

[12] 发明专利申请公开说明书

[21] 申请号 99811616.5

[43] 公开日 2001 年 11 月 21 日

[11] 公开号 CN 1323435A

[22] 申请日 1999.10.1 [21] 申请号 99811616.5

[30] 优先权

[32] 1998.10.2 [33] US [31] 60/102,957

[32] 1999.1.27 [33] US [31] 60/117,595

[86] 国际申请 PCT/US99/22925 1999.10.1

[87] 国际公布 WO00/21075 英 2000.4.13

[85] 进入国家阶段日期 2001.3.30

[71] 申请人 国际商业机器公司

地址 美国纽约

[72] 发明人 斯蒂芬·H·梅斯

波纳尼·格帕拉克里世南

[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

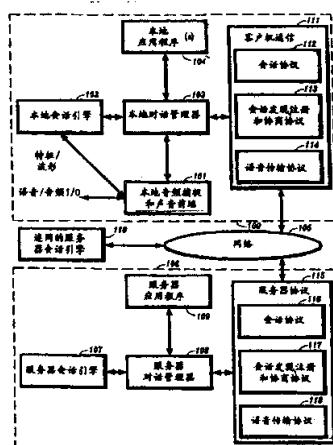
代理人 王以平

权利要求书 4 页 说明书 23 页 附图页数 5 页

[54] 发明名称 提供网络协同会话服务的系统和方法

[57] 摘要

一个在网络连接的服务器和设备及它们相应的应用程序之间提供自动和协同共享会话资源，例如功能和主目，的系统和方法。一方面，一个提供自动和协同共享会话资源的系统包含：包含第一（100）和第二（106）网络设备的网络；此第一（100）和第二（106）网络设备每个包含一组会话资源（102、107）、一对话管理器（103、108）用来管理一会话和执行调用请求会话服务，以及一通信栈（111、115）用以使用会话协议在网络上传递消息，其中会话协议在第一和第二设备的对话管理器之间建立协同的网络通信来自动地共享第一和第二网络设备的一组会话资源，在需要时，执行它们各自所请求的会话服务。



权 利 要 求 书

1. 一种提供会话资源的自动和协同共享的系统，包括：
一个包含至少第一和第二网络设备的网络；
第一和第二网络设备每个包括：
一组会话资源；
一对话管理器，用于管理会话，并执行调用请求会话服务；及
一通信栈，使用会话协议在网络上传递消息，其中使用会话协议
传递的消息，在第一和第二网络设备的对话管理器之间建立协同的网
络通信，以自动地共享第一和第二网络设备的一组会话资源，当需要
时，执行它们各自请求的会话服务。
2. 按照权利要求 1 的系统，其中所说的第一和第二网络设备的一
组会话资源集包括至少一语音识别引擎、一说话者识别引擎、一文本
到语音合成 TTS 引擎、一自然语言理解 NLU 引擎、一自然语言产生 NLG
引擎、一个音频捕捉和压缩/解压缩引擎、一主题识别引擎、一音频/
多媒体索引和查找引擎，及其组合。
3. 按照权利要求 1 的系统，其中的会话协议包括协同协议，以允
许第一和第二设备的对话管理器交换包括它们各自的会话状态、主目
和上下文信息，并交换对话组件。
4. 按照权利要求 3 的系统，其中的协同协议用于对第一和第二设
备对话管理器之间的主/从和对等网络通信中的一个进行协同。
5. 按照权利要求 1 的系统，其中的会话协议包括发现协议，以允
许第一和第二设备发现明白会话的设备及在网络上的应用程序。
6. 按照权利要求 5 的系统，其中的发现协议实现“广播和收听”
方法。
7. 按照权利要求 6 的系统，其中的发现协议被执行，以至少在第
一和第二网络设备之间建立动态和自发的网络。
8. 按照权利要求 1 的系统，其中的会话协议包括注册协议，用来
交换关于会话资源、能力、及需求的信息。

9. 按照权利要求 8 的系统，其中的会话协议包括协商协议，用来交换信息，以便根据它们各自会话资源、能力，在第一和第二网络设备之间建立一网络配置。

10. 按照权利要求 9 的系统，其中的网络配置包括主/从网络和对等网络之一。主/从网络中第一和第二设备之一的对话管理器控制第一和第二设备两者的会话资源；对等网络中第一和第二设备的对话管理器协商以控制会话资源。

11. 按照权利要求 1 的系统，其中的会话协议包括语音传输协议，用于在第一和第二设备之间传送压缩的语音波形，压缩的语音特征及压缩的结果之一。

12. 一种提供会话资源的自动和协同共享的系统，包括：

一客户机，包含本地会话资源和一对话管理器，其中对话管理器用以管理本地会话资源、处理对会话服务请求，并判断是否会话服务请求能用本地会话资源实现；及

一包含服务器会话资源的服务器，其中如果使用本地会话资源不能执行请求的会话服务，客户机的对话管理器将自动地访问服务器，以便用服务器的会话资源处理。

13. 一种在网络设备之间提供会话资源的自动和协同共享的方法，包括以下步骤：

由第一网络设备接收一会话服务请求；

由第一网络设备判断用哪一种方式：是本地地用第一网络设备的会话资源，还是远程地用至少一第二网络设备的会话资源，或是本地地和远程地用本地和远程两者的会话资源去处理请求的会话服务；及

如果判断出会话服务至少一部分要远程地使用至少第二网络设备的会话资源而被处理，则自动地同至少第二网络设备通信。

14. 按照权利要求 13 的方法，还包括使用会话协议传输消息的步骤，以便在第一和至少第二网络设备之间建立协同的网络通信，以共享会话资源。

15. 按照权利要求 13 的方法，其中的判断步骤包括判断是否本地

的会话资源可用来处理请求的会话服务的步骤；并且该方法还包括步骤：

如果判断本地的会话资源可用来处理请求的会话服务，使用本地的会话资源执行此请求的会话服务；

判断本地处理的结果是否可接受；及

如果确定本地处理的结果不可接受，自动地访问至少第二网络设备，远程地处理此请求的会话服务。

16. 按照权利要求 13 的方法，其中的判断步骤包括判断是否至少第二网络设备是由第一网络设备预先指派，以处理会话服务的步骤。

17. 按照权利要求 13 的方法，其中的判断步骤是基于网络连接的可用性和第一网络设备和至少第二网络设备之间的网络通信量之一。

18. 按照权利要求 14 的方法，其中的自动访问步骤包括以下步骤：

通过使用会话协议传输的消息，同至少第二网络设备自动建立网络连结；及

把压缩的语音特征和压缩的波形之一传送到至少第二网络设备。

19. 一种通过机器可读的程序存储器设备，确实地嵌入机器可执行的程序指令，以实现在网络设备之间提供会话资源的自动和协同共享的方法步骤，这些方法步骤包括：

由第一网络设备接收一会话服务请求；

由第一网络设备判断用哪一种方式：是本地地用第一网络设备的会话资源，还是远程地用至少一第二网络设备的会话资源，或是本地地和远程地用本地和远程两者的会话资源去处理请求的会话服务；及

如果判断出会话服务至少一部分要远程地使用至少第二网络设备的会话资源而被处理，自动地同至少第二网络设备通信。

20. 按照权利要求 19 的程序存储器设备，还包括按会话协议执行传输消息的步骤的指令，以便在第一和至少第二网络设备之间建立协同的网络通信，共享会话资源。

21. 按照权利要求 19 的程序存储器设备，其中执行判断步骤的指令包括执行判断是否本地的会话资源可用来处理请求的会话服务的指

令；并且其中程序存储器设备还包括执行下面步骤的指令：

如果判断本地的会话资源可用来处理请求的会话服务，使用本地的会话资源执行此请求的会话服务；

判断本地处理的结果是否可接受；及

如果确定本地处理的结果不可接受，自动地访问至少第二网络设备，远程地处理此请求的会话服务。

22. 按照权利要求 19 的方法，其中执行判断步骤的指令包括执行判断是否至少第二网络设备是由第一网络设备预先指派，以处理此会话服务的步骤的指令。

23. 按照权利要求 19 的程序存储器设备，其中的判断步骤是基于网络连接的可用性和第一网络设备和至少第二网络设备之间的网络通信量之一。

24. 按照权利要求 20 的方法，其中执行自动访问步骤的指令包括执行下列步骤的指令：

使用会话协议，同至少第二网络设备自动建立网络连结；及
把压缩的语音特征和压缩的波形之一传送到至少第二网络设备。

说 明 书

提供网络协同会话服务的系统和方法

本申请基于 1998 年 10 月 2 日递交的第 60/102,957 号美国临时申请和 1999 年 1 月 27 日递交的第 60/117,595 号美国临时申请。

本申请一般地涉及会话系统，并特别涉及一种用以在网络连接的设备、服务器和应用程序之间自动及协同共享会话功能/资源的系统和方法。

传统的会话系统（即，单纯具有语音 I/O 的系统或者具有语音 I/O 的多模式系统）一般地被限于个人计算机（PC）和具有合适结构和足够处理能力的本地机器。另一方面，对于电话技术应用，会话系统通常位于一服务器上（例如，IVR 服务器）并可以经由传统电话和蜂窝电话而能访问。尽管这种会话系统变得更为流行，有代表性的是所有的会话处理或者在客户机端或者在服务器端执行（即，所有的配置或者是完全本地的或者是完全客户机/服务器式）。

随着普及计算的出现，期望数十亿计的低资源客户设备（例如，个人数据助理 PDA，智能电话等）将被连网在一起。由于客户设备尺寸的减小和用户所期望这种设备所执行任务的复杂性的增加，传统的图形用户界面在这种小型的用户设备上不切实际，所以用户界面成为一个关键性的问题。因此，期望会话系统将作为用户界面的主要单元来提供纯语音/音频 I/O 或带有语音/音频 I/O 的多模式 I/O。

