(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 102243870 A
(43)申请公布日 2011.11.16

(21)申请号 201110123670.9

(22)申请日 2011.05.13

(30)优先权数据
12/780402 2010.05.14 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 J.M. 斯蒂芬 G. 塔尔瓦
R. 琴加尔瓦拉延

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

(51)Int.Cl.
G10L 13/02(2006.01)

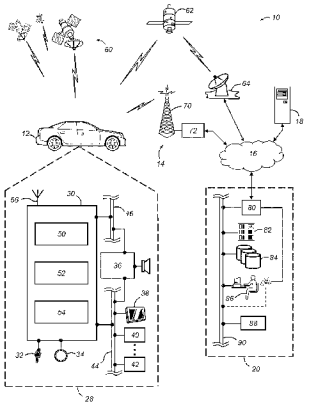
权利要求书 1 页 说明书 13 页 附图 3 页

(54)发明名称

语音合成中的语音调节

(57)摘要

本发明涉及语音合成中的语音调节。一种用于语音合成的方法和系统，第一文本输入和第二文本输入接收在文本至语音系统中，并且使用所述系统的处理器处理为分别与第一说话者和第二说话者的所存储语音相应的各自的第一语音输出和第二语音输出。第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。



CN 102243870 A

1. 一种语音合成的方法，其包括步骤：

(a) 在文本至语音系统中接收第一文本输入和第二文本输入；

(b) 使用所述系统的处理器将第一文本输入和第二文本输入处理为与分别来自第一说话者和第二说话者的所存储语音相应的各自的第一语音输出和第二语音输出；以及

(c) 使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

2. 如权利要求 1 所述的方法，还包括步骤：

(d) 输出第一说话者的第一语音输出；以及

(e) 输出第二说话者的经调解的第二语音输出。

3. 如权利要求 2 所述的方法，其中，所述第一语音输出是导航指令，而所述第二语音输出是导航变量。

4. 如权利要求 3 所述的方法，其中，所述导航指令是定向调动，而所述导航变量是街道名称。

5. 如权利要求 2 所述的方法，还包括步骤：(f) 修改结合处理来自第二说话者的所存储语音使用的模型。

6. 如权利要求 5 所述的方法，其中，步骤 (f) 包括修改隐马尔可夫模型。

7. 如权利要求 1 所述的方法，其中，步骤 (c) 包括：

(c1) 对于第一说话者的至少一个说话者特定特性分析第一语音输出的声学特征；

(c2) 基于第一说话者的所述至少一个说话者特定特性，调整用于滤波来自第二语音输出的声学特征的声学特征滤波器；以及

(c3) 使用步骤 (c2) 中调整的滤波器滤波来自第二语音输出的声学特征。

8. 如权利要求 7 所述的方法，其中，步骤 (c3) 包括：调整梅尔频率倒频谱滤波器的至少一个参数，数包括滤波器组中心频率、滤波器组截止频率、滤波器组带宽、滤波器组形状或滤波器增益的至少一个。

9. 一种计算机程序产品，其包括在计算机可读介质上且由语音合成系统的计算机处理器可执行以使系统实施下述步骤的指令，所述步骤包括：

(a) 在文本至语音系统中接收第一文本输入和第二文本输入；

(b) 使用所述系统的处理器将第一文本输入和第二文本输入处理为与分别来自第一说话者和第二说话者的所存储语音相应的各自的第一语音输出和第二语音输出；以及

(c) 使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

10. 一种语音合成系统，其包括：

第一文本源；

第二文本源；

第一语音数据库，包括预先记录的来自第一说话者的语音；

第二语音数据库，包括预先记录的来自第二说话者的语音；

预处理器，其将文本转换成能够合成的输出；

处理器，其将来自第一文本源和第二文本源的第一文本输入和第二文本输入转换为与分别来自第一说话者和第二说话者的预先记录的语音相应的各自的第一语音输出和第二语音输出；以及

后处理器，其使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

语音合成中的语音调节

技术领域

[0001] 本发明总体涉及语音信号处理，且更具体地，涉及语音合成。

背景技术

[0002] 语音合成是通过人工方法从文本产生语音。例如，文本至语音（TTS）系统从文本合成语音，以向传统计算机到人可视输出设备，例如计算机监视器或显示器提供替换物。存在多种 TTS 合成的变体，包括共振峰 TTS 合成以及拼接调整 TTS 合成。共振峰 TTS 合成不输出记录的人类语音，而是输出计算机产生的音频，其听起来是人工的或机器人的。在拼接调整 TTS 合成中，存储的人类语音段被拼接，并输出以产生听起来更加平滑的，更加自然的语音。

[0003] TTS 系统可以包括下面的基本元素。原始文本源包括将被合成到语音的词语、数字、符号、缩略语和 / 或标点。语音数据库包括来自一个或多个人的预先记录的语音。预处理器将原始文本转换为与书写的词语等同的输出。合成引擎按照发音转录预处理器输出，并且将预处理器输出转换为适当的语言单元，例如，句子、从句和 / 或短语。单元选择器从语音数据库选择与来自合成引擎的语言单元最好地相应的语音单元。声学接口将选择的语音单元转换为音频信号，并且扬声器将语音信号转换为可听语音。

[0004] TTS 合成遇到的一个问题是一些应用会使用从具有明显不同声音的不同人记录的语音。例如，TTS 使能的车辆导航系统使用具有多个部分语法的声音导航，可包括定向调动话语（例如，“执行合法的到…的掉头”）和街道名称话语（例如，“North Telegraph Road”）。调动话语可以由导航服务提供者的第一发话者产生，街道名称话语可以由地图数据提供者的第二发话者产生。当在语音导航期间将话语一起播放时，组合的话语会使用户听起来不舒服。例如，用户可能察觉到从调动话语到街道名称话语的转变，例如，由于发话者之间语调的差异。

发明内容

[0005] 根据本发明的一方面，提供一种语音合成的方法。所述方法包括步骤 (a) 在文本至语音系统中接收第一文本输入和第二文本输入 ；(b) 使用所述系统的处理器将第一文本输入和第二文本输入处理为分别与第一说话者和第二说话者的所存储语音相应的各自的第一语音输出和第二语音输出 ；以及 (c) 使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

[0006] 根据本发明的另一方面，提供一种计算机程序产品，其包括在计算机可读介质上并且由文本至语音系统的计算机处理器可执行以使系统实施上述步骤的指令。

[0007] 根据本发明的额外方面，提供一种语音合成系统，包括 ：第一文本源 ；第二文本源 ；第一语音数据库，包括预先记录的来自第一说话者的语音 ；第二语音数据库，包括预先记录的来自第二说话者的语音 ；和预处理器，将文本转换成能够合成的输出。所述系统还包括处理器，将来自第一文本源和第二文本源的第一文本输入和第二文本输入转换为分别

与第一说话者和第二说话者的预先记录的语音相应的各自的第一语音输出和第二语音输出；以及后处理器，使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

