(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 102693725 A
(43)申请公布日 2012.09.26

(21)申请号 201210081427.X

(22)申请日 2012.03.26

(30)优先权数据
13/072003 2011.03.25 US

(71)申请人 通用汽车有限责任公司
地址 美国密执安州

(72)发明人 G. 塔尔瓦 X. 赵

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 刘桢 杨楷

(51)Int.Cl.
G10L 15/26(2006.01)
G10L 15/28(2006.01)
G10L 15/08(2006.01)

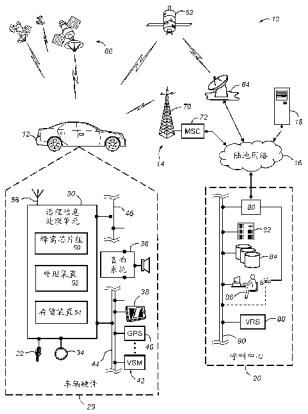
权利要求书 2 页 说明书 17 页 附图 4 页

(54)发明名称

依赖于文本信息语境的语音识别

(57)摘要

本发明涉及依赖于文本信息语境的语音识别，提供了一种自动语音识别方法。通过麦克风从用户接收对文本信息的回复话语，所述麦克风将所述回复话语转换为语音信号。使用至少一个处理器处理所述语音信号，以从所述语音信号提取声音数据。使用与所述文本信息相关的会话语境从多个声音模型中识别一个声音模型，以解码所述声音数据。使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设。



CN 102693725 A

1. 一种自动语音识别方法，包括下列步骤：

a) 通过麦克风从用户接收对文本信息的回复话语，所述麦克风将所述回复话语转换为语音信号；

b) 使用至少一个处理器来预处理所述语音信号，以从所述语音信号提取声音数据；

c) 使用与所述文本信息相关的会话语境来识别多个声音模型中的一个声音模型，以解码所述声音数据；以及

d) 使用识别的声音模型来解码所述声音数据，以产生用于所述回复话语的多个假设。

2. 如权利要求 1 的方法，还包括如下步骤：

e) 后处理所述多个假设，以将所述假设中的一个识别为所述回复话语。

3. 如权利要求 2 的方法，还包括如下步骤：

f) 将所述识别的假设展现给用户；

g) 从用户寻求所述识别的假设为正确的确认；以及

h) 如果用户确定所述识别的假设是正确的，那么输出所述识别的假设作为回复文本信息的至少一部分。

4. 如权利要求 3 的方法，还包括如下步骤：

i) 处理所述文本信息，利用会话语境特定的语言模型识别对应于所述文本信息的会话语境，以及利用情感语境特定的语言模型识别对应于所述文本信息的情感语境，其中所述语言模型存储在客户端装置上；以及

j) 使用所述情感语境来完善所述声音模型的识别。

5. 如权利要求 2 的方法，还包括如下步骤：

f) 利用识别的假设来改编所述多个声音模型，用以随时间完善语音识别性能。

6. 如权利要求 5 的方法，其中步骤 a) 和 b) 在语音识别客户端装置上执行，步骤 c) 至 f) 在语音识别服务器上执行。

7. 如权利要求 6 的方法，其中改编步骤 f) 还包括使用所述识别的假设改编存储在所述服务器上的多个语境特定的语言模型，和将所述多个语境特定的语言模型从所述服务器发送至所述客户端装置，以更新存储在所述客户端装置上的语言模型，用以随时间改善文本信息语境分类。

8. 如权利要求 6 的方法，还包括如下步骤：

在所述语音识别客户端装置接收文本信息；

处理所述文本信息，利用会话语境特定的语言模型来识别对应于所述文本信息的会话语境，和利用情感语境特定的语言模型来识别对应于所述文本信息的情感语境，其中所述语言模型存储在客户端装置上；以及

发送所述识别的会话和情感语境至所述语音识别服务器。

9. 一种自动语音识别方法，包括如下步骤：

a) 在语音识别客户端装置接收文本信息；

b) 使用所述客户端装置的至少一个处理器通过存储在所述客户端装置上的会话语境特定的语言模型来处理所述文本信息，以识别对应于所述文本信息的会话语境；

c) 从所述文本信息合成语音；

d) 通过所述客户端装置的扬声器将所述合成的语音发送至所述客户端装置的用户；

e) 通过所述客户端装置的麦克风从用户接收回复话语，所述麦克风将所述回复话语转换为语音信号；

f) 使用至少一个处理器来预处理所述语音信号，以从所述接收的语音信号提取声音数据；

g) 将所述提取的声音数据和识别的会话语境发送至语音识别服务器；

h) 使用识别的会话语境来识别存储在所述服务器上的多个声音模型中的一个声音模型，以解码所述声音数据；

i) 使用识别的声音模型来解码所述声音数据，以产生用于所述回复话语的多个假设；以及

j) 后处理所述多个假设，以将所述假设之一识别为所述回复话语。

10. 一种自动语音识别方法，包括如下步骤：

a) 在语音识别客户端装置接收文本信息；

b) 使用所述客户端装置的至少一个处理器通过存储在所述客户端装置上的会话语境特定的语言模型来处理所述文本信息，以识别对应于所述文本信息的会话语境；

c) 从所述文本信息合成语音；

d) 通过所述客户端装置的扬声器发送所述合成的语音至所述客户端装置的用户；

e) 通过所述客户端装置的麦克风从用户接收回复话语，所述麦克风将所述回复话语转换为语音信号；

f) 使用至少一个处理器预处理所述语音信号，以从接收的语音信号提取声音数据；

g) 使用识别的与所述文本信息相关的会话语境，识别所述多个声音模型中的一个声音模型，以解码所述声音数据；

h) 使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；

i) 确定与所述回复话语的多个假设的至少一个相关的信任值是否大于或小于信任阈值；

j) 如果所述信任值被确定为小于所述信任阈值，那么将提取的声音数据和会话语境发送至语音识别服务器，否则后处理所述多个假设，以将所述假设之一识别为所述回复话语，并从所述客户端装置输出所述识别的假设作为回复的文本信息的至少一部分；

k) 使用识别的会话语境在所述服务器识别存储在所述服务器的多个声音模型中的一个声音模型，以解码所述声音数据；

l) 使用在所述服务器识别的所述声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；

m) 后处理所述多个假设，以将所述假设之一识别为所述回复话语；以及

n) 从所述服务器输出所述识别的假设作为回复的文本信息的至少一部分。

依赖于文本信息语境的语音识别

技术领域

[0001] 本发明总地涉及语音信号处理。

背景技术

[0002] 通常，语音信号处理包括处理电气和 / 或电子信号，用于识别或合成语音。语音合成是通过人工方式从文本到语音的产物，文本到语音（TTS）系统给传统的计算机到人的视觉输出装置（例如计算机监视器或显示器）提供了替代方案。相反，自动语音识别（ASR）技术使配备了麦克风的计算装置能够转译语音，从而为传统的人到计算机的触觉输入装置（例如键盘或键区）提供了替代方案。

[0003] 在某些环境下，TTS 和 ASR 技术被组合以为用户提供与系统交互的免提音频。例如，车辆中的远程信息处理系统可接收文本信息、使用 TTS 技术将信息以音频形式展现给驾驶员、接收驾驶员的回复话语、并将回复转至服务器，该服务器识别所述回复并产生和发送相应的文本信息响应。语音识别通常是困难的任务，特别是在处理现代文本信息的陌生缩略语和其它口语特色时。

发明内容

[0004] 根据本发明的一个实施例，提供了一种自动语音识别方法，包括下列步骤：

a) 通过麦克风从用户接收对文本信息的回复话语，所述麦克风将所述回复话语转换为语音信号；

b) 使用至少一个处理器预处理所述语音信号，以从所述语音信号提取声音数据；

c) 使用与所述文本信息相关的会话语境识别多个声音模型中的一个声音模型，以解码所述声音数据；以及

d) 使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设。

[0005] 根据本发明的另一实施例，提供了一种自动语音识别方法，包括下列步骤：

a) 在语音识别客户端装置接收文本信息；

b) 使用所述客户端装置的至少一个处理器通过存储在所述客户端装置上的特定会话语境语言模型处理所述文本信息，以识别对应于所述文本信息的会话语境；

c) 从所述文本信息合成语音；

d) 通过所述客户端装置的扬声器发送所述合成的语音至所述客户端装置的用户；

e) 通过所述客户端装置的麦克风从用户接收回复话语，所述麦克风将所述回复话语转换为语音信号；

f) 使用至少一个处理器预处理所述语音信号，以从接收的所述语音信号提取声音数据；

g) 将提取的声音数据和识别的会话语境发送至语音识别服务器；

h) 使用识别的会话语境识别存储在所述服务器上的多个声音模型中的一个声音模型，以解码所述声音数据；

i) 使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；以及

j) 后处理所述多个假设，以将所述假设之一识别为所述回复话语。

[0006] 根据本发明的另一实施例，提供了一种自动语音识别方法，包括下列步骤：

a) 在语音识别客户端装置接收文本信息；

b) 使用所述客户端装置的至少一个处理器通过存储在所述客户端装置上的特定会话语境语言模型处理所述文本信息，以识别对应于所述文本信息的会话语境；

c) 从所述文本信息合成语音；

d) 通过所述客户端装置的扬声器发送所述合成的语音至所述客户端装置的用户；

e) 通过所述客户端装置的麦克风从用户接收回复话语，所述麦克风将所述回复话语转换为语音信号；

f) 使用至少一个处理器预处理所述语音信号，以从接收的所述语音信号提取声音数据；

g) 使用与所述文本信息相关的识别的会话语境，识别所述多个声音模型中的一个声音模型，以解码所述声音数据；

h) 使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；以及

i) 确定与所述回复话语的多个假设的至少一个相关的信任值是否大于或小于信任阈值；

j) 如果所述信任值被确定为小于所述信任阈值，那么将提取的声音数据和会话语境发送至语音识别服务器，否则后处理所述多个假设，以将所述假设之一识别为所述回复话语，并从所述客户端装置输出所述识别的假设作为回复的文本信息的至少一部分；

