焦点的概率分布上的思想来实施关于通信目的的推理，并且向用户的交谈对方提供筛选选择以及供感知的图像、文本和/或言语翻译。推断将用户输入处理为推理出思想、情形、目标和/或对其消除歧义时的证据的一部分。可将系统对通信的理解返回给用户以便确认。可提供识别和信息收集组件的上下文敏感聚焦，它可使用从先前或当前用户话语识别出的单词来进一步聚焦推断。

100



1. 一种有助于言语翻译的系统，包括：

言语识别组件，处理所感测的当前上下文的数据并基于所感测的数据有助于言语识别过程；

历史活动组件，存储与所述言语识别过程相关联的历史数据；以及

语言机会组件，通过向用户推行一个或多个词项的训练会话来改进所述言语识别过程，当在所述言语识别过程期间使用所述一个或多个词项时，所述训练增加了成功的可能性。

2. 如权利要求 1 所述的系统，其特征在于，还包括感测系统，所述感测系统包括话筒、图像捕捉子系统和定位子系统的至少其中之一，并且输出表示这些的所感测的数据。

3. 如权利要求 1 所述的系统，其特征在于，还包括向所述用户推行解决歧义性的请求的机会组件。

4. 如权利要求 1 所述的系统，其特征在于，还包括被创建成识别所述用户的言语的语言模型。

5. 如权利要求 4 所述的系统，其特征在于，还包括基于所述语言模型来理解所述用户的音位模块。

6. 如权利要求 1 所述的系统，其特征在于，还包括用于训练用户如何讲外语的模拟器组件。

7. 如权利要求 6 所述的系统，其特征在于，所述模拟器组件用用户语言提示所述用户，并且输出相应的外语翻译。

8. 如权利要求 1 所述的系统，其特征在于，所述历史活动组件包括用户数据的集中式数据存储和本地用户数据的本地化数据存储的至少其中之一。

9. 如权利要求 1 所述的系统，其特征在于，所述言语识别组件是自动且自适应的。

10. 如权利要求 1 所述的系统，其特征在于，还包括机器学习和推理组件，所述组件使用基于概率和/或统计的分析来预测或推断用户期望自动执行的动作。

11. 一种计算机可读介质，其上存储了用于实现权利要求 1 所述的系统的计算机可执行指令。

12. 一种便携式计算设备，它使用权利要求 1 所述的系统。

13. 一种服务器，它使用权利要求 1 所述的系统。

14. 一种有助于在不同语言的用户之间进行言语翻译的计算机实现方法，包括：

在言语识别过程期间，接收用户的言语信号；

基于对所述言语信号的分析，计算对用户思想和用户上下文的至少其中之一的推断；

根据所述推断修改所述言语识别过程；

与所述用户交互以解决歧义性言语；以及

向讲外语的说话者提供经翻译的言语。

15. 如权利要求 14 所述的方法，其特征在于，还包括基于所述交互的动作生成和更新用户音位模型及用户语言模型的至少其中之一的动作。

16. 如权利要求 14 所述的方法，其特征在于，所述交互的动作包括经由言语和机械交互的至少其中之一来用用户反馈响应以解决所述歧义性言语的动作。

17. 如权利要求 14 所述的方法，其特征在于，还包括通过处理除所述言语信号之外的所感测的数据来校正所述推断的动作。

18. 如权利要求 14 所述的方法，其特征在于，还包括基于来自所述讲外语的说话者的响应而修改所述言语识别过程的动作。

19. 一种有助于不同语言的用户之间的通信的系统，包括：

用于在言语识别过程期间接收用户和讲外语的说话者的至少其中之一的言语信号的装置；

用于基于对所述言语信号的分析来计算对思想和上下文的至少其中之一的推断的装置；

用于与所述用户或讲外语的说话者的至少其中之一交互以解决歧义性言语的装置；

用于根据所述推断修改所述言语识别过程的装置；以及

用于向所述讲外语的说话者呈现经翻译言语、表示所述言语的符号以及表示所述言语的图像的至少其中之一的装置。

20. 如权利要求 19 所述的系统，其特征在于，还包括用于在本地分布式数据存储中存储用户交互数据和讲外语的说话者交互数据的至少其中之一的装置。

用于不同语言的说话者之间的增强的交互和理解的上下文敏感通信和翻译方法

背景

诸如因特网的全球通信网络的出现已在便携式计算设备中充当起会聚计算能力和服务的催化剂。例如，在过去几年中，诸如蜂窝电话和个人数据助理（PDA）的便携式设备已使用分别用于语音通信和个人信息存储的单独功能。现在，可在单个便携式设备中找到这些功能，例如经由硬件和软件中增加的计算能力使用多模功能的蜂窝电话。这些设备更通常地被称为“智能电话”。

随着手持式和便携式设备中的技术进步，持续并日益需要最大化这些持续发展的技术的益处。给定这些便携式无线计算设备的存储和计算能力的进步，它们现在能够处理各种各样的不同数据类型，比如图像、视频剪辑、音频数据和用于管理和组织各种 PIM（个人信息管理器）数据的电子组织器。这些数据通常被单独地用于特定目的。

因特网已通过使上百万的网络用户经由移动设备（例如电话）、电子邮件、web 站点等彼此联系而带来国际化，这些设备中的一些可提供某种程度的文本翻译。例如，用户可选择他们的浏览器来安装语言插件，当用户访问外国的 web 站点时，这些语言插件有助于在某种程度上从一种语言文本翻译到另一种。然而，世界也在变得更移动。越来越多的人为了商业和享受而旅行。这出现了人们彼此之间面对面的情形和/或处在语言障碍成问题的外国中的情形。对于大量多语言移动助理的情况，言语翻译是极高的门槛。需要一种可采用便携式设备增加的计算能力以便增强用户翻译体验的机制。

概述

以下呈现了简要的概述以便提供对所公开发明的一些方面的基本理解。此概述并非广泛的概观，并且它并非旨在标识重要/关键的元素或者刻划出其范围。其唯一目的是以简要方式呈现一些概念作为在以下呈现的更详细描述的前言。

所公开的本发明是与不同语言的一个用户或多个用户交互以增强之间的言语翻译的架构。此处所述的是这样一个概念：捕捉和/或感测和/或集中所识别的思想或情形，并且使用对用户在与讲另一种语言的其他人共事时的通信目提供推断的简

化的设备来对混合发起的用户交互消除歧义。使用这种方法，基于当前注意力的焦点或者可能的注意力的焦点的概率分布上的思想和情形来实施关于通信目的的推理，并且向用户或用户的交谈对方提供了适当的筛余选择，以及供审阅和/或感知的图像、符号、文本和/或言语翻译。

这些推论还可将来自用户的话语或其它输入作为输入以作为推理出思想、情形、目的和/或对其消除歧义的证据的一部分。然后，系统对通信症结上的问题、需求或意图的最佳理解或再阐述可被返回给用户以便确认。

由系统进行推断可提供基于监听的识别和信息收集组件的上下文敏感聚焦，并且可使用从先前或当前的用户话语中识别出的单词来进一步聚焦推断。另外，可提供图形和/或言语识别暗道（backchannel）来用于用户输入、确认和/或由所定址的个人进行指点，从而更好地理解应答。

此架构有助于识别关键上下文线索、用户注意力焦点处的思想、和用户输入和指导以便针对与注意力焦点推理出用户意图、需求和目标，然后，基于注意力焦点和上下文构造一组候选话语。如有需要，它们可用于在适当的位置自动将注意力焦点处的思想以及可能的上下文融入该语言，然后使用户能够提炼或选择话语、文本串和/或图像来向讲外语的说话者传递。移动设备可任选地向其他用户提供输入信息或选择话语以便将向设备所有者传回响应的装置。

在其另一个方面，提供了使用基于概率和/或统计的分析来预测或推断用户期望要自动执行的动作的机器学习和推理组件。

为了实现前述及相关目的，本文将结合以下说明书和附图来描述所公开发明的特定说明性方面。然而，这些方面仅指示可使用本文所公开的原理的各种方式的少数几种，但是旨在包括所有这些方面以及它们的等效方案。根据以下详细描述并结合附图考虑时，其它优点和新颖特征将变得显而易见。

附图简述

图 1 示出了有助于不同语言的用户之间的言语翻译的一种系统。

图 2 示出了有助于不同语言的用户之间的言语的翻译的一种方法。

图 3 示出了根据本发明的另一个方面的自适应言语识别处理的方法。

图 4 示出了根据本发明的另一个方面的自适应言语识别处理的替换方法。

图 5 示出了根据本发明的一个方面、与用户交互以便确认上下文和/或思想数据的一种方法。

图 6 示出了根据另一个方面、处理所感测的输入数据以寻找改进对用户思想、基于上下文确定的推断的线索一种方法。

图 7 示出了根据本发明的一个方面的、有助于推断处理的数据存储网络。

图 8 示出了根据本发明的言语翻译处理的、同步化分布式和集中式数据库的一种方法。

图 9 示出了为生成用户音位模型和用户语言模型而训练自适应 ASR 的一种方法。

图 10 示出了使用用户和 FLS 两者的言语识别训练的一种方法。

图 11 示出了有助于使用不同语言的至少两个用户之间的翻译的一个示例性翻译组件 1100 的一个示意性框图。

图 12 示出了根据另一个方面的、使用模拟器教学的一种系统。

图 13 示出了根据另一个方面的、对使用外语的用户进行教学的一种方法。

图 14 示出了有助于自动化根据本发明的一个或多个特征、使用机器学习和推理组件的一种系统。

图 15 示出了根据本发明的一个方面的、有助于翻译的便携式无线设备的示意性框图。

图 16 示出了根据一个方面的、用户与 FLS 接收者之间的一种设备对设备翻译系统。

图 17 示出了根据本发明另一个方面的、可用在用户与 FLS 之间的单个设备翻译系统。

图 18 示出了可用于执行所公开的翻译架构的计算机的框图。

图 19 示出了根据另一个方面的、有助于言语翻译的示例性计算环境的一个示意性框图。

具体实施方式

现在参照附图来描述本发明，其中在全文中相同的附图标记用于指相同的要素。在以下描述中，为说明的目的，阐述了许多特定细节以便提供对其的全面理解。然而，在没有这些特定细节的情况下，也可实践本发明，这是显而易见的。在其它示例中，以框图形式示出了众所周知的结构和设备，以便有助于描述。

如本申请所用术语“组件”和“系统”旨在指计算机相关实体，或是硬件、硬件和软件的组合、软件或者是执行中的软件。例如，组件可以是但不限于：运行

于处理器上的进程、处理器、硬盘驱动器、多个存储驱动器（光学和/或磁性存储介质的）、对象、可执行代码、执行的线程、程序和/或计算机。作为示例，运行于服务器上的应用程序和该服务器都可以是组件。一个或多个组件可驻留在进程和/或执行的线程中，并且组件可位于一个计算机上和/或分布在两个或多个计算机之间。

如本文所用术语“推断”或“推论”通常指根据如经由事件和/或数据所捕捉到的一组观测来推理或推断系统、环境和/用户的状态的过程。可使用推断来例如识别特定上下文或动作，或者可生成状态上的概率分布。推断可以是概率的——即，基于数据和事件的考虑计算感兴趣的状态上的概率分布。推断还可指用于从一组事件和/数据构成较高层的事件的技术。这些推断导致从一组已观测到的事件和/或已存储的事件数据构成新事件或动作，无论这些事件是否在时间上紧密相关，也无论这些事件和数据是来自一个还是多个事件和数据源。

一开始参看附图，图 1 示出了有助于不同语言的用户之间的言语翻译的系统 100。所公开的发明是与不同语言的一个用户或多个用户交互以便增强他们之间的翻译的架构。本文所述的是这样一个概念：捕捉和/或感测和/或集中所识别的思想或情形，并且使用对用户在与讲另一种语言的其他人共事时的通信目提供推断的简化的设备来对混合发起的用户交互消除歧义。使用这些方法，基于当前注意力的焦点的思想或情形来实施关于通信目的的推理，向用户提供了适当筛余的选择，并呈现文本和/或言语翻译以供感知。

推断还可将从用户的话语作为输入以作为推理出思想、情形、目标和/或对其消除歧义的证据的一部分。然后，系统对问题、需求或意图的理解或再阐述被返回给用户以便确认。

由系统进行的推断可提供基于监听的深度聚焦，并且还可使用从用户的话语中识别出的单词来进一步聚焦推断。另外，可提供图形和/或言语识别暗道来用于用户输入、确认和/或由所定址的个人进行指点，从而更好地理解应答。

因此，系统 100 包括自适应自动言语识别（ASR）组件 102，该组件处理当前上下文和/或思想所感测到的数据并有助于基于所感测的数据进行言语识别过程。系统 100 的历史活动组件 104 存储与言语识别过程相关联的历史数据。换言之，当用户与系统交互作用时，这些交互数据被存储在数据存储中作为将来分析和推断的基础。系统 100 还可包括语言机会组件 106，该组件通过向用户推行一个或多个词项的训练会话来改进言语识别过程，当在将来的言语识别过程中使用这一个或多个

词项时，训练会话提高了成功的可能性。

图 2 示出了有助于不同语言的用户之间的言语翻译的一种方法。虽然为了简化说明的目的，这里例如以流程图或流程示图的方式示出的一种或多种方法，被示出并描述成一连串动作，但是应当明白和理解，本发明不受这些动作的次序的限制，因为一些动作可根据需要以不同次序发生和/或与本文所示和所述的其它动作同时发生。例如，本领域技术人员应当明白和理解，方法可替换表示成诸如状态图中的一系列相关状态或事件。此外，并非所有所示动作都是对实现根据本发明的方法所必需的。