[0008] 本发明还提供如下方案：

1. 一种语音合成的方法，其包括步骤：

(a) 在文本至语音系统中接收第一文本输入和第二文本输入；

(b) 使用所述系统的处理器将第一文本输入和第二文本输入处理为与分别来自第一说话者和第二说话者的所存储语音相应的各自的第一语音输出和第二语音输出；以及

(c) 使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

[0009] 2. 如方案 1 所述的方法，还包括步骤：

(d) 输出第一说话者的第一语音输出；以及

(e) 输出第二说话者的经调解的第二语音输出。

[0010] 3. 如方案 2 所述的方法，其中，所述第一语音输出是导航指令，而所述第二语音输出是导航变量。

[0011] 4. 如方案 3 所述的方法，其中，所述导航指令是定向调动，而所述导航变量是街道名称。

[0012] 5. 如方案 2 所述的方法，还包括步骤：(f) 修改结合处理来自第二说话者的所存储语音使用的模型。

[0013] 6. 如方案 5 所述的方法，其中，步骤 (f) 包括修改隐马尔可夫模型。

[0014] 7. 如方案 1 所述的方法，其中，步骤 (c) 包括：

(c1) 对于第一说话者的至少一个说话者特定特性分析第一语音输出的声学特征；

(c2) 基于第一说话者的所述至少一个说话者特定特性，调整用于滤波来自第二语音输出的声学特征的声学特征滤波器；以及

(c3) 使用步骤 (c2) 中调整的滤波器滤波来自第二语音输出的声学特征。

[0015] 8. 如方案 7 所述的方法，其中，步骤 (c3) 包括：调整梅尔频率倒频谱滤波器的至少一个参数，数包括滤波器组中心频率、滤波器组截止频率、滤波器组带宽、滤波器组形状或滤波器增益的至少一个。

[0016] 9. 如方案 7 所述的方法，其中，所述至少一个说话者特定特性包括声道或鼻腔相关特性中的至少一个。

[0017] 10. 如方案 9 所述的方法，其中，所述特性包括长度、形状、转换函数、格式或音调频率中的至少一个。

[0018] 11. 一种计算机程序产品，其包括在计算机可读介质上且由语音合成系统的计算机处理器可执行以使系统实施下述步骤的指令，所述步骤包括：

(a) 在文本至语音系统中接收第一文本输入和第二文本输入；

(b) 使用所述系统的处理器将第一文本输入和第二文本输入处理为与分别来自第一说话者和第二说话者的所存储语音相应的各自的第一语音输出和第二语音输出；以及

(c) 使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

[0019] 12. 如方案 11 所述的产品，其中，步骤 (c) 包括：

(c1) 对于第一说话者的至少一个说话者特定特性分析第一语音输出的声学特征；

(c2) 基于第一说话者的所述至少一个说话者特定特性，调整用于滤波来自第二语音输出的声学特征的声学特征滤波器；以及

(c3) 使用步骤 (c2) 中调整的滤波器滤波来自第二语音输出的声学特征。

[0020] 13. 一种语音合成系统，其包括：

第一文本源；

第二文本源；

第一语音数据库，包括预先记录的来自第一说话者的语音；

第二语音数据库，包括预先记录的来自第二说话者的语音；

预处理器，其将文本转换成能够合成的输出；

处理器，其将来自第一文本源和第二文本源的第一文本输入和第二文本输入转换为与分别来自第一说话者和第二说话者的预先记录的语音相应的各自的第一语音输出和第二语音输出；以及

后处理器，其使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

[0021] 14. 如方案 13 所述的系统，还包括：

声学接口，其将语音输出转换为音频信号；以及

扬声器，其将音频信号转换为可听语音。

[0022] 15. 如方案 14 所述的系统，其中，所述扬声器输出第一说话者的第一语音输出，并输出经调节的第二说话者的第二语音输出。

[0023] 16. 如方案 13 所述的系统，其中，所述后处理器修改结合处理来自第二说话者的所存储的语音使用的模型。

[0024] 17. 如方案 13 所述的系统，其中，所述后处理器对于第一说话者的至少一个说话者特定特性分析第一语音输出的声学特征，基于第一说话者的所述至少一个说话者特定特性调整用于滤波来自第二语音输出的声学特征的声学特征滤波器，以及使用调整的滤波器滤波来自第二语音输出的声学特征。

[0025] 18. 如方案 17 所述的系统，其中，所述后处理器调整梅尔频率倒频谱滤波器的至少一个参数，包括滤波器组中心频率、滤波器组截止频率、滤波器组带宽、滤波器组形状或滤波器增益的至少一个。

附图说明

[0026] 下面将结合附图描述本发明的一个或多个优选的示例性实施例，在附图中，相似的标识表示相似的元件，并且其中：

图 1 是描绘能够利用在此公开的方法的通信系统的示例性实施例的框图；

图 2 是示出能够与图 1 的系统一起使用且用于实施语音合成的示例性方法的 TTS 系统的示例性实施例的框图；以及

图 3 是示出 TTS 方法的示例性实施例的流程图。

具体实施方式

[0027] 下面的描述描绘了示例通信系统、能够与通信系统一起使用的示例文本至语音

(TTS)系统以及能够与上述系统中的一个或两个一起使用的一个或多个示例方法。下述方法可由车辆远程信息处理单元(VTU)使用作为合成语言以便输出到 VTU 的用户的一部分。尽管下述方法是在程序执行或运行期间可以在导航背景中被实施用于 VTU,但是将理解,它们可以用于任何类型的 TTS 系统和其他类型 TTS 系统,以及用于除了导航背景之外的背景。在一个特定示例中,所述方法不仅可以用在程序运行期间,而且可以或替代地在用户激活系统或程序使用之前在训练 TTS 系统中使用。

[0028] 通信系统 :

参照图 1,示出包括移动车辆通信系统 10 且可以用于实施在此公开的方法的示例性操作环境。通信系统 10 大体包括车辆 12、一个或多个无线载波系统 14、地面通信网络 16、计算机 18 和呼叫中心 20。应该理解,公开的方法可以与任何数量的不同系统一起使用,并且不被具体地限制到在此示出的操作环境。另外,系统 10 的架构、结构、设置和操作以及其各个组件在本领域通常是公知的。因此,下面的段落仅提供一种这样的示例性系统 10 的简要概述,然而,在此没有示出的其他系统也可以采用所公开的方法。

[0029] 在示出的实施例中将车辆 12 描述为客车,但是应该理解,也可以使用任何其他交通工具,包括摩托车、卡车、运动型多功能车(SUV)、休闲车(RV)、船只、航空器等。图 1 中大体示出一些车辆电子设备 28,其包括远程信息处理单元 30、麦克风 32、一个或多个按钮或其他控制输入 34、音频系统 36、可视显示器 38 和 GPS 模块 40 以及多个车辆系统模块(VSM)42。这些设备的一些可以直接连接到远程信息处理单元,例如,麦克风 32 和(多个)按钮 34,而其他使用一个或多个网络连接诸如通信总线 44 或娱乐总线 46 间接连接到远程信息处理单元。适当网络连接的示例包括控制器区域网络(CAN)、媒体导向系统传输(MOST)、本地互连网络(LIN)、局域网(LAN)和其他适当连接,诸如符合已知 ISO、SAE 和 IEEE 标准和规范的以太网或其他,仅列出一些。

[0030] 远程信息处理单元 30 是 OEM 安装的设备,其能够通过无线载波系统 14 和通过无线联网进行无线语音和 / 或数据通信,使得车辆能够与呼叫中心 20、其他远程信息处理使能的车辆或一些其他实体或设备进行通信。远程信息处理单元优选地使用无线电传输来建立与无线载波系统 14 的通信信道(语音信道和 / 或数据信道),使得能够通过信道发送和接收语音和 / 或数据传输。通过提供语音和数据通信,远程信息处理单元 30 使得车辆能够提供多种不同服务,包括与导航、电话、紧急援助、诊断、信息娱乐等相关的服务。可以使用本领域中已知的技术通过数据连接诸如通过数据信道的包数据传输或者通过语音信道发送数据。对于包括语音通信(例如,在呼叫中心 20 使用在线指导或语音响应单元)和数据通信(例如,以向呼叫中心 20 提供 GPS 位置数据或车辆诊断数据)的组合服务,系统可以使用通过语音信道的单独呼叫以及按照需要在语音信道上进行语音和数据传输之间的切换,并且这可以使用本领域技术人员公知的技术来实施。

