



(12)发明专利申请

(10)申请公布号 CN 106030699 A

(43)申请公布日 2016.10.12

(21)申请号 201580008549.8

(22)申请日 2015.09.29

(30)优先权数据

62/061,830 2014.10.09 US

14/675,932 2015.04.01 US

(85)PCT国际申请进入国家阶段日

2016.08.12

(86)PCT国际申请的申请数据

PCT/US2015/052860 2015.09.29

(87)PCT国际申请的公布数据

W02016/057268 EN 2016.04.14

(71)申请人 谷歌公司

地址 美国加利福尼亚州

(72)发明人 马修·谢里菲

(74)专利代理机构 中原信达知识产权代理有限
责任公司 11219

代理人 李宝泉 周亚荣

(51)Int.Cl.

G10L 15/22(2006.01)

G10L 15/08(2006.01)

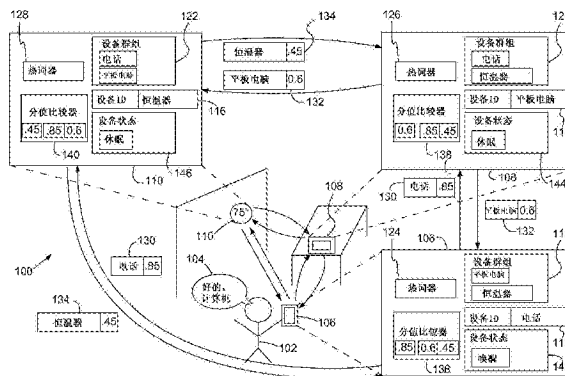
权利要求书3页 说明书12页 附图3页

(54)发明名称

多个设备上的热词检测

(57)摘要

公开了多个设备上的热词检测的方法、系统和装置,包括在计算机存储介质上进行编码的计算机程序。在一个方面,一种方法包括由第一计算设备接收对应于话语的音频数据的动作。该动作进一步包括确定与该话语包括热词的可能性相对应的第一值。该动作进一步包括接收与该话语包括热词的可能性相对应的由第二计算设备确定的第二值。该动作进一步包括将该第一值与该第二值进行比较。该动作进一步包括基于对该第一值与该第二值进行比较,发起对该音频数据的语音识别处理。



1. 一种计算机实现的方法,包括:
 - 由第一计算设备接收对应于话语的音频数据;
 - 确定与所述话语包括热词的可能性相对应的第一值;
 - 接收与所述话语包括所述热词的可能性相对应的第二值,所述第二值由第二计算设备确定;
 - 将所述第一值与所述第二值进行比较;以及
 - 基于对所述第一值与所述第二值进行比较,发起对所述音频数据的语音识别处理。
2. 根据权利要求1所述的方法,包括:
 - 确定所述第一值满足热词分值阈值。
3. 根据权利要求1所述的方法,包括:
 - 将所述第一值传送至所述第二计算设备。
4. 根据权利要求1所述的方法,包括:
 - 基于对所述第一值与所述第二值进行比较来确定所述第一计算设备的激活状态。
5. 根据权利要求4所述的方法,其中,基于对所述第一值与所述第二值进行比较来确定所述第一计算设备的激活状态包括:
 - 确定所述激活状态为活动状态。
6. 根据权利要求1所述的方法,包括:
 - 由所述第一计算设备接收对应于附加话语的附加音频数据;
 - 确定与所述附加话语包括所述热词的可能性相对应的第三值;
 - 接收与所述话语包括所述热词的可能性相对应的第四值,所述第四值由第三计算设备确定;
 - 将所述第三值与所述第四值进行比较;以及
 - 基于对所述第三值与所述第四值进行比较,确定所述第一计算设备的激活状态为非活动状态。
7. 根据权利要求3所述的方法,其中:
 - 将所述第一值传送至所述第二计算设备包括:
 - 通过本地网络或者通过短距离无线电来将所述第一值传送至服务器,以及
 - 接收与所述话语包括所述热词的可能性相对应的由第二计算设备确定的第二值包括:
 - 通过所述本地网络或者通过所述短距离无线电从所述服务器接收由第二计算设备确定的第二值。
8. 根据权利要求1所述的方法,包括:
 - 识别所述第二计算设备;以及
 - 确定所述第二计算设备被配置为对包括所述热词的话语作出响应。
9. 根据权利要求3所述的方法,其中:
 - 将所述第一值传送至所述第二计算设备包括:
 - 传送所述第一计算设备的第一标识符,以及
 - 接收与所述话语包括所述热词的可能性相对应的由第二计算设备确定的第二值包括:
 - 接收所述第二计算设备的第二标识符。
10. 根据权利要求5所述的方法,其中,确定所述激活状态为活动状态包括:

确定自从接收到与所述话语相对应的所述音频数据起已经流逝特定时间量。

11. 根据权利要求5所述的方法, 包括:

基于确定所述激活状态为活动状态来在特定时间量内继续传送所述第一值。

12. 一种计算设备, 包括:

存储指令的一个或多个存储设备, 所述指令在被所述计算设备执行时能够操作为使得所述计算设备执行操作, 包括:

由第一计算设备接收对应于话语的音频数据;

确定与所述话语包括热词的可能性相对应的第一值;

接收与所述话语包括所述热词的可能性相对应的第二值, 所述第二值由第二计算设备确定;

将所述第一值与所述第二值进行比较; 以及

基于对所述第一值与所述第二值进行比较, 发起对所述音频数据的语音识别处理。

13. 根据权利要求12所述的系统, 其中, 所述操作进一步包括:

确定所述第一值满足热词分值阈值。

14. 根据权利要求12所述的系统, 其中, 所述操作进一步包括:

将所述第一值传送至所述第二计算设备。

15. 根据权利要求12所述的系统, 其中, 所述操作进一步包括:

基于对所述第一值与所述第二值进行比较来确定所述第一计算设备的激活状态。

16. 根据权利要求15所述的系统, 其中, 基于对所述第一值与所述第二值进行比较来确定所述第一计算设备的激活状态包括:

确定所述激活状态为活动状态。

17. 根据权利要求12所述的系统, 其中, 所述操作进一步包括:

由所述第一计算设备接收对应于附加话语的附加音频数据;

确定与所述附加话语包括所述热词的可能性相对应的第三值;

接收与所述话语包括所述热词的可能性相对应的第四值, 所述第四值由第三计算设备确定;

将所述第三值与所述第四值进行比较; 以及

基于对所述第三值与所述第四值进行比较, 确定所述第一计算设备的激活状态为非活动状态。

18. 根据权利要求14所述的系统, 其中:

将所述第一值传送至所述第二计算设备包括:

通过本地网络或者通过短距离无线电来将所述第一值传送至服务器, 以及

接收与所述话语包括所述热词的可能性相对应的由第二计算设备确定的第二值包括:

通过所述本地网络或者通过所述短距离无线电从所述服务器接收由第二计算设备确定的第二值。

19. 根据权利要求12所述的系统, 其中, 所述操作进一步包括:

识别所述第二计算设备; 以及

确定所述第二计算设备被配置为对包括所述热词的话语作出响应。

20. 一种存储软件的非暂时性计算机可读介质, 所述软件包括能够由一个或多个计算

机执行的指令,所述指令在这样的执行时使得所述一个或多个计算机执行操作,包括:

由第一计算设备接收对应于话语的音频数据;

确定与所述话语包括热词的可能性相对应的第一值;

接收与所述话语包括所述热词的可能性相对应的第二值,所述第二值由第二计算设备确定;

将所述第一值与所述第二值进行比较;以及

基于对所述第一值与所述第二值进行比较,发起对所述音频数据的语音识别处理。

多个设备上的热词检测

技术领域

[0001] 本说明书一般地涉及用于识别人正在说的词语的系统和技术,其还被称作语音识别。

背景技术

[0002] 支持语音的住所或其它环境—也就是用户在其中仅需要出声地讲出查询或命令并且基于计算机的系统将处理并应答该查询和/或使得命令得以执行的住所或其它环境—的现实正在到来。支持语音的环境(例如住所、工作场所、学校等)能够使用分布在该环境的各个房间或区域各处的被连接的麦克风设备的网络来实现。通过这样的麦克风的网络,用户有能力从基本上该环境中的任何地方口头查询系统而无需在他/她的前方或者甚至在其附近具有计算机或其它设备。例如,当在厨房中烹饪时,用户可以询问系统“三杯是多少毫升?”,并且作为响应而从系统接收到回答,例如以合成语音输出的形式的回答。替选地,用户可以询问系统诸如“离我最近的加油站什么时候关门”或者在准备离开房子时询问“我今天应当穿外套吗?”。