h) 使用识别的会话语境在所述服务器识别存储在所述服务器的多个声音模型中的一个声音模型，以解码所述声音数据；

i) 使用在所述服务器识别的所述声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；

j) 后处理所述多个假设，以将所述假设之一识别为所述回复话语；以及

k) 从所述服务器输出所述识别的假设作为回复的文本信息的至少一部分。

[0007] 本发明提供下列技术方案。

[0008] 技术方案 1：一种自动语音识别方法，包括下列步骤：

a) 通过麦克风从用户接收对文本信息的回复话语，所述麦克风将所述回复话语转换为语音信号；

b) 使用至少一个处理器来预处理所述语音信号，以从所述语音信号提取声音数据；

c) 使用与所述文本信息相关的会话语境来识别多个声音模型中的一个声音模型，以解码所述声音数据；以及

d) 使用识别的声音模型来解码所述声音数据，以产生用于所述回复话语的多个假设。

[0009] 技术方案 2：如技术方案 1 的方法，还包括如下步骤：

e) 后处理所述多个假设，以将所述假设中的一个识别为所述回复话语。

[0010] 技术方案 3：如技术方案 2 的方法，还包括如下步骤：

f) 将所述识别的假设展现给用户；

g) 从用户寻求所述识别的假设为正确的确认；以及

h) 如果用户确定所述识别的假设是正确的，那么输出所述识别的假设作为回复文本信息的至少一部分。

[0011] 技术方案 4：如技术方案 3 的方法，还包括如下步骤：

i) 处理所述文本信息，利用会话语境特定的语言模型识别对应于所述文本信息的会话语境，以及利用情感语境特定的语言模型识别对应于所述文本信息的情感语境，其中所述语言模型存储在客户端装置上；以及

j) 使用所述情感语境来完善所述声音模型的识别。

[0012] 技术方案 5：如技术方案 2 的方法，还包括如下步骤：

f) 利用识别的假设来改编所述多个声音模型，用以随时间完善语音识别性能。

[0013] 技术方案 6：如技术方案 5 的方法，其中步骤 a) 和 b) 在语音识别客户端装置上执行，步骤 c) 至 f) 在语音识别服务器上执行。

[0014] 技术方案 7：如技术方案 6 的方法，其中改编步骤 f) 还包括使用所述识别的假设改编存储在所述服务器上的多个语境特定的语言模型，和将所述多个语境特定的语言模型从所述服务器发送至所述客户端装置，以更新存储在所述客户端装置上的语言模型，用以随时间改善文本信息语境分类。

[0015] 技术方案 8：如技术方案 6 的方法，还包括如下步骤：

在所述语音识别客户端装置接收文本信息；

处理所述文本信息，利用会话语境特定的语言模型来识别对应于所述文本信息的会话语境，和利用情感语境特定的语言模型来识别对应于所述文本信息的情感语境，其中所述语言模型存储在客户端装置上；以及

发送所述识别的会话和情感语境至所述语音识别服务器。

[0016] 技术方案 9：如技术方案 1 的方法，所述识别和解码步骤 c) 和 d) 起初使用语音识别客户端来执行。

[0017] 技术方案 10：如技术方案 1 的方法，还包括如下步骤：

确定与用于所述回复话语的多个假设中的至少一个相关的信任值是否大于信任阈值；以及

如果所述信任值被确定为小于所述信任阈值，那么将所述提取的声音数据和所述会话语境发送至语音识别服务器；否则

后处理所述多个假设，以将所述假设中的一个识别为所述回复话语；以及

从所述客户端装置输出所述识别的假设作为回复文本信息的至少一部分。

[0018] 技术方案 11：一种自动语音识别方法，包括如下步骤：

a) 在语音识别客户端装置接收文本信息；

b) 使用所述客户端装置的至少一个处理器通过存储在所述客户端装置上的会话语境特定的语言模型来处理所述文本信息，以识别对应于所述文本信息的会话语境；

c) 从所述文本信息合成语音；

d) 通过所述客户端装置的扬声器将所述合成的语音发送至所述客户端装置的用户；

e) 通过所述客户端装置的麦克风从用户接收回复话语，所述麦克风将所述回复话语

转换为语音信号；

f) 使用至少一个处理器来预处理所述语音信号，以从所述接收的语音信号提取声音数据；

g) 将所述提取的声音数据和识别的会话语境发送至语音识别服务器；

h) 使用识别的会话语境来识别存储在所述服务器上的多个声音模型中的一个声音模型，以解码所述声音数据；

i) 使用识别的声音模型来解码所述声音数据，以产生用于所述回复话语的多个假设；以及

j) 后处理所述多个假设，以将所述假设之一识别为所述回复话语。

[0019] 技术方案 12：如技术方案 11 的方法，还包括如下步骤：

k) 使用识别的假设来改编所述多个声音模型，用于随时间完善语音识别性能。

[0020] 技术方案 13：如技术方案 12 的方法，其中所述改编步骤还包括使用识别的假设来改编存储在所述服务器上的多个语境特定的语言模型，并将所述多个语境特定的语言模型从所述服务器传送至所述客户端装置，以更新存储在所述客户端装置上的语言模型，用于随时间完善文本信息语境分类。

[0021] 技术方案 14：如技术方案 12 的方法，还包括如下步骤：

l) 使用所述客户端装置的至少一个服务器通过存储在所述客户端装置上的情感语境特定的语言模型来处理所述文本信息，以识别对应于所述文本信息的情感语境；以及

m) 将识别的情感语境发送至所述语音识别服务器。

[0022] 技术方案 15：如技术方案 14 的方法，其中所述识别步骤还使用识别的情感语境来执行，以完善所述声音模型的识别。

[0023] 技术方案 16：如技术方案 14 的方法，还包括如下步骤：

n) 将识别的假设展现给用户；

o) 从用户寻求所述识别的假设为正确的确认；

p) 如果用户确认所述识别的假设为正确，那么输出该识别的假设作为回复文本信息的至少一部分；否则

q) 使用所述情感语境来完善所述声音模型的识别，并重复步骤 e) 至 p)。

[0024] 技术方案 17：一种自动语音识别方法，包括如下步骤：

a) 在语音识别客户端装置接收文本信息；

b) 使用所述客户端装置的至少一个处理器通过存储在所述客户端装置上的会话语境特定的语言模型来处理所述文本信息，以识别对应于所述文本信息的会话语境；

c) 从所述文本信息合成语音；

d) 通过所述客户端装置的扬声器发送所述合成的语音至所述客户端装置的用户；

e) 通过所述客户端装置的麦克风从用户接收回复话语，所述麦克风将所述回复话语转换为语音信号；

f) 使用至少一个处理器预处理所述语音信号，以从接收的语音信号提取声音数据；

g) 使用识别的与所述文本信息相关的会话语境，识别所述多个声音模型中的一个声音模型，以解码所述声音数据；

h) 使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；

i) 确定与所述回复话语的多个假设的至少一个相关的信任值是否大于或小于信任阈值；

j) 如果所述信任值被确定为小于所述信任阈值，那么将提取的声音数据和会话语境发送至语音识别服务器，否则后处理所述多个假设，以将所述假设之一识别为所述回复话语，并从所述客户端装置输出所述识别的假设作为回复的文本信息的至少一部分；

k) 使用识别的会话语境在所述服务器识别存储在所述服务器的多个声音模型中的一个声音模型，以解码所述声音数据；

l) 使用在所述服务器识别的所述声音模型解码所述声音数据，以产生用于所述回复话语的多个假设；

m) 后处理所述多个假设，以将所述假设之一识别为所述回复话语；以及

n) 从所述服务器输出所述识别的假设作为回复的文本信息的至少一部分。

附图说明

[0025] 下面结合附图描述本发明的一个或多个优选实施例，其中相同的标记表示相同的元件：

图 1 为示出能够利用本文所公开的方法的通信系统的示例性实施例的框图；

图 2 为示出可与图 1 的系统一起使用、用于实施示例性语音合成方法和 / 或改进语音辨别的文本到语音（TTS）系统的示例性实施例的框图；

图 3 为示出可与图 1 的通信系统和图 2 的 TTS 系统一起使用、用于实施示例性语音辨别方法和 / 或改进语音假设的自动语音辨别（ASR）系统的示例性实施例的框图；以及

图 4 为示出可由图 1 的通信系统、图 2 和图 3 的 TTS 和 ASR 系统执行的自动语音辨别方法的示例性实施例的流程图。

具体实施方式

[0026] 下面的内容描述了实例通信系统、可与该通信系统一起使用的实例 TTS 和 ASR 系统、以及可与前述系统一起使用的一个或多个实例方法。下面描述的方法可被车辆远程信息处理单元（VTU）使用，作为辨别 VTU 用户发出的语音的一部分。尽管下述方法是它们可执行用于 VTU，但是应当清楚，它们可用在任意类型的车辆语音辨别系统和其它类语音辨别系统。例如，该方法可执行在 ASR 启用的移动计算装置或系统、个人电脑等中。

[0027] 通信系统 -

参考图 1，示出了包括移动车辆通信系统 10 的示例性操作环境，其可用于执行本文所公开的方法。通信系统 10 通常包括车辆 12、一个或多个无线载波系统 14、陆地通信网络 16、计算机 18 和呼叫中心 20。应当理解，所公开的方法可与任意数量的不同系统一起使用，且不具体限于这里所示的操作环境。并且，系统 10 的架构、结构、设置和操作及其各个部件通常是本领域内公知的。因此，下面的段落简单地提供了一个这种示例性系统 10 的简要概述；然而，这里未示出的其它系统也可利用所公开的方法。

[0028] 车辆 12 在所示实施例中被描述为轿车，但是应当清楚，包括摩托车、卡车、运动型旅行车（SUV）、旅游汽车（RV）、海洋舰船、飞行器等的任意其它交通运输工具也可使用。图 1 中总地示出了一些车辆电气 28，包括远程信息处理单元 30、麦克风 32、一个或多个按钮或

其它控制输入 34、音响系统 36、可视显示器 38、和 GPS 模块 40 以及多个车辆系统模块（VSM）42。这些装置中的一些可直接连接至远程信息处理单元，例如麦克风 32 和按钮 34，而其它使用一种或多种网络连接间接地连接，例如通信总线 44 或以及娱乐总线 46。适当的网络连接的例子包括控制器局域网（CAN）、多媒体导向系统传输（MOST）、局域互联网络（LIN）和其它适当的连接，例如以太网或与已知 ISO、SAE 和 IEEE 标准和规定等相符的其它连接，仅举几个例子。