[0031] 根据一个实施例,远程信息处理单元 30 使用根据 GSM 或 CDMA 标准的蜂窝通信,因此包括用于语音通信(例如,免提通话)的标准蜂窝芯片组 50、用于数据通信的无线调制解调器、电子处理设备 52、一个或多个数字存储设备 54 和双天线 56。应该理解,可以通过存储在远程信息处理单元中且通过处理器 52 执行的软件实现调制解调器,或者调制解调器可以是位于远程信息处理单元 30 内部或外部的分立硬件组件。调制解调器可以使用任何数量的不同标准和协议诸如 EVDO、CDMA、GPRS 和 EDGE 来运行。也可以使用远程信息处理单

元 30 实施车辆与其他联网的设备之间的无线联网。为此，远程信息处理单元 30 可以被配置为根据一个或多个无线协议诸如 IEEE 802.11 协议、WiMAX 或蓝牙进行无线通信。当用于诸如 TCP/IP 的分组交换数据通信时，远程信息处理单元可以配置有静态 IP 地址或者能够设置为自动接收来自网络上的另一设备诸如路由器或者来自网络地址服务器的所分配的 IP 地址。

[0032] 处理器 52 可以是能够处理电子指令的任何类型的设备，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)。其可以是仅用于远程信息处理单元 30 的专用处理器，或者可以与其他车辆系统共享。处理器 52 执行各种类型的数字存储指令，诸如存储器 54 中存储的软件或固件程序，其使远程信息处理单元能够提供多种类型的服务。例如，处理器 52 能够执行程序或处理数据，以实施在此讨论的方法的至少一部分。

[0033] 远程信息处理单元 30 可以用于提供多元化的车辆服务，包括来自车辆的无线通信和 / 或到车辆的无线通信。这些服务包括 :结合基于 GPS 的车辆导航模块 40 提供的转向和其他导航相关服务 ;结合一个或多个碰撞传感器接口模块诸如车身控制模块(未示出)提供的安全气囊展开通知和其他紧急或路边援助相关的服务 ;使用一个或多个诊断模块的诊断报告 ;以及娱乐信息相关服务，其中，音乐、网页、电影、电视节目、视频游戏和 / 或其他信息通过娱乐信息模块(未示出)下载且存储用于当前或以后播放。上述列出的服务不是远程信息处理单元 30 的所有功能的详尽列表，而是仅是远程信息处理单元 30 能够提供的一些服务的列举。此外，应该理解，上述模块的至少一部分可以按照所存储的内部于或外部于远程信息处理单元 30 的软件指令的形式来实施，它们可以是位于远程信息处理单元 30 内部或外部的硬件组件，或者它们可以彼此之间或者与车辆内的其他系统集成和 / 或共享，仅阐述了几种可能性。在将模块实施为位于远程信息处理单元 30 外部的 VSM 42 的情况下，它们可以使用车辆总线 44 以与远程信息处理单元交换数据和命令。

[0034] GPS 模块 40 从 GPS 卫星的星座 60 接收无线电信号。根据这些信号，模块 40 可以确定车辆位置，用于向车辆驾驶员提供导航和其他位置相关服务。导航信息可以在显示器 38 (或者车辆内的其他显示器)上呈现，或者可以口头表示，诸如当提供转向导航时这样做。可以使用专用的车辆中设置的导航模块(其可以是 GPS 模块 40 的一部分)提供导航服务，或者可以通过远程信息处理单元 30 完成部分或全部导航服务，其中，为了向车辆提供导航地图、地图标注(感兴趣的点、餐馆等)、路线计算等，向远程位置发送位置信息。为了其他目的，诸如车队管理，位置信息可以提供给呼叫中心 20 或其他远程计算机系统，诸如计算机 18。另外，可以通过远程信息处理单元 30 将新的或更新的地图数据从呼叫中心 20 下载到 GPS 模块 40。

[0035] 除了音频系统 36 和 GPS 模块 40 之外，车辆 12 可以包括电子硬件组件形式的其他车辆系统模块(VSM)42，其位于车辆内且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、监控、控制、报告和 / 或其他功能。优选地，每个 VSM 42 通过通信总线 44 连接到其他 VSM 以及连接到远程信息处理单元 30，并且可以被编程以运行车辆系统和子系统诊断测试。作为示例，一个 VSM 42 可以是引擎控制模块(ECM)，其控制引擎操作的各个方面，诸如燃料点火和点火正时，另一 VSM 42 可以是动力系统控制模块，其调整车辆动力系统的一个或多个组件的操作，而另一 VSM 42 可以是车身控制模块，其管理车辆内的各个

电子组件，例如，车辆的电动门锁和前灯。根据一个实施例，引擎控制模块配备有车载诊断(OBD)特征件，其提供诸如从包括车辆排放传感器的各种传感器接收到的各种实时数据，并且提供标准化的一系列诊断故障码(DTC)，其允许技术人员快速识别和修理车辆内的故障。如本领域的技术人员所知，上述 VSM 仅是可以在车辆 12 中使用的一些模块的示例，许多其他模块也是可行的。

[0036] 车辆电子设备 28 还包括多个车辆用户接口，其向车辆占用者提供用于提供和 / 或接收信息的装置，包括麦克风 32、(多个)按钮 34、音频系统 36 和可视显示器 38。如在此使用，术语“车辆用户接口”广泛地包括任何适当形式的电子设备，包括硬件和软件组件，其位于车辆上且使车辆用户能够与车辆的组件通信或者通过车辆的组件进行通信。麦克风 32 向远程信息处理单元提供音频输入，以使驾驶员或其他占用者能够通过无线载波系统 14 提供语音命令和实施免提呼叫。为此，其可以利用本领域中已知的人机接口(HMI)技术连接到车载自动语音处理单元。(多个)按钮 34 允许到远程信息处理单元 30 的手动用户输入，以启动无线电话呼叫和提供其他数据、响应或控制输入。分立的按钮可以使用以便向呼叫中心 20 发起紧急呼叫和常规服务援助呼叫。音频系统 36 向车辆占用者提供音频输出，并且可以是专用独立系统或者是主车辆音频系统的一部分。根据在此示出的具体实施例，音频系统 36 可操作地连接到车辆总线 44 和娱乐总线 46，并且能够提供 AM、FM、卫星无线电、CD、DVD 和其他多媒体功能。可以结合或者独立于上述娱乐信息模块提供此功能。可视显示器 38 优选地是图形显示器，诸如仪表板上的触摸屏或者挡风玻璃反射的抬头显示器，并且可以用于提供多种输入和输出功能。也可以使用各种其他车辆用户接口，因为图 1 的接口仅是一种具体实施方式的示例。