[0003] 此外,用户可以询问系统的查询和/或发出命令,所述查询或命令涉及用户的个人信息。例如,用户可以询问系统“我与John何时会面?”或者命令系统“当我回家时提醒我给John打电话。”

发明内容

[0004] 对于支持语音的系统,用户与系统交互的方式被设计为主要—如果不是专门地—凭借语音输入来进行。因此,可能拾取到周围环境中所发出的包括并非指向系统的那些话语在内的所有话语的系统必须以某些方式来辨别任何给定的话语何时是指向该系统而不是例如指向出现在环境中的个人。一种实现该目的的方式是使用热词,该热词通过环境中的用户间的协定被预留作为预定的词语,其被讲出以唤起系统的注意。在一个示例环境中,用来唤起系统的注意的热词是词语“OK computer(好的,计算机)”。因此,每次词语“OK computer”被讲出时,其就被麦克风所拾取,传达至系统,该系统执行语音识别技术来确定热词是否被讲出,并且如果是,则等待随后的命令或查询。因此,指向系统的话语采用[热词][查询]的一般形式,其中本示例中的“热词”是“OK computer”并且“查询”能够是任何问题、命令、声明、或者能够由系统单独或者经由网络与服务器结合来进行语音识别、解析、以及按照其实施动作的其它请求。

[0005] 根据本说明书中所描述的主题的一个创新方面,一种用户设备接收用户所讲出的话语。该用户设备确定该话语是否包括热词并且计算指示该话语包括热词的热词置信度分值。该用户设备将该分值传送至在附近近邻区域中的其它用户设备。该其它用户设备很可能接收到相同的话语。该其它用户设备计算热词置信度分值并且将它们的分值传送至该用户设备。该用户设备比较热词置信度分值。如果该用户设备具有最高的热词置信度分值,则该用户设备保持活动并且准备处理附加音频。如果该用户设备不具有最高热词置信度分值,

则该用户设备不处理附加音频。

[0006] 大体上,本说明书中所描述的主题的另一个创新方面可以包含在一种方法中,所述方法包括以下动作:由第一计算设备接收对应于话语的音频数据;确定与该话语包括热词的可能性相对应的第一值;接收与该话语包括热词的可能性相对应的由第二计算设备确定的第二值;将该第一值与该第二值进行比较;以及基于对该第一值与该第二值进行比较,发起对该音频数据的语音识别处理。

[0007] 这些和其它实施例均能够可选地包括以下特征的一个或多个。该动作进一步包括:确定该第一值满足热词分值阈值。该动作进一步包括:将该第一值传送至第二计算设备。该动作进一步包括:基于对该第一值与该第二值进行比较来确定第一计算设备的激活状态。基于对该第一值与该第二值进行比较来确定第一计算设备的激活状态的所述动作进一步包括:确定该激活状态为活动状态。该动作进一步包括:由第一计算设备接收对应于附加话语的附加音频数据;确定与该附加话语包括热词的可能性相对应的第三值;接收与话语包括热词的可能性相对应的第四值,该第四值由第三计算设备确定;将该第一值与该第二值进行比较;并且基于对该第一值与该第二值进行比较,确定该第一计算设备的激活状态为非活动状态。

[0008] 将该第一值传送至第二计算设备的所述动作进一步包括:通过本地网络或者通过短距离无线电来将该第一值传送至服务器。接收与该话语包括热词的可能性相对应的由第二计算设备确定的第二值的所述动作进一步包括:通过本地网络或者通过短距离无线电从该服务器接收由第二计算设备确定的第二值。该动作进一步包括:识别该第二计算设备;以及确定该第二计算设备被配置来对包括热词的话语作出响应。将该第一值传送至该第二计算设备的所述动作进一步包括:传送该第一计算设备的第一标识符。该接收与该话语包括热词的可能性相对应的由第二计算设备确定的第二值的所述动作进一步包括:接收该第二计算设备的第二标识符。确定该激活状态为活动状态的所述动作进一步包括:确定自从接收到对应于话语的音频数据起已经流逝特定时间量。该动作进一步包括:基于该激活状态为活动状态来继续在特定时间量内继续传送该第一值。