在 200，接收到自适应 ASR 组件，它将用户语言的言语翻译成外语以便由讲外语的说话者（FLS）来感知，并且将 FLS 的语言译回用户的语言。在 202，系统基于包括但不限于用户与 FLS 之间的言语通信的所感测的输入数据来推断上下文和/或思想信息。这些数据还例如可包括将在下文中更详细描述图像、位置信息、音频信息、手势识别和搜索信息。在 204，ASR 基于推断出的上下文和/或思想信息来适应于当前言语识别过程。在 206，系统基于所感测的数据（例如言语）检测当前言语翻译交换中的一个或多个不准确或歧义的地方。在 208，系统基于需要改进处理的歧义单词和/或词项来向用户和/或 FLS 推行单词和/或词项训练和/或澄清。此训练返回解决后的结果。在 210，解决后的结果被反馈到 ASR 以便在当前言语处理会话中进行自适应处理。在 212，解决后的结果以及用户/FLS 交互各方面被存储在历史数据存储中用于将来的访问和推断过程。

图 3 示出了根据本发明的另一个方面的自适应言语识别处理的一种方法。在 300 处，自适应 ASR 处理用户与 FLS 之间的言语翻译会话，并且基于所感测的输入数据推断上下文和/或思想。在 302，系统检测言语识别过程期间的歧义性。在 304，系统中断会话。在 306，系统向用户和/或 FLS 提示所建议的无歧义数据。在 308，系统接收并处理所感测的用户和/或 FLS 的输入数据。这可以是言语和/或其它所感测的输入数据形式的。在 310，系统确定歧义性是否被解决。如果是，则在 312，系统根据用户和/或 FLS 的响应来适应言语识别过程。在 314，与用户会话相关联地存储响应和/或所感测的输入数据，用于将来的访问和分析。如果在 312，歧义性未被解决，则前进到 316，其中系统得到替换的无歧义数据。然后，流程返回到 306 以向用户和/或 FLS 呈现该替换数据。该系统/用户/FLS 的交互可继续直至解决歧义性。

图 4 示出了根据本发明的另一个方面的自适应言语识别处理的替换方法。在

400, 自适应 ASR 处理用户与 FLS 之间的言语翻译会话, 并且基于所感测的输入数据来推断上下文和/或思想。在步骤 402, 系统检测言语识别过程期间的歧义性。在 404, 系统中断会话并向用户和/FLS 描述歧义性。在 406, 用户和/或 FLS 用所建议的无歧义数据作出响应。在 408, 系统接收并处理所感测的用户和/或 FLS 的输入数据。这可以是言语和/或其它所感测的输入数据形式的。在 410, 系统确定歧义性是否被解决。如果是, 则在 412, 系统根据用户和/或 FLS 的响应来适应言语识别过程。在 414, 与该用户相关联地存储响应和/或所感测的输入数据, 用于将来的访问和分析。如果在 412, 歧义性未被解决, 则前进到 416, 其中用户和/或 FLS 输入替换无歧义数据。然后, 流程返回到 406 以向系统呈现该替换数据以便进行处理。该系统/用户/FLS 的交互可继续直至解决歧义性。

图 5 示出了根据本发明的一个方面、与用户交互以便确认上下文和/或思想数据的一种方法。在 500, 接收到一感测系统, 它可捕捉图像、言语、声音等, 并且对它们执行分析以推断用户上下文和/或思想。在 502, 在用户与 FLS 之间启动通信。在 504, 系统确定通信是否成功进行。如果不成功, 则在 506, 系统启动对上下文和/或思想数据的感测。在 508, 一旦从中接收到所感测的输入数据, 则可推断上下文/或思想以及基于所感测的输入数据推断附加的通信词项。在 510, 系统还可请求和/或不经请求地接收用于解决歧义词项的用户反馈。在 512, 如果歧义性未被解决, 则流程再次返回到 506, 执行上下文和/或思想感测, 并且重复歧义性解决方案。如果会话结束, 则在 516 进行核对, 系统停止翻译过程, 并等待下一用户/FLS 输入。

然而, 如果会话未结束, 则流程从 516 进行到 518 继续通信, 并且流程继续返回到 504 以基于新的输入来确定通信是否成功。如果在 504, 通信成功进行, 则流程从 504 进行到 516 以确定通信是否完成。在任一情况中, 流程按上述进行。

图 6 示出了根据另一个方面的、处理所感测的输入数据以寻找改进对用户思想的、基于上下文确定的推断的线索的一种方法。在 600, 系统启动用户注意力焦点处用户思想的确定。这可包括完全基于用户和/或 FLS 的言语及其词项来作出关于用户思想的推断。在 602, 系统基于从感测系统所感测的数据来执行上下文确定。这可包括关于用户凝视的方向、手势识别、文本识别等来分析所感测的数据。在 604, 处理所感测的输入数据来寻找线索。在 606, 从上下文线索推断出思想。

在 608, 系统进行核对以便确定所推断的思想是否正确。这可通过向用户直接或明显地提示、通过将所推断的思想应用到自适应 ASR 来执行, 其结果可指示所

推断的思想是否基本上正确。如果正确,则在 610,处理思想数据来实现由自适应 ASR 进行通信所使用的那些词项。在 612,翻译词项并将其呈现给 FLS。如果所推断的思想被认为不正确,则流程从 608 进行到 614 来获得用户反馈。这些反馈可影响在随后的上下文确定的轮次内可使用哪些感测系统组件。

图 7 示出了根据本发明的一个方面的、有助于推断处理的数据存储网络 700。用户和/或 FLS 通常在与许多分布式数据存储器 702 (表示为分布式₁(DISTRIBUTED₁)、分布式₂(DISTRIBUTED₂)、...、分布式_N(DISTRIBUTED_N))之一相关联的位置上交互。例如,这里所述的历史活动(例如响应于设备提示来解决歧义性、用户在教学期间的交互、...)最初可存储在交互位置本地,诸如在有助于用户与现场的 FLS 之间的言语翻译的便携式无线设备(PWD)(未示出)中。这可由作为 PWD 的一部分并存储向其传递的数据的第一分布式数据存储 704 来表示。最后,对其更新的内容可上传到可以是区域性数据存储的中央数据存储 706。其重要性之一是当第一用户在使用数据存储系统 700 的外国旅行,在包含其它分布式数据存储 702 的热门区域之间旅行时,有助于上传如由这些另外位置的上下文交互所记录的第一用户的数据。这还有助于使用在第一用户到达之前就被存储在这些位置上的其它用户交互数据,从而进一步提高第一用户在这些位置的翻译体验。

图 8 示出了根据本发明的言语翻译处理、同步化分布式和集中式数据库的一种方法。在 800,接收到数据存储的本地化分布式系统和集中式系统。在 802,基于用户在此位置的动作以用户数据来更新第一本地分布式数据存储。这可包括言语、所访问的位置、用户在每个位置的动作以及言语、手势等。在 804,当数据本地存储时,和/或在同步化对系统带宽影响最小时,从分布式数据存储向集中式数据存储上传用户数据。在 806,集中式数据存储向其它分布式数据存储同步这些更新,从而如果访问这些位置,则可在这些位置提高用户翻译体验。

图 9 示出了为生成用户音位模型和用户语言模型而训练自适应 ASR 的一种方法。在 900,系统和/或用户启动训练会话。在 902,系统指导用户说出一个或多个单词。这可通过系统来选择,以捕捉被系统认为为解决诸如由于用户的发音、声调、音调等而产生的歧义性重要的许多方面。在 904,系统接收所说的一个或多个单词。在 906,系统处理接收到的单词并确定它们是否与先前对系统指导的相匹配。如果不匹配,则流程转到 908,其中系统请求用户重复单词。这可包括当说出所请求的单词时,系统打断用户。当在 910 核对时,如果单词匹配,则流程转到 912,其中系统记录并存储用户信息,并更新音位和/或用户语言模型。在 908,如果存在匹配,

则流程转到 912 以执行存储和更新过程。在 910, 如果用户的第二次发音不匹配, 则系统可返回到 908 重复此过程, 例如使用可获得相同或类似结果的其它单词或词项, 直至获得满意的结果。

图 10 示出了使用用户和 FLS 两者的言语识别训练的一种方法。在 1000, 当存在用户和 FLS 时, 启动训练会话。这可以是系统启动或用户启动的。在 1002, 系统指导用户说出一个或多个预订的单词和/或句子。在 1004, 系统处理所说的单词和/或句子, 并且将它们翻译成 FLS 的语言。在 1006, 系统将所翻译的数据输出给 FLS。这可以是使用机器说出的信号、显示器上作为文本的呈现和/或音频及文本两者的。在 1008, FLS 提供反馈以确认单词和/或句子的准确性。在 1010, 在任何情况中, 与训练会话相关联的数据被存储以用于分析, 并且对 ASR 自适应更新以用于将来的言语翻译。

图 11 示出了有助于使用不同语言的至少两个用户之间的翻译的一个示例性翻译组件 1100 的一个示意性框图。翻译组件 1100 包括感测系统子组件 1102, 该子组件有助于能够对所感测的、表示至少一个用户上下文 1104 和/或用户思想的输入数据进行感测、捕捉和处理。在其支持下, 翻译组件 1100 包括一般表示用于收集关于用户上下文和/或思想的多个不同感测子系统的感测系统子组件 1102 (或模态)。例如, 成像组件 (或子系统) 1106 (例如摄像机或成像仪) 可用于获取菜单 (用于内容)、路标或可提供关于用户思想和/或上下文 1104 的线索的任何其它项目的图片快照。视频组件 1108 (例如摄像机) 有助于捕捉上下文 1104 的视频剪辑。之后, 视频处理可用于提供关于用户思想和/或上下文 1104 的线索。音频组件 1110 (例如话筒和附带的电子产品) 有助于记录对特定上下文 1104 为本地的音频数据。

言语组件 1112 处理用户或其他人的言语输入以便识别。因而, 此组件 1112 还可包括有助于识别输入言语的语言的语言组件库 (未示出)。这种言语可被转换成带错误或不带错误的文本。光学识别组件 (OCR) 1114 处理由例如成像组件 1106 和视频组件 1108 捕捉到的图像化文本。如上所示, 用户可通过捕捉各种数据以寻找关于上下文和内容的线索来提高翻译准确性。在一个示例中, 用户可获取路标、企业标记、菜单的快照, 这些快照的文本随后被处理以改进翻译成果从而可理解地输出。

感测子系统 1102 还可包括搜索组件 1116, 可手动或自动地激活该组件以执行对关于上下文和/或内容信息的数据的搜索。此搜索可以是对存储在本地数据存储 (未示出) 中的数据本地搜索。另外, 或者与其组合, 搜索可使用网络搜索引

擎在无线和/或有线范围（例如因特网）上进行。在任何情况中，搜索组件 1116 可接收来自感测子系统 1102 的组件中任一个、用于查询的搜索词项。在另一个实现中，可从接收并处理选自感测子系统 1102 的输出数据的另一个子系统（未示出）生成搜索词项。

感测子系统 1102 还可包括手势组件 1118，该组件有助于处理如由成像组件 1106 和/或视频组件 1108 捕捉到的手势。手势识别可用于提高例如输入识别、紧急事件和/或情感交互。定位组件 1120 有助于确定用户的地理位置。此组件 1120 可包括全球定位系统（GPS）技术和/或其它合适的三角测量技术，诸如 IEEE 802.11、Wi-Fi 信号以及经由一种或多种技术得到的关于如设备是否在运行（经由加速计或 GPS 感测）、设备附近的环境温度、是否检测到谈话、是否识别或标识其他人等的这些上下文信息。或者，定位组件 120 可协同一个或多个其它感测子系统 1102 的来起作用以得到用户位置。例如，基于成像组件 1106 的图像处理、OCR 组件 1114 的 OCR 输出以及搜索组件 1116 的结果，可在某种程度上成功推导出位置为 X。

翻译组件 1100 还可包括推断组件 1122，该组件处理来自感测系统 1102 的所感测的数据以得到可能的线索并发展或生成关于用户思想和/或上下文 1104 的推断。翻译组件 1100 还可包括音位模型组件 1124，该组件存储并更新用户和/或 FLS 在言语交流期间的言语特性。另外，可生成针对用户个人的用户语言模型组件 1126，该组件定义了任何给定情形和/或位置下的用户语言。最后，翻译组件 1100 有助于在用户与 FLS 之间可理解地翻译单词、词项、短语和/或句子。

现在参看图 12，它示出了根据另一个方面的、使用模拟器教学的系统 1200。在本发明的另一个方面中，此架构可使用向用户提供外语的教学的模拟组件 1202。模拟组件 1202 可连接自适应 ASR 组件 102、历史活动组件 104 和推行语言机会组件 106 中的任何一个或者与其结合来进行工作。模拟组件 1202 用于用用户语言来提示用户用外语说出所提示的词项、单词、短语和/或句子，响应于此，用户用外语讲出相应的翻译。然后，系统 1200 处理用户响应，并且可提示改变语调、单词用法、句子结构等作为用户教育过程的一部分。