从而，在便携式客户设备上，语音嵌入的会话应用正得到发展并进入成熟。不幸的是，由于资源有限，预期这种客户设备可能不能执行复杂的会话服务，例如语音识别（尤其是当词汇量非常大或特殊或者需要领域特定/应用特定的语言模型或语法时）、NLU（自然语言理解）、NLG（自然语言产生）、TTS（文本到语音合成）、音频捕捉和压缩/解压缩、重放、对话产生、对话管理、说话者识别、主题识别、音频/多媒体索引和查找等。例如一个设备的存储器和 CPU（和其他资源）

的限制可以限制这种设备能提供的会话能力。

此外，即使一已联网设备功能足够强大（在 CPU 和存储器方面）以便执行所有这些会话任务，但该设备可能没有合适的会话资源（例如，引擎）或者会话主目（argument, 即被引擎用到的数据文件）（例如，语法、语言模型、词汇表文件、语法分析、标记、声纹和 TTS 规则等）来执行合适任务。确实，一些会话功能可能太特殊而适于特定的服务，因此要求仅能从网络上的其它设备或者机器才可得到的后端信息。例如，因为用来产生对话需要的整组会话主目或者功能（例如语法分析器、标记器、翻译器等）或者需要很大量存储器来存储（在客户设备上不能得到）或者太广泛（依据通信带宽）以至于不能传递到客户机端，所以一个客户设备上的 NLU 和 NLG 服务典型地需要服务器端支持。这个问题在多语种应用时被进一步加重，当一个客户设备或者本地应用程序没有足够的存储器或者处理能力来存储和处理这些主目，这些主目是需要的，以便按多语种来处理语音和执行会话功能。代之，用户必须手动连接到远程的服务器上来执行这种任务。

同时，与客户机和服务器间的分布式结构和分布式处理有关的问题需要新的会话连网方法。这种方法包含了通信量和分布在网上的资源的管理，以保证参与该网络上会话交互的每一个用户的合适的对话流。

因而，允许一个有有限资源的网络设备自动利用连网资源、以一种对用户来说自动的和透明的方式来执行复杂的会话任务的系统和方法是非常需要的。

本发明针对一种在网络连接的服务器和设备（和它们相应的应用程序）之间提供自动的和协同的共享会话资源的系统和方法。根据本发明的一个实施例，系统包含多个连网的服务器、设备和/或应用程序，通过使用会话网络协议（或者方法）传达消息，使得相互之间“明白会话”，协议自动地允许每一明白会话的网络设备自动地并以协同和同步的方式共享会话资源，以便通过某一网络设备之一的界面来提供一个无缝的会话界面。

根据本发明的一个方面，一个提供自动的和协同的共享会话资源的系统，包含：

一个网络，它包含至少第一和第二网络设备；

这第一和第二网络设备每个包含

一组会话资源；

一个对话管理器用以管理会话并执行调用来请求会话服务；以及一使用会话协议在整个网络上传达消息的通信栈，在此通过会话协议传达的消息在第一和第二设备的对话管理器之间建立协同的网络通信，以自动地共享第一和第二网络设备成组的会话资源，当需要时，去执行它们各自请求的会话服务。

本发明允许一个低资源的客户设备透明地本地执行简单的任务，又以二进制或模拟方式与一具有更复杂会话能力的服务器（或者别的设备）相连接透明地执行复杂任务。服务器端的功能（例如语音识别）可以通过一个常规的 IP 网络或者 LAN 网络，也可以经由传统的电话线路或包交换网络数字传输，或者在无线网络上经由任何传统的无线数据协议而能被执行。

有利的是，本发明在任何有限 CPU、存储器和处理能力（也含有限的会话资源）的设备（如盛行的嵌入式设备）上提供一完全成熟的会话用户界面，这使利用低资源的客户设备不需要下载，例如来自一网络服务器所必需的会话主目，便可提供复杂的会话服务。本地能力允许用户利用本地设备而不需请求连接，例如无线电话供应商的外部覆盖。同时，持续连接的花费也减小，并且当持续连接丢失时恢复的难度能够减轻。

在接下来的优选实施例的详细描述中，本发明的这些和其他方面、特征及优点将会被描述并变得明了。

图 1 是根据本发明的一个实施例在连网的设备之间通过自动的和协同的共享会话资源来提供会话服务的一个系统框图；

图 2 是根据本发明的一个方面在连网的设备之间通过自动的和协同的共享会话资源来提供会话服务的一个方法流程图；

图 3 是根据本发明的另一个方面在连网的设备之间通过自动的和协同的共享会话资源来提供会话服务的一个方法流程图；

图 4 是根据本发明采用会话浏览器的另一个实施例提供会话服务的分布式系统的一个框图；以及

图 5 是根据本发明采用会话浏览器的另一个实施例提供会话服务的分布式系统的一个框图。

将会了解到本发明可以被以不同形式的硬件、软件、固件、专用处理器或者它们的组合实现。更可取地，本发明作为一个包含程序指令的应用程序以软件实现，它们确实地嵌入在程序存储器设备上（例如软盘、RAM、CDROM、ROM 和闪存器），并且可以被任何设备或者包含合适结构的机器执行，如一个或者多个中央处理器（CPU）、一随机存取存储器（RAM）和音频输入/输出（I/O）接口。

将会进一步了解，因为在附图中描述的组成系统的某些部件和方法步骤优选以软件实现，系统组件之间的实际联系（或者处理步骤）随本发明被编程的方式而可能不同。此间给出的说明，一个相关领域的一般技术人员将能设想到本发明的这些和相似的实现方法或结构配置。

现参看图 1，框图阐明了根据本发明的一个具体范例在连网设备之间通过自动的和协同的共享会话资源和会话主目（数据文件）来提供会话服务的一个系统。该系统包含一本本地客户设备 100，客户设备 100 包含一个声学前端 101 来处理音频/语音输入和输出由客户设备 100 产生的音频/语音。例如，客户设备 100 可以是一个智能电话或任何能处理语音的 PDA（个人数字助理）。客户设备 100 还包含一个或者多个本地会话引擎 102 用来处理声特征和/或由声学前端 101 产生和/或捕捉的波形并产生对话输出给用户。本地会话引擎 102 可以包含，例如，一个嵌入式语音识别，一个说话者识别引擎，一个 TTS 引擎，一个 NLU 和 NLG 引擎和一个音频捕捉和压缩/解压缩引擎及任何其它类型的会话引擎。

客户设备 100 还包含一个本地对话管理器 103 它执行任务管理并

控制和协同一经由系统调用 (API 或者协议调用) 请求的会话服务 (或者本地的或者经由网络设备) 的执行, 同时也管理本地的和与连网设备的会话。更特别地, 如下所详细描述的, 对话管理器 103 确定是否一个给定的会话服务将在客户 100 上本地地或者是在远程网络连接的服务器 (或者设备) 上处理和执行。这个决定是基于如客户 100 的会话能力和其它网络连接的设备能力的比较等因素, 也基于为处理这一请求的会话服务可能必需的可用资源和会话主目。其它因素包括从连网的设备接收结果时网络通信量和预期的延迟。对话管理器 103 执行任务管理和资源管理任务, 如装载管理和资源分配, 同时也管理本地会话引擎 102 和能处理语音的本地应用程序 104 之间的对话。

如图 1 的示例, 客户设备 100 经由网络 105 与服务器 106 连网, 服务器 106 包含服务器应用程序 109, 也包含向客户设备 100 (或者任何其它网络设备或者应用程序) 在需要时提供会话服务的服务器会话引擎 107。和本地会话服务器引擎 102 一样, 服务器引擎 107 可以包含如一个嵌入式语音识别、一个 TTS 引擎、一个 NLU 和 NLG 引擎和一个音频捕捉和压缩/解压缩引擎及任何其它类型的会话引擎。服务器 106 包含一个服务器对话管理器 108, 服务器对话管理器 108 操作方式和上面描述的本地对话管理器 103 相同。例如, 服务器对话管理器 108 确定是否一个从本地对话管理器 103 来的一个会话服务请求被服务器 106 或者是远程网络连接的服务器或者设备处理和执行。此外, 服务器对话管理器 108 管理服务器会话引擎 107 和能处理语音的服务器应用程序 109 之间的对话。

图 1 的系统进一步描述了客户设备 100 和远程服务器 106 被连网到一个有会话引擎和/或能被客户 100 和服务器 106 在需要时存取的会话主目的服务器 110 上。网络 105 可以是如因特网、LAN (局域网)、公司内部网、PSTN (公共交换电话网)、或者无线网 (经由 RF (射频)、或者 IR (红外线) 无线通信)。可以理解, 尽管图 1 描述了一个客户机/服务器系统, 其中术语可以被擅长该技术的人理解, 图 1 的系统可以包含许多连网的服务器、设备和相互“明白会话”的应用程序来提

供自动的和协同的共享会话功能、主目和资源。如下文更详细的说明，这种“明白会话”可以通过利用会话网络协议（或者方法）传输被各自的对话管理器处理的消息来达到，使连网的设备以一种自动的和同步的方式共享会话资源和功能。这种会话协同提供了一个无缝的会话界面，通过网络设备之一的界面来访问远程服务器、设备和应用程序。