[0009] 该方面的其它实施例包括对应的系统、装置、以及记录在计算机存储设备上的计算机程序,它们均被配置为执行该方法的操作。

[0010] 本说明书中所描述的主题的特定实施例能够实现以便实现以下优势中的一种或多种。多个设备能够检测热词并且仅一个设备将对热词作出响应。

[0011] 在附图和以下的描述中阐述了本说明书中所描述主题的一个或多个实施例的细节。该主题的其它特征、方面、和优势从说明书、附图、和权利要求书将变得显而易见。

附图说明

[0012] 图1是用于热词检测的示例系统的示图。

[0013] 图2是用于热词检测的示例过程的示图。

[0014] 图3示出了计算设备和移动计算设备的示例。

[0015] 各图中相似的附图标记和名称指示相似的要素。

具体实施方式

[0016] 在不久的将来,有可能许多设备可以连续地收听热词。当单个用户具有被训练以对他们的语音作出响应的多个设备(例如,电话、平板计算机、电视等)时,可能期望在用户不太可能向其说话的设备上抑制对热词作出响应。例如,当用户朝向一个设备讲出热词时,如果所述用户的任何其它设备在附近,则它们也将很可能触发语音搜索。在许多情况下,这并不是用户的意图。因此,如果仅有单个设备特别是用户正在对其话语的设备会触发,则可能是有利的。本说明书解决了选择正确的设备以用于对热词作出反应并且在其它设备上抑制对该热词的响应的问题。

[0017] 图1是用于热词检测的示例系统100的示图。总体上,系统100图示了讲出话语104的用户102,该话语104被计算设备106、108、和110的麦克风检测到。计算设备106、108、和110对话语104进行处理以确定话语104包括热词的可能性。计算设备106、108、和110均互相传送指示话语104包括热词的可能性的数据。计算设备106、108、和110均对该数据进行比较,并且计算出话语104包括热词的最高可能性的计算设备发起对话语104的语音识别。没有计算出话语104包括热词的最高可能性的计算设备发起对话语104之后的语音的语音识别。

[0018] 在向另一个计算设备传送指示话语104对应于热词的数据之前,位于彼此附近的计算设备互相进行识别。在一些实施方式中,计算设备通过在本地网络上搜索被配置为对热词作出响应的其它设备来互相识别。例如,计算设备106可以在局域网搜索被配置为对热词作出响应的其它设备并且识别出计算设备108和计算设备110。

[0019] 在一些实施方式中,计算设备通过识别登录到每个设备的用户来识别附近被配置为对热词作出响应的其它计算设备。例如,用户102登录到计算设备106、108、和110。用户102在其手中具有计算设备106。计算设备108放在桌上,而计算设备110则位于附近的墙上。计算设备106检测到计算设备108和110并且每个计算设备共享与登录到该计算设备的用户相关的信息,诸如用户标识符。在一些实施方式中,计算设备可以通过利用说话者识别对被配置为对热词作出响应的计算设备进行识别来识别被配置为在相同用户讲出该热词时作出响应的其它附近的计算设备。例如,用户102配置了:计算设备106、108和110在用户102讲出热词时均对用户102的语音作出响应。该计算设备通过向彼此提供用户102的用户标识符来共享说话者识别信息。在一些实施方式中,计算设备可以通过短距离无线电来识别被配置为对热词作出响应的其它计算设备。例如,计算设备106可以通过短距离无线电传送信号,所述信号搜索被配置为对热词作出响应的其它计算设备。计算设备可以采用这些技术中的一个或者它们的组合来识别被配置为对热词作出响应的其它计算设备。

[0020] 一旦计算设备106、108、和110已经识别出被配置为对热词作出响应的其它计算设备,则计算设备106、108、和110共享并存储所识别计算设备的设备标识符。该标识符可以基于设备的类型、设备的IP地址、MAC地址、用户对设备所给出的名称,或者任何类似的唯一标识符。例如,计算设备106的设备标识符112可以是“电话”。计算设备108的设备标识符114可以是“平板计算机”。设备110的设备标识符116可以是“恒温器”。计算设备106、108、和110存储被配置为对热词作出响应的其它计算设备的设备标识符。每个计算设备具有该计算设备将设备标识符存储在其中的设备群组。例如,计算设备106具有设备群组118,其将“平板电脑”和“恒温器”列为将接收到如计算设备106所计算的音频数据包括热词的可能性的两个设备。计算设备108具有设备群组120,其列出“电话”和“恒温器”作为将接收到如计算设备