[0029] 远程信息处理单元 30 可为能够经无线载波系统 14 通过无线网络进行无线语音和 / 或数据通信的安装了（嵌入了）OEM 的装置或零件市场装置，使得车辆可与呼叫中心 20、其它能够进行远程通信的车辆、或某些其它实体或装置通信。远程信息处理单元优选使用无线电传输来与无线载波系统 14 建立通信通道（声音通道和 / 或数据通道），使得可通过该通道收发声音和 / 或数据传输。通过提供声音和数据通信，远程信息处理单元 30 使车辆能够提供许多不同服务，包括与导航、电话、紧急援助、诊断、娱乐等相关的那些服务。数据可使用本领域已知的技术通过数据连接（例如通过经数据通道的数据包传输）或通过声音通道发送。对于包括声音通信（例如，与呼叫中心 20 处的人工顾问或语音响应单元）和数据通信（例如，给呼叫中心 20 提供 GPS 定位数据或车辆诊断数据）的组合服务，该系统可利用经由声音通道的一个呼叫，若需要还有声音通道上的声音与数据传输之间的开关，并且这可使用本领域技术人员已知的技术来进行。

[0030] 根据一个实施例，远程信息处理单元 30 利用根据 GSM 或 CDMA 标准的蜂窝通信，因此包括用于声音通信（例如，非手持式呼叫）的标准蜂窝芯片 50、用于数据传输的无线调制解调器、电子处理装置 52、一个或多个数字存储装置 54、和双天线 56。应当清楚，调制解调器可通过存储在远程信息处理单元并由处理器 52 执行的软件来实施，或者可以是位于远程信息处理单元 30 内部或外部的单独硬件部件。调制解调器可使用任意多种不同标准或协议来操作，例如 EVDO、CDMA、GPRS 和 EDGE。车辆与其它联网装置之间的无线网络也可使用远程信息处理单元 30 来实施。为此目的，远程信息处理单元 30 可构造成根据一种或多种无线协议而无线地通信，例如 IEEE 802.11 协议、WiMAX 或蓝牙中的任意一种。当用于分组交换数据通信（例如 TCP/IP）时，远程信息处理单元可构造有静态 IP 地址，或者可设置成从网络上的另一装置（例如路由器）或从网址服务器自动接收分配的 IP 地址。

[0031] 处理器 52 可为能够处理电子指令的任意类型的装置，包括微处理器、微控制器、主处理器、控制器、车辆通信处理器和专用集成电路（ASIC）。其可为仅用于远程信息处理单元 30 的专用处理器，或者可与其它车辆系统共享。处理器 52 执行各种类型的数字存储指令，例如存储在存储器 54 中使远程信息处理单元能够提供大量服务的软件或固件程序。例如，处理器 52 可执行程序或处理数据，以执行本文所述方法的至少一部分。

[0032] 远程信息处理单元 30 可用于提供不同范围的车辆服务，包括至车辆和 / 或来自车辆的无线通信。这种服务包括 ：联合基于 GPS 的车辆导航模块 40 提供的建议线路规划指示及其它与导航相关的服务 ；联合一个或多个碰撞传感器接口模块如车身控制模块（未示出）提供的气囊展开通知及其它紧急或与路侧援助相关的服务 ；使用一个或多个诊断模块的诊断汇报 ；以及与娱乐相关的服务，其中音乐、网页、电影、电视节目、视频游戏和 / 或其它信息通过娱乐模块（未示出）下载并被存储，用于当前或后期回放。上述服务并不是远程信息处理单元 30 全部能力的详尽列表，而是远程信息处理单元能够提供的一些服务的简单列

举。另外，应当理解，上述模块至少一部分能够以存储在远程信息处理单元 30 内部或外部的软件指令的形式来实施，它们可为位于远程信息处理单元 30 内部或外部的硬件部件，或者它们可彼此或与位于车辆上的其它系统集成和 / 或共享，等等。在模块被实施为位于远程信息处理单元 30 外部的 VSM 42 的情形下，它们可利用车辆总线 44 与远程信息处理单元交换数据和指令。

[0033] GPS 模块 40 从 GPS 卫星星群 60 接收无线电信号。从这些信号，模块 40 可确定用于向车辆驾驶员提供导航和其它与位置相关的服务的车辆位置。导航信息可展现在显示器 38 上（或车辆内的其它显示器），或可以声音的方式展现，例如当提供建议线路规划指示时这样做。导航服务可使用车内的专用导航模块（可为 GPS 模块 40 的一部分）提供，或者一部分或全部导航服务可通过远程信息处理单元 30 来进行，其中位置信息被发送至远程位置，用于给车辆提供导航地图、地图注解（关注点、饭店等）、路线计算等等。位置信息可被提供给呼叫中心 20 或其它远程计算机系统，例如计算机 18，用于其它目的，例如车队管理。并且，新的或更新的地图数据可通过远程信息处理单元 30 从呼叫中心 20 下载至 GPS 模块 40。

[0034] 除音响系统 36 和 GPS 模块 40 之外，车辆 12 可包括为电子硬件部件形式的其它车辆系统模块（VSM）42，其位于车辆上，并通常从一个或多个传感器接收输入，且使用感测的输入执行诊断、监测、控制、报告和 / 或其它功能。VSM 42 中每个都优选通过通信总线 44 连接至其它 VSM 以及远程信息处理单元 30，并可编程为运行车辆系统和子系统诊断测试。例如，一个 VSM 42 可为控制发动机操作各方面（例如燃料点火和点火正时）的发动机控制模块（ECM），另一 VSM 42 可为调节车辆动力系的一个或多个部件的操作的动力系控制模块，另一 VSM 42 可为管理位于车辆上的各种电子部件（例如车辆动力门锁和车灯）的车身控制模块。根据一个实施例，发动机控制模块配备有车载诊断（OBD）特征，其提供无数实时数据，例如从包括车辆排放传感器的各种传感器接收的数据，并提供允许技术人员快速识别和修正车辆内的故障的一系列标准化诊断故障代码（DTC）。如本领域的技术人员所清楚的，上述 VSM 只是可在车辆 12 中使用的一些模块的例子，还可使用许多其它模块。

[0035] 车辆电气 28 还包括给车辆乘客提供用来提供和 / 或接收信息的许多车辆用户接口，包括麦克风 32、按钮 34、音响系统 36 和可视显示器 38。如这里所使用的，术语“车辆用户接口”广义地包括任何适当形式的电气装置，包括位于车辆上并使车辆用户能够与车辆部件通信或通过其通信的硬件和软件部件。麦克风 32 向远程信息处理单元提供音频输入，使驾驶员或其它乘客能够通过无线载波系统 14 提供声音指令和实施免提呼叫。为此目的，可利用本领域内已知的人机交互（HMI）技术将麦克风连接至车载自动声音处理单元。按钮 34 允许人工用户输入进远程信息处理单元 30，以开始无线电话呼叫并提供其它数据、响应或控制输入。相对于到呼叫中心 20 的常规服务援助呼叫，可使用单独的按钮来启动紧急呼叫。音响系统 36 给车辆乘客提供音频输出，并可以是专用的独立系统或为主车辆音响系统的一部分。根据这里示出的特定实施例，音响系统 36 可操作地联接至车辆总线 44 和娱乐总线 46，并可提供 AM、FM 和卫星无线电、CD、DVD 及其它多媒体功能。该功能可联合上述娱乐模块或独立于该模块来提供。可视显示器 38 优选为图形显示器，例如仪表板上的触摸屏或挡风玻璃上反射的头顶显示器，并可用于提供许多输入和输出功能。还可使用各种其它车辆用户接口，例如图 1 的接口仅仅是一种特定实施方案的例子。

[0036] 无线载波系统 14 优选为蜂窝电话系统，包括多个单元塔 70（仅示出了一个）、一个

或多个移动交换中心(MSC)72、以及将无线载波系统 14 与陆地网络 16 连接所需的任何其它网络部件。每个单元塔 70 都包括收发天线和基站，不同单元塔的基站或者直接或者通过中间设备(例如基站控制器)连接至 MSC 72。蜂窝系统 14 可采用任何适当的通信技术，例如包括，模拟技术(例如 AMP)或新型数字技术(例如 CDMA (如 CDMA2000)或 GSM/GPRS)。如本领域技术人员所清楚的，各种单元塔 / 基站 /MSC 布置都是可能的，可与无线系统 14 一起使用。例如，基站和单元塔可共同位于同一地点，或者它们可彼此远离，每个基站都可负责一个单元塔，或者一个基站可服务多个单元塔，并且多个基站可联接至一个 MSC，等等。

[0037] 除使用无线载波系统 14 之外，可使用卫星通信形式的不同无线载波系统来提供与车辆的单向或双向通信。这可使用一个或多个通信卫星 62 和上行发射站 64 来进行。单向通信可为例如卫星无线电服务，其中节目内容(新闻、音乐等)由发射站 64 接收、被打包上载、然后发送至卫星 62，卫星 62 将节目广播给订阅者。双向通信可为例如使用卫星 62 来中转车辆 12 与站 64 之间的电话通信的卫星电话服务。如果使用，该卫星电话可被附加使用或替代无线载波系统 14。

[0038] 陆地网络 16 可为传统的陆基远程通信网络，其连接至一个或多个地面通信线电路，并将无线载波系统 14 连接至呼叫中心 20。例如，陆地网络 16 可包括例如用于提供硬连线电话的公共交换电话网络(PSTN)、封包交换数据通信、和互联网基础设施。一段或多段陆地网络 16 可通过标准有线网络、光纤或其它光学网络、电缆网、电线、其它无线网络(例如无线局域网(WLAN)或提供宽带无线访问(BWA)的网络)或它们的任意组合来实施。另外，呼叫中心 20 不需要通过陆地网络 16 连接，但是可包括无线电话设备，使得它可与无线网络(例如无线载波系统 14)直接通信。