这仅作为其中可执行用户训练的许多不同情况的一个示例。例如，模拟组件 1202 可用外语提示用户，响应于此，用户提供用户语言的翻译。在另一个示例中，模拟组件 1202 提示用户用外语重复与系统输出相同的词项、短语和/或句子，响应于此，用户用外语说出相同的词项、短语和/或句子。然后，系统 1200 分析用户言

语，并提供更好地反映合适的外语口语的语调、重音、音调等的形式的校正。然后，这些教学数据可被存储在历史数据存储中，并用于进一步开发用户音位和语言模型，以及用于用户将来的翻译交流。

图 13 示出了根据另一个方面的、对用户教学外语的一种方法。在 1300，可启动模拟器训练。这可由用户手动或由系统自动启动。在 1302，如果是首次使用的用户，则系统训练其自身以便学习用户的言语特性并开发个性化的用户音位和言语模型。在 1304，系统提示用户讲出一个单词、一组单词、短语和/或句子。在 1306，系统接收用户响应并至少基于用户音位和语言模型来处理该响应。在 1308，系统计算任何错误、歧义性和不准确性，并且来回地与用户进行言语识别来解决这些问题。在 1310，在一个用户学习会话中，系统向用户提供供用户感知的翻译（例如经由计算机说出的信号、文本显示和/或两者）。在另一个学习会话中，系统请求用户以相应的外语翻译回答。在 1312，系统向用户提供关于用户翻译的准确性的反馈。

图 14 示出了使用有助于自动化根据本发明的一个或多个特征的机器学习和推理（MLR）组件 1402 的一种系统 1400。本发明（例如结合选择）可使用各种基于人工智能的方案来实现其各个方面。例如，可经由自动分类器系统和过程促进用于确定在系统和/或用户训练中使用哪些词项、词项组、短语和/或句子的过程。

分类器是将输入属性矢量 $x = (x_1, x_2, x_3, x_4, x_n)$ 映射到类标签 $class(x)$ 的函数。分类器还可输出该输入属于一个类的置信度，即 $f(x) = confidence(class(x))$ 。此分类可使用基于概率和/或统计的分析（例如分解成分析效用和成本）来预测或推断用户期望自动执行的动作。

支持矢量机（SVM）是可使用的分类器的一个示例。通过在以最佳形式将触发输入事件与非触发事件分隔开的可能输入的空间中寻找超曲面来操作 SVM。直观地，这使分类对接近、但与训练数据不同的测试数据进行校正。其它有向和无向模型分类方法包括例如朴素贝叶斯（naïve Bayes）、贝叶斯（Bayesian）网络、决策树、神经网络、模糊逻辑模型和提供可使用的不同独立模式的概率分类模型。如本文所用的分类还包括用于开发优先级模型的统计回归。

如可容易地从本说明书理解的，本发明可使用经显式训练（例如经由一般训练数据）以及隐式训练（例如经由观察用户行为、接收外来信息）的分类器。例如，经由分类器构造器和特征选择模型内的学习或训练阶段来配置 SVM。因而，可使用分类器来自动学习和执行大量功能。

为了连接到自适应 ASR 组件 102、历史活动组件 104 和推行机会语言组件 106，自动学习和执行可包括但不限于：根据预定标准来确定何时用户讲出的单词或词项是被认为是准确或无歧义的。在另一个示例中，MLR 组件 1402 可有助于确定应当对给定用户使用多少交互来达到所说出的外语的期望准确度。这仅是可经由 MLR 组件 1402 来学习和自动地执行的许多特征中很少的一部分。

图 15 示出了根据本发明的一个方面的、有助于翻译的 PWD 1500（例如蜂窝电话）的示意性框图。设备 1500 包括连接到一个或多个内部组件以便对数据和指令进行控制和处理的处理器 1502。处理器 1502 可被编程为控制和操作设备 1500 内的各种组件以便实现本文所述的各种功能。处理器 1502 可以是多种合适的处理器（例如，DSP 数字信号处理器）中的任一种，并且可以是多处理器子系统。

存储器和存储组件 1504 连接到处理器 1502 并用于存储程序代码，并且还用作对诸如数据、应用程序、服务、元数据、设备状态等的信息的存储装置。存储器和存储组件 1504 可包括适于存储从感测系统和/或传感器获得的至少一整组所感测的输入数据的非易失性存储器。因而，存储器 1504 可包括供处理器 1502 高速访问的 RAM 或闪存和/或海量存储器，例如能够存储包括文本、图像、音频和/或视频内容的千兆字节数据的微驱动器。根据一个方面，存储器 1504 具有足够的存储容量来存储与不同的服务相关的多组信息，并且处理器 1502 可包括有助于在与不同服务相对应的各组信息之间交替或循环的程序。

显示器 1506 可经由显示器驱动器子系统 1508 耦合到处理器 1502。显示器 1506 可以是彩色液晶显示器（LCD）、等离子显示器、触摸屏显示器等。显示器 1506 用于呈现数据、图形或其它信息内容。另外，显示器 1506 可呈现用户可选择并提供对设备 1500 的控制和配置的多种功能。在触摸屏示例中，显示器 1506 可显示有助于用户交互以便控制和/或配置的触摸可选择图标。

可通过板上电源系统 1510（例如电池组或燃料电池）向处理器 1502 和形成设备 1500 的其它板上组件供电。在电源系统 1510 无法供电或与设备 1500 断开时，备用电源 1512 可用于向处理器 1502 和其它组件（例如传感器、图像捕捉设备…）供电，并且如果是可充电工艺，则可对板上电源系统 1510 充电。例如，备用电源 1512 可有助于经由功率变换器连接到外部并网。处理器 1502 可被配置成提供电源管理服务，以便例如引起减小当前电流消耗的休眠模式，或者基于对预期电源故障的检测而启动设备 1500 的有序关机。

设备 1500 包括具有数据通信端口 1516 的数据通信子系统 1514，该端口 1516

用于将设备 1500 连接到远程计算系统、服务器、服务等。端口 1516 可包括诸如通用串行总线 (USB) 和/或 IEEE 1394 的、提供串行通信能力的一个或多个串行接口。其它技术也可被包括,但不限于:例如使用红外通信端口的红外通信以及无线分组通信(例如蓝牙™、Wi-Fi 和 Wi-Max)。如果是智能电话,数据通信子系统 1514 可包括 SIM(用户身份模块)数据以及蜂窝注册和网络通信所需的信息。

设备 1500 还可包括用于与处理器 1502 通信的射频(RF)发射机部分 1518。RF 部分 1518 包括 RF 接收机 1520,该接收机经由天线 1522 从远程设备或系统接收 RF 信号,并且可解调信号以获得其中调制的数字信息。RF 部分 1518 还包括用于响应于用户经由用户输入设备 1526(例如小键盘)的手动输入、或者自动响应于对进入通信范围或者其它预定和编程的准则的检测和/或对离开的预测来向远程设备或系统发送信息(例如数据、服务)的 RF 发射机 1524。

设备 1500 还可包括由处理器 1502 控制并处理来自话筒或类似音频输入设备(未示出)的语音输入的音频 I/O 子系统 1528。音频子系统 1528 还有助于经由扬声器或类似音频输出设备(未示出)来呈现言语和音频输出信号。

设备 1500 还可包括连接到处理器并有助于向设备本身和/或远程系统认证用户的认证组件 1530。处理器 1502 还连接到感测子系统块 1532,该感测子系统块有助于捕捉和输入 OCR 数据、语音数据、手写数据和图像/视频数据,例如以便确定用户上下文和/或思想。另外,设备 1500 包括一个或多个应用程序 1534(例如,成像程序、视频呈现程序、OCR 程序、搜索引擎...),这些程序可被启动以启用感测系统用于根据本发明架构接收所感测的输入数据的操作。这些应用程序还可包括用于音位和语言模型的创建和更新以及推断分析和选择的程序。

设备 1500 还包括允许直接物理连接到另一个系统(例如经由连接器)而非在其间使用无线通信或电缆通信的物理接口子系统 1536。

图 16 示出了根据一个方面的、用户与 FLS 接收者之间的设备对设备翻译系统 1600。用户 1602 使用便携式无线设备(PWD) 1604(其包括本发明的翻译架构)来经由接收设备 1608(其也包括本发明的翻译架构)与 FLS 接收者 1606 无线地通信。用户 1602 向用户 PWD 1604 输入言语信号,然后,这些信号被处理成经翻译的输出并将其无线地发送到接收设备 1608。接收设备 1608 将用户言语翻译成用户文本,这些文本可被显示在接收设备 1608 上,和/或作为经翻译的用户言语向接收者 1606 输出。类似地,用户设备 1604 将接收者言语翻译成接收者文本,这些文本被显示在用户设备 1604 上,和/或作为经翻译的接收者言语向用户 1602 输出。

如果用户 1602 与接收者 1606 位于几乎相同的上下文中，则设备 1604 或/和 1608 的任一个或两者可执行上述的上下文和/或思想处理以增强翻译。因而，1604 或/和 1608 的任一个或两者的模态 1610 和 1612 可分别用于生成所期望的翻译。

图 17 示出了根据本发明另一个方面的、可用在用户与 FLS 接收者之间的单个设备翻译系统 1700。用户 1702 和 FLS 接收者 1704 例如，是面对面的，但无法有效地通信。用户对设备 1706 讲话，然后，该设备可将用户言语处理成可由接收者 1704 解释或理解的文本、符号和/其它媒体（例如图像、声音、视频...）。文本、符号和/其它媒体可被显示在设备 1706 上，并且文本以用户语言和接收者语言两种方式显示。非文本媒体（例如符号、图像、声音、视频）的使用可用在例如接收者不识字的情况中。应当理解，在本发明的上下文中，无论何时在描述示例时用到文本或言语，其还可包括可以某种形式在各方之间输出或通信的符号和/或媒体。或者，或与其结合，可向接收者 1704 呈现作为经翻译的用户言语的用户言语。类似地，设备 1704 将接收者言语翻译成接收者文本，这些文本可显示在设备 1704 上，和/或向用户 1702 输出作为经翻译的接收者言语。根据本发明的一个方面，设备 1706 还包括感测组件 1708 形式的模态，可在设备 1704 中提供一个或多个这些模态以便有助于改进翻译。

现在参看图 18，示出了可用于执行所公开的翻译架构的计算机的框图。另外，计算机可用作上述的本地分布式数据存储和/或集中式数据存储系统。为了向其各个方面提供附加上下文，图 18 和以下讨论旨在提供其中可实现本发明的各个方面的合适的计算环境 1800 的简要、一般描述。虽然以上在可运行于一个或多个计算机上的计算机可执行指令的一般上下文中进行了描述，但是本领域技术人员应当理解，本发明可结合其它程序模块和/或作为硬件和软件的组合来实现。

通常，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、数据结构等。此外，本领域技术人员应当理解，本发明的方法也可使用其它计算机系统配置来实践，包括：单处理器或多处理器计算机系统、小型计算机、大型计算机、以及个人计算机、手持型计算设备、基于微处理器或可编程消费电子产品等，它们的每一个都可用于耦合到一个或多个相关联的设备。

也可在其中由经由通信网络链接的远程处理设备执行特定任务的分布式计算环境中实践本发明的所示各个方面。在分布式计算环境中，程序模块可位于本地或远程存储器存储设备中。

计算机一般包括各种计算机可读介质。计算机可读介质可以是可由计算机访

间的任何可用介质，包括易失性和非易失性介质、可移动和不可移动介质。作为示例而非限制，计算机可读介质可包括计算机存储介质和通信介质。计算机存储介质包括以存储诸如计算机可读指令、数据结构、程序模块或其它数据的信息的任何方法和技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括但不限于：RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字通用盘(DVD)或其它光盘存储、磁带盒、磁带、磁盘存储器或其它磁存储设备、或者可被用来存储所需信息并可由计算机访问的任何其它介质。

通信介质通常以诸如载波或其它传送机制的已调制数据信号的方式来体现计算机可读指令、数据结构、程序模块或者其它数据，并且包括任何信息传输介质。术语“已调制数据信号”是指以在信号中编码信息的方式设置或改变其特征中的一个或多个的信号。作为示例而非限制，通信介质包括诸如有线网络或直接连接的有线介质、以及诸如声波、RF、红外线和其它无线介质的无线介质。以上介质的任何组合应当也可包括在计算机可读介质的范围内。

再次参看图 18，用于实现各个方面的示例性环境 1800 包括计算机 1802，该计算机 1802 包括：处理单元 1804、系统存储器 1806、以及系统总线 1808。系统总线 1808 将包括但不限于系统存储器 1806 的系统组件耦合到处理单元 1804。处理单元 1804 可以是各种可购买的处理器中的任一种。双微处理器和其它多处理器架构也可用作处理单元 1804。

系统总线 1808 可以是可使用各种可购买的总线架构中的任一种进一步互连到存储器总线（使用或不使用存储器控制器）、外围总线以及局域总线的若干类型总线结构中的任一种。系统存储器 1806 包括只读存储器（ROM）1810 和随机存取存储器（RAM）1812。基本输入/输出系统（BIOS）被存储在诸如 ROM、EPROM、EEPROM 的非易失性存储器 1810 中，该 BIOS 包含有助于诸如在启动期间在计算机 1802 内的元件之间传递信息的基本例程。RAM 1812 还可包括诸如用于高速缓存数据的静态 RAM 的高速 RAM。