特别是，在连网设备之间提供会话协同来共享它们的会话功能、资源和主目，每一连网设备利用会话协议（或者方法）传递消息来交换关于它们会话能力和需求的信息。例如，如图 1 所示，客户设备 100 包含一个利用会话协议 112、会话发现、注册和协商协议 113 和语音传输协议 114（或者会话编码协议）用以传送和接收消息的通信栈 111。同样，服务器 106 包含一个服务器通信栈 115，服务器通信栈 115 包含会话协议 116、会话发现、注册和协商协议 117 和语音传输协议 118。这些协议（方法）在同时递交的、名称为“Conversational Computing Via Conversational Virtual Machine”的专利申请（IBM Docket NO. Y0999-111p）中相对于 CVM（会话虚拟机器）得以详细讨论，此专利已一般转让，此处加以引用。

简单的说，会话协议 112、116（或者说是在 Y0999-111P 中称为“分布式会话协议”）是允许连网设备（例如，客户机 100 和服务器 106）或者应用程序与其它网络设备的对话管理器传输消息，以便注册它们的会话状态、主目和上下文的协议（或者方法）。会话协议 112、116 也允许设备交换其它信息如 Java 小应用程序、ActiveX 构件和其他可执行的编码，这些编码允许设备或者相关的应用程序来协调按照如主/从或对等会话网络配置的这些设备之间的会话。分布式会话协议 112、116 允许信息交换来协同包含多设备或者应用程序，包括主/从会话网络、对等会话网络和匿名伙伴的会话。在连网设备之间利用分布式会话协议可以交换的信息包含指向数据文件（主目）的指针、数据文件和其他会话主目的传输（如果需要）、输入通知、输出事件和识别结果、会话引擎 API 调用和结果、状态通知和上下文改变及其他系统事件、注册更新：注册协商更新的握手信号：协商握手信号、及当

一请求的资源丢失时的发现更新。

本（分布式）会话协议也包含对话管理器（DM）协议，此协议允许对话管理器去分配服务、行为和会话应用、I/O 和引擎 API 如在 IBM Docket NO. Y0999-111P 中所描述的。例如，DM 协议允下列信息被交换：(1) DM 体系结构注册(例如，每一 DM 可以是本地 DM 的汇集); (2) 与元信息关联的指针(用户、设备能力、应用需求等); (3) DM 网络拓扑协商(例如，主/从、对等); (4) 数据文件(会话主目)如果可用，即，如果引擎被利用则被一个主 DM 控制; (5) I/O 事件通知，如用户输入、输出到多个用户以传输到引擎和/或加到上下文; (6) 识别事件通知; (7) 传输被处理的输入从引擎到一主 DM; (8) 传输主 DM 的职责到注册的 DM; (9) DM 处理结果事件; (10) DM 异常; (11) 可信度和模糊性结果的传输，建议的反馈和输出，建议的期望状态，建议的操作，建议的上下文改变，建议的新的对话状态; (12) 通知决定，上下文更新，动作更新，状态更新等; (13) 完成、失败或者被中断动作通知; (14) 上下文改变通知；和/或者 (15) 由于动作引起的数据文件、上下文和状态的更新。

例如，在主-从网络配置中在任何时候仅仅连网设备之一驱动会话。特别是，主设备（即主设备的对话管理器）管理和协调网络设备间的会话并决定哪一设备来执行给定的会话服务或功能。这个决定可以基于每一设备或者应用程序提供的关于它们会话能力的信息。这个决定也可以基于主设备决定那一个从设备（具备所需的会话能力）能最佳地执行给定的会话功能。例如，主设备可以要求多个从设备来执行语音识别并提供结果给主设备。主设备然后选择最佳的结果。可以理解这里在语音识别水平上所描述的是分布式对话管理器之间在 DM (对话管理器) 协议水平上的机制（如在 Y0999-111P 中所描述）。确实，当在多个对话管理器之间发生对话时，主设备将获得每一对话管理器结果得分的度量，从而作出决定看好哪一个对话管理器继续进行输入，不但要根据语音识别的准确性，并且也要根据对话（含意）、上下文和历史，也考虑其他项目，如用户的优先选择、历史和应用的偏

好。

在对等连接中，每一设备将试图确定自己能执行的功能并记录一个请求去执行之。可接受到任务的设备将执行这个任务然后对其执行性能评分。然后依据它们的评分设备协商哪一个设备将执行这个任务。

在一个实施例中，分布式会话协议 112、116 经由 RMI（远程方法调用）和 RPC（远程过程请求）系统调用，以完成应用程序和整个网络上不同会话引擎之间的调用来实现。行内人都知道，RPC 是允许一个应用通过网络从另一个应用请求一个服务的协议。同样地，RMI 是在分布式网络上对象可以交互的一种方法。RMI 允许一个或者多个对象随同请求被传递。此外，信息也可以被存储在一个对象中，该对象经由 CORBA 或者 DCOM 被交换，或者以说明性的方式呈现（例如经由 XML）。如在上面插入的专利申请 IBM Docket NO. Y0999-111P 中所讨论，会话协议（方法）（或者分布式协议）可以被用来经由会话 API 在会话应用和 CVM 命外壳之间或者经由会话引擎 API 在 CVM 和会话引擎之间获得由一个 CVM（会话虚拟机器）外壳支持的会话功能的分布式实现。会话引擎 API 是核心引擎和应用程序之间的界面，通过这个界面和协议与核心引擎（本地的或者连网的）通信。会话 API 提供一个 API 层来挂钩或者发展明白会话的应用，包括建立会话用户界面的基础类和组件。

同样地，根据本发明一对话管理器可经由 API 与应用程序和引擎（本地的或者连网的）通信。通过这种方式，一对话管理器可以对来自所有远程过程的结果和回调起作用（对远程引擎和应用程序的过程调用），如同它是一个本地的应用，以至于，例如，在应用和资源（本地的或者连网的）之间仲裁、区分优先次序和确定激活的应用，以及确定那一个结果认为是有效的。

会话发现、注册和协商协议 113、117 是被用作“发现”本地的或者网络的明白会话的系统（即“讲”会话协议的应用程序或者设备）的网络协议（或者方法）。注册协议允许设备或者应用程序注册它们的会话功能、状态和主目。协商协议允许设备协商主-从、对等或者匿名

伙伴网络。

在一个实施例中，发现协议实现一个“广播和收听”方法来触发一个从其它“广播和收听”设备来的反应。这能允许如网络动态地和自发地创建（例如下面讨论的蓝牙和 Hopping 网络）。在另一个实施例中，一个缺省的服务器（可能是主设备）装置可以被利用它注册不同网络设备的“地址”。在这个实施例中，发现总计网络上与服务器通信的各个设备来核对注册的设备列表，以便确定那些设备与这些设备连接。经由发现协议交换的信息包括如下：（1）握手信号的广播请求或者收听请求；（2）设备标识符交换；（3）第一次注册时的句柄/指针交换；和（4）第一次协商时的句柄交换。

在实现注册协议的一个实施例中依据连接，设备可以以一预定的协议（例如，TTS 英文、任何文本、语音识别、500 字+FSG 语法、无说话者识别等）通过交换一组标志或一设备特性对象来交换关于它们会话能力的信息。同样，应用程序可以交换引擎需求列表。通过一个主/从网络配置，主对话管理器可以编辑所有列表并使功能和需要与会话能力相匹配。在没有主设备（对话管理器）的情况下，一个普通服务器可以被用来传输会话信息到网络上每一个机器或者设备。注册协议允许以下信息被交换：（1）能力和装载消息包括定义和更新事件；（2）引擎资源（是否一个给定设备包括 NLU、DM、MLG、TTS、说话者识别、语音识别压缩、编码、存储等）；（3）I/O 能力；（4）CPU、存储器和装载能力；（5）数据文件类型（范围指定、字典、语言模型、语种等）；（6）网络地址和特征；（7）关于用户的信息（定义和更新事件）；用户对设备、应用或对话的偏好；（9）用户化；（10）用户体验；（11）帮助；（12）每一应用（和应用状态）（定义和更新事件）的能力需求；（13）CUI 服务和行为的元信息（帮助文件、编目、会话优先权等）（定义和更新事件、一般通过指向表格的指针）；（14）协议握手；和/或（15）拓扑协商。

注册可以使用传统的通信协议如 TCP/IP、TCP/IP29、X-10 或者 CEBus 和设备之间的套接口实现。设备利用一个分布式会话结构与它

们相关的会话引擎和对话管理器及它们的会话主目（例如，激活的词汇表、语法和语言模型、语法分析和翻译/标记模型、声波纹、合成规则、基本格式（发音规则）和声型）通信。这个信息或者以文件或者以流的形式传递给对话管理器和会话引擎或者作为 URL。此外，上下文信息可以通过指示通路或者指到设备的上下文栈/历史，或控制器可以访问和添加到其上下文栈的应用而能被传递。设备也传输关于它们多模式 I/O 和 UI 能力（屏幕/无屏幕、音频入和出的能力、键盘等）的信息。会话主目允许会话引擎基于当前状态和上下文估计 NLU 引擎有关的新询问。

至于协商协议，依据各个网络设备的注册需求和能力，在协商期间网络设备和应用程序可以对一给定的暂时性配置进行投票。如果一个应用利用此配置（即拓扑），则决定被自动地强加。否则它可以是请求成为一个主设备或者从设备或者对等机。根据请求计数，一个优选的配置被决定并与所有的设备和应用程序通信（保持在每个设备/应用程序的可用资源表格中）。无论什么时候一个系统改变它的状态和需求时，它将会与其它连网的对话管理器/资源通信来开始一个新的协商并沟通新的状态和上下文信息。