[0037] 无线载波系统 14 优选地是蜂窝电话系统，其包括多个蜂窝塔(cell tower)70(仅示出一个)，一个或多个移动交换中心(MSC)72 以及将无线载波系统 14 与地面网络 16 连接所需的任何其他联网组件。每个蜂窝塔 70 包括发送和接收天线以及基站，其中，来自不同蜂窝塔的基站直接地连接到 MSC 72 或者通过诸如基站控制器的中间设备连接到 MSC72。蜂窝系统 14 可以实施任何适当的通信技术，例如，包括诸如 AMPS 的模拟技术，或者诸如 CDMA(例如，CDMA2000)或 GSM/GPRS 的更新的数字技术。如本领域的技术人员所理解，各种蜂窝塔 / 基站 /MSC 布置是可行的，并且可以与无线系统 14 一起使用。例如，基站和蜂窝塔可以共同位于相同站点，或者他们可以彼此远离，每个基站可以负责单个蜂窝塔或者单个基站可以服务于各个蜂窝塔，以及各个基站可以连接到单个 MSC，仅列出一些可行布置。

[0038] 除了使用无线载波系统 14 之外，可以使用卫星通信形式的不同的无线载波系统，以向车辆提供单向或双向通信。这可以使用一个或多个通信卫星 62 和上行链路发射站 64 来实施。例如，单向通信可以是卫星无线电服务，其中，节目内容(新闻、音乐等)由发射站 64 接收、被打包用于上载、然后发送到卫星 62，卫星 62 向用户广播节目。例如，双向通信可以是使用卫星 62 的卫星电话服务，以在车辆 12 与站 64 之间中继电话通信。如果使用，则额外于无线载波系统 14 或者代替无线载波系统 14，可以使用这种卫星电话。

[0039] 地面网络 16 可以是常规的基于地面的电信网络，其连接到一个或多个有线电话并且将无线载波系统 14 连接到呼叫中心 20。例如，地面网络 16 可以包括公共交换电话网(PSTN)，诸如用于提供硬线电话、分组交换数据通信和因特网基础设施。可以通过使用标准有线网络、光纤或其他光学网络、电缆网络、电源线、诸如无线局域网(WLAN)的其他无线网

络或者提供宽带无线接入(BWA)的网络或者其任意组合来实施一段或多段地面网络 16。此外,呼叫中心 20 不需要通过地面网络 16 连接,而是可以包括无线电话设备,从而它可以与无线网络诸如无线载波系统 14 直接通信。

[0040] 计算机 18 可以是通过私有或公共网络诸如因特网可访问的多个计算机之一。每个这种计算机 18 可以用于一种或多种目的,诸如通过远程信息处理单元 30 和无线载波 14 可由车辆访问的 web 服务器。例如,其他这种可访问的计算机 18 可以是 :服务中心计算机,其中,可以通过远程信息处理单元 30 从车辆上载诊断信息和其他车辆数据 ;客户机计算机,其可由车辆拥有者或其他用户使用以便如访问或接收车辆数据或者设置或配置用户喜好或控制车辆功能的目的 ;或者第三方存储器,无论通过与车辆 12 或呼叫中心 20 或这两者通信,车辆数据或其他信息被提供到所述第三方存储器或从所述第三方存储器提供。计算机 18 还可以用于提供因特网连接,诸如 DNS 服务或者作为网络地址服务器,其使用 DHCP 或其他适当协议以向车辆 12 分配 IP 地址。

[0041] 呼叫中心 20 被设计成向车辆电子设备 28 提供多个不同系统后端功能,并且根据在此示出的示例性实施例,大体包括一个或多个交换机 80、服务器 82、数据库 84、在线指导者 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。这些各种呼叫中心组件优选地通过有线或无线局域网 90 彼此连接。交换机 80,其可以是专用交换(PBX)交换机,路由进入信号,使得语音传输通常通过常规电话发送到在线指导者 86 或者使用 VoIP 发送到自动语音响应系统 88。在线指导者电话也可以使用 VoIP,如图 1 的虚线所指示。通过交换机 80 的 VoIP 和其他数据通信通过在交换机 80 与网络 90 之间连接的调制解调器(未示出)来实施。数据传输通过调制解调器到服务器 82 和 / 或数据库 84。数据库 84 可以存储账户信息,诸如用户认证信息、车辆标识、个人资料记录、行为模式和其他相关用户信息。还可以通过无线系统,诸如 802.11x、GPRS 等进行数据传输。尽管所示实施例被描述为它通过利用在线指导者 86 结合人工呼叫中心 20 而使用,但是将明白,呼叫中心可以使用 VRS 88 作为自动指导者,或者可以使用 VRS 88 与在线指导者 86 的组合。

[0042] 语音合成系统 :

现在转到图 2,示出能够使用当前所公开方法的文本至语音(TTS)系统 210 的示例性架构。通常,用户或车辆占用者可以与 TTS 系统交互,以从应用例如车辆导航应用、免提呼叫应用等的菜单提示接收指令或收听菜单指示。通常,TTS 系统从文本源提取输出词语或标识符,将输出转换成适当的语言单元,选择与语言单元最好地对应的所存储的语音单元,将选择的语音单元转换成音频信号,并且输出音频信号作为与用户交互的可听语音。

[0043] TTS 系统通常对于本领域的技术人员是已知的,如在背景技术部分所描述。但是,图 2 示出根据本公开的改进 TTS 系统的示例。根据一个实施例,系统 210 的部分或全部可以驻留在图 1 的远程信息处理单元 30 上,并且使用图 1 的远程信息处理单元 30 进行处理。根据可选示例性实施例,系统 210 的部分或全部可以驻留在远离车辆 12 的位置中的计算机设备例如呼叫中心 20 上,并且使用该计算机设备处理。例如,语言模型、声学模型等可以存储在呼叫中心 20 的服务器 82 之一的存储器和 / 或数据库 84 中,且被通信到远程信息处理单元 30 用于车辆内置的 TTS 处理。类似地,可以使用呼叫中心 20 的服务器 82 之一的处理器处理 TTS 软件。换句话说,TTS 系统 210 可以驻留在远程信息处理单元 30 中,或者按照任何期望的方式分布在呼叫中心 29 和车辆 12。

[0044] 系统 210 可以包括一个或多个文本源 212a, 212b 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 212a, 212b 的文本且存储 TTS 软件和数据。系统 210 还可以包括处理器，例如，远程信息处理器 52，处理文本，并且与存储器一起以及结合下面的系统模块运行。预处理器 214 从文本源 212a, 212b 接收文本，并且将文本转换成适当的词语等。合成引擎 216 将来自预处理器 214 的输出转换成适当的语言单元，例如，短语、从句和 / 或句子。一个或多个语音数据库 218a, 218b 存储记录的语音。单元选择器 220 从语音数据库 218a, 218b 选择与来自合成引擎 216 的输出最好地对应的所存储的语音单元。后处理器 222 修改或调节一个或多个选择的所存储语音的单元。一个或多个语言模型 224 用作为到合成引擎 216 的输入，一个或多个声学模型 226 用作为到单元选择器 220 的输入。系统 210 还可以包括 :声学接口 228，以将选择的语音单元转换成音频信号 ;以及扬声器 230，例如远程信息处理音频系统的扬声器，以将音频信号转换成可听语音。系统 210 还可以包括麦克风，例如，远程信息处理麦克风 32，以及声学接口 232，以将语音数字化成声学数据用作为到后处理器 222 的反馈。