计算机 1802 还包括：内部硬盘驱动器（HDD）1814（例如 EIDE、SATA），该内部硬盘驱动器 1814 还可被配置成在合适的机箱（未示出）中外部使用；软磁盘驱动器（FDD）1816（例如对可移动磁盘 1818 读或写）；以及光盘驱动器 1820（例如读取 CD-ROM 盘 1822 或者对诸如 DVD 的其它大容量光学介质读或写）。硬盘驱动器 1814、磁盘驱动器 1816 和光盘驱动器 1820 可分别通过硬盘驱动器接口 1824、磁盘驱动器接口 1826 和光学驱动器接口 1828 连接到系统总线 1808。用

于外部驱动器实现的接口 1824 包括通用串行总线 (USB) 和 IEEE 1394 接口技术的至少其中之一或两者。其它外部驱动器连接技术落在本发明的考虑范围内。

驱动器以及与它们相关联的计算机存储介质提供了对数据、数据结构、计算机可执行指令等的非易失性存储。对于计算机 1802, 驱动器和介质适于存储任何合适的数字格式的数据。尽管以上对计算机可读介质的描述指 HDD、可移动磁盘以及诸如 CD 或 DVD 的可移动光学可读介质, 但是本领域技术人员应当理解, 诸如 zip 驱动器、磁带盒、闪存卡、盒式磁带等的可通过计算机读取的其它类型的介质也可用在示例性操作环境中, 另外, 任何这样的介质可包含用于执行所公开的发明的方法的计算机可执行指令。

大量程序模块可被存储在驱动器和 RAM 1812 中, 包括操作系统 1830、一个或多个应用程序 1832、其它程序模块 1834 和程序数据 1836。操作系统、应用程序、模块和/或数据的全部或一部分还可被高速缓存在 RAM 1812 中。应当理解, 本发明可使用各种可购买的操作系统或操作系统的组合来实现。

用户可通过诸如键盘 1838 和诸如鼠标 1840 的定点设备的一个或多个有线/无线输入设备来向计算机 1802 输入命令和信息。其它输入设备 (未示出) 可包括话筒、IR 遥控器、操纵杆、游戏手柄、指示笔、触摸屏等。这些或其它输入设备通常经由耦合于系统总线 1808 的输入设备接口 1842 连接到处理单元 1804, 但也可通过诸如并行端口、IEEE 1394 串行端口、游戏端口、USB 端口、IR 接口等的其它接口连接。

监视器 1844 或其它类型的显示设备也可经由诸如视频适配器 1846 的接口连接到系统总线 1808。除监视器 1844 之外, 计算机通常包括其它外围输出设备 (未示出), 诸如扬声器、打印机等。

计算机 1802 可工作在使用经由无线和/或无线通信到诸如远程计算机 1848 的一个或多个远程计算机的逻辑连接的网络化环境中。远程计算机 1848 可以是工作站、服务器计算机、路由器、个人计算机、便携式计算机、基于微处理器的娱乐设备、对等设备或其它公共网络节点, 并且通常包括以上就计算机 1802 所描述的许多或所有元件, 尽管为了简便的目的, 仅示出存储器/存储设备 1850。所示的逻辑连接包括对局域网 (LAN) 1852 和/或例如广域网 (WAN) 1854 的较大网络的有线/无线地连接。这些 LAN 和 WAN 网络环境在办公室和公司中是常见的, 并且有助于诸如内联网的企业范围计算机网络, 它们全都可连接到例如因特网的全球通信网络。

当用在 LAN 网络环境中时，计算机 1802 经由有线和/或无线通信网络接口或适配器 1856 连接到局域网 1852。适配器 1856 可有助于有线或无线通信到 LAN 1852，该适配器还包括其上设置的用于与无线适配器 1856 通信的无线接入点。

当用在 WAN 网络环境中时，计算机 1802 可包括调制解调器 1858、或连接到 WAN 1854 上的通信服务器、或具有诸如经由因特网的用于在 WAN 1854 上建立通信的其它装置。可以为内置式或外置式以及为有线或无线设备的调制解调器 1858 可经由串行端口接口 1842 连接到系统总线 1808。在网络化环境中，关于计算机 1802 所述的程序模块或其一部分可被存储在存储器/存储设备 1850 中。应当理解，所示网络连接仅是示例性的，并且也可使用在计算机之间建立通信链接的其它装置。

计算机 1802 可用于与可操作地设置于无线通信中的任何无线设备或实体通信，例如打印机、扫描仪、台式计算机和/或便携式计算机、便携式数据助理、通信卫星和与无线可检测标签相关联的任何一种装置或位置（例如书报电话亭、报摊、休息室）以及电话。这包括至少 Wi-Fi 和蓝牙™无线技术。因而，通信可以是与常规网络一样的预定结构，或仅是至少两个设备之间的自组织通信。

Wi-Fi 或无线保真度在不用电线的情况下，允许从家中的长椅、旅馆房间中的床或工作的会议室连接到因特网。Wi-Fi 是类似于用在蜂窝电话中的无线技术，该技术使比如计算机的设备在户内或户外、在基站的范围内的任何位置发送和接收数据。Wi-Fi 网络使用被称为 IEEE 802.11（a、b、g 等）的无线电技术提供安全、可靠、快速的无线连接。Wi-Fi 网络可用于将计算机彼此连接、连接到因特网以及连接到有线网络（其使用 IEEE 802.3 或以太网）。Wi-Fi 网络以 11 Mbps（802.11a）或 54 Mbps（802.11b）的数据率工作在未经许可的 2.4 和 5 GHz 无线电频带中，例如或者使用包含两个频带（双频）的设备来操作，从而该网络可提供类似于用在许多办公室中的基本 10BaseT 有线以太网网络的真实世界性能。

现在参看图 19，它示出了根据另一个方面的、有助于言语翻译的示例性计算环境 1900 的一个示意性框图。系统 1900 包括一个或多个客户机 1902（例如用于输入和输出的 PWD）。客户机 1902 可以是硬件和/或软件（例如线程、进程、计算设备）。客户机 1902 可以通过例如使用本发明来容纳 cookie 和/或相关联的上下文信息。

系统 1900 还包括一个或多个服务器 1904（例如本地分布式数据存储服务器和/或集中式数据存储服务器）。服务器 1904 还可以是硬件和/或软件（例如线程、进程、计算设备）。服务器 1904 可通过例如使用本发明来容纳用于执行变换的线

程。客户机 1902 与服务器 1904 之间的一种可能的通信可以是适于在两个或多个计算机进程之间传输的数据分组形式的。数据分组可例如包括 cookie 和/或相关联的上下文信息。系统 1900 包括可用来有助于客户机 1902 与服务器 1904 之间的通信的通信架构 1906（例如诸如因特网的全球通信网络）。

可经由有线（包括光纤）和/或无线技术来便于通信。客户机 1902 可用于连接到一个或多个客户机数据存储 1908，这些数据存储可用于存储客户机 1902 的本地信息（例如 cookie 和/或相关联的上下文信息）。类似地，服务器 1904 可用于连接到一个或多个服务器数据存储 1910，这些数据存储可用于存储服务器 1904 的本地信息。

上述包括所公开的发明的示例。当然，描述组件和/或方法的每一种可想象的组合是不可能的，但是本领域的普通技术人员应当理解，许多其它组合和排列是可能的。因此。本发明旨在包括落在所附权利要求书的精神和范围内的所有这些改变、修改和变化。此外，就术语“包括”被用在详细描述或权利要求而言，此术语旨在以类似于术语“包含”用作权利要求书中的过渡词时的方式为包含性的。