语音传输协议 114、118 允许设备传送和接收压缩的语音或者本地处理结果到/从网络上的其它设备和应用程序。会话引擎 102、107 更适宜包括压缩/解压缩引擎用以传输时压缩语音（或者结果）和为了本地处理对通过网络从另一设备或应用程序获得的压缩语音（或者结果）进行解压缩。语音传输协议被在设备中执行的语音传输的客户机利用来向/从其它连网设备、系统或者应用程序传送/接收被压缩的语音进行处理。设备的语音传输客户机与压缩、解压缩和重建引擎联合操作，利用合适的压缩硬件，处理通过网络传输的语音。语音编码器提供感知可接受的或可理解的被压缩语音的重建和优化的会话性能（如，字错率）。语音在各自的连网设备上利用声信号处理引擎（音频子系统）和合适的音频硬件被捕捉（和转换成特征）。此外，压缩的语音文件格式在处理语音的设备之间可以被传输和接收。特别是，语音传输协议

允许设备向/从网络上的其它设备和应用程序传输和接收压缩的语音或本地的处理结果。在一个实施例中，当在一个传输设备和一个接收设备之间握手处理后，一数据流（基于信息包）被发送到接收器。包的头最好指出为编码此语音（或者结果）所利用的编码方案和编码主旨（即，抽样频率、特征特性、维数、前端应用的变换、前端的性质等）。此外，纠错信息也可以被引入（例如，如果前一个包丢失或者延迟，前一个包的最后特征矢量纠正差动解码器）或者合适的消息来恢复（重发）这些丢失的包。

此外，对话管理器可以经由对话管理器协议或者 DM 协议通信。（如在上面引用的 IBM Docket No. Y0999-111P 中所讨论的）。DM 协议可以被用来在多个对话管理器之间协商哪个对话管理器被激活，或者哪个对话管理器应该接收该输入。因为在本实施例中，服务器资源仅仅当被真正需要时被“轮询”，DM 协议提供一种变化：本地对话管理器进行一预先测试以决定是否此功能应该被远程执行。在有错误产生或者存在疑惑的情况下，对话管理器可以等待本地引擎的推测并在仍存有疑惑时决定轮询一服务器来比较。

因此，根据上面的讨论，应该明白网络协议提供在连网设备之间的协同（或者一个协同界面）来共享会话服务和功能。术语协同界面意味着一单个会话可以在不同的参加者（设备/应用程序）之间被把握，如同它们都理解整个会话并合适地知道在任意给定的时间谁被访问。每一会话系统的行为或者应用可以被对话管理器（例如，主-从模式中的主设备）、应用（此应用可以确立谁是主、从或者对等）、系统对话管理器（如果有）、组织和协商（在对等模式中）控制/管理，在合适的系统上对用户透明地执行每一会话功能。对于一个客户设备上的用户来说，提供了一个无缝的会话界面（即，所有的会话交互似乎是经由一个单一的会话系统），尽管某些会话功能、系统和资源可能被提供几个连网的设备（例如，一蜂窝式电话、一呼机和一 PDA）。

协同的会话系统的一个例子（上面描述的公开在前面引用的的 IBM Docket No. Y0999-111P 中）是称作 UCA（通用会话设备）的远程控制。

UCA 发现明白会话的设备。每个会话连接的设备将发送它的会话主目（词汇表和语法）到 UCA。UCA 作为这种设备的主设备并且当一会话交互随同用户结果成为一命令到此设备时将更新适当的设备。相反地，依据命令的执行或者每次设备改变状态时，它将发送一个更新到远程控制。一个没有其他会话能力（除了相应主目）的会话设备是那个被称作“匿名伙伴”的。

在另一个实施例中，一个服务器或者基站，除音频捕捉压缩和由远程控制（或 UCA）执行的传送外可以执行所有的会话功能。远程控制也可以提供一些 UI 给用户来通知他/她不同设备的状态。这可以经由语音、GUI 或者这些形式（或者其它）的任何会话的组合完成。

尽管会话网络拓扑可以与图 1 的系统联系起来使用，一个最优的网络拓扑是提供一种自发的动态连网（即，在一定的通信范围内的设备之间自发的建立的一网络）。这种自发的连网可以通过应用当前新兴的例如描述在 <http://www.bluetooth.com> 上的“蓝牙”连网协议实现。简单地说，蓝牙是给网络协议的一个代号，蓝牙网络协议提供特别的无线网络的连通性。更特别的是，蓝牙是在一个特定的范围内的设备（例如，智能电话、蜂窝式电话、呼机、PDA、便携式电脑、移动设备等）之间在特定的范围内提供短程无线射频链路来动态地和自发地在这样的设备之间建立一个网络（或是大家知道的“piconet”（微网））。一个微网指的是在网络连接的剩余期间与在每个微网中扮演主设备的一个结点以特殊方式连接的蓝牙允许设备（结点）的集合。两个或更多的微网可以被网络连接以构成所谓的一个“scatternet”（分散网）。

看得出，依照本发明任何自发的动态连网协议可以被实现。例如，图 1 的网络拓扑可以依照在美国专利申请序列号 No. 09/198,378 描述的“hopping”通信网络实现，该专利申请 1998 年 11 月 24 日递交、名称为“Automated Traffic Mapping”，此专利已一般转让，此处引用以做参考。

现参看图 2，图 2 是描述根据本发明的一个方面在连网设备之间提供自动和协同共享会话功能的一种方法的流程图。特别是，图 2 进

一步详细描述了图 1 系统操作的一种方法。开始，用户发出一个口头命令（或者否则发出一个询问）到本地客户设备，这个口头命令或询问通过数字化或抽取数字化语音信号的相关特征而被预处理过（步骤 200）。作为选择，本地对话管理器可以接收从本地应用程序 104 传来的请求产生合成语音（TTS）以输出到该用户（步骤 200）。

（经由本地对话管理器 103）决定本地处理是否可用（步骤 201），例如，是否语音识别或者语音合成可在本地执行。对于这个决定可以看出，本地对话管理器 103 可以明确地预先确定识别/合成必须在其上发生的一个远程服务器，（例如，一 IP 地址对于套接口连接的、一 URL 地址对于经由小服务程序（servlet）基于服务器登记，或者一电话号码对于直接连接或连接到一个 IVR 的）。此外，本地设备没有可以利用的资源或者主目来执行（或者有效地执行）一任务的确定，也可由对话管理器 103 依据本地应用程序 104 的执行根据由该应用程序在头文件中指示的资源需求而作出。另外，某些命令或者请求的功能可以引起对话管理器自动地连接到一个远程服务器。例如，安全性应用程序（如，说话者确认）可以自动地被切换到服务器端处理，所以声波纹不被分配到客户机。再者，本地汽车导航系统可以利用电话或者基于套接口的服务器能被自动地切换到远程服务器，以致于本地设备不必存储大量的导航信息。

如果确定本地处理是可行的（肯定的确定在步骤 201），则处理将经由本地引擎 102 在本地执行（步骤 202）。另一方面，如果确定本地处理是不可行的（否定的确定在步骤 201），那么相关的特征/波形/信息被自动传输给一个远程网络连接的服务器（步骤 204）（经由 IP、LAN、蓝牙、IR、RF 或者经由电话或者 IP 电话），在其中远程处理（例如，语音识别/合成）被执行（步骤 205）（可能与某个用户/服务器交互）。

显然从本地客户机到远程网络连接的服务器的语音传输（反之亦然）可以利用不同的技术实现。例如，可以当作文件、流或者信息包流直接传送波形。另外，一个被压缩的波形可以利用传统的方法如

ADPCM 和 APC 传输。而且，一个特征的流可以依据 G. Ramaswamy 等人在 “Compression Of Acoustic Features For Speech Recognition In Network Environments,” Vol. 2, pp. 977-980, Proc. ICASSP, 1998 公开的方法传输，此处引入以供参考。此方法允许在接收端识别（语音识别、说话者识别或者 NLU）而不重建信号。此外，语音传输可以利用任何编码方法或者基于压缩特征和音调估计的方案，允许语音信号以足够可理解的质量和平滑舒缓的重放（用以重放、校正、进一步人性处理、或者归档）而被重建。这种编码方案应该提供低到 4kbits/s 和 5kbits/s 之间的数据率而不降低识别性能。其结果是以后端（服务器）资源支持的的交互式交换，甚至通过无线调制解调器或者无线数据链路，就可以实时执行。可以理解，利用相似的编码方案若要提供非常高质量的重放，其它的方案必须被采用。另外，任何允许 cepstra 特征和音调压缩、允许在服务器端识别（语音、说话者、NLU）且在接收端没有退化并重建信号的方法可与本发明结合使用。此重建对于以后从服务器的重放或者从客户机（若是本地存储）的重放和处理，为随后的校对抄本、错误校正或者人的处理监控是有用的。可以理解任何合适的压缩方案（编码协议）可以被利用。

可以理解，压缩或者编码方案（传送协议或编码协议）在不同的设备之间可以不同。例如，从音频捕捉系统（客户机的）传送输入语音到连网资源的编码可能不同于从连网资源（服务器）到音频输出（客户机）用于传送输出语音（例如提示、重放或者 TTS）的编码协议。确实，在第一种情况下，编码应该被优化以在服务器端提供良好的识别性能，回放的重建固然重要但并不引人注目地重要。当然比特率（压缩率）是重要的。压缩率的折衷可在鲁棒性（错误率-特征失真）和知觉质量之间作出调整以达到和保持一个目标比特率。同样，某些方案可以被选择以给某些信道或者背景失真增加鲁棒性。另一方面，对于后面的任务（输出信号），应该为了可理解或者知觉质量和舒适性，或者保持声音或音频某些特殊的特征而优化编码。经过本地处理（步骤 202）或远程处理（步骤 205）后，一个关于该处理结果是否被接受的