[0039] 计算机 18 可为可通过私人或公共网络(例如互联网)访问的多个计算机中的一个。每个这种计算机都可用于一个或多个目的，例如可由车辆通过远程信息处理单元 30 和无线载波系统 14 访问的网络服务器。其它这类可访问计算机 18 可为例如 ：可通过远程信息处理单元 30 从车辆上载诊断信息和其它车辆数据的服务中心计算机 ；为访存取或接收车辆数据的目的或者为了建立或配置订阅者喜好或控制车辆功能，由车辆所有者或其它订阅者使用的客户端计算机 ；或者第三方库，向该库或从该库提供车辆数据或其它信息，无论是通过与车辆 12 或呼叫中心 20 或两者通信。计算机 18 还可用于提供互联网连接，例如 DNS 服务，或作为使用 DHCP 或其它适当协议以给车辆 12 分配 IP 地址的网址服务器。

[0040] 呼叫中心 20 被设计成给车辆电气 28 提供许多不同系统后端功能，根据这里所示的示例性实施例，呼叫中心 20 通常包括一个或多个交换机 80、服务器 82、数据库 84、人工顾问 86 以及自动语音响应系统(VRS)88，所有这些都是本领域已知的。这些各种呼叫中心部件优选通过有线或无线局域网络 90 彼此联接。可为专用带宽交换(PBX)交换机的交换机 80 传送输入的信号，使得声音输送通常由常规电话发送至人工顾问 86 或使用 VoIP 发送至自动语音响应系统 88。人工顾问电话也可使用 VoIP，如图 1 中虚线所示。经由交换机 80 的 VoIP 及其它数据通信通过连接在交换机 80 与网络 90 之间的调制解调器(未示出)实施。数据传输通过调制解调器送至服务器 82 和 / 或数据库 84。数据库 84 可存储帐户信息，例如订阅者认证信息、车辆标识、外形记录、行为类型及其它相关订阅者信息。数据传输还可由无线系统(例如 802.11x、GPRS 等)进行。尽管所示实施例已经描述成使用人工顾问 86 联合人工呼叫中心 20 一起使用，但是应清楚，呼叫中心还可利用 VRS 88 作为自动顾问，或者

可使用 VRS 88 与人工顾问 86 的组合。

[0041] 语音合成系统 -

现在参考图 2，示出了用于文本到语音（TTS）系统 210 的示例性架构，其可用于执行目前公开的方法。通常，用户或车辆乘客可与 TTS 系统交互，以从应用程序（例如，车辆导航应用程序、免提操作呼叫应用程序等）的菜单提示接收指令或听从提示。通常，TTS 系统从文本源提取输出词语或标志、将该输出转换为适当的语言单位、选择与该语言单位最佳对应的存储的语音单位、将选择的语音单位转换为语音信号、并输出语音信号为用来与用户交互的可听语音。

[0042] TTS 系统通常对本领域的技术人员是已知的，如背景技术部分所描述的。但是图 2 示出了根据本公开的改进 TTS 系统的例子。根据一个实施例，系统 210 的一部分或全部可常驻在图 1 的远程信息处理单元 30 上并利用其进行处理。根据另一示例性实施例，TTS 系统 210 的一部分或全部可常驻在处于车辆 12 远程位置的计算设备（例如呼叫中心 20）上并利用其进行处理。例如，语言模型、声音模型等可存储在呼叫中心 20 的服务器 82 和 / 或数据库 84 之一的存储器中，并被发送至车辆远程信息处理单元 30，用于车内 TTS 处理。类似地，TTS 软件可使用呼叫中心 20 中的一个服务器 82 的处理器来处理。换句话说，TTS 系统 210 可常驻在远程信息处理单元 30 中或以任意期望方式分布在呼叫中心 20 和车辆 12 中。

[0043] 系统 210 可包括一个或多个文本源 212 和存储器，例如远程信息处理存储器 54，用于存储来自文本源 212 的文本及存储 TTS 软件和数据。系统 210 还可包括处理器，例如远程信息处理器 52，以便与存储器并联合下面的系统模块来处理文本和功能。预处理器 214 从文本源 212 接收文本，并将该文本转换为适当的词语等。合成引擎 216 将来自预处理器 214 的输出转换为适当的语言单位，例如短语、子句和 / 或句子。一个或多个语音数据库 218 存储记录的语音。单位选择器 220 从数据库 218 选择最佳对应于合成引擎 216 的输出的存储语音的单位。后处理器 222 修改或改编一个或多个选择的存储语音单位。一个或多个语言模型 224 被用作合成引擎 216 的输入，一个或多个声音模型 226 被用作单位选择器 220 的输入。系统 210 还可包括将所选的语音单位转换为音频信号的声音接口 228、和将音频信号转换为可听语音的例如远程信息音响系统的扬声器 239。系统 210 还可包括麦克风（例如远程信息处理麦克风 32）和声音接口 232，以将语音数字化为声音数据，用作后处理器 222 的反馈。

[0044] 文本源 212 可为任意适当的介质，可包括任何适当的内容。例如，文本源 212 可为一个或多个扫描文档、文本文件或应用程序数据文件、或任何其它适当的计算机文件等。文本源 212 可包括要被合成进语音并输出至文本转换器 214 的词语、数字、符号和 / 或标点。可使用任意适量和类型的文本源。

[0045] 预处理器 214 将来自文本源 212 的文本转换为词语、标志等。例如，在文本为数值形式的情形下，预处理器 214 可将数值转换为对应的词语。在另一实例中，当文本为标点时，通过大写字母或其它特殊字符（例如指示适当重要性的元音变化和声调、下划线或黑体）来强调，预处理器 214 可将该文本转换为适于被合成引擎 216 和 / 或单位选择器 220 使用的输出。

[0046] 合成引擎 216 从文本转换器 214 接收输出，并可将该输出布置成语言单位，可包括一个或多个句子、分句、短语、词语、子词等。引擎 216 可使用语言模型 224 来辅助语言单位

的最可能布置的协调。语言模型 224 提供将来自文本转换器 214 的输出布置成语言单位的规则、语法和 / 或语义。模型 224 还可定义系统 210 在任意给定时间以任意给定 TTS 模式预计的语言单位领域、和 / 或可提供管理语言单位的类型的规则等、和 / 或可逻辑地遵循其它类语言单位的韵律学和 / 或形成自然发声语音的韵律学。所述语言单位可包括语音对应物，例如音素字符串等，并可为音素 HMM 的形式。

[0047] 语音数据库 218 包括从一人或多人预先记录的语音。语音可包括预先存储的句子、分句、短语、词语、预存词语的子词等。语音数据库 218 还可包括与预先记录的语音相关联的数据，例如识别用来供单位选择器 220 使用的记录语音段的元数据。可使用任何适当类型和数量的语音数据库。

[0048] 单位选择器 220 将合成引擎 216 的输出与存储的语音数据作比较，并选择与合成引擎输出最佳对应的存储语音。单位选择器 220 选择的语音可包括预先记录的句子、分句、短语、词语、预先记录词语的子词等。选择器 220 可使用声音模型 226 来辅助存储语音的最可能或最佳对应备选的比较和选择。声音模型 226 可联合选择器 220 使用来比较和对比合成引擎输出的数据与存储的语音数据，评估其间的差别或相似性，最终使用确定的逻辑来识别最匹配的存储的语音数据并输出对应的存储的语音。

[0049] 通常，最佳匹配的语音数据为与合成引擎 216 的输出的不相似性程度最小或有最有可能为合成引擎 216 的输出的数据，如本领域技术人员所知各种技术任意一种所确定的。这类技术可包括动态时间规整分类器、人工智能技术、神经网络、自由音素识别器、和 / 或概率图形匹配器如隐马尔可夫模型（HMM）引擎。HMM 引擎对于制造多个 TTS 模型备选或假设的领域内的技术人员是公知的。在通过语音的声音特征分析来最终识别和选择表示合成引擎输出的最可能正确诠释的存储语音数据时考虑所述假设。更具体地，HMM 引擎产生语言单位假设“N 最佳”列表形式的统计模型，所述模型根据例如通过应用贝叶斯定理给出一个或其它语言单位的声音数据的观察次序的 HMM 计算信任值或可能性分级。

[0050] 在一个实施例中，单位选择器 220 的输出可直接传送给声音接口 228，或通过后处理器 222，没有后处理。在另一实施例中，后处理器 222 可从单位选择器 220 接收输出，用于进一步处理。

[0051] 在任一种情形下，声音接口 228 都将数字音频数据转换为模拟音频信号。接口 228 可为数模转换装置、电路和 / 或软件等。扬声器 230 为电子声音变换器，其将模拟音频信号转换为用户可听见的并可被麦克风 32 接收的语音。

[0052] 自动语音识别系统 -

现在参考图 3，示出了可用于执行当前所公开方法的 ASR 系统 310 的示例性架构。通常，车辆乘客与自动语音识别系统（ASR）口头地交互，用于下列基本目的中的一个或多个：训练系统理解车辆乘客的特定声音；存储不相关的语音，例如口语标志或口语控制词，如数字或关键词；或识别用于任意适当的目的（例如语音拨号、菜单导航、抄写、服务请求、车辆装置或装置功能控制等）的车辆乘客的语音。通常，ASR 从人语音提取声音数据、并将声音数据与存储的子词数据作比较和对比、选择与其它所选子词相关联的适当子词、并输出相关子词或词语，用于后处理，例如听写或抄写、地址薄拨号、存储至存储器、训练 ASR 模型或适应参数等。

[0053] ASR 系统通常是本领域技术人员已知的，图 3 只示出了一个具体的示例性 ASR 系统

310。尽管图示和描述了嵌在车辆远程信息处理单元 30 中，但是本领域的技术人员会认识到，可在呼叫中心部署或在车辆与呼叫中心之间分布类似的系统。系统 310 包括接收语音的装置（例如远程信息麦克风 32）和具有将语音数字化为声音数据的模数转换器的声音接口 33（例如远程信息处理单元 30 的声卡）。系统 310 还包括用于存储声音数据并存储语音识别软件和数据库的存储器（例如远程信息存储器 54）、以及处理声音数据的处理器（例如远程信息处理器 52）。处理器通过存储器并联合下列模块来运行：一个或多个前端处理器、预处理器或预处理器软件模块 312，用于将语音的声音数据流分析为参数表示，例如声音特征；一个或多个解码器或解码软件模块 314，用于将声音特征解码以获得对应于输入语音话语的数字子词或词语输出数据；和一个或多个后端处理器、后处理器或后处理器软件模块 316，用于为任意适当目的使用解码器模块 314 的输出数据。