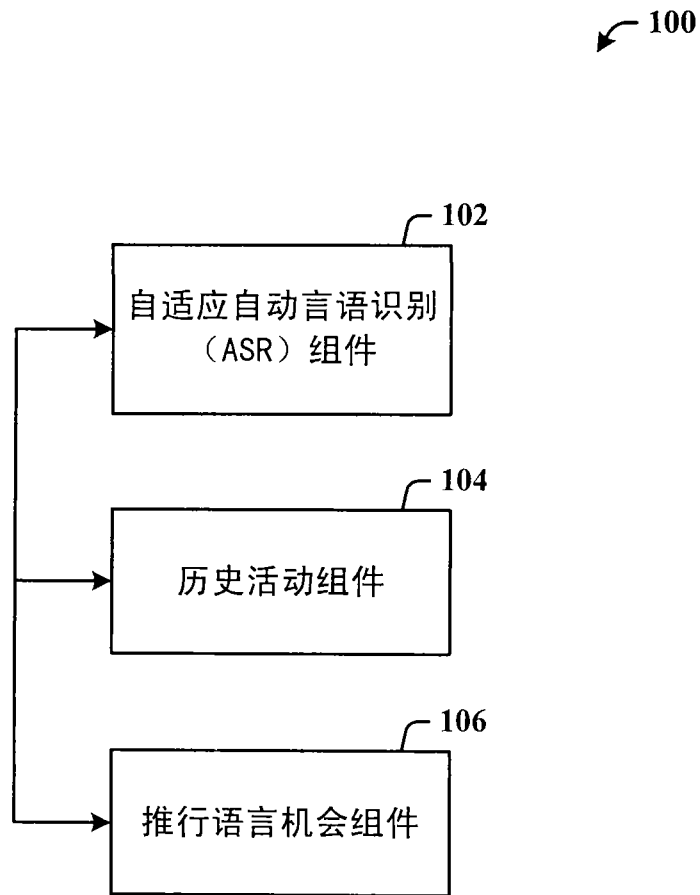


图 1

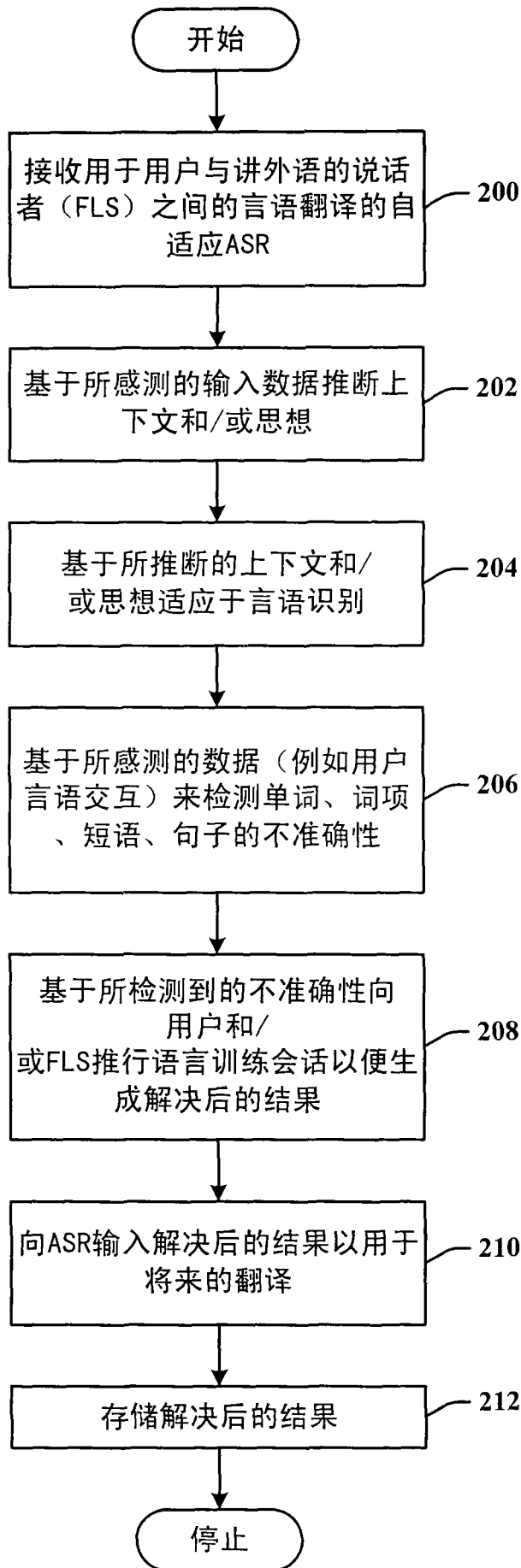


图 2

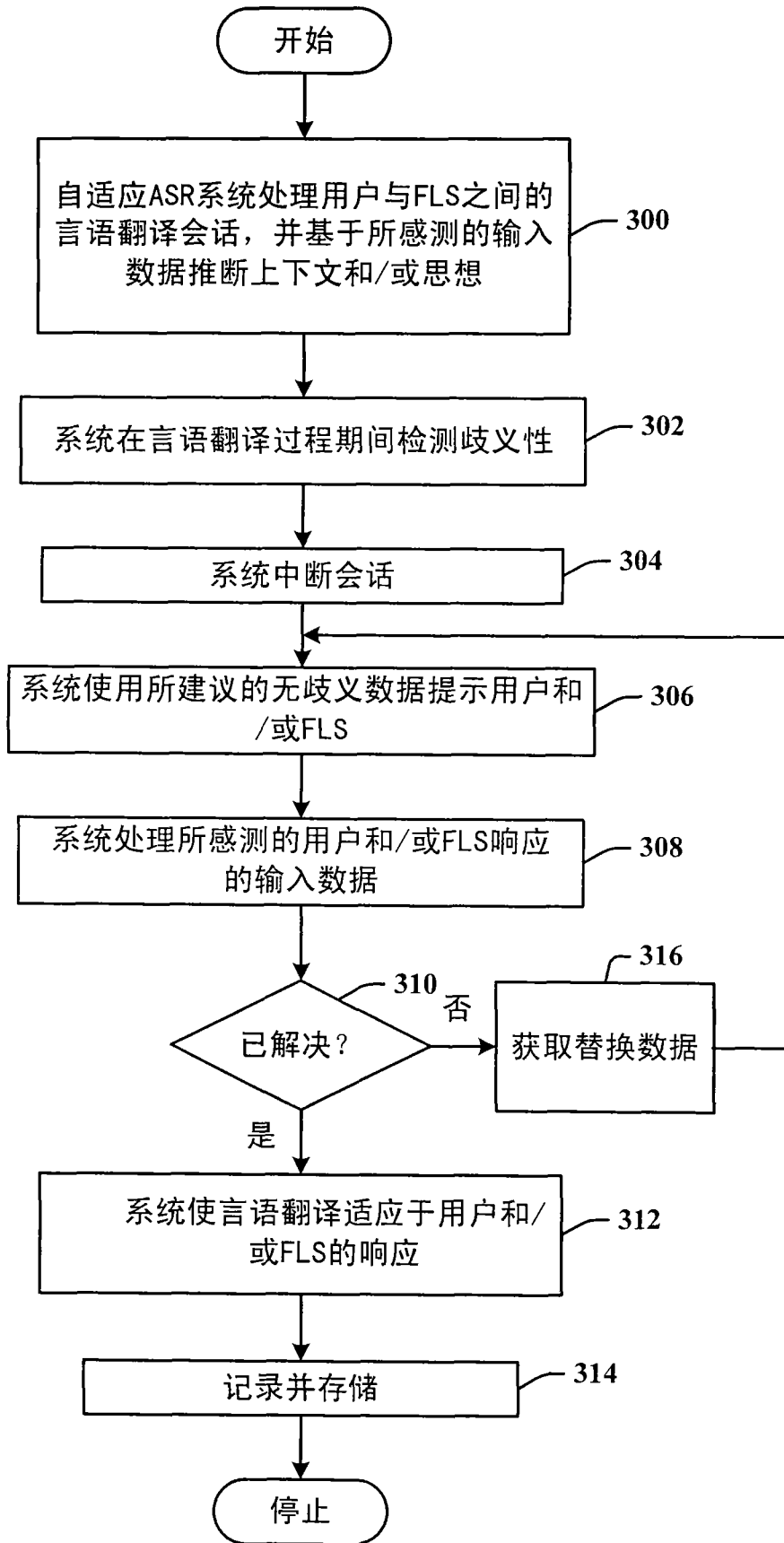


图 3

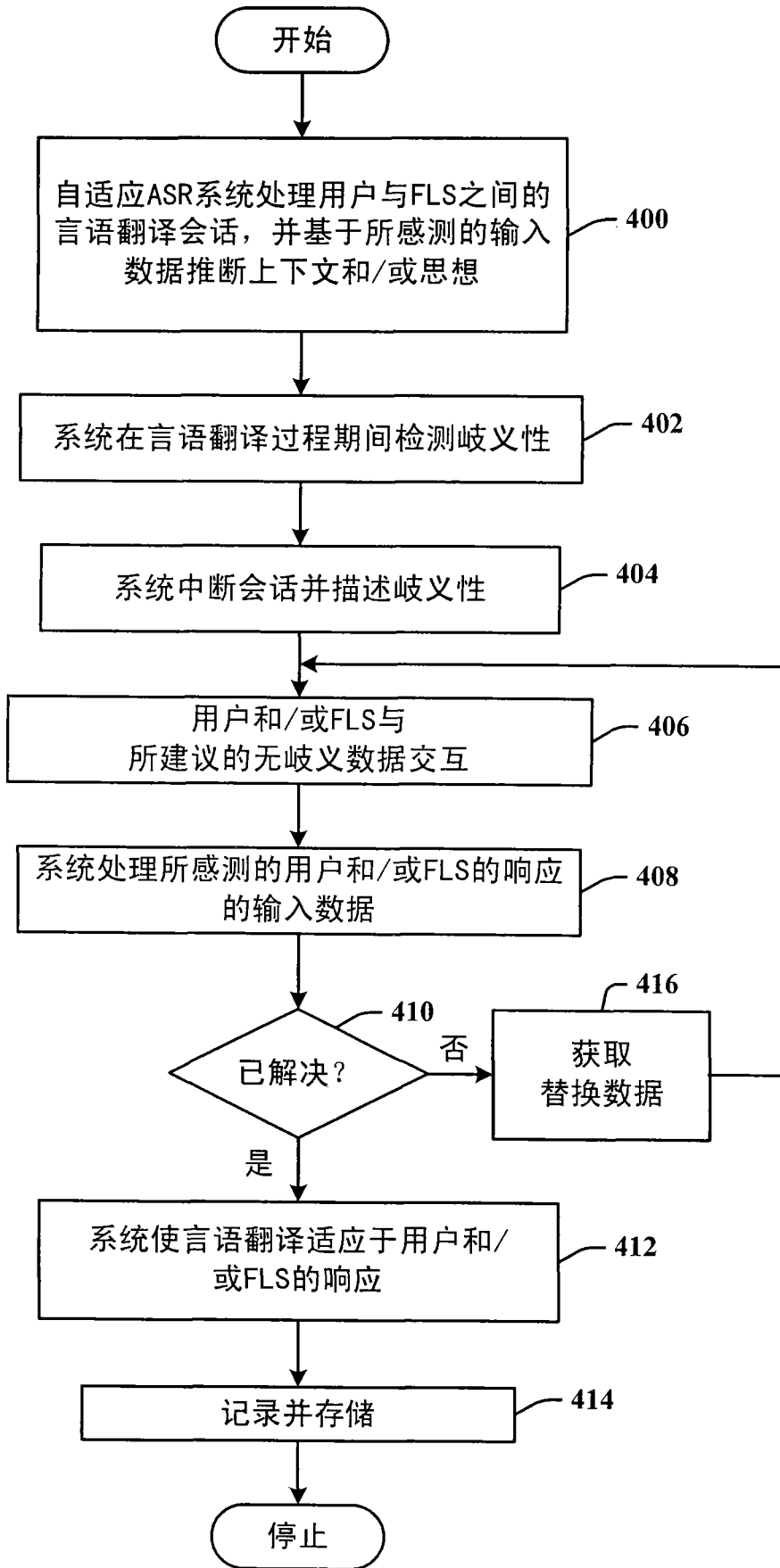


图 4

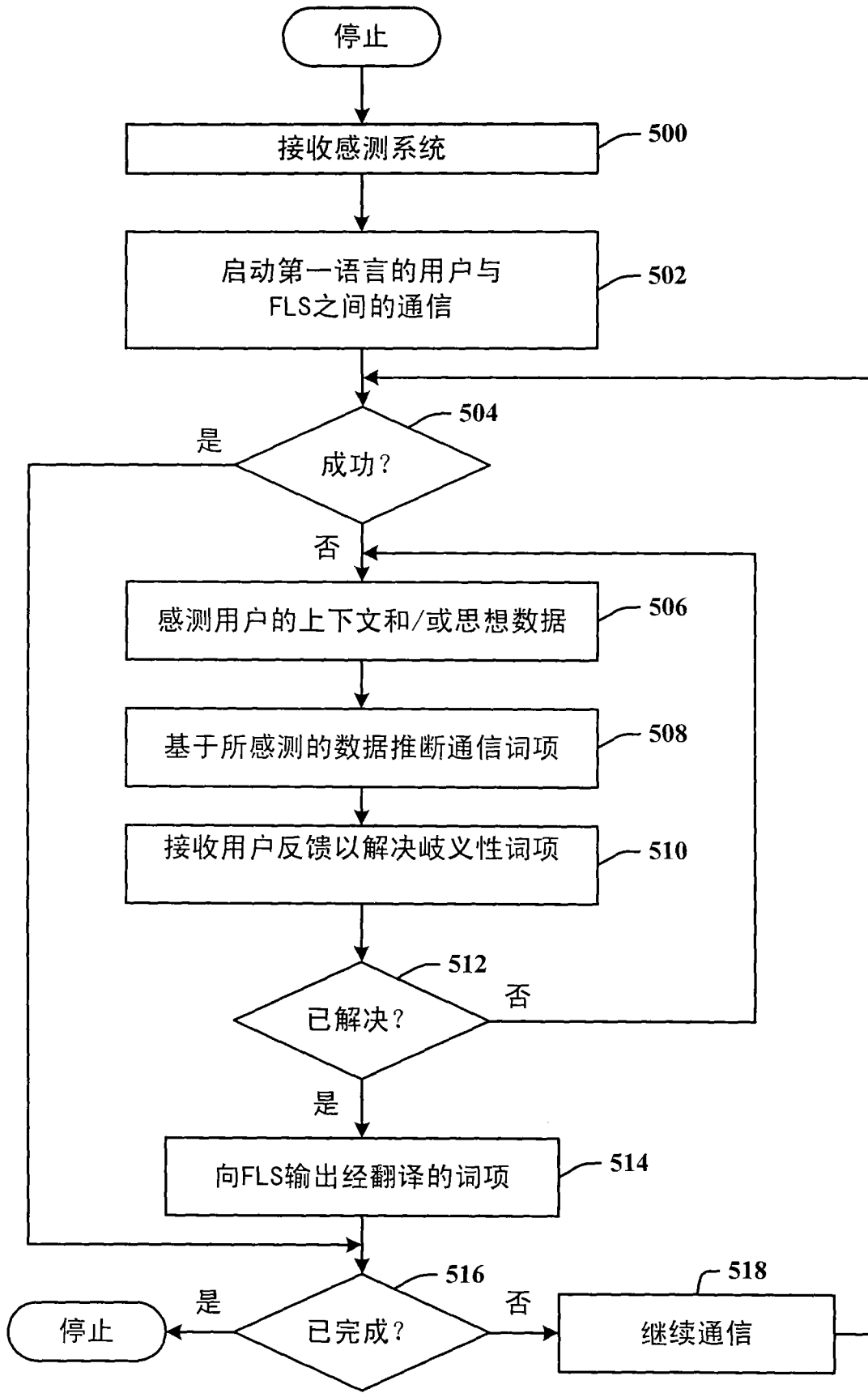