决定（步骤 203 和 206）被作出（经由本地对话管理器 103 或者服务器对话管理器 108）。如果确定处理的结果是不可接受的（否定决定在步骤 203 和 206），则本地客户机或者远程服务器将自动地（经由 IP、LAN、蓝牙、IR、RF 或者经由电话或者 IP 电话连接）转发该特征或者波形给能执行这种处理的服务器（步骤 204 和步骤 207）。例如，如果这种结果是未知或者没有被识别或者是模糊不清（或者基于与每个资源（本地的或服务器）的对话管理器相关的可信度度量），则对语音识别结果或 NLU 的拒绝可能发生。更可取的是，从本地的或者远程系统到一个服务器系统的自动连接可以基于声音或者被返回的 LM（语言模型）得分的水平上，该得分由本地语音解码器应用如在美国专利申请 5,937,383, Ittycheriah 等人、名称为 “Apparatus and Methods For Speech Recognition Including Individual or Speaker Class Dependent Decoding History Caches For Fast Word Acceptance or Rejection” 中所教导的技术，此专利已一般转让，此处引入以供参考（例如，当这些得分被判定低于一个给定的阈值时远程服务器被连接）。可以理解，估计可信度或识别说话和询问（在其期间或者对话以后）的任何合适的度量或者方法可以被用来决定一个被会话系统获得的结果是否可以接受（在这种情况下另一个系统被考虑）。

相似地对 TTS，本地的和远程的对话管理器 103、108 可以检查一个文本的复杂性来决定是否此 TTS 将被本地或者远程地执行。例如，当一个字的发音规则不知道时或者当文本需要复杂的句法分析时 TTS 将被远程执行。另一个例子是如果此 TTS 必须被用一个不同的口音、方言或者一种不同的语种发音或者假定模仿某人特定的句子时。

当处理被远程执行后，结果被发送回本地客户机（步骤 208）（经由电话、IP 地址、MAC（媒体存取控制）地址等）。很明显，输出（即，给用户输出的语音）可以被本地地或者在服务器上合成。如果合成是在服务器上进行，被合成的语音可以以压缩格式传输（利用上面讨论的语音传输协议）给用户在本地解压缩。可以理解，编码方案可以和用于从客户机到服务器传输语音特征的方案相同或者不同。作为选择，

例如，在另外的模拟 PSTN 线路上，用通过从客户机到服务器的电话呼叫（回叫）建立的联接，语音可以通过服务器被直接“广播”。

最近的努力开始发展识别语音的合适的可信性度量。例如，在“LVCSR Hub5 Workshop”，1996 年 4.29 – 5.1，MITAGS, MD，由 NIST 和 DARPA 组织，提出不同的方法来把一个可信度级别加于一个字一个可信度级别上。一个利用决策树的方法在字相关特征（训练说话的数量、最小和平均三音素 (triphone) 事件、语言模型训练事件、音素 /lefemes 数量、持续时间、声音得分（快速匹配或者精细匹配）、语音非语音），句子相关特征（信号噪音比、讲话速率估计：每秒钟的字数或 lefemes 数或者元音数、语言模型提供句子的似然、似然率、每帧规范化平均似然、语言模型中的三字母组事件），上下文中的字特征（语言模型中三字母组事件）及说话者外貌特征（口音、方言、性别、年龄、讲话速率、身份、音频质量、SNR 等...）上进行训练。对于此树的每一个叶在训练数据上都计算了错误的概率。构造这样一个树的算法被 Breiman 等人在“Classification and Regression Trees”，Chapman & Hall, 1993 中讨论过。识别方面，所有的或这些特征中的一些在识别过程中被度量，并且对于每一个字，决策树走到提供一个可信度级别的叶。此外，参考 Neti 等人的标题为“Word Based Confidence Measure As A Guide For Stack Search In Speech Recognition”，ICASSP97, 慕尼黑，德国，1997.4 一文，描述了一完全依靠 IBM 栈解码器返回的得分的方法（使用对数似然—实际上用的是平均增量对数似然、精细匹配、快速匹配）。

在 LVCSR 处理中，使用通过线性回归的预测器估计可信度级别的另一种方法被执行。被利用的预测器是字的持续时间、语言模型得分、每一帧平均声音得分（最好得分）和 NBEST 列表中同顶部选择相同的字的一部分。显然根据本发明的一个实施例，两种方法（经由决策树度量的可信度级别和经由线性预测器度量的可信度级别）结合起来，在任何转换处理中系统地提取可信度级别，并不局限于语音识别。

基于过去的改进和该领域的飞速发展，现在我们可以说对于几种

转换可以连系一个可信度值到正被转换的组件上，例如，从 0 到 1，0 意味着无转换被执行，1 意味着转换无疑，此处组件可以是要转换的文本、短语、字和更一般地任何材料的逻辑块。如上描述的线性预测器和决策树的结合是本发明优选采用的一种方法。确实，作为例子，由说话者识别引擎返回的得分累计（快速匹配得分和精细匹配得分及背景模型和同伴的得分）可以被用来构建可信度级别的决策树和/或一线性预测器，这样说话者真正被正确鉴别。实际上，在说话者识别的情况下，这个执行验证的总量与识别阶段获得的相等。

可以看出远程服务器可以发送信息例如 TTS 规则或者基本格式、语法等给本地客户机以存储在高速缓存中，这样本地设备随后可以应用这种信息本地处理类似的请求。因为本地设备由于缺乏所需要的资源而不能执行某些任务，由服务器对话管理器 108 发送这种处理信息给本地客户机的决定可以根据由本地设备在与远程服务器连接时关于它的会话能力向远程服务器的注册（经由上面所讨论的注册协议）而被作出。

显然本发明可以在这种情况下实现，此时由一设备（经由它的对话管理器）执行的会话功能量，设备是不能提供必须的资源用于这些功能的即时执行（例如，IVR 有太多的被系统同时使用的端口）。因而，对话管理器可以实现提供会话系统管理和装载管理，借此，在一个特定的功能执行时对话管理器可以决定利用另一个会话系统来继续处理被请求的功能。特别是，参看图 3，开始，用户发出一个口头命令给本地客户设备，它是被预处理过的，例如，被数字化和提取数字化信号的相关的特征（步骤 300）。作为另一种选择，本地对话管理器可以从一本本地应用程序 104 接收一请求来产生一个合成语音（TTS）输出给用户（步骤 300）。对话管理器将决定是否本地处理应该被执行（步骤 301）（例如，是否要语音识别、对话管理或者语音合成）。这个决定根据的不仅是本地会话能力、主目和资源（如上面所讨论的），而且也根据由于网络交通拥塞网络引起的延迟与利用可用的但被约束的本地资源（假定本地的和远程的设备可以执行相同的功能）执行此会话功能

可能引起的延迟相比较的评价。因而，例如，当命令和控制功能在本地/通过网络受到威胁被延迟时可以远程地/本地地执行来减小延迟。确实，可以掌握更长延迟的查询（例如，由于和后端功能相联合可以适应延迟如因特网或视听查询），可以在一个优化其资源或花费的系统上执行。

此外，万一网络连接暂时不可用或者缺乏网络资源，则所有可以被本地执行的功能将被执行。其他功能可以被细分成可以延期的模式执行的功能（在以后连接时可被重新建立）以及不共存的功能。典型的例子是更新地址薄、经由口述或者大体上的口述应答 e-mail 或者消息。再者，最好应用程序可以决定是否该命令是本地的或者是延期的。也可以考虑一种带有延期模式的对等，在此一延期模式管理器和一本本地引擎决定是否该功能是本地的或者是延期的。

再参看图 3，如果对话管理器确定本地处理是合适的（肯定决定在步骤 301），则对话管理器将分配必需的会话引擎给端口（步骤 302）。一旦此会话引擎被分配给该端口，如果会话引擎当前没有被最初分配的端口所使用，则对话管理器可以分配那个引擎给其它端口（步骤 303）（例如，当前说话者没有说话而只是在听）。当本地引擎再次被最初被分配的端口所需要时，如果此本地端口不可得到，另一个可用的引擎（本地或者在一个远程设备上）可以被利用（步骤 304）。这种动态分配过程与传统的装载管理大不相同，在传统的装载管理中对话管理器决定并在功能调用的整个持续时间，分配会话引擎到每个端口。

不难看出管理和决定传输语音例如到一个网络服务器或者设备不但可以根据系统管理/装载平衡的水平（通过客户机或者服务器上的对话管理器），而且还根据网络的通信量。例如，如果一个连接（特别是在因特网上基于 TCP/IP 的网络连接）被认为是超载（步骤 305），则基于通信量一个新的服务器或者设备被选择（步骤 306）。这个决定可以在传统协议如 VoIP 协议（因特网上的声音协议）顶端作出，象 RSVP 协议（资源预订协议）一样，借此，当需要一个信道时，连接可以随同相关质量服务的合适预订一起建立。否则，如上所述远程处理将被

执行（步骤 307）并返回结果。

显然这里所描述的系统和方法可以实现以用于各种允许的语音和会话应用程序。本发明对于满足在嵌入的和流行的计算世界的不断增长的要求及 NLU/NLG 会话系统方面特别有用。然而，可以理解，本发明可以为不同的应用而被发展，并不局限于嵌入式系统。下面作为范例的实施例将说明本发明的优点。