[0054] 系统 310 还可从任何其它的适当音源 31 接收语音，该音源 31 可与预处理器软件模块 312 直接通信，如实线所示，或者可通过声音接口 33 与其间接通信。音源 31 可包括例如电话音源（例如语音邮件系统）或其它类电话服务。

[0055] 可使用一个或多个模块或模型作为解码器模块 314 的输入。首先，语法和 / 或辞典模型 318 可提供管理哪些词能逻辑地跟随其它词以形成有效句子的规则。从广义来说，辞典或语法可定义系统 310 在任何给定时间在任意给定 ASR 模式中期望的词语范畴。例如，如果系统 310 为用来训练指令的训练模式，那么辞典或语法模型 318 可包括系统 310 所知和所用的全部指令。在另一实例中，如果系统 310 为主菜单模式，那么有效的辞典或语法模型可包括系统 310 所预期的所有主菜单指令，例如呼叫、拨号、退出、删除、号码薄等。第二，声音模型 320 辅助与来自预处理器模块 312 的输入相对应的最可能的子词或词语的选择。第三，词语模型 322 和句子 / 语言模型 324 提供将所选子词或词语放入词语或句子上下文中的规则、句法和 / 或语义。并且，句子 / 语言模型 324 可定义系统 310 在任意给定时间在任意给定 ASR 模式下预期的句子范畴，和 / 或可提供管理哪些句子能够逻辑地跟随其它句子以形成有效扩展语音的规则等。

[0056] 根据另一示例性实施例，ASR 系统 310 的一部分或全部可常驻在位于车辆 12 远程位置的计算设备（例如呼叫中心 20）上并被其处理。例如，语法模型、声音模型等可存储在呼叫中心 20 中的服务器 82 和 / 或数据库 84 之一的存储器内，并被发送至车辆远程信息处理单元 30，用于车内语音处理。类似地，语音识别软件可使用呼叫中心 20 中的其中一个服务器 82 的处理器来处理。换句话说，ASR 系统 310 可常驻在远程信息处理单元 30 中，或以任意期望的方式分布在呼叫中心 20 和车辆 12 中。

[0057] 首先，声音数据从人的语音中提取，其中车辆乘客对麦克风 32 说话，麦克风 32 将话语转换成电信号，并将该信号发送至声音接口 33。由于空气压力的变化，麦克风 32 中的声音响应元件捕获乘客的语音话语，并将该话语转换为模拟电信号的相应变化，例如直流或电压。声音接口 33 接收模拟电信号，该电信号先被取样，使得模拟信号的值在离散的时刻被捕获，然后被量化，使得模拟信号的幅度在每个取样时刻被转换为连续的数字语音数据流。换句话说，声音接口 33 将模拟电信号转换为数字电信号。数字数据为二进制位，其缓存在远程信息存储器 54 中，然后被远程信息处理器 52 处理，或可在它们被处理器 52 开始接收时被实时处理。

[0058] 第二，预处理器模块 312 将连续的数字语音数据流转换为离散的声音参数序列。

更具体地，处理器52执行预处理器模块312，以将数字语音数据分段成具有例如10-30毫秒持续时间的重叠的语音或声音帧。所述帧对应于声音子词，例如音节、半音节、音素、双音、音位等。预处理器模块312还执行语音分析，以从每帧内的乘客语音提取声音参数，例如变时特征向量。乘客语音内的话语可表现为这些特征向量的序列。例如，如本领域技术人员所知的，特征向量可被提取且可包括例如音调、能线图、频谱属性和/或倒谱系数，它们可通过执行帧的傅里叶变换并使用余弦变换解相关声谱来获得。声帧和覆盖语音特定持续时间的相应参数被连接成要被解码的语音的未知测试图形。

[0059] 第三，处理器执行解码器模块314，以处理各测试图形的输入特征向量。解码器模块314也称为识别引擎或分类器，并使用存储的已知语音基准图形。类似测试图形，基准图形被定义为相关声帧和相应参数的连接。解码器模块314将要被识别的子词测试图形的声音特征向量与存储的子词基准图形作比较和对比，评估其间的区别或相似性的量，最终使用判定逻辑来选择最佳匹配的子词作为识别的子词。通常，最佳匹配的子词为对应于所存储的已知基准图形的子词，该已知基准图形与本领域内技术人员所知各种技术之一所确定的用于分析和识别子词的测试图形具有最小不相似性或者最有可能为该测试图形。这类技术可包括动态时间规整分类器、人工智能技术、神经网络、自由音素识别器、和/或概率图形匹配器如隐马尔可夫模型(HMM)引擎。

[0060] HMM引擎对于制造声音输入的多个语音识别模型假设领域内的技术人员是已知的。在最终识别和选择识别输出中考虑所述假设，该识别输出表示通过语音的特征分析的声音输入的最可能正确解码。更具体地，HMM引擎产生子词模型假设“N最佳”列表形式的统计模型，所述模型根据例如通过应用贝叶斯定理给出一个或另一个子词的声音数据的观察次序的HMM计算信任值或可能性分级。

[0061] 贝叶斯HMM过程识别与声音特征向量的给定观察次序的最可能的发音或子词次序相对应的最佳假设，并且其信任值依赖于许多因素，包括与输入声音数据相关的声音信号相对于噪声的比率。HMM还可包括称为对角高斯混合的统计分布，其获得每个子词的每个观察特征向量的可能性得分，该得分可用于记录假设的N最佳列表。HMM引擎还可识别和选择其模型的可能性得分是最高的子词。

[0062] 通过类似的方式，可连接一连串子词的各自HMM，以建立一个或多个词语HMM。其后，可产生并进一步评估一个或多个词语基准图形和相关参数值的N最佳列表。

[0063] 在一个例子中，语音识别解码器314使用适当的声音模型、语法和算法来处理特征向量，以产生基准图形的N最佳列表。如本文所使用的，术语“基准图形”可与模型、波形、模板、富信号模型、样本、假定或其它类基准互相交换。基准图形可包括表示一个或多个词语或子词的一系列特征向量，其可基于特定的说话者、说话类型和声音环境条件。本领域的技术人员会认识到，可通过ASR系统的适当基准图形训练来产生基准图形，并且该基准图形存储在存储器中。本领域的技术人员还会认识到，可控制存储的基准图形，其中基准图形的参数值基于ASR系统的基准图形训练与实际使用之间的语音输入信号的差别而被改编。例如，基于来自不同车辆乘客或不同声音条件的有限训练数据量，针对一个车辆乘客或特定声音条件训练的一组基准图形可被改编并存储为用于不同车辆乘客或不同声音条件的另一组基准图形。换句话说，所述基准图形可不必是固定的，且可在语音识别期间进行调节。

[0064] 使用词汇语法及任意适当的解码算法和声音模型，处理器从存储器中存取解释测试图形的几个基准图形。例如，处理器可产生 N 最佳词汇结果或基准图形连同相应参数值的列表，并将其存储到存储器中。示例性参数值可包括词汇的 N 最佳列表中各基准图形的信任分数及相应段持续时间、可能性分数、信噪比（SNR）值等。词汇的 N 最佳列表可按照参数值的降序排列。例如，具有最高信任分数的词汇基准图形为第一最佳基准图形，等等。一旦建立了一串识别的子词，它们就可用于与来自词语模型 322 的输入来构造词语，和与来自语言模型 324 的输入来构造句子。

[0065] 最后，后处理器软件模块 316 从解码器模块 314 接收输出数据，用于任意适当目的。在一个例子中，后处理器软件模块 316 可从一个或多个词语基准图形的 N 最佳列表识别或选择其中一个基准图形作为识别的语音。在另一例子中，后处理器模块 316 可用于将声音数据转换为文本或数字，以便与 ASR 系统或其它车辆系统的其它方面一起使用。在另一例子中，后处理器模块 316 可用于提供解码器 314 或预处理器 312 的训练反馈。更具体地，后处理器 316 可用于训练解码器模块 314 的声音模型，或训练预处理器模块 312 的自适应参数。

[0066] 方法 -

现在参考图 4，示出了自动语音识别方法 400，其在车辆远程信息处理单元 30 的操作环境内可使用图 2 的 TTS 系统 210 和 / 或图 3 的 ASR 系统 310 的适当编程以及使用图 1 中所示其它部件的适当硬件和编程来执行。例如，语音识别硬件、固件和软件可常驻在计算机 18 和 / 或呼叫中心 20 的其中一个服务器 82 上。换句话说，ASR 系统 310 可常驻在远程信息处理单元 30 中或以任意期望方式分布在车辆 12 和计算机 18 和 / 或呼叫中心 20 上。

[0067] 基于上面的系统描述以及下面结合其余附图描述的方法，上述这种编程和硬件的使用对本领域的技术人员是清楚的。本领域的技术人员还会认识到，该方法在其它操作环境下可使用其它 ASR 系统来执行。该方法的步骤可以连续地处理也可以不连续地处理，本发明可包含这类步骤的任意次序、重叠或并行处理。

[0068] 通常，根据下列步骤，语音信号处理方法 400 改进了自动语音识别 ：通过麦克风从用户接收对文本信息的回复话语，所述麦克风将所述回复话语转换为语音信号 ；使用至少一个处理器预处理所述语音信号，以从所接收的语音信号提取声音数据 ；使用与所述文本信息相关的会话语境识别多个声音模型中的一个声音模型，以解码所述声音数据 ；以及使用识别的声音模型解码所述声音数据，以产生用于所述回复话语的多个假设。

[0069] 更具体地，参考图 4 并间或参考图 1-3，方法 400 以任意适当的方式开始于步骤 402。例如，车辆用户优选地通过下述开始与远程信息处理单元 30 的用户接口交互，按下用户接口按钮 34 来开始用户输入语音指令的对话，当操作于语音识别模式时，该语音指令由远程信息处理单元 30 翻译。使用音响系统 36，远程信息处理单元 30 可通过针对用户或乘客的指令播放声音或提供口头请求来确认按钮激活。