图 5

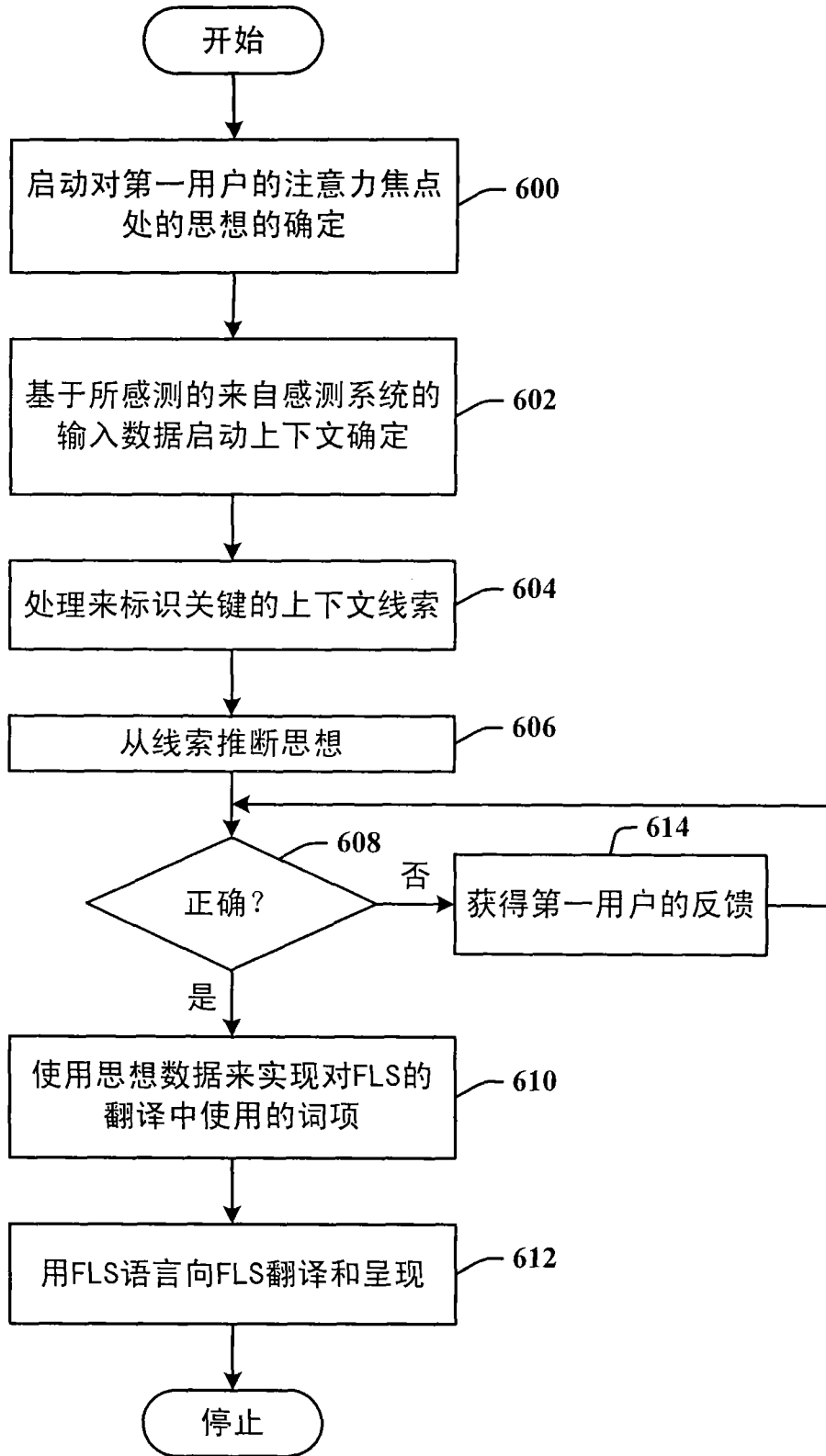


图 6

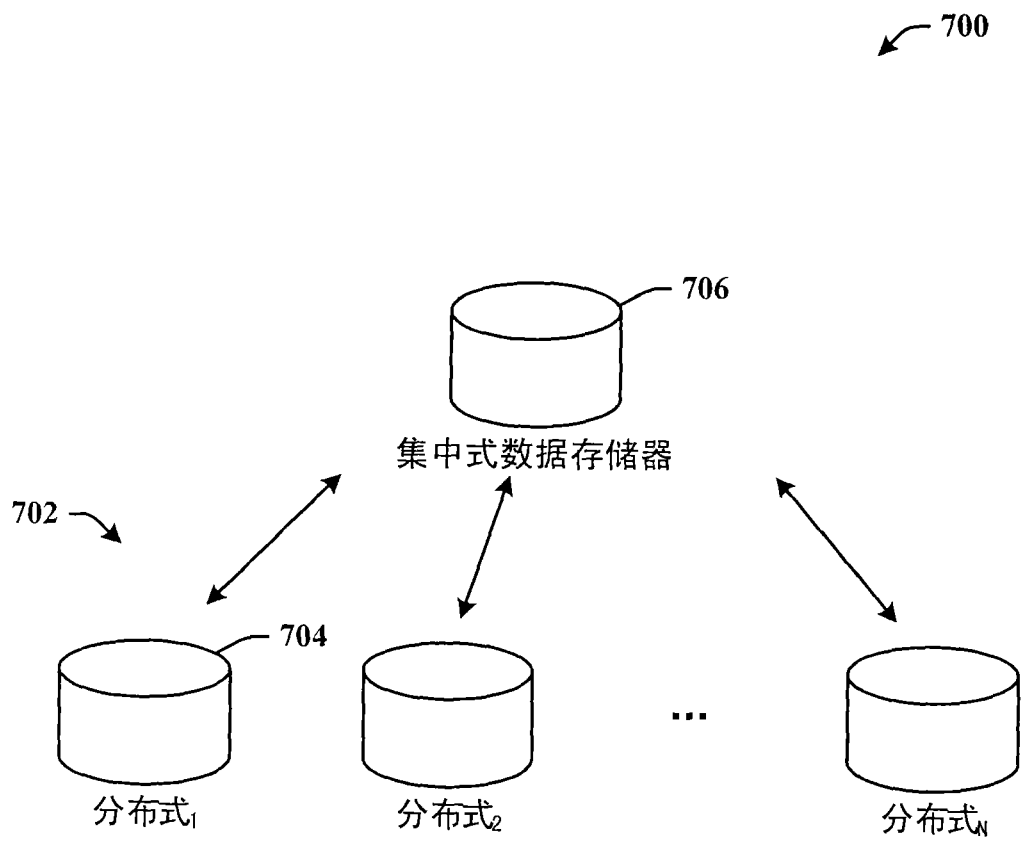


图 7

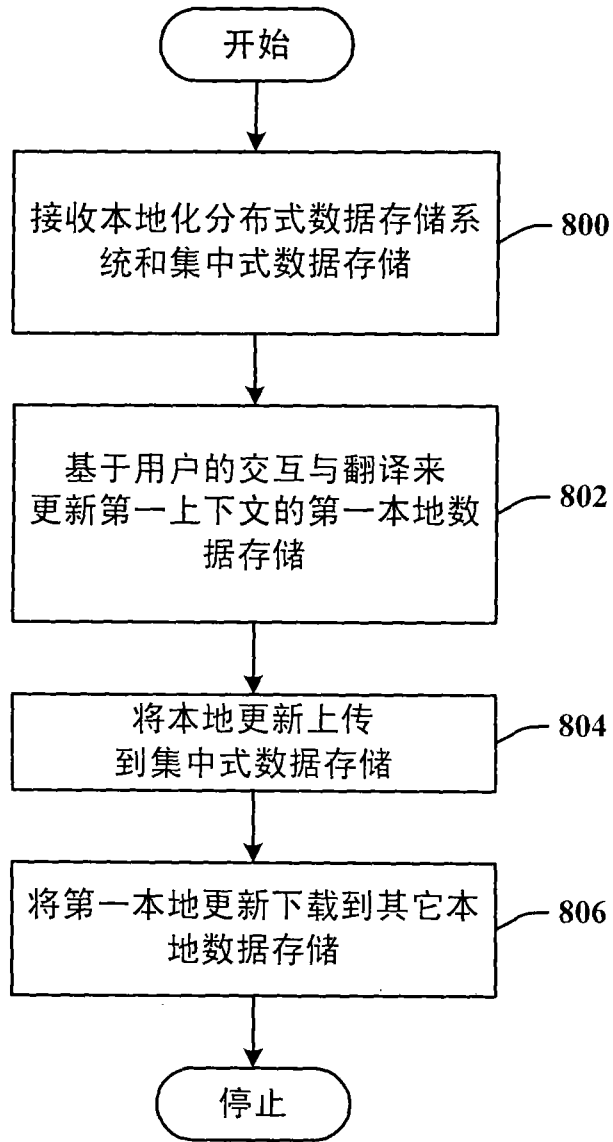


图 8

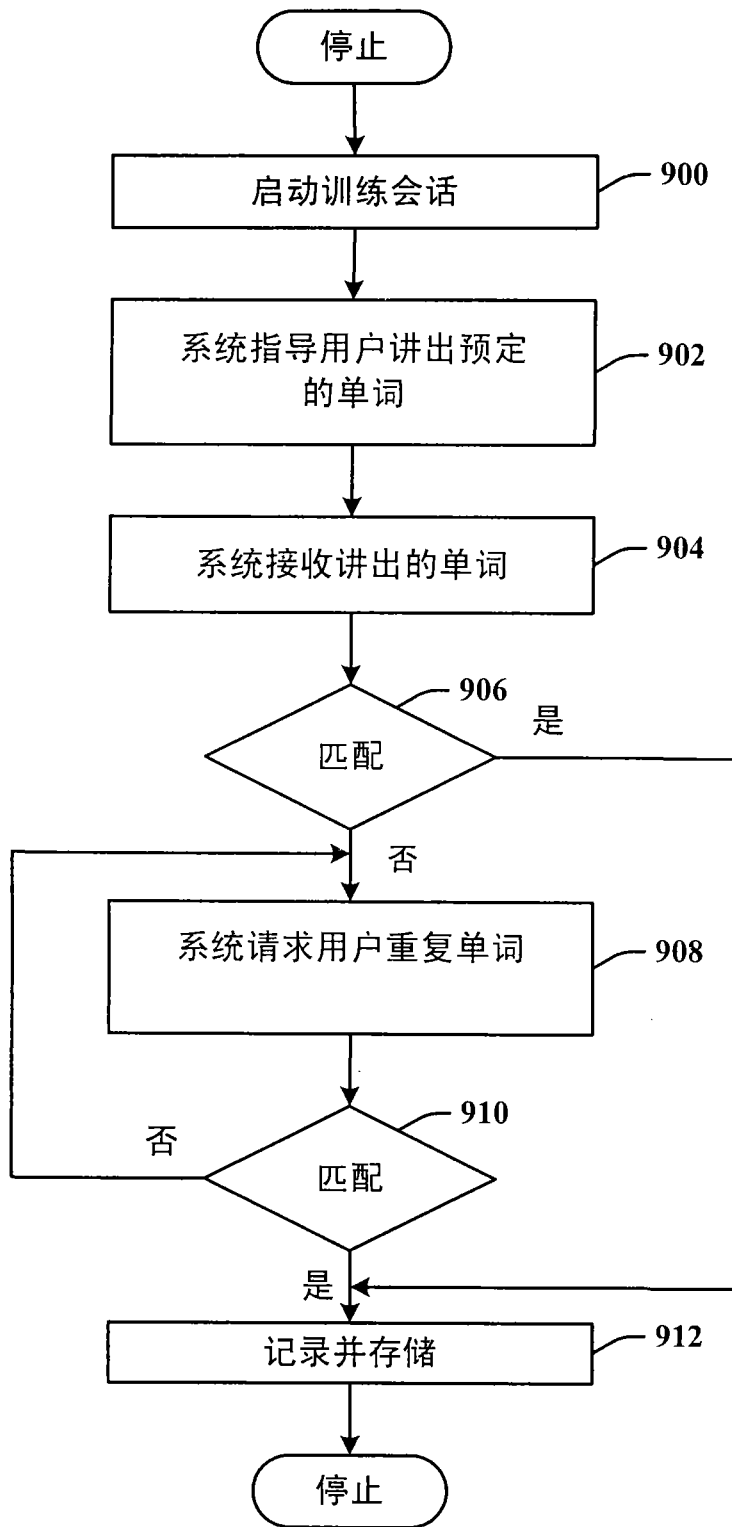


图 9

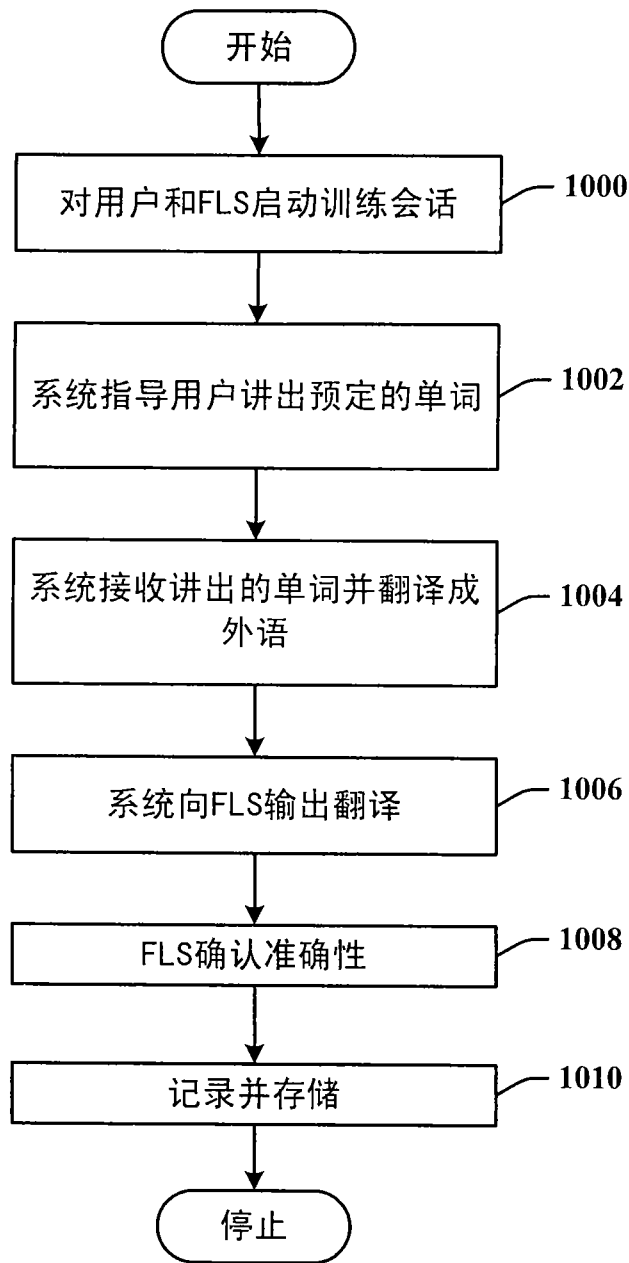