例如，在智能电话上应用的商业可用嵌入式的姓名拨号器（例如，一个具有 PDA（个人数字助理）能力的无线电话）是一个典型的应用。例如，假定客户设备 100 是一个具有姓名拨号器本地应用的智能电话。用户将在智能电话的电子地址簿中本地存储一个所希望的姓名和地址的表。然后用户可以发出一个命令如 “dial first name last name”（拨姓和名）at... 可能的限定符（家庭、办公室、蜂窝式电话），通过对命令的识别/理解（经由本地会话引擎 102），智能电话将自动地拨地址簿中与此人相关的电话号码（经由本地应用程序 104）。另一方面，当发出的姓名不在地址簿中时（因而不被识别/理解），但该姓名存在于一个更大的公司的（或者公用的）目录中（如包含在远程服务器 106 中），则这个请求可以被（以特征或者波形）保存和传输给一个远程服务器 106 来识别。然后拨号可以直接被远程服务器执行或者依据通过远程服务器接收的合适信息，由智能电话来执行。作为替代，在第一个情况下用户可以被连接到一个远程的服务器并建立一个对话，或重新请求要拨号的姓名或者要求进一步的信息（在白页或者黄页服务类型的情况下）。

本发明的另一种有用的应用涉及个人信息系统，例如商业可用的 PointCast（参见 <http://www.pointcast.com>），这个 PointCast 允许用户根据预定的用户优先选择得到例如，股票报价、关于某一主题特定的信息和关于此主题最近公布的信息。应用根据本发明构成的个人信息系统，如果用户希望获得关于一股票（例如，IBM）或者一主题（例如，预报在 Kent 的绿豆生产）的信息，则用户可以发出一个语音请求给客户设备 100。如果 “IBM” 是在本地词汇表（用户概况）中，则它

将立即被解码，用户得到最新报价，例如获得最近一次更新（PointCast）。另一方面，如果用户关于“绿豆”的请求没有被本地客户设备 100 理解，则该请求被作为特征流自动向前传递给（内容提供者的）远程服务器 106，在此处服务器可以投入更多的资源来解码这种请求和检索相关信息（无论如何都得做），然后传输这种信息给本地系统。如果远程服务器遵循一“推销方法（push approach）”，则上面这些在下次更新时就被完成（例如，PointCast）。

客户设备 100 也可以是一个允许使用语音的 PVA（个人交通工具助理），用以提供如会话的汽车导航。例如，如果用户不想在此系统应用 CD-ROM（因为缺乏空间、能源要求、重量、成本、防震等），用户可以决定存储有限的信息，例如有关用户当前所处位置、用户最近所处位置、及用户想出游的区域/位置的词表和地图。在这个例子中，无论何时，当用户的请求与本地词表或者地图设置不匹配时，这个请求可以被自动发送给一个远程服务器 106，并解码（甚至以提示返回给用户去缩小范围查找）来获得路线、地图（或者更新地图）下载到车上。再者，这种操作（即使花费高代价的下载）对用户来说是基本上透明的，仅开始需要本地道路。

此外，一个 NLU/FSG 系统可以根据本发明被设计，以便如果用户的请求需要 FSG（限定状态语法），则该请求可以被本地执行，除非该请求更复杂并处在蒙昧状态，因而需要向前传递给一个远程服务器来识别。

公司的姓名-拨号器服务器提供另一种有意思的特点。一个公司为它的雇员们保留一个电话号码的有效数据库。这个数据库总是最新的。用户可以选择周期性地来使它本地的信息与公司数据库保持同步。这是一个传统的概念。然而，当用户利用姓名拨号器并且需要经由 TCP/IP 连接到服务器上时，此同步可以实现（在语音识别期间），这样本地拨号信息总被更新。同样，当用户请求导航到一个新的其信息没有包括在用户本地地图中时，系统可以下载位于本地 PVA 上的声信息及用户所希望出游地区的一组导航信息。

图 1 的系统也可以用会话浏览器系统实现，会话浏览器系统描述于 IBM Docket No. Y0998-392P，当前同此发明一起被递交，名称为“Conversational Browser and Conversational System”，并已被一般转让，引入此处以供参考，其中从一个内容提供器（服务器）传输（并被会话浏览器处理）的 CML（会话标记语言）页在概念上类似用于可视显示的 HTML（超文本标记语言）页，用来描述要被呈现给用户的一个会话 UI。在这个例子中，会话浏览器可以是客户设备 100 的本地应用程序 104 和/或远程(IVR)服务器 106 中的服务器应用程序 109。可由内容提供者或应用开发者（或者代理/代码转换器）来决定用户应该提供的一个给定的项目（例如，NLU，或 FSG 输入表单或经口授填充一空表单）必须在服务器 106 上被识别，而不是提供所有的数据给客户设备 100 在本地识别（因为此任务对于本地资源来说太复杂或者因为通过网络必须发送太多的信息）。其实现是通过例如在一个 CML 文件中提供一个 URL（统一资源定位器）和标记来指示服务器其中一个处理将发生，或者通过在 CML 页中装载一小应用程序、一 Active X 组件或者一个捕捉音频的插件（或者它的任何变化），实现可能的一些会话功能，并为其它功能将其传输给其它设备（这是由页面的制作者作出的典型的决定）。这个决定可以自动地被代码转换器和注册机制实现，如在 IBM Docket No. Y0998-392P 中所描述，由此，浏览器向它取出 CML 页的服务器明白地描述它的能力。当代码转换器另考虑浏览器的能力并使内容适合这种能力时（这种能力即是所提到的会话代理），现在，根据浏览器的能力，代码转换器可以增加服务器的 URL(s) 以重定向服务。在这种情况下，被客户设备 100 收集到的语音可以被作为一个波形（被压缩或者没有）或者作为特征流发送给远程服务器 106 或者识别发生其上的网络服务器 110（或者 NLU/NLG）。然后识别的结果可以被发送回客户设备 100 或者 CML 提供者服务器（远程服务器 106）来决定操作的下一个行动或者进一步处理。再者如上面所述的，这可以被能直接结合资源/引擎/服务器的 URL 的应用程序决定或者要被用来识别一个给定的输入、菜单表格或者对话的本地设备决

定。此外，本发明在一个 CML 页必须重放/合成声音或者文本对于本地设备 102 的本地会话引擎 102 来说太复杂的情况下是有用的。太复杂的部分可以作为特征流或者被压缩波形从任一个特定的服务器（这个服务器可能是或者不是提供 CML 页的服务器）上得到。而且，对于多语言系统，如果一个 CML 页包含一种不同的语种，则没有合适能力的本地客户设备 100 可以请求一个远程服务器来执行按那种语言的会话功能。

显然，会话协同可以在有会话浏览器的分布式应用中使用。例如，参看图 4，一个分布式系统有一个表示服务器 400、一个引擎服务器 401、和一具有会话浏览器 403 的客户机 402(如在上面参考 Y0998-392P 中所讨论)。该浏览器 403 从表示服务器 400 接收 CML 的页并处理此 CML 页。CML 页可包含允许浏览器 403 决定发送语音到何处进行处理的信息。一个语音服务器位于引擎服务器 401 上。假定 CML 页需要引擎服务器 401 处理语音，浏览器 403 可以经由 HTTP(或者套接口或者 RMI) 与语音服务器联系 (传输呼叫) 来传输音频给语音服务器并发送合适的数据文件指令和引擎呼叫。客户浏览器 403 被假定有一些本地处理能力来执行语音处理 (经由语音 API 和语音识别引擎 405)。如上面提到，在本地语音处理和服务器端语音处理之间的转移由从表示服务器 400 接收到的 CML 页决定。这个决定可以被内容提供者编码或者适应此设备 (客户 402 决定它能执行此任务并把该任务发送给一个已知的或者已被发现的服务器或者代码转换器)。

在图 5，浏览器 403 位于被客户 402 存取的一个浏览器服务器 404 上 (浏览器服务器 404 在客户机 402 和表示服务器 400 之间充当中介物)。再者，浏览器 403 决定是否执行本地或者服务器端的处理。如此处所描述，音频是可以利用会话编码被传输。

本发明允许一个低资源客户设备透明地本地执行简单的任务，又以二进制或模拟方式与一具有更复杂会话能力的服务器（或者别的设备）相连接透明地执行复杂任务。服务器端功能（例如语音识别）可以通过一个常规的 IP 网络或者 LAN 网络实现，也可以经由传统的电话

线路或包交换网络数字传输，或者在无线网络上经由任何传统的无线数据协议而能被执行。模拟/数字连接实施例描述了至少两个方案。第一，它是等价于一个调制解调器实现的二进制连接，所有的功能是数据传递功能。此外，当一个服务器/分布式资源被包含时，系统可以调用一个电话服务器作为资源，且声音在网络上被发送（而不是波形数据或者它的变换如 cepstra）。这一方案的一个例子是与本地语音识别功能（姓名拨号器和数字拨号器）的一无线连接，此本地语音识别功能经由一个常规的无线连接到一个具有 IVR 的电话服务器上来获取其它功能，象语音浏览因特网、获得股票/共有基金报价和通过语音执行金融交易。这种机制在今天可以利用现存的蜂窝式电话设备上装备一些语音识别能力。

另外，不同的机制可以被利用来管理流通量和在网络上分布的资源，以保证在网络上保持一合适的会话交互的对话流。这种机制包括：会话协议（如上面所讨论的），音频：RecoVC（识别兼容 VoCoder）（允许回放重建带有音调的编码协议），应用程序和元信息：分布式应用协议、发现、注册、协商、维护对话流的服务器负载管理、保持对话流的流量平衡和路由选择、基于任务特征和能力需求及会话主目可用性（数据文件）的引擎服务器选择，会话主目分配：存储、流量/路由选择和高速缓存。