108所计算的音频数据包括热词的可能性的两个设备。计算设备110具有设备群组122,其将“电话”和“平板电脑”列为将接收到如计算设备110所计算的音频数据包括热词的可能性的两个设备。

[0021] 当用户102讲出话语104“OK computer(好的,计算机)”时,具有处于用户102邻近区域中的麦克风的每个计算设备都检测到该话语104并对其进行处理。每个计算设备通过诸如麦克风的音频输入设备来检测话语104。每个麦克风将音频数据提供至相应的音频子系统。相应的音频子系统对该音频数据进行缓冲、滤波、和数字化。在一些实施方式中,每个计算设备可以对音频数据执行端点指示(endpointing)和说话者识别。该音频子系统将经处理的音频数据提供至热词器。该热词器将经处理的音频数据与已知热词数据进行比较并且计算出指示话语104对应于热词的可能性的置信度分值。热词器可以从经处理的音频数据中提取音频特征,诸如滤波器组能量或Mel频率倒谱系数。热词器可以使用分类窗口来处理这些音频特征,诸如通过使用支持向量机或神经网络来进行处理。基于对音频特征的处理,热词器124计算出置信度分值0.85,热词器126计算出置信度分值0.6,并且热词器128计算出置信度分值0.45。在一些实施方式中,置信度分值可以被标准化到0至1的范围,其中越高的数字指示话语104包括热词的置信度越大。

[0022] 每个计算设备将相应的置信度分值数据分组传送至设备群组中的其它计算设备。每个置信度分值数据分组包括相应的置信度分值以及计算设备的相应设备标识符。例如,计算设备106将包括置信度分值0.85和设备标识符“电话”的置信度分值数据分组130传送至设备群组118中的计算设备—计算设备108和110。计算设备108将包括置信度分值0.6和设备标识符“平板电脑”的置信度分值数据分组132传送至设备群组120中的计算设备—计算设备106和110。计算设备110将包括置信度分值0.45和设备标识符“恒温器”的置信度分值数据分组134传送至设备群组118中的计算设备—计算设备106和108。

[0023] 在一些实施方式中,如果置信度分值满足热词分值阈值,则计算设备可以传送置信度分值数据分组。例如,如果热词分值阈值为0.5,则计算设备110将不会向设备群组122中的其它计算设备传送置信度分值数据分组134。计算设备106和108仍然将分别向计算设备群组118和120中的计算设备传送置信度分值数据分组130和132。

[0024] 在一些实施方式中,传送置信度分值数据分组的计算设备可以向其它计算设备直接传送置信度分值数据分组。例如,计算设备106可以通过短距离无线电来向计算设备108和110传送置信度分值数据分组130。两个计算设备之间的通信协议可以是通用即插即用。在一些实施方式中,传送置信度分值数据分组的计算设备可以广播该置信度分值数据分组。在此实例下,置信度分值数据分组可以被设备群组中的计算设备接收以及被其它计算设备接收。在一些实施方式中,传送置信度分值数据分组的计算设备可以将置信度分值数据分组传送至服务器,并且之后该服务器将该置信度分值数据分组传送至数据群组中的计算设备。服务器可以位于计算设备的局域网内或者可通过互联网访问。例如,计算设备108将置信度分值数据分组132以及设备群组120中的计算设备的列表发送至服务器。该服务器将该置信度分值数据分组132传送至计算设备106和110。在计算设备正向另一个计算设备传送置信度分值数据分组的实例中,进行接收的计算设备可以发回进行接收的计算设备接收到置信度分值数据分组的确认。

[0025] 每个计算设备使用分值比较器来比较该计算设备已经接收到的热词置信度分值。

例如,计算设备106计算出热词置信度分值0.85并且接收到热词置信度分值0.6和0.45。在此实例中,分值比较器136对三个分数进行比较并且识别出分值0.85为最高。对于计算设备118和110而言,分值比较器138和140得出类似的结论,识别出与计算设备106相对应的分值0.85为最高。