图 10

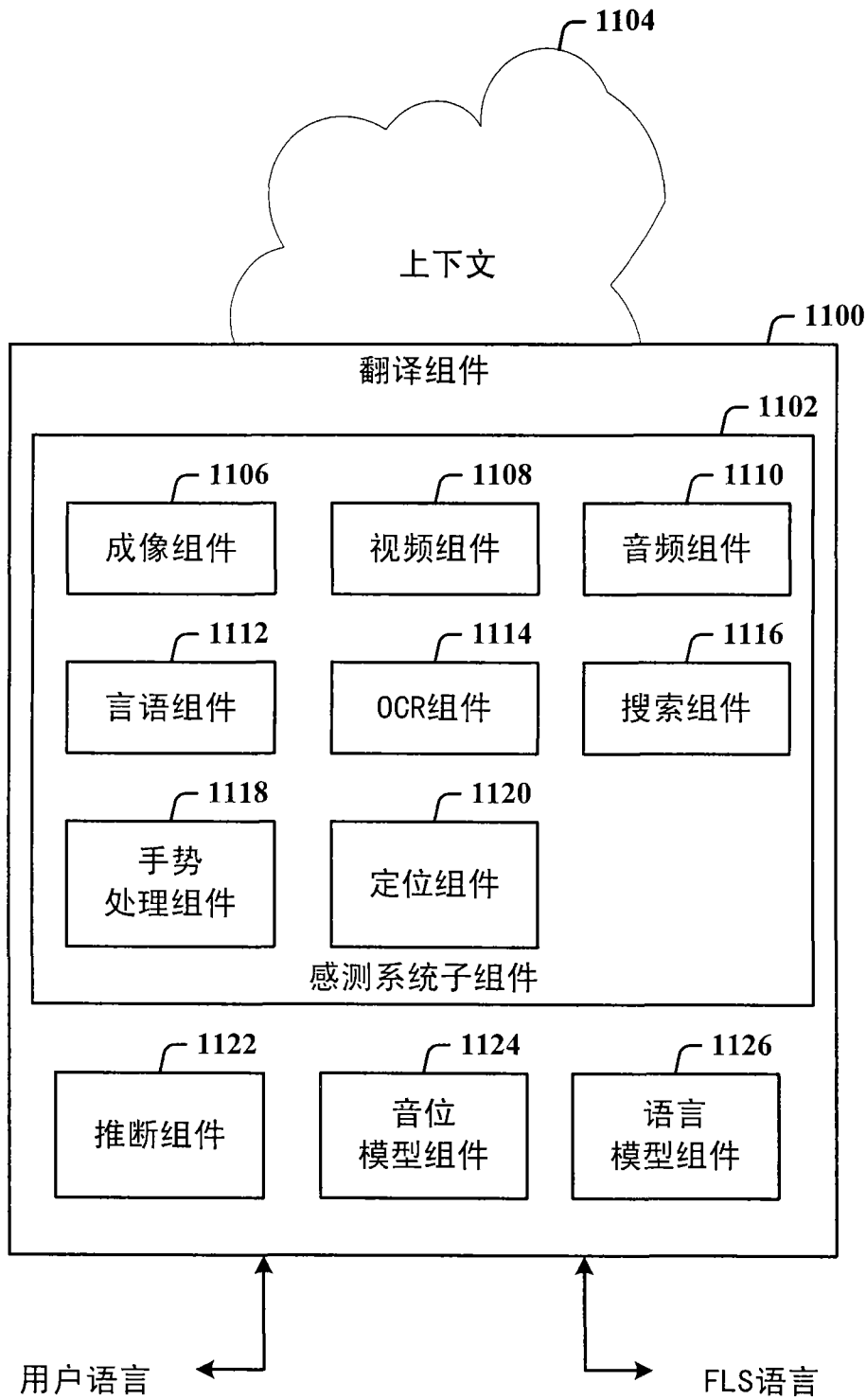


图 11

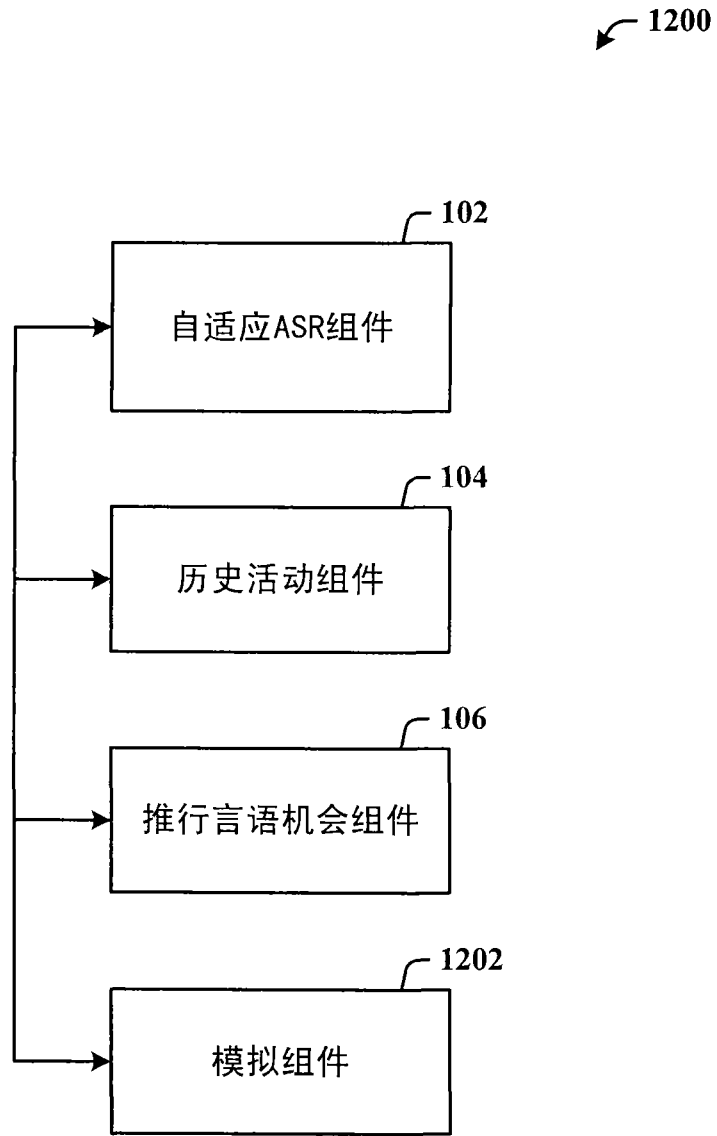


图 12

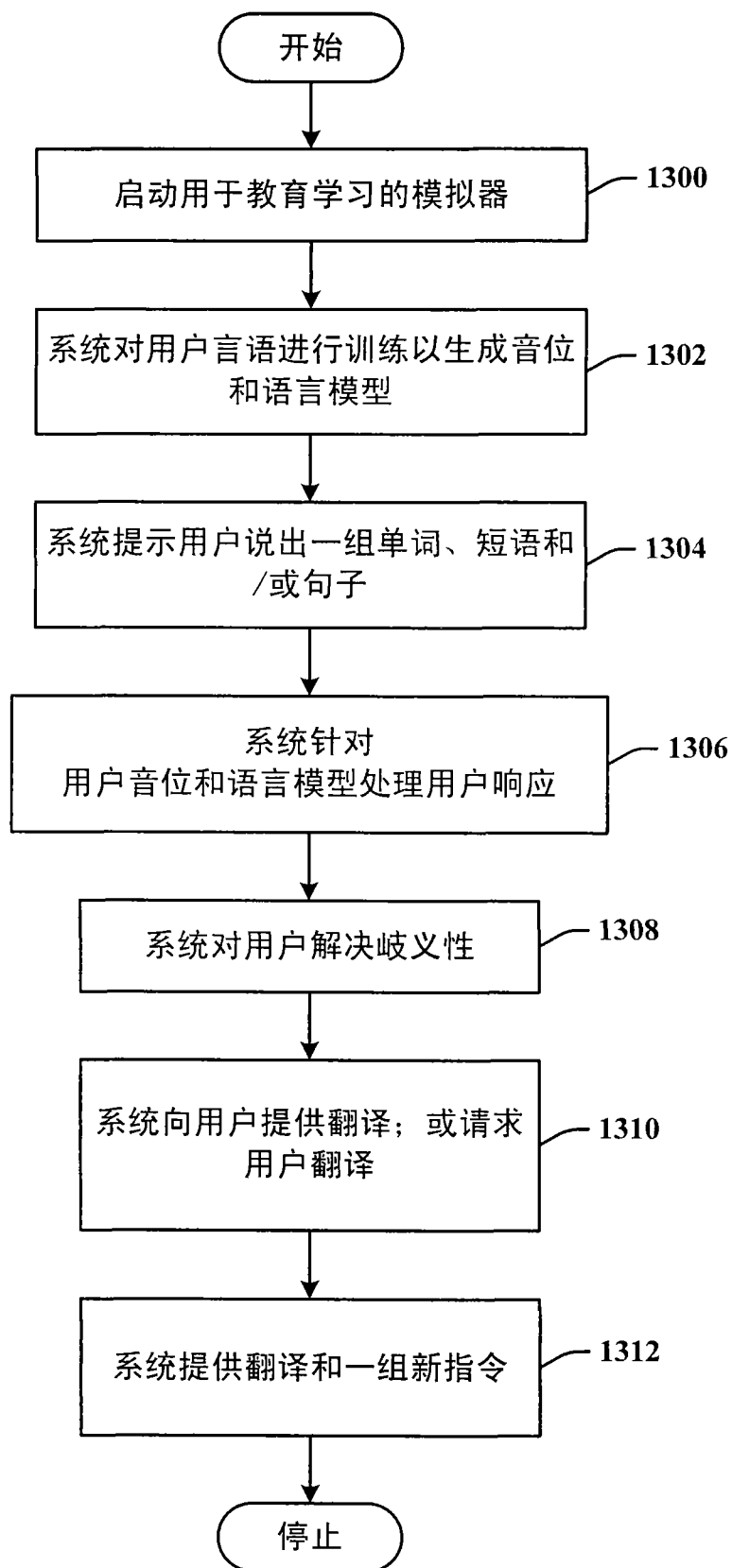


图 13

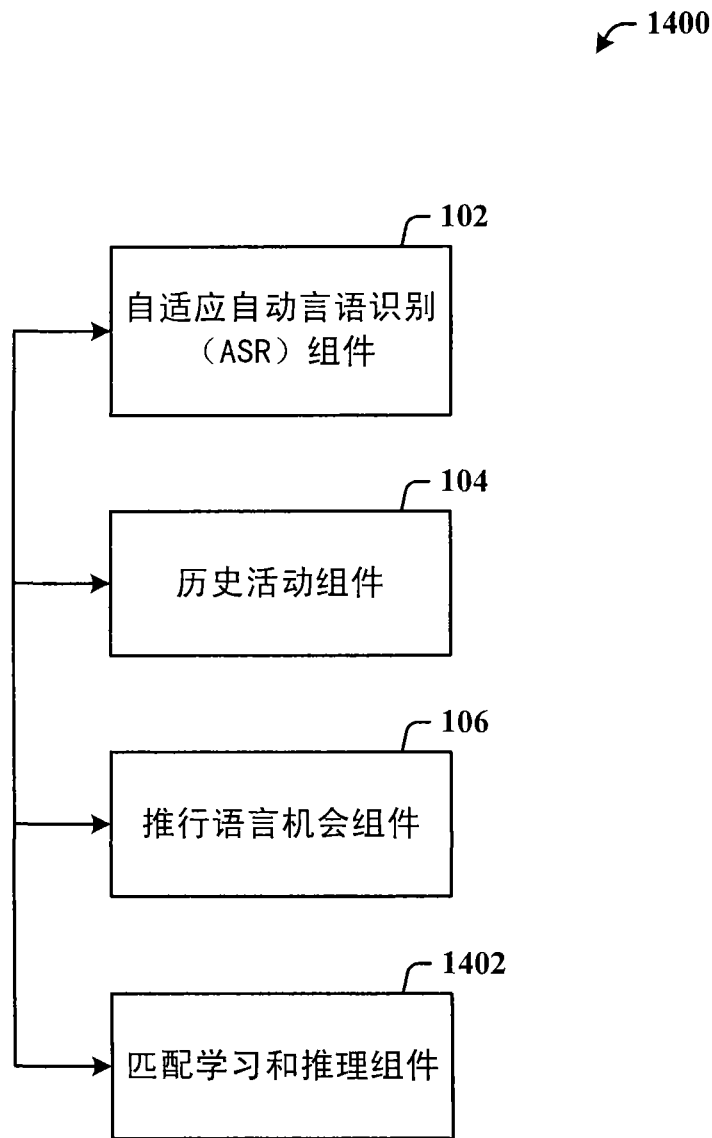


图 14