[0045] 文本源 212a, 212b 可以是任何适当的介质，且可以包括任何适当的内容。例如，文本源 212a, 212b 可以是一个或多个扫描文档、文本文件或应用数据文件、或者任何其他适当的计算机文件等。文本源 212a, 212b 可以包括将被合成为语音的词语、数字、符号和 / 或标点，并且用于输出到文本转换器 214。可以使用任何适当量的文本源。但是在一个示例性实施例中，第一文本源 212a 可以来自第一服务提供者，第二文本源 212b 可以来自第二服务提供者。例如，第一服务提供者可以是导航服务提供者，第二服务提供者可以是地图数据服务提供者。

[0046] 预处理器 214 将来自文本源 212 的文本转换成词语、标识符等。例如，在文本是数字格式的情况下，预处理器 214 可以将数字转换为相应的词语。在另一示例中，在文本是标点、具有帽 (cap)、下划线或粗体的强调的情况下，预处理器 214 可以将其转换成适合于合成引擎 216 和 / 或单元选择器 220 使用的输出。

[0047] 合成引擎 216 从文本转换器 214 接收输出，并且将该输出布置为语言单元，其可以包括一个或多个句子，从句、短语、词语、子词等。引擎 216 可以使用语言模型 224，以便辅助协调最可能的语言单元排列。在将来自文本转换器 214 的输出排列为语言单元时，语言模型 224 提供规则、语法和 / 或语义。语言模型 224 还可以限定处于任意给定时间任何给定 TTS 模式系统 210 期望的语言单元的领域，和 / 或可以提供规则等，从而管理哪种类型的语言单元和 / 或语调可以在逻辑上跟随其他类型语言单元和 / 或语调，以形成自然的发声语音。语言单元可以包括语音等同物，例如，音素串等，并且可以是音素 HMM 的形式。

[0048] 语音数据库 218a, 218b 包括从一个或多个人预先记录的语音。语音可以包括预先记录的句子、从句、短语、词语、预先记录的词语的子词等。语音数据库 218a, 218b 还可以包括与预先记录的语音相关联的数据，例如，元数据，以识别所记录的语音段，以便由单元选择器 220 使用。可以使用任何适当量的语音数据库。但是在一个示例性实施例中，第一语音数据库 218a 可以来自第一服务提供者，第二语音数据库 218b 可以来自第二服务提供者。在此实施例中，第二文本源 212b 和第二语音数据库 218b 中的一个或两个可以是系统 210 的集成部分，或者分别连接到系统 210，如相对于第二语音数据库 218b 所示，并且可以是独立于 TTS 系统 210 的产品的一部分，例如，来自地图提供者的地图数据库产品 215。

[0049] 单元选择器 220 将来自合成引擎 216 的输出与存储的语音数据进行比较，并且选择与合成引擎输出最好地对应的存储的语音。由单元选择器 220 选择的语音可以包括预先记录的句子、从句、短语、词语、预先记录的词语的子词等。选择器 220 可以使用声学模型 226，以便辅助比较和选择最可能或最好地对应的存储语音的候选。可以结合选择器 220 使用声学模型 226，以比较和对比合成引擎输出的数据与存储的语音数据，评估它们之间的差异或类似度的幅度，并且最终使用决策逻辑来识别最佳匹配的所存储的语音数据并输出相应的所记录的语音。

[0050] 通常，最佳匹配的语音数据是与合成引擎 216 的输出具有最小差异或最大可能为合成引擎 216 的输出，如通过对本领域的技术人员所知的多种技术中的任何一种所确定。这些技术可以包括动态时间 - 规整（time-warping）分类器、人工智能技术、神经网络、自由音素识别器、和 / 或概率模式匹配器，诸如隐马尔可夫模型（HMM）引擎。HMM 引擎是本领域的技术人员公知的用于产生多个 TTS 模型候选或假设。可以在通过语音的声学特征分析最终识别和选择表示合成引擎输出的最可能的正确解释的所存储的语音数据中考虑所述假设。更具体地，HMM 引擎例如通过应用贝叶斯定理根据给定一个或另一个语言单元的声学数据的所观察序列的经 HMM 计算的信任值或概率所排序的语言单元假设的“N 最佳”列表的形式产生静态模型。

[0051] 在一个实施例中，来自单元选择器 220 的输出可以直接通过到达声学接口 228 或者不经过后处理而通过后处理器 222。在另一实施例中，后处理器 222 可以接收来自单元选择器 220 的输出用于进一步处理。

[0052] 在这两种情况下，声学接口 228 将数字音频数据转换成模拟音频数据。接口 228 可以是数字 - 模拟转换设备、电路和 / 或软件等。扬声器 230 是将模拟音频数据转换成用户可听的且麦克风 32 可接收的语音的电声换能器。

[0053] 在一个实施例中，麦克风 32 可以用于将来自扬声器 230 的语音输出转换成电信号，并且将此信号通信到声学接口 232。声学接口 232 接收模拟电信号，该模拟电信号首先被采样，使得模拟信号值在离散时刻被捕获，然后被量化，从而在每个采样点将模拟信号的幅度转换成数字语音数据的连续流。换句话说，声学接口 232 将模拟电信号转换成数字电信号。数字数据是二进制比特，其在存储器 54 中缓冲，然后由处理器 52 进行处理，或者在它们由处理器 52 初始接收时被实时处理。

[0054] 类似地，在此实施例中，后处理器模块 222 可以将来自接口 232 的数字语音数据的连续流变换成声学参数的离散序列。更具体地，处理器 52 可以执行后处理器模块 222，以将数字语音数据分段成例如持续时间为 10-30 ms 的重叠语音或声学帧。所述帧对应于声学子词，诸如音节、半音节、单音、双音、音素等。后处理器模块 222 还可以执行语音分析，以从每帧内的数字化的语音（诸如时间变化特征的向量）提取声学参数表示。语音内的话语可以被表示为这些特征向量的序列。例如，并如本领域的技术人员所知，特征向量可以提取，并且例如，可以包括通过执行帧的傅里叶变换和使用余弦变换对声学谱解相关所获得的音高、能源轮廓、谱特性和 / 或倒频谱系数。可以存储和处理覆盖特定语音持续时间的声学帧以及相应参数。

[0055] 在优选实施例中，后处理器 222 可以按照任何适当的方式修改存储的语音。例如，存储的语音可以被修改，从而使从一个说话者记录的语音调节成听起来类似于从另一说话

者记录的语音，或者使从说话者的一种语言记录的语音调节成听起来类似于从相同说话者的另一种语言记录的语音。后处理器 222 可以将来自一个说话者的语音数据与来自另一说话者的语音数据转换。更具体地，对于一个说话者的说话者具体特性，后处理器 222 可以从该说话者提取或以其他方式处理倒频谱的声学特征，并且对那些特征进行倒频谱分析。在另一示例中，对于一个说话者的说话者具体特性，后处理器 222 可以从该说话者提取声学特征，并且对那些特征进行归一化变换。如在此使用，术语一个说话者和另一说话者或者两个不同说话者，可以包括两个不同人说相同语言或者一个人说两种不同语言。

[0056] 另外，在此实施例中，后处理器 222 可以用于适当地特征滤波第二说话者的语音。然而，在执行这种特征滤波之前，第一说话者的说话特定特征用于调整在第二说话者的语音的声学特征滤波中使用的滤波器组的一个或多个参数。例如，可以在基于人耳的心理声学模型模拟频率范围的一个或多个滤波器组的频率规整中使用说话者特定特性。更具体地，频率规整可以包括梅尔频率倒频谱滤波器组的中心频率的调整，改变到这种滤波器组的上截止频率和下截止频率，修改这些滤波器组的形状(例如，抛物线形，梯形)，调整滤波器增益等。一旦已经修改滤波器组，它们就用于滤波来自第二说话者的语音的声学特征。当然，从第二说话者的语音所滤波的声学特征从其在没有滤波器组修改的情况下被修改，因此，可以促进来自第二说话者的调节语音的输出，和 / 或调节或再训练 HMM，以便用于选择或处理第二说话者的语音。