[0070] 在步骤 404，接收文本信息。例如，可通过通信系统在远程信息处理单元 30 接收文本信息。文本信息可为短信服务（SMS）类信息、扩展信息服务、移动瞬时信息和 / 或任意其它适当类型的信息服务，并使用标准邮件协议、例如通过 TCP/EP 的 SMTP、会话启动协议、专有协议和 / 或任何其它适当协议。

[0071] 在步骤 406，文本信息由会话语境特定的语言模型 407 来处理，以确认与文本信息

相对应的会话语境。例如，来自步骤 404 的文本信息可使用存储在客户端装置（例如远程信息处理单元 30）上的语言模型并使用客户端装置的至少一个处理器（例如处理器 52）来处理。并且，会话语境可包括幽默会话的“幽默”、或用于有关用餐计划的会话的“用餐”、或用于情爱会话的“浪漫”、或用于闲话聊天的“闲聊”、或用于邀请或相关回复的“邀请”、或用于介绍类会话的“问候”。会话语境可包括上述所有例子中的一种或多种，和 / 或任意其它适当类型的会话语境。在一个实施例中，每个语言模型 407 都对应于一种会话语境，并且在语音识别运转之前可以任何适当的方式通过多个扬声器来展开和训练。在另一实施例中，如果统计语言模型未遇到起初训练资料中的文本，那么可使用任意新的文本输入来以任何适当方式来更新该统计语言模型。

[0072] 在步骤 408，文本信息还可使用特定情感语境语言模型 409 来识别对应于文本信息的情感语境。例如，来自步骤 406 的文本信息可使用存储在客户端装置（例如远程信息处理单元 30）上的语言模型并使用客户端装置的至少一个处理器（例如处理器 52）来处理。并且，情感语境可包括用于不友善对话的“生气”、或用于高兴对话的“快乐”、或用于不高兴会话的“悲伤”、或“困惑”等。情感语境可包括前述所有例子的任意一个或多个和 / 或任意其它适当类型的情感语境。在一个实施例中，每个语言模型 409 都对应于一种情感语境，并且在语音识别运转之前可以任意适当的方式通过多个扬声器来展开和训练。另外，情况语境可用于产生恰当的 TTS 翻译，例如，以向用户 / 听众可听地表达情感语境。

[0073] 在步骤 410，语音由文本信息合成。例如，来自步骤 406 和 / 或 408 的文本信息被预处理，以将文本转换为适于语音合成的输出。更具体地，TTS 预处理器 214 可将文本信息转换为词语、标志等，以便被 TTS 合成引擎 216 使用。然后，该输出可配置成语言单位。例如，TTS 合成引擎 216 可从文本转换器 214 接收输出，并可通过语言模型 224 将输出配置成语言单位，该语言单位可包括一个或多个句子、分句、短语、词语、子词等。语言单位可包括语音对应物，例如音素字符串等。其后，语言单位可与存储的语音数据作比较，选择与语言单位最佳对应的语音作为表示文本信息的语音。例如，单位选择器 220 可使用 TTS 声音模型 228 将从合成引擎 216 输出的语言单位与存储在第一语音数据库 218 中的语音数据作比较，并选择具有与合成引擎输出最佳对应的相关数据的存储语音。

[0074] 在步骤 412，合成的语音被发送给用户。例如，由选择器 220 从数据库 218 选择的预先记录的语音可通过远程信息处理单元 30 的接口 228 和扬声器 230 输出。

[0075] 在步骤 414，从用户接收语音回复。例如，可通过麦克风从用户接收话语，麦克风将话语转换成语音信号。更具体地，远程信息麦克风 32 可用于将用户语音话语转换为电信号用于发送给声音接口 33，声音接口 33 可将该语音数字化成声音数据。

[0076] 在步骤 416，语音信号被预处理，从该语音信号提取声音数据。例如，语音信号可使用 ASR 处理器 312 或任何其它适当的远程通信预处理器或任意类型的处理装置来预处理。该语音信号通过预处理器从语法上分析成参数表示流，例如声音特征或声音特征向量等。例如，来自声音接口 33 的声音数据可通过上述 ASR 系统 310 的预处理器模块 312 预处理。

[0077] 在步骤 417，对于声音数据的下游解码，使用识别的会话语境来识别存储在客户端装置的多个声音模型中的一个声音模型。在第一实施例中，仅使用所述会话语境。在第二实施例中，还可使用情感语境。在第一实施例中，每个声音模型都可特定于其中一种会话语境。在第二实施例中，多个模型可包括会话 / 情感模型的置换矩阵。例如，模型可包括“用

餐”/“快乐”声音模型、“用餐”/“生气”声音模型、“闲聊”/“困惑”声音模型等。

[0078] 还是在步骤 417，产生的声音特征向量由识别的模型来解码，以为接收的语音产生多个假设。例如，所述多个假设可为假设的 N 最佳列表，ASR 系统 310 的解码器模块 314 可用于解码声音特征向量。会话和 / 或情感语境特定的统计语言模型 407、409 可用于以任意适当的方式辅助解码。

[0079] 在步骤 418，确定与回复话语的多个假设中的至少一个相关的信任值是否大于信任阈值。信任值和计算对于本领域的技术人员是公知的，可以任意适当的方式来计算，包括使用监督学习技术、神经网络等。对于任意给定应用，特定阈值可凭经验来确定，因此可为适于给定环境和情形的任意值。例如，信任阈值的示例可为 75% 的信任水平。更具体地，可设定可接受的信任阈值，且如果与一个或多个假设相关的信任分数小于该阈值，那么从步骤 416 提取的声音数据及从步骤 406 和 / 或 408 提取的会话语境可被发送至语音识别服务器并因而在步骤 420 接收。否则，该方法可进行至步骤 419。

[0080] 在步骤 419，所述多个假设可被后处理成将其中一个假设识别为回复话语，且识别的假设可使用任何适当的文本信息技术和协议从客户端装置直接输出作为回复文本信息的至少一部分。例如，ASR 后处理器 216 可包括任意适当的装置或模块来构成文本信息。并且，后处理器可通过与远程信息处理单元 30 的配合，直接通过蜂窝通信输出文本信息，或者可通过例如蓝牙连接间接输出文本信息至用户电话等，进而可通过蜂窝通信输出文本信息。

[0081] 在另一实施例中，可省略步骤 417-419，其中来自步骤 316 的语音信号的声音数据及在步骤 406 和 / 或 408 中识别的语境可直接发送至语音识别服务器。这些数据可以任意适当的方式打包，并通过数据连接发送，例如，通过经数据通道（例如私人或公共分组交换网络（PSN））的分组数据传输，或使用车辆 12 内车载的和在呼叫中心 20 和 / 或计算机中的调制解调器经蜂窝声音通道的数据、或以任何其它的适当方式。在使用经由声音通道的调制解调器通信的情形下，数据可使用任何适当的声码器从车辆 12 发送，该声码器可包含在蜂窝芯片 50 中。

[0082] 在步骤 420，发送的声音数据在服务器 18 和 / 或 82 接收。例如，数据可经由分组数据传输、经由通过声音协议的数据、和 / 或经由任意其它适当的方式来接收。所述数据可保存在任何适当的位置。

[0083] 在步骤 422，确定是否在识别相应声音模型的下游步骤中使用情感语境。在所示实施例中，例如，如果声音数据第一次在服务器 18 和 / 或 82 被处理，那么可省略情感语境。在另一实施例中，然而，情感语境可用于从一开始补充会话语境。

[0084] 在步骤 424，对于声音数据的下游解码，使用会话语境识别存储在服务器 18 和 / 或 82 的多个声音模型 425 中的一个声音模型。在第一实施例中，只使用会话语境。在第二实施例中，还可使用情感语境。在第一实施例中，每个声音模型 425 可特定于其中一种会话语境。在第二实施例中，多个模型 425 可包括会话 / 情感模型的置换矩阵。例如，模型可包括“用餐”/“快乐”声音模型、“用餐”/“生气”声音模型、“闲聊”/“困惑”声音模型等。

[0085] 在步骤 426，产生的声音特征向量由识别的模型来解码，以为接收的语音产生多个假设。例如，所述多个假设可为假设的 N 最佳列表，ASR 系统 310 的解码器模块 314 可用于解码声音特征向量。会话和 / 或情感语境特定的统计语言模型 427、429 可用于以任意适当

的方式辅助解码。

[0086] 在步骤428，所述多个假设可被后处理，以将其中一个假设识别为回复话语。例如，ASR系统310的后处理器316可使用信任阈值等对步骤426的假设进行后处理，以将多个假设中的一个识别为接收的语音。可使用会话和/或情感语境特定的统计语言模型427、429来以任意适当方式来辅助解码。后处理器316还可用于产生对应于识别的语音的文本数据。

[0087] 在步骤430，可响应于步骤428的识别的语音来改编一个或多个模型。例如，可改编或训练声音模型425和/或统计语言模型427、429，使得该模型反映最新或新近接收的文本信息缩略语和图形，从而可更加精确地识别出文本信息回复。声音模型改编和训练是本领域的技术人员公知的，可使用任何适当的技术。该步骤还可包括将多个语境特定的语言模型427、429从服务器发送至客户端装置，以更新存储在客户端装置上的语言模型407、409，用于随着时间改进文本信息语境分类。

[0088] 在一个实施例中，该处理方法可直接进行至步骤440，以输出或发送对应于识别的语音的文本数据。在另一实施例中，该处理方法进行至步骤432，以便用户确认和/或重新输入。

[0089] 在步骤432和434，步骤428的假设可发送至车辆12并由其接收。例如，用于假设或与其相关的数据可通过分组数据连接、经由蜂窝语音通道的数据或以任意其它适当的方式发送，并由远程信息处理单元30接收且在其上存储于任意适当的存储器内。在一个实施例中，所述假设为文本数据格式。在另一实施例中，所述假设为可为声音文件格式的声音数据。

[0090] 在步骤436，用户对文本信息的答复的假设可以任意适当的方式展现给用户。例如，所述假设可通过远程信息处理单元用户接口的显示屏以可视文本的方式展现，和/或通过远程信息处理单元用户接口的扬声器以可听的方式展现。