尽管图例的实施例在此参考附图已被描述，应该理解，本系统和方法并不局限于这些刻板的实施例，在不脱离本发明的范围和精神的情况下，技术上熟悉的人可作各种不同的改变和修改。所有这种改变和修该都应包括在所附加的权利要求书中定义的范围内。

说 明 书 附 图

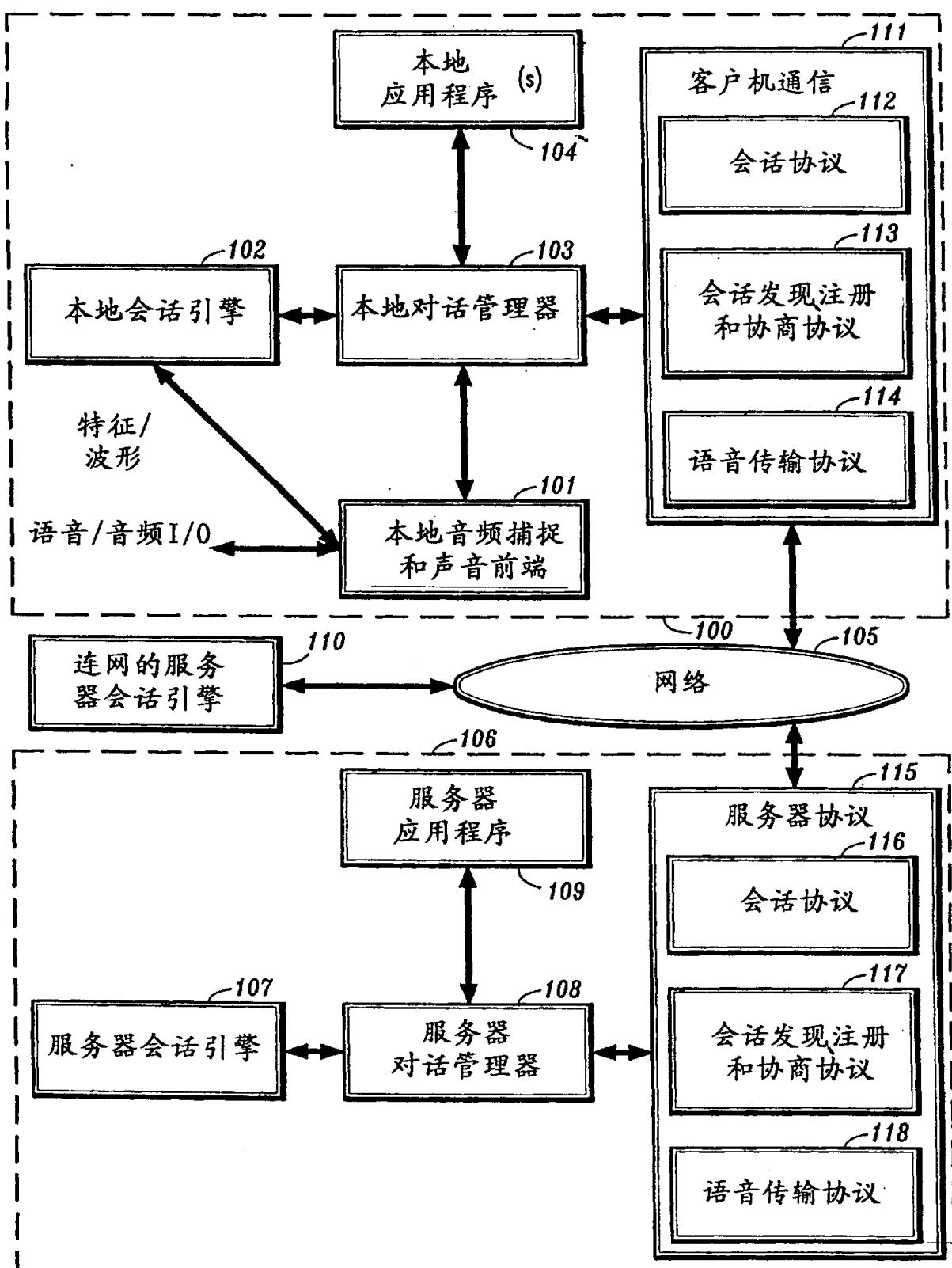


图1