[0057] 方法 ：

现在转到图 3，示出语音合成方法 300。可以使用在车辆远程信息处理单元 30 的操作环境内适当编程的图 2 的 TTS 系统 210 以及使用适当硬件和对图 1 所示的其他组件编程来实施图 3 的方法 300。基于上述系统描述以及下面结合其他附图描述的方法讨论，本领域的技术人员将知道任何特定实施方式的这些特征。本领域的技术人员还将认识到，可以使用其他操作环境内的其他 TTS 系统实施所述方法。

[0058] 通常，方法 300 包括在 TTS 系统中接收第一和第二文本输入，使用系统处理器将第一和第二文本输入处理为与分别来自第一和第二说话者的所存储的语音相应的各自的第一和第二语音输出，并且使第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。

[0059] 再参照图 3，方法 300 在步骤 305 以任何适当的方式开始。例如，车辆用户开始与远程信息处理单元 30 的用户接口交互，优选地，通过按下用户接口按钮 34，以开始会话，其中，在 TTS 模式下操作的同时用户从远程信息处理单元 30 接收 TTS 音频。在一个示例性实施例中，方法 300 可以作为远程信息处理单元 30 的导航路由应用的一部分而开始。

[0060] 在步骤 310，在 TTS 系统中接收第一文本输入。例如，第一文本输入可以包括来自 TTS 系统 210 的第一文本源 212a 的导航指令。导航指令可以包括定向调动，例如，IN 500’ TURN RIGHT ONTO (在 500’，右转到) …。

[0061] 在步骤 315，对第一文本输入进行预处理，以将文本转换成适合于语音合成的输出。例如，预处理器 214 可以将从文本源 212a 接收的文本转换成词语、标识符等，以便供合成引擎 216 使用。更具体地，可以将来自步骤 310 的示例导航指令转换成“在 500 英尺，右转到…”。

[0062] 在步骤 320，来自步骤 315 的输出被排列为语言单元。例如，合成引擎 216 可以从

文本转换器 214 接收输出，并且使用语言模型 224 可以将输出排列为语言单元，所述语言单元可以包括一个或多个句子、从句、短语、词语、子词等。语言单元可以包括语音等同物，例如，音素串等。

[0063] 在步骤 325，将语言单元与存储的语音数据进行比较，选择与语言单元最好地对应的语音被选择作为代表输入文本的语音。例如，单元选择器 220 可以使用声学模型 228，以将从合成引擎 216 输出的语言单元与存储在第一语音数据库 218a 中的语音数据进行比较，并且选择具有与合成引擎输出最好地对应的相关联数据的所存储的语音。步骤 320 和 325 一起可以构成使用所存储的来自第一说话者的语音将第一文本输入处理或合成为第一语音输出的示例。

[0064] 在步骤 330，在 TTS 系统中接收第二文本输入。例如，第二文本输入可以包括来自 TTS 系统 210 的第二文本源 212b 的导航变量。导航变量可以包括街道名称，例如，“S. M-24”。

[0065] 在步骤 335，将第二文本输入进行预处理，以将文本转换成可合成输出或适合于语音合成的输出。例如，预处理器 214 可以将从第二文本源 212b 接收的文本转换成词语、标识符等，以便供合成引擎 216 使用。更具体地，来自步骤 330 的示例导航变量可以被转换成“(向南 M 二十四) Southbound M Twenty Four ”。导航指令和变量一起可以构成 TTS 塑造提示。

[0066] 在步骤 340，将来自步骤 335 的输出排列为语言单元。例如，合成引擎 216 可以从文本转换器 214 接收输出，并且使用语言模型 224 可以将输出排列为语言单元，所述语言单元可以包括一个或多个句子、从句、短语、词语、子词等。语言单元可以包括语音等同物，例如，音素串等。

[0067] 在步骤 345，将语言单元与存储的语音数据进行比较，并且与语言单元最好地对应的语音被选择作为代表输入文本的语音。例如，单元选择器 220 可以使用声学模型 228，以将从合成引擎 216 输出的语言单元与存储在第二语音数据库 218b 中的语音数据进行比较，并且选择具有与合成引擎输出最好地对应的相关联数据的所存储的语音。步骤 340 和 345 一起可以构成使用存储的来自第二说话者的语音将第二文本输入处理或合成为第二语音输出的示例。

[0068] 在步骤 350，第二说话者的第二语音输出调节成听起来像第一说话者的第一语音输出。例如，对于第一说话者的一个或多个说话者特定特性可分析第一语音输出的声学特征，然后可以基于第一说话者的(多个)说话者特定特性调整用于从第二语音输出滤波声学特征的声学特征滤波器，其后，可使用调整的滤波器对来自从第二语音输出的声学特征进行滤波。

[0069] 在一个实施例中，可以通过调整梅尔频率倒频谱滤波器的一个或多个参数来调整滤波器。所述参数可以包括滤波器组中心频率、滤波器组截止频率、滤波器组带宽、滤波器组形状、滤波器增益等。说话者特定特性包括声道或鼻腔相关特性中的至少一个。更具体地，所述特性可包括长度、形状、转换函数、格式、音调频率等。

[0070] 在一个实施例中，可以从预先记录的语音预先提取第一语音输出的声学特征，且将该声学特征与该语音相关联地存储在例如语音数据库 218a, 218b 中。在另一实施例中，可以通过后处理器 222 从 TTS 系统 210 内的选择的预先记录的语音提取声学特征。在另一

实施例中，可以在声学特征从扬声器 230 输出、由麦克风 32 接收且经由接口 232 反馈到后处理器 222 之后，从选择的预先记录的语音提取声学特征。通常，声学特征提取对于本领域的普通技术人员来说是公知的，并且声学特征可以包括梅尔频率倒频谱系数（MFCC），相关频谱变换 - 感知线性预测特征（RASTA-PLP 特征），或者任何其他合适声学特征。

[0071] 在步骤 355，输出来自第一说话者的第一语音输出。例如，可以通过接口 228 和扬声器 230 输出由选择器 220 从数据库 218a 选择的来自第一说话者的预先记录的语音。

[0072] 在步骤 360，输出来自第二说话者的经调节的第二语音。例如，可以通过接口 228 和扬声器 230 输出由选择器 220 从数据库 218b 选择的并且通过后处理器 222 调节的来自第二说话者的预先记录的语音。

[0073] 在步骤 365，可以修改与处理来自第二说话者的所存储语音结合使用的模型。例如，声学模型 226 可以包括可以按照任何适当方式调节的 TTS 隐马尔可夫模型（HMM），使得来自第二说话者的随后的语音听起来越来越像来自第一说话者的。如这里相对于 TTS 系统 21 先前所述，后处理器 222 可用于以任何适当的方式修改存储的语音。如虚线所示，经调节的 TTS HMM 可以反馈上游以改善随后的语音的选择。

[0074] 在步骤 370，方法可以以任何适当的方式结束。

[0075] 与用于在说话者声音听起来不同的 TTS 系统中输出来自多个不同说话者的语音的现有技术相比，当前公开的语音合成方法，使得来自说话者之一的语音调节成听起来像说话者中另一个的语音。