[0091] 在步骤438，确定用户是否将该假设确认为用户意欲对文本信息的回复。如果用户确认，那么所述方法进行至步骤440，而如果用户拒绝该假设，那么该方法进行至步骤442。

[0092] 在步骤440，使用任意适当的文本信息技术和协议，对应于识别的语音的文本数据被输出或发送为回复文本信息。该步骤可由车辆12的发送来触发，并由以任意适当方式通信的任意适当指令的服务器18和/或82接收。

[0093] 在步骤442，确定用户在步骤438的拒绝对于当前分析的回复是否是第一次。如果是，那么该方法可循环回步骤422，然后可使用情感语境，特别是如果之前未使用情感语境来分析当前的回复。该步骤可由车辆12的发送来触发，并由以任意适当方式通信的任意适当指令的服务器18和/或82接收。然而，如果不是，那么该方法可进行至步骤444。

[0094] 在步骤444，可请求用户重复或重述对在步骤404接收的文本信息的回复。例如，可以任意适当的方式使用远程信息处理单元30，以可视地或可听地发送信息，例如“请重复或重述您对文本信息的回复”。其后，该方法循环回步骤414，以接收用户重复或重述的对文本信息回复。

[0095] 在步骤446，方法400可以任意适当的方式结束。

[0096] 该方法或其一部分可在计算机程序产品中执行，包括承载在计算机可读介质上以便被一个或多个计算机的一个或多个处理器使用从而执行一个或多个方法步骤的指令。计

算机程序产品可包括：一个或多个软件程序，该软件程序包括为源代码、目标代码、可执行代码或其它格式的程序指令；一个或多个固件程序；或硬件描述语言（HDL）文件；以及任何与程序相关的数据。所述数据可包括数据结构、查寻表或任意其它适当格式的数据。程序指令可包括程序模块、例行程序、程序、对象、组件等。计算机程序可在一个计算机或彼此通信的多个计算机上执行。

[0097] 程序可嵌在计算机可读介质上，该介质可包括一个或多个存储装置、制品等。示例性计算机可读介质包括计算机系统存储器（例如 RAM（随机存取存储器）、ROM（只读存储器））、半导体存储器（例如，EPROM（可擦写可编程 ROM）、EEPROM（电可擦写可编程 ROM））、闪存）、磁盘或光盘或磁带或光带等等。计算机可读介质还可包括计算机到计算机的连接，例如，当数据通过网络或其它通信连接（有线或者无线或者它们的组合）传输或提供时。上述例子的任意组合也包含在计算机可读介质的范围内。因此应当理解，该方法可被能够执行对应于所公开方法的一个或多个步骤的指令的任意电子制品和 / 或装置至少部分地执行。

[0098] 应当理解，前面的内容是本发明一个或多个优选实施例的描述。本发明不限于本文所公开的特定实施方式，而是仅由所附权利要求限定。另外，前面描述中所含的内容涉及特定实施例，并不构成对本发明范围或权利要求中所用术语定义的限制，除非上面明确定义了的术语或短语。各种其它实施方式以及对所公开实施例的各种改变和修改对本领域的技术人员是清楚的。例如，本发明可应用于语音信号处理的其它领域，例如移动远程通信、通过互联网协议应用的话音等。所有这些其它实施方式、改变和修改都落在所附权利要求的范围内。

[0099] 如在本说明书和权利要求书中所使用的，当结合一个或多个部件或其它项目的列表使用时，术语“例如”、“譬如”、“诸如”和“如同”、动词“包含”、“具有”和“包括”及它们的其他动词形式均构造为开放式的，意味着所述列表并不认为排除其它、另外的部件或项目。其它术语使用它们最广义的合理含义来解释，除非它们被用在需要不同诠释的上下文中。

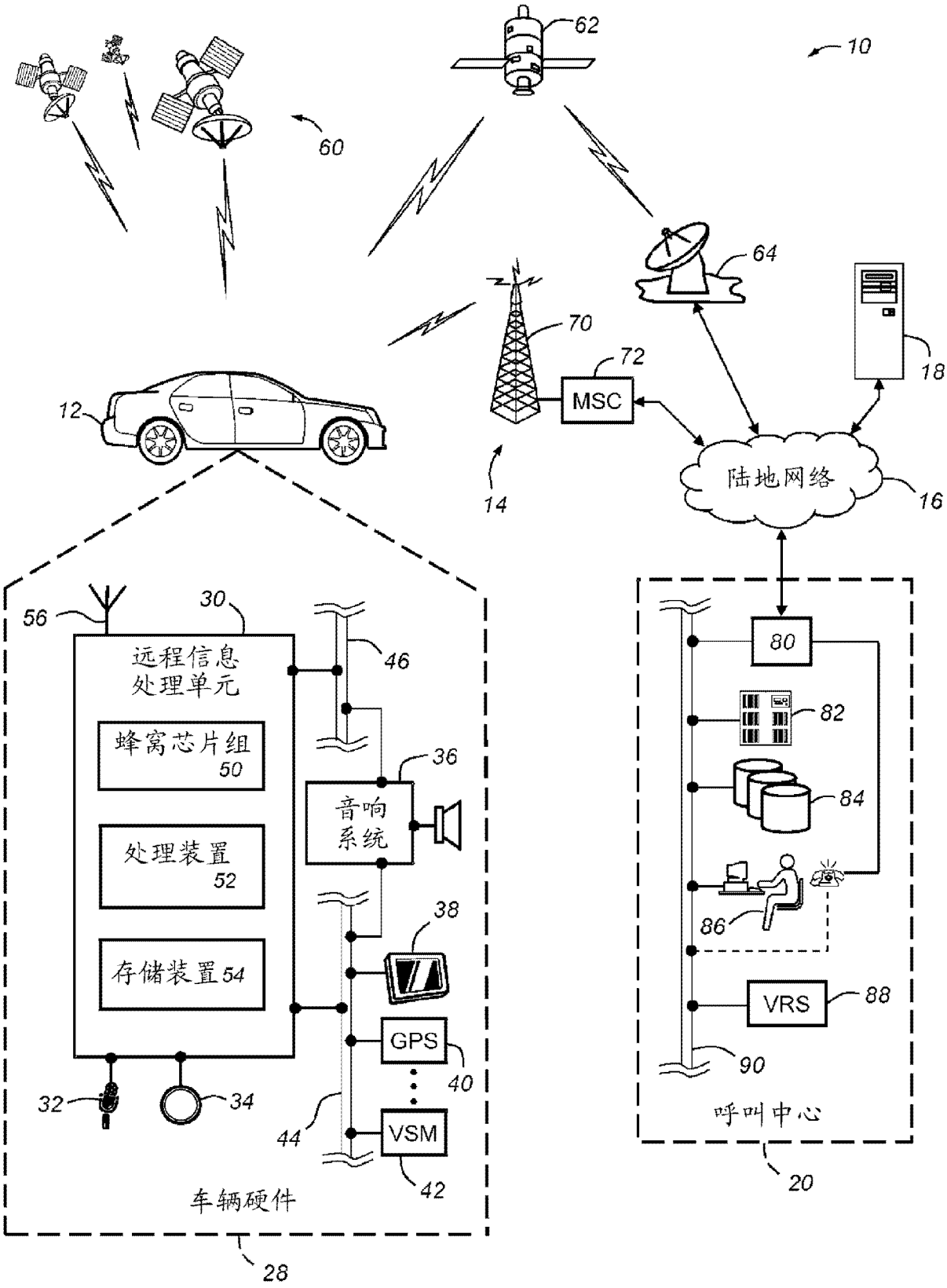

62
10
60
64
70
72
MSC
18
12
14
陆地网络
16
56
30
远程信息
处理单元
蜂窝芯片组
50
处理装置
52
存储装置54
46
36
音响
系统
38
GPS
40
32
34
44
VSM
42
车辆硬件
28
80
82
84
86
VRS
88
90
呼叫中心
20