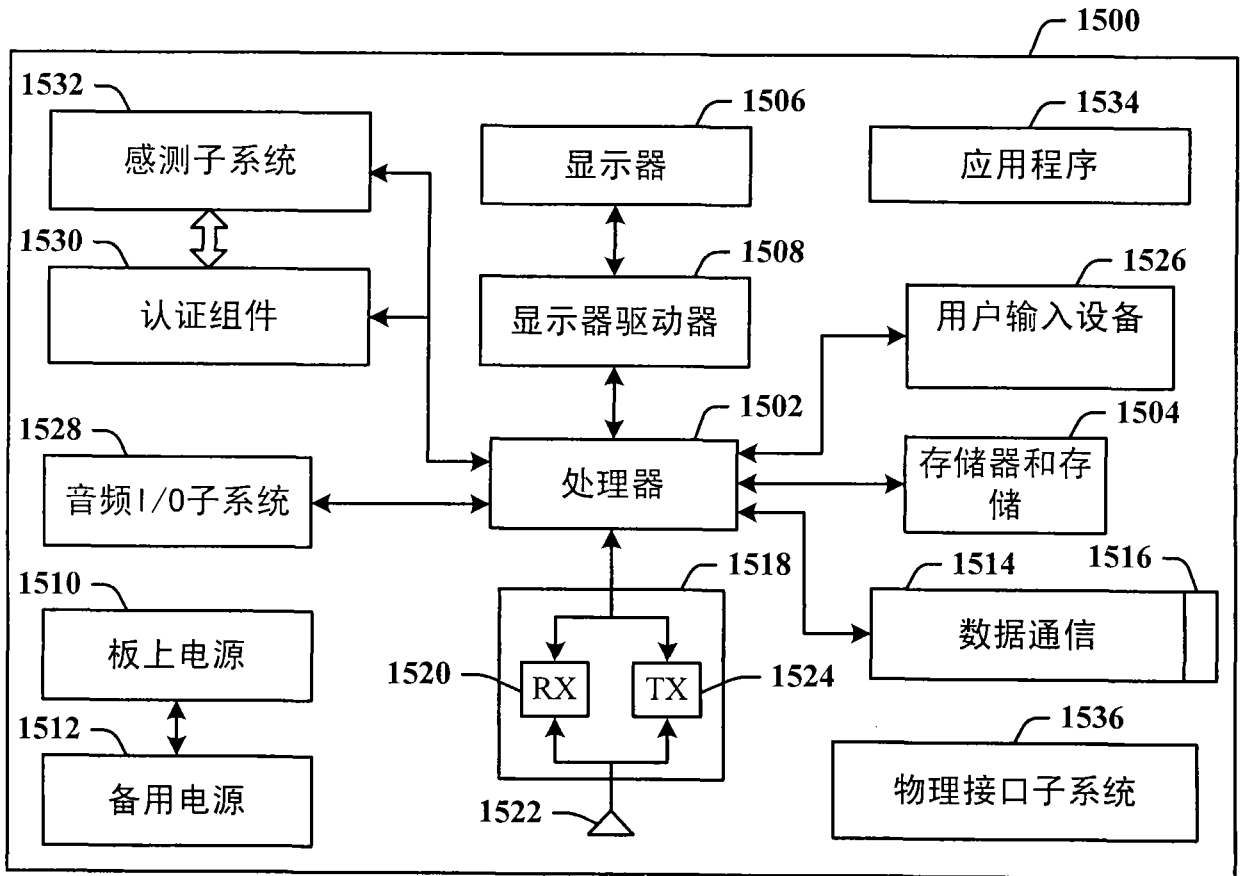


图 15

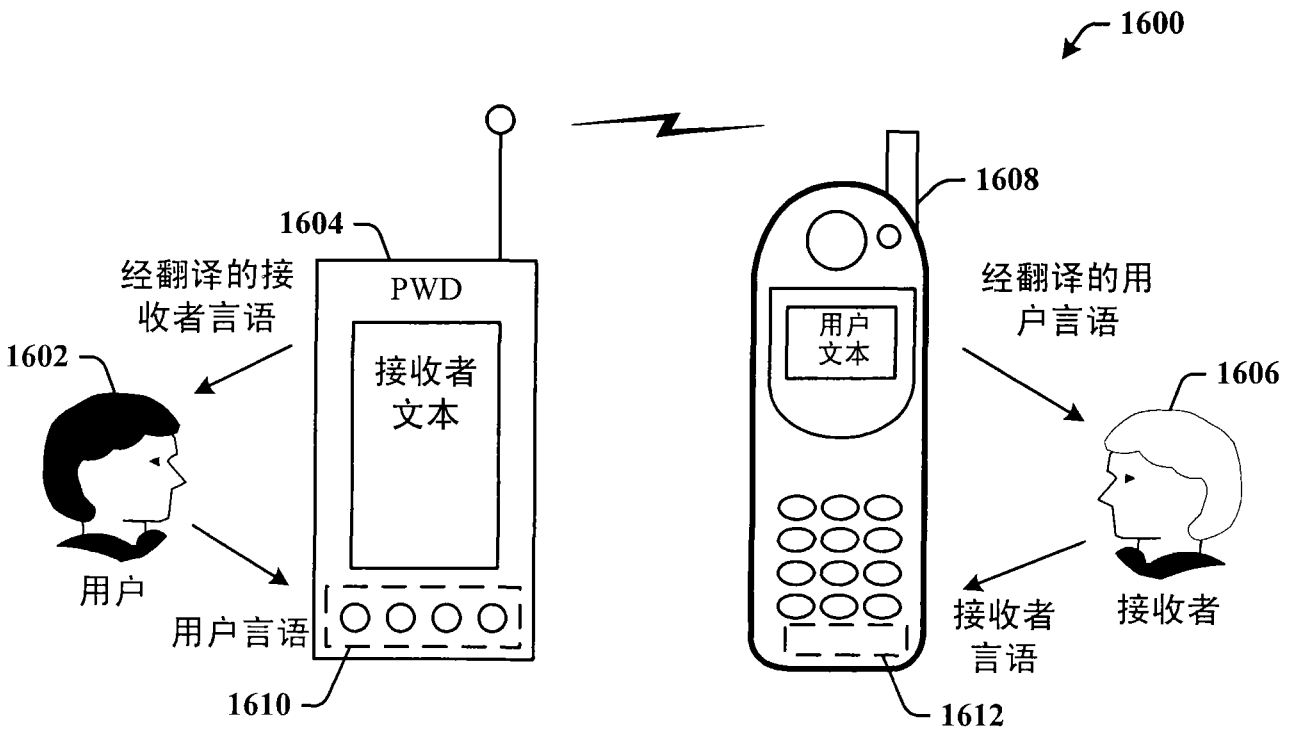


图 16

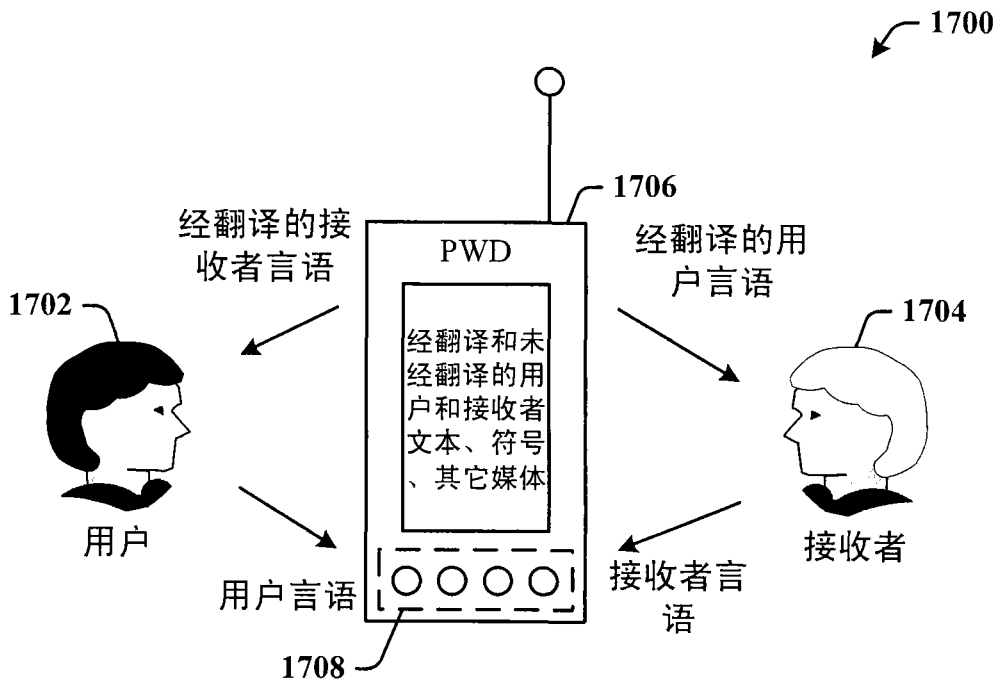


图 17

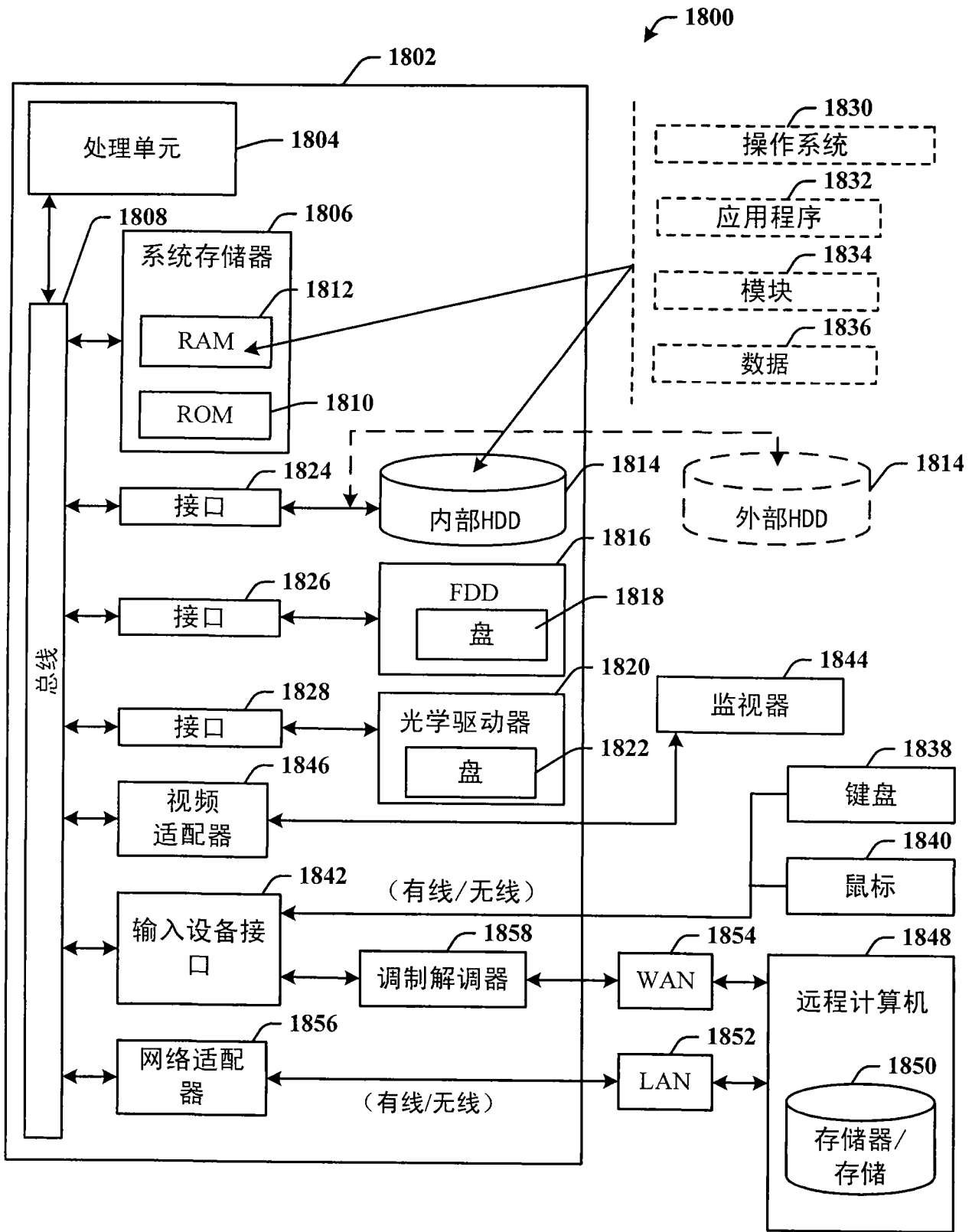


图 18

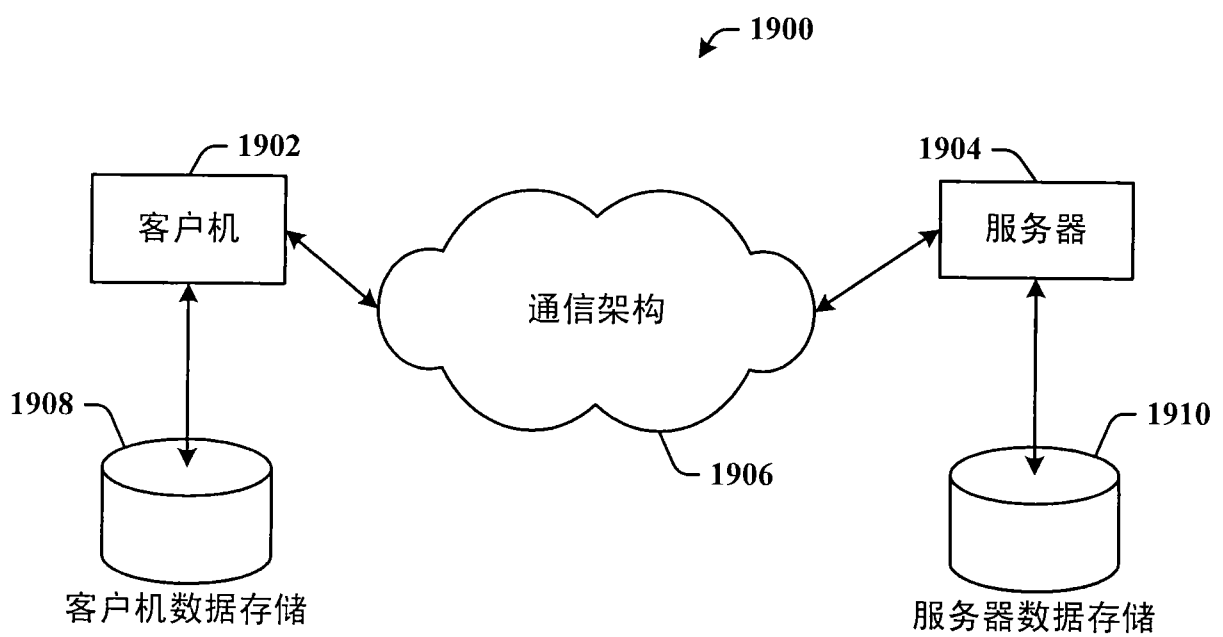


图 19