[0076] 尽管结合在导航背景中的示例塑造提示（sculpted prompt）或指令描述了当前公开的方法，但是，可以在任何其他适当背景中使用所述方法。例如，可以在免提呼叫背景中使用所述方法，以使存储的标签调节成听起来像发音的命令，或者反之亦然。在其他示例中，可以在自动语音菜单、语音控制设备等中在调节来自不同说话者的指令时使用所述方法。

[0077] 所述方法或其一部分可以在包括计算机可读介质上实施的指令的计算机程序产品中来实施，以便由一个或多个计算机的一个或多个处理器使用来实施一个或多个所述方法步骤。计算机程序产品可以包括一个或多个软件程序，包括源代码、目标代码、可执行代码或其他格式的程序指令；一个或多个固件程序；或者硬件描述语言（HDL）文件；以及任何程序相关数据。所述数据可以包括数据结构、查找表、或任何其他适当格式的数据。所述程序指令可以包括程序模块、例程、程序、对象、分量等。可以在一台计算机上或者在彼此通信的多台计算机上执行计算机程序。

[0078] （多个）程序可以体现在计算机可读介质上，其可以包括一个或多个存储设备、制造物品等。示例性计算机可读介质包括计算机系统存储器，例如，RAM（随机访问存储器）、ROM（只读存储器）；半导体存储器，例如，EPROM（可擦除可编程 ROM）、EEPROM（电可擦除可编程 ROM）、闪存；磁或光盘或带等。计算机可读介质还可以包括计算机到计算机连接件，例如，当通过网络或另一通信连接（有线、无线或其组合）传递或提供数据时。上述示例的任何组合也包括在计算机可读介质的范围内。因此，应理解，可以通过能够执行与所公开方法的一个或多个步骤相应的指令的任何电子物品和 / 或设备至少部分地执行所述方法。

[0079] 应理解，上面是本发明的一个或多个优选示例性实施例的描述。本发明不限于在此公开的（多个）具体实施例，而是仅由所附权利要求限定。此外，上述描述中包含的陈述

涉及具体实施例，并且不被解释为限制本发明的范围或者限制权利要求中使用的术语的定义，除非其中上面明确限定术语或短语。各种其他实施例以及对所公开的（多个）实施例的各种改变和修改对于本领域的技术人员将是明显的。例如，本发明可应用于语音信号处理的其他领域，诸如移动通信、通过因特网协议应用的语音等。所有这些其他实施例、改变和修改意在落入所附权利要求的范围内。

[0080] 如在此说明书和权利要求中所使用，当结合一个或多个组件或其他项的列表使用时，术语“例如”，“比如”，“诸如”和“等”以及动词“包括”，“具有”，“包含”以及其他动词形式，每个被解释为开放式，意味着所述列表不被认为是排除其他额外组件或项。其他术语被解释为使用它们的最广泛的合理含义，除非它们使用在需要不同的解释的背景中。

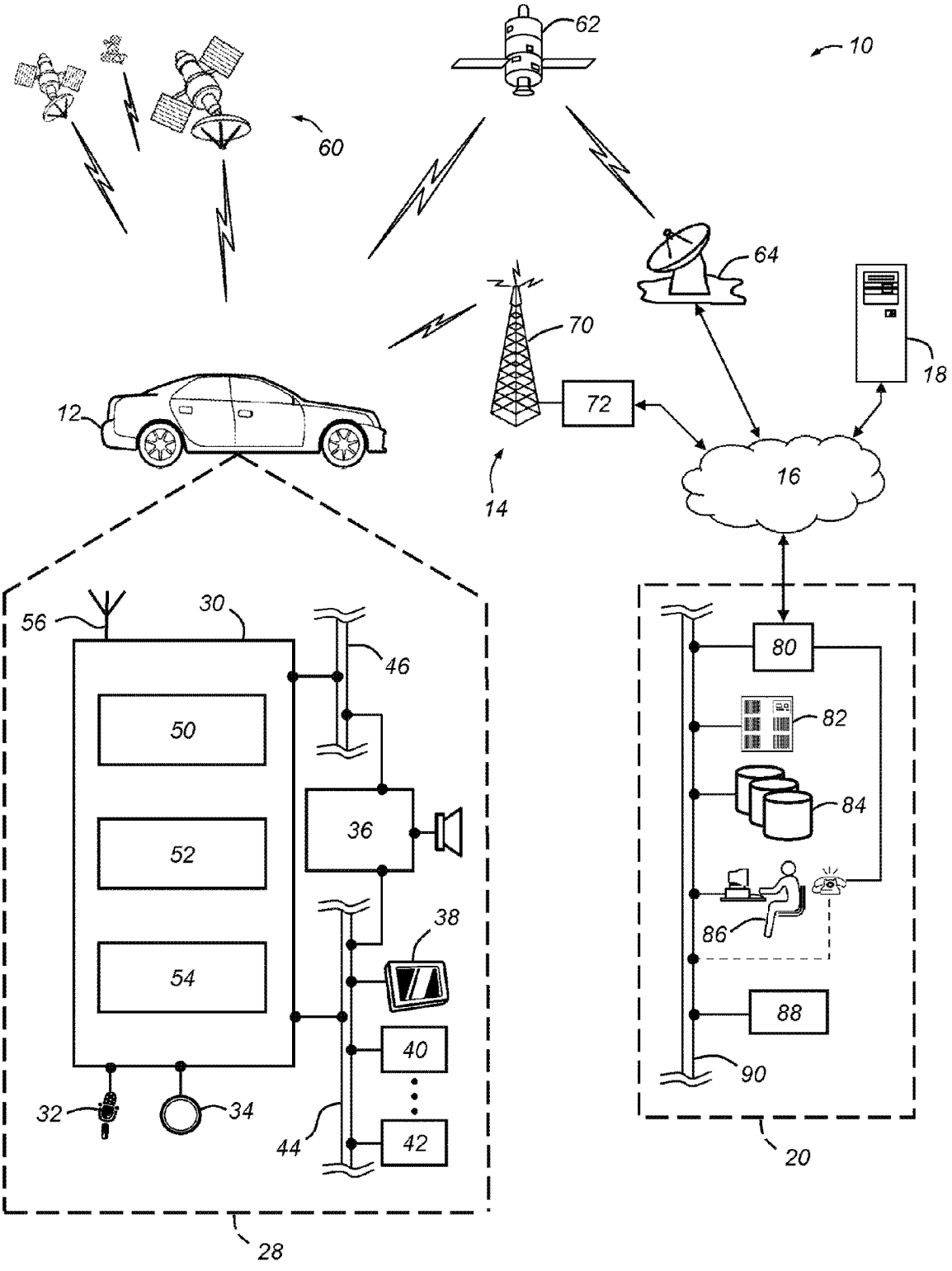

62
10
60
64
70
18
72
12
16
14
56
30
46
50
52
54
36
38
40
42
32
34
44
28
80
82
84
86
88
90
20