图 1

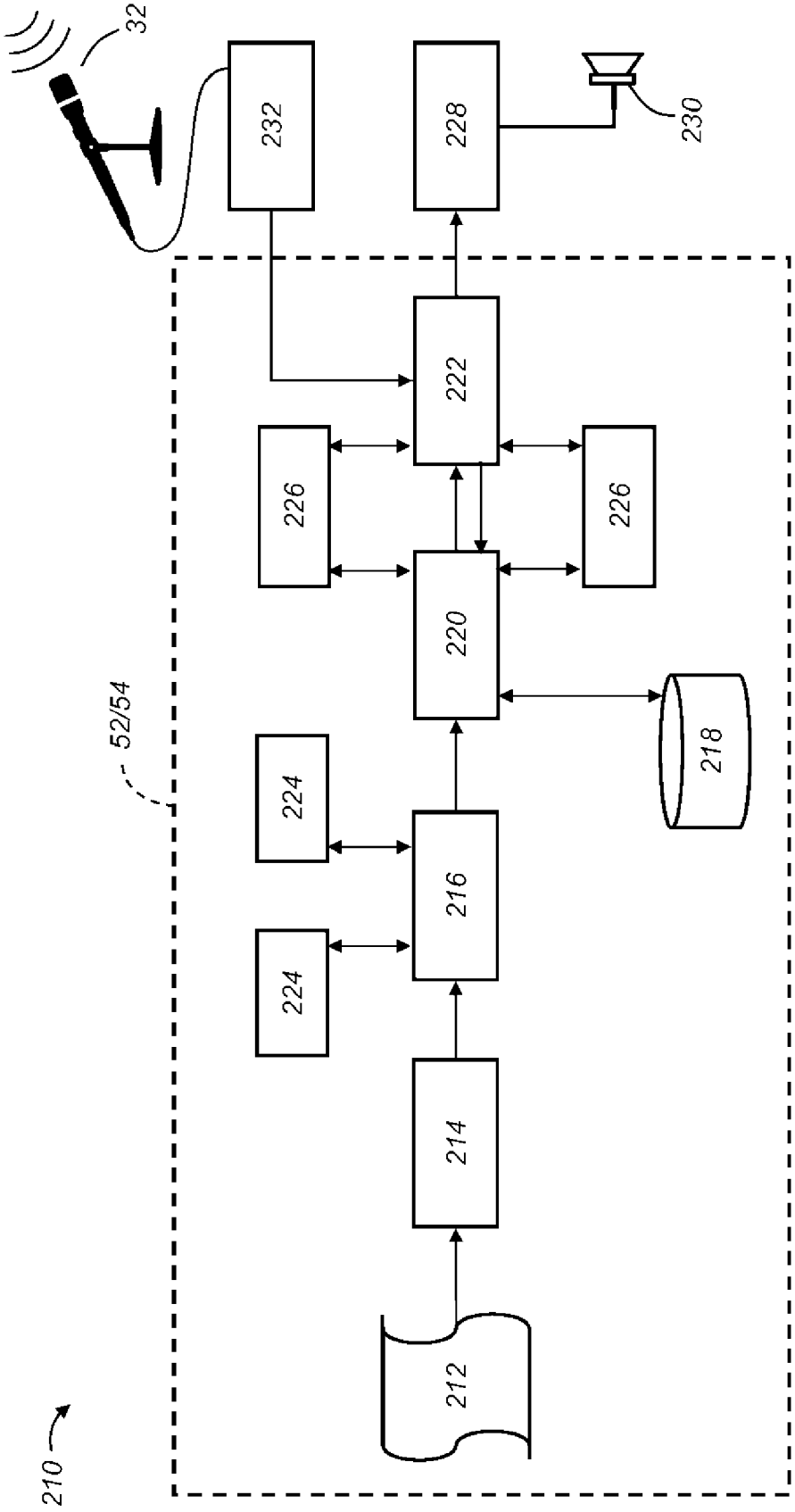


图 2

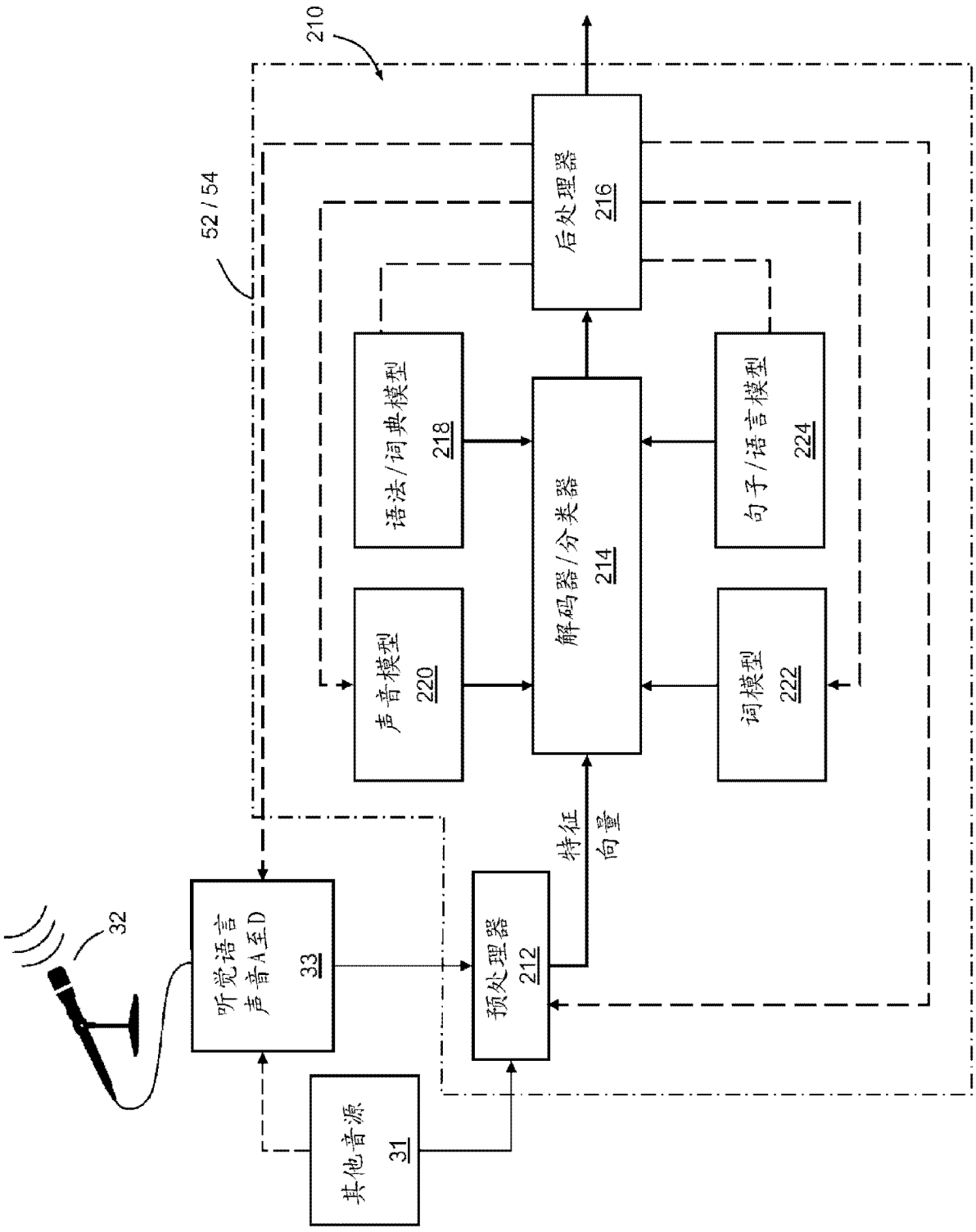

210
52 / 54
后处理器
216
语法/词典模型
218
解码器/分类器
214
句子/语言模型
224
声音模型
220
词模型
222
特征
向量
预处理器
212
听觉语言
声音A至D
33
32
其他音源
31

图 3

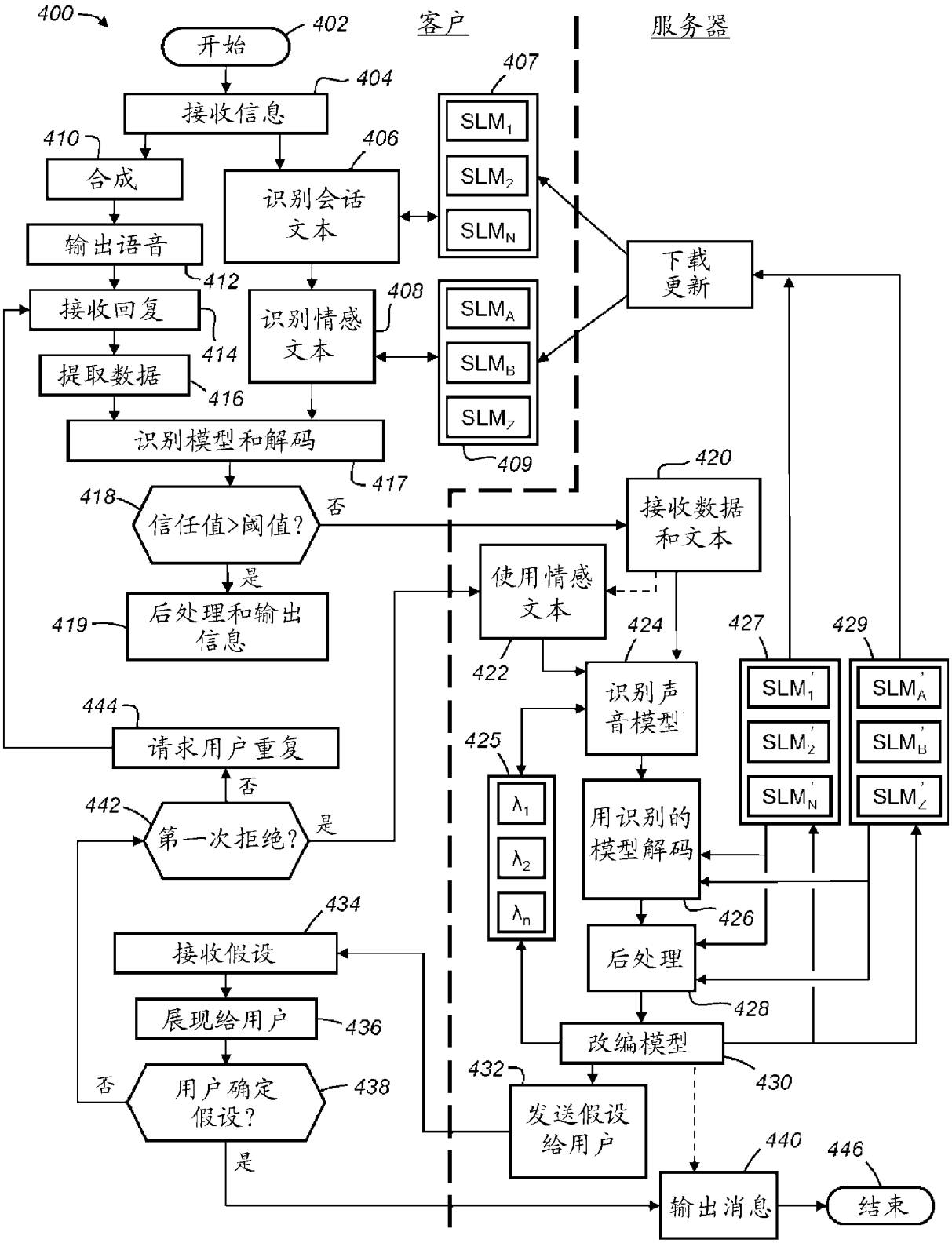

400
开始
402
客户
服务器
接收信息
404
407
SLM1
SLM2
SLMN
410
合成
406
识别会话文本
输出语音
412
接收回复
414
识别情感文本
408
SLMA
SLMB
SLMZ
下载更新
提取数据
416
识别模型和解码
417
409
420
418
信任值>阈值?
否
接收数据和文本
是
使用情感文本
后处理和输出信息
419
422
424
427
429
识别声音模型
SLM'1
SLM'2
SLM'N
SLM'A
SLM'B
SLM'Z
444
请求用户重复
425
λ1
λ2
λn
否
442
第一次拒绝?
是
用识别的模型解码
426
434
接收假设
后处理
428
展现给用户
436
改编模型
否
用户确定假设?
438
432
发送假设给用户
430
440
446
是
输出消息
结束

图 4