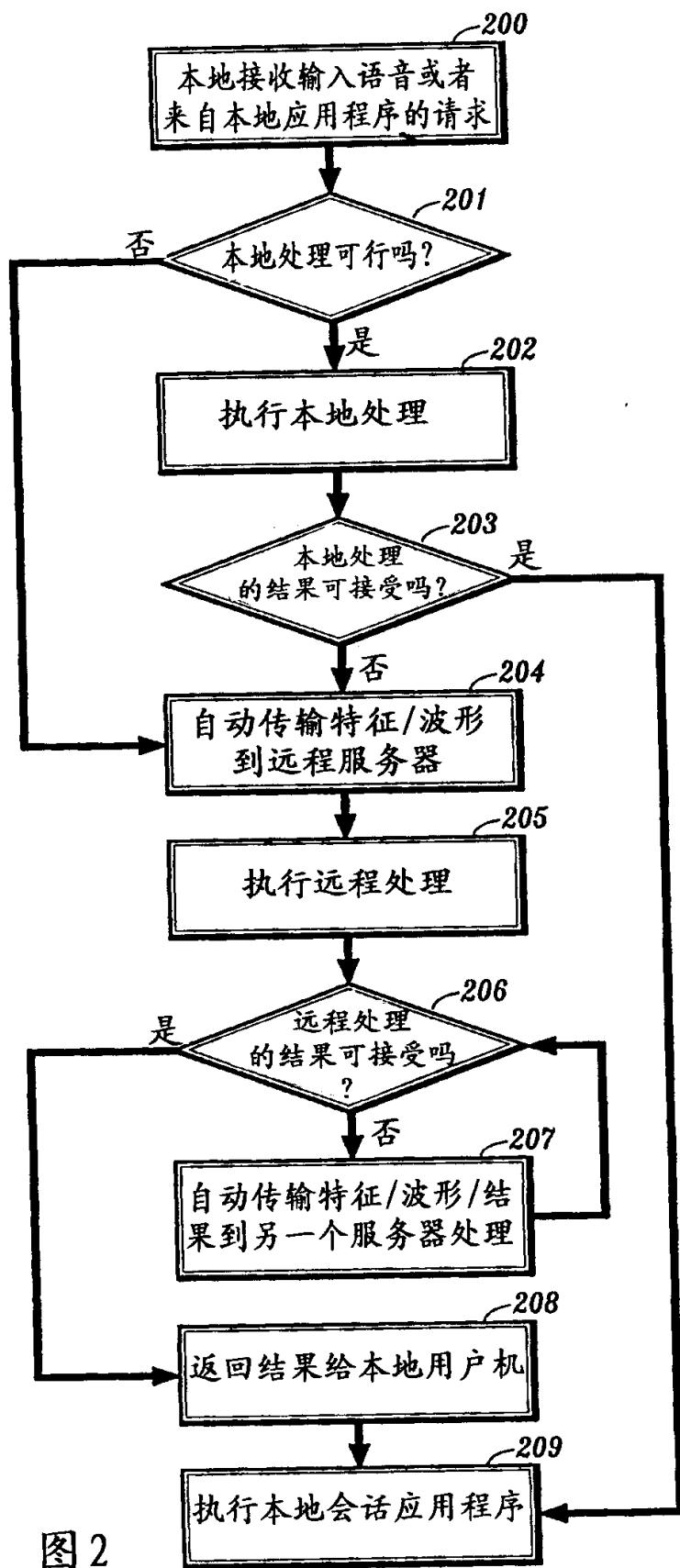


图 2

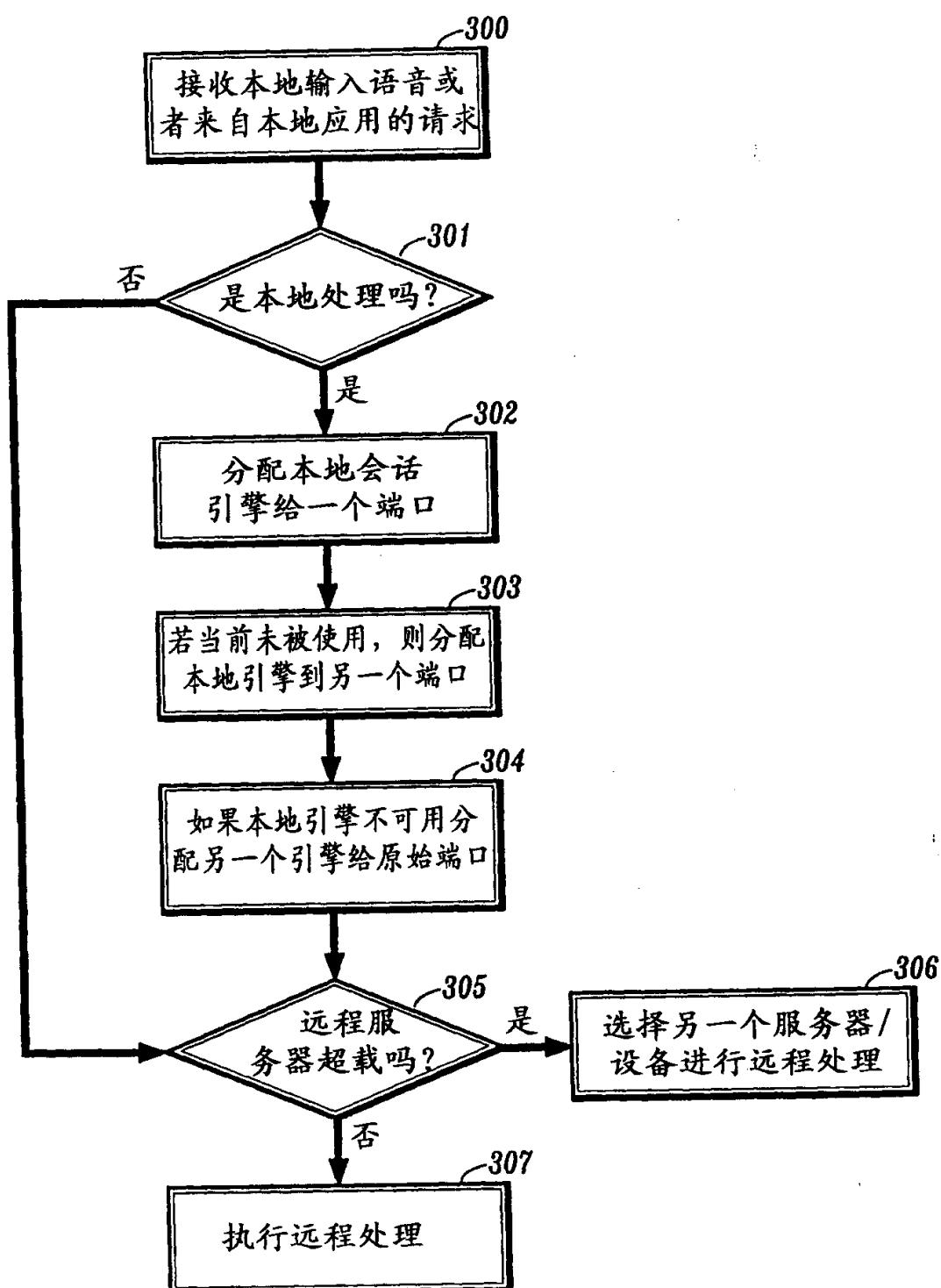
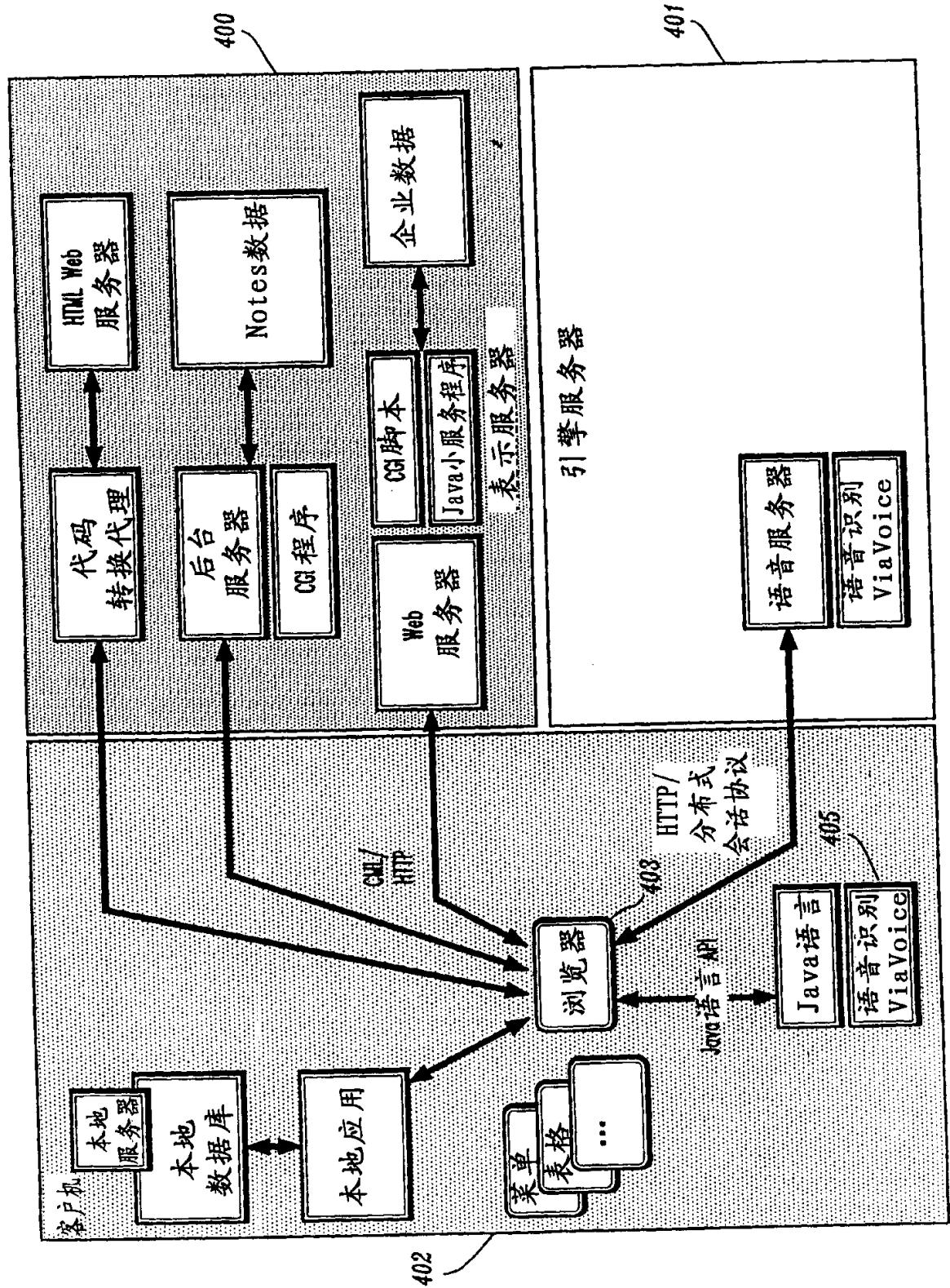


图 3

图 4



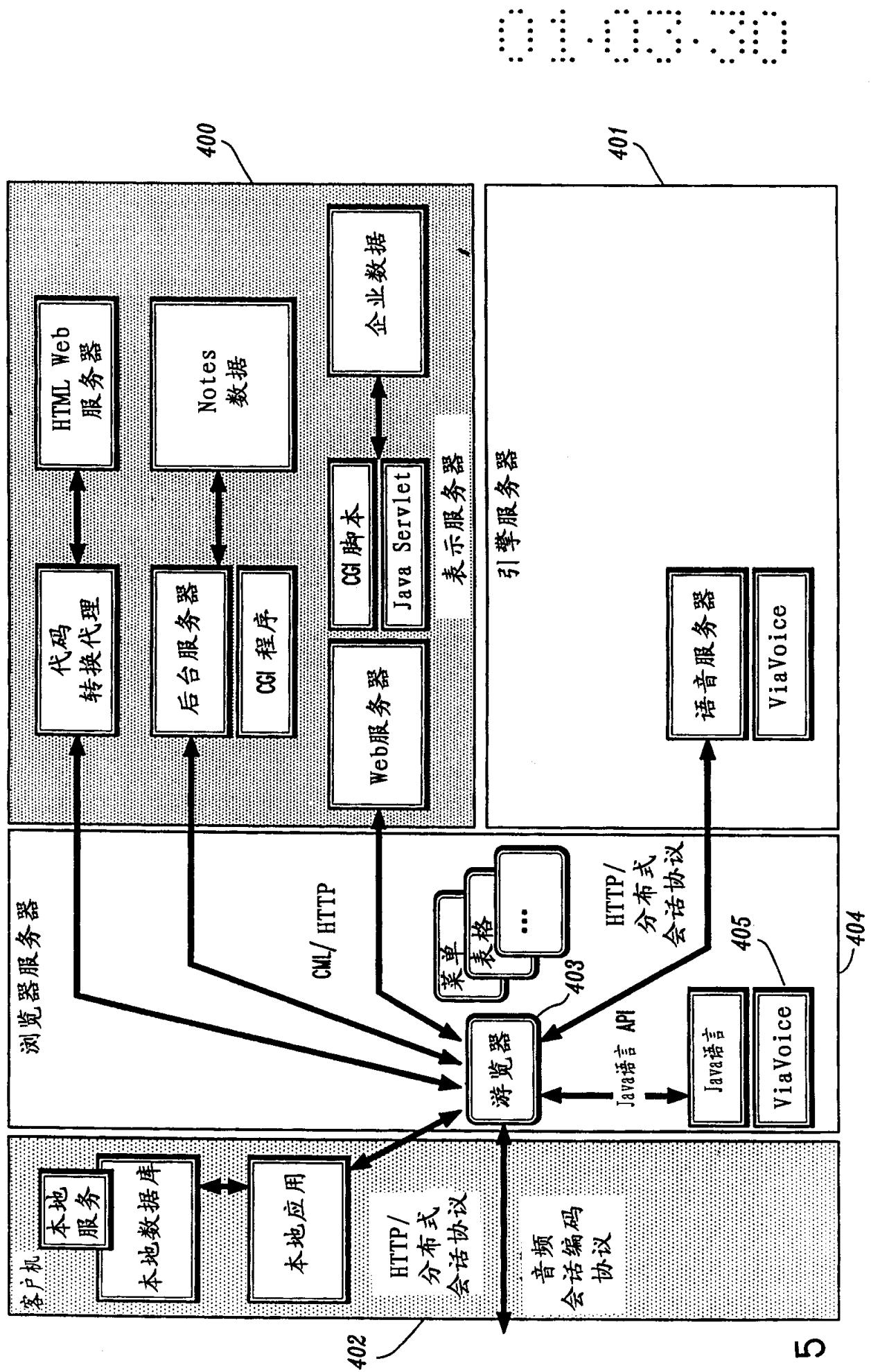


图 5