图 1

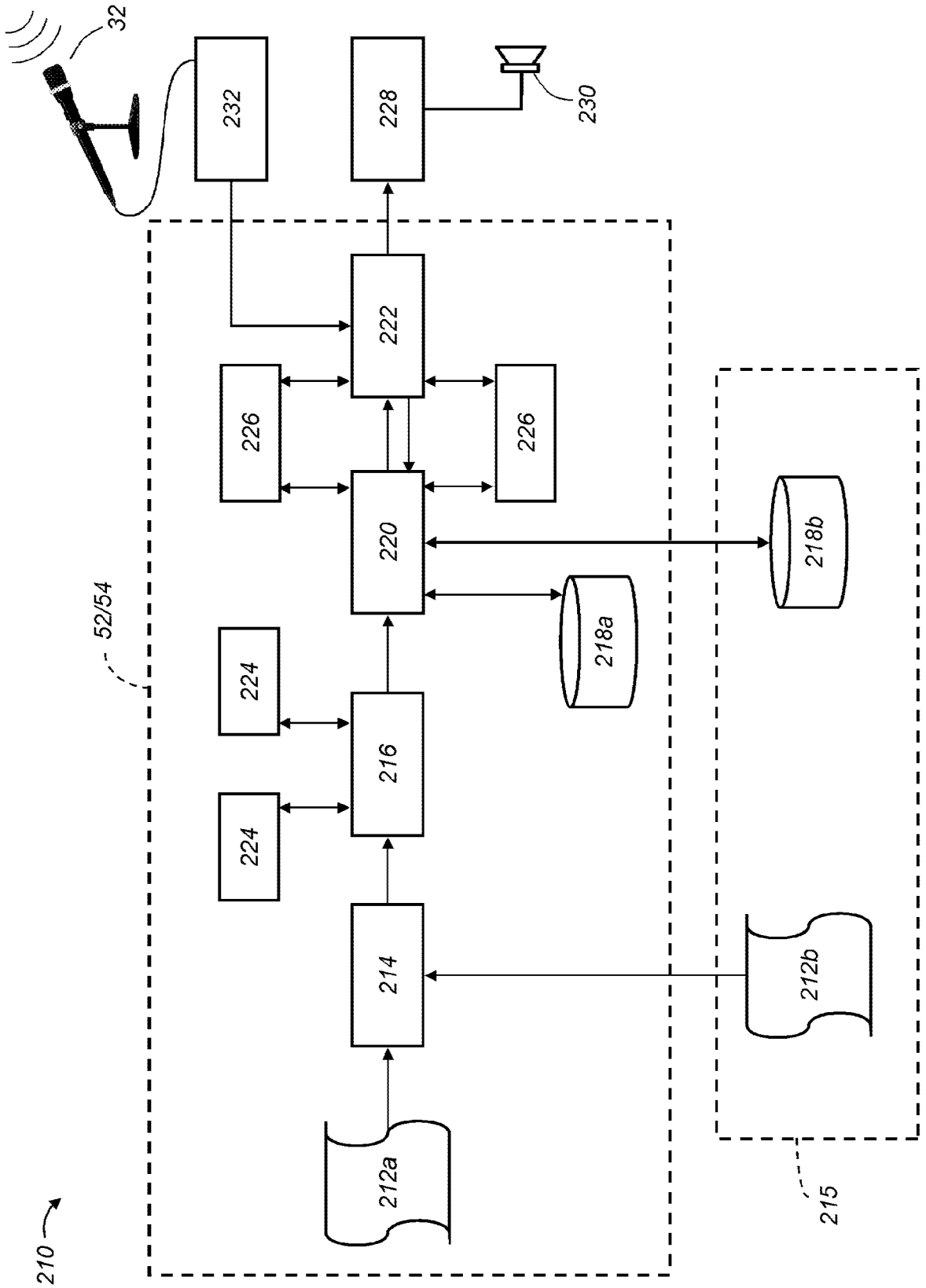

32
232
228
230
222
226
226
220
218b
218a
52/54
224
216
224
214
212b
212a
210
215

图 2

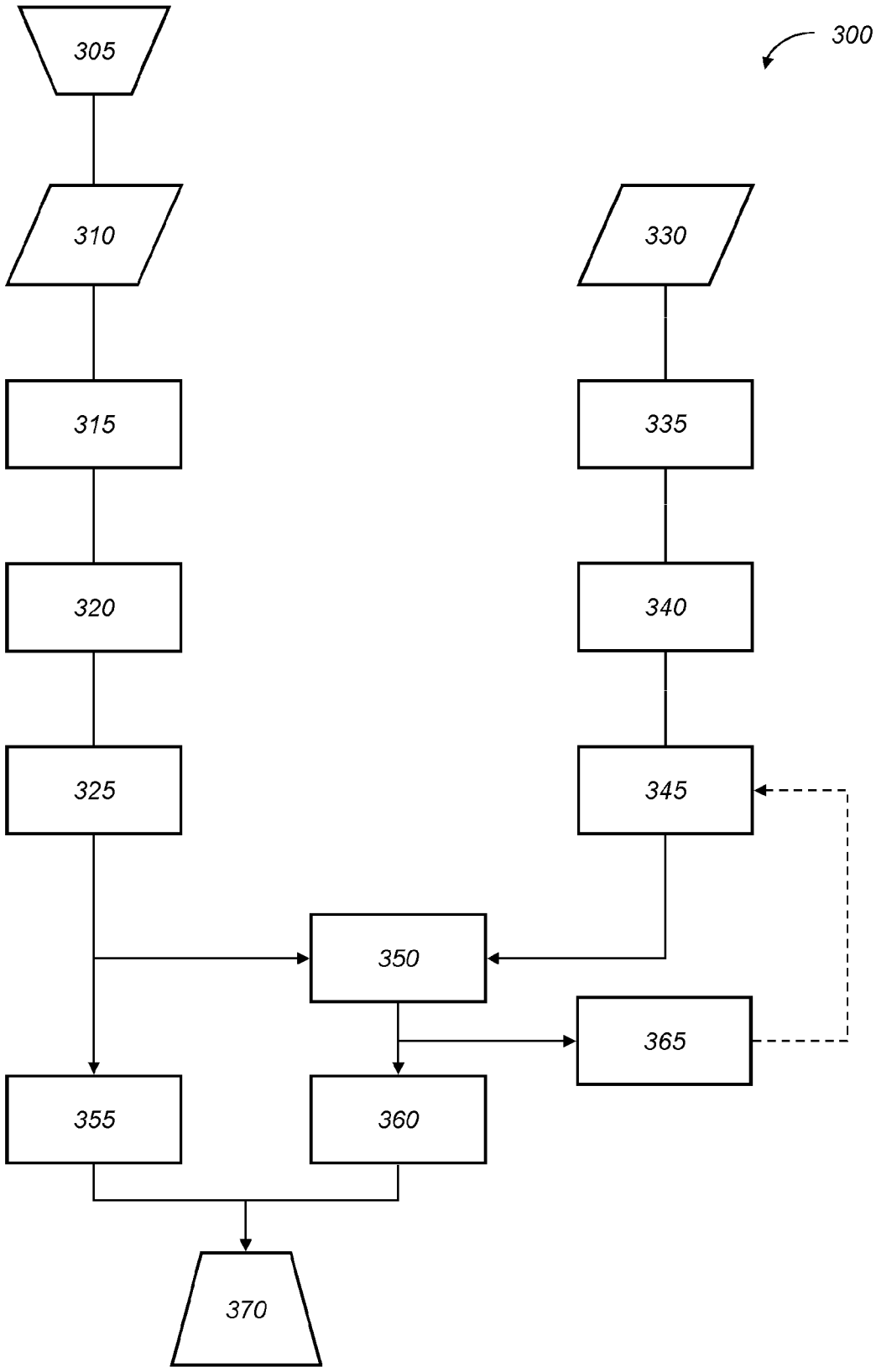


图 3