[0026] 确定它自己的热词置信度分值为最高的计算设备发起对热词话语之后的语音数据的语音识别。例如,用户可以讲出“OK computer”并且计算设备106可以确定其具有最高的热词置信度分值。计算设备106将发起对在热词之后所接收到的音频数据的语音识别。如果用户讲出“call Alice(给Alice打电话)”,则计算设备106将对该话语进行处理并且执行适当命令。在一些实施方式中,接收到热词可以使得接收到该热词的计算设备从休眠状态激活。在实例中,具有最高的热词置信度分值的计算设备保持处于唤醒状态,而不具有最高热词置信度分值的其它计算设备则不处理热词话语之后的语音数据并且进入休眠状态。

[0027] 如图1所示,分值比较器136识别出与计算设备106相对应的热词置信度分值为最高。因此,设备状态142为“唤醒”。分值比较器138和140也识别出与计算设备106相对应的热词置信度分值为最高。因此,设备状态138和140“休眠”。在一些实施方式中,计算设备的激活状态可以不受影响。例如,用户102可能在计算设备108上观看电影并且在其手中具有计算设备106。当用户102讲出“OK computer”时,计算设备106依靠具有最高的热词置信度分值来发起对该热词之后的音频数据的语音识别。计算设备108不发起对该热词之后的音频数据的语音识别,并且继续播放电影。

[0028] 在一些实施方式中,确定其具有最高热词置信度分值的计算设备在开始对热词之后的语音执行语音识别之前等待特定时间量。这样做允许计算出最高热词置信度分值的计算设备在没有等待到更高热词置信度分值的情况下开始对热词之后的语音执行语音识别。为了进行说明,计算设备106的分值比较器136分别从计算设备108和110接收到热词置信度分值0.6和0.45,以及来自热词器124的热词置信度分值0.85。从热词器124计算“OK computer”音频数据的热词置信度分值的时间起,计算设备106在对该热词之后的语音执行语音识别之前等待五百毫秒。在分值比较器接收到更高分值的实例中,计算设备可以在将设备状态设置为“休眠”之前不等待特定时间量。例如,计算设备108的热词器126计算出热词置信度分值0.6并且接收到热词置信度分值0.85和0.45。一旦计算设备108接收到热词置信度分值0.85,则计算设备108能够将设备状态144设置为“休眠”。这假设了:计算设备108在热词器126计算出热词置信度分值0.6之后的特定时间量内接收到热词置信度分值0.85。

[0029] 在一些实施方式中,当计算设备具有最高的热词置信度分值时,该计算设备可以在特定时间量内继续广播置信度分值数据分组以确保其它计算设备接收到该置信度分值数据分组。此策略在计算设备从另一个计算设备接收到置信度分值数据分组时确实发回确认的实例中能够得到最多应用。因此,如果计算设备106向数据群组118中的计算设备传送置信度分值数据分组130并且在诸如五百毫秒的特定时间量之前接收到确认,则计算设备106可以开始对热词之后的语音执行语音识别。在计算设备广播其置信度分值数据分组而且并不期望确认的实例中,计算设备可以在诸如五百毫秒的特定时间量内继续广播其热词置信度分值,或者继续广播其热词置信度分值直至计算设备接收到更高的热词置信度分值为止,以较早者为准。例如,计算设备110计算出热词置信度分值0.45并且开始广播置信度分值数据分组134。在三百毫秒之后,计算设备110接收到置信度分值数据分组130并且停止

广播置信度分值数据分组134,因为来自置信度分值数据分组130的热词置信度分值0.85比热词置信度分值0.45更高。作为另一个广播示例,计算设备106计算出热词置信度分值0.45并且开始广播置信度分值数据分组130。在五百毫秒之后,计算设备106停止广播置信度分值数据分组130并且开始对热词之后的语音执行语音识别。计算设备106可以在已经流逝五百毫秒之前接收到置信度分值数据分组132和134,但是因为置信度分值数据分组132和134中的热词置信度分值比0.85更低,所以该计算设备继续等待直至已经流逝五百毫秒。

[0030] 在一些实施方式中,计算设备可以开始对热词之后的语音执行语音识别直至该计算设备接收到更高的热词置信度分值。热词器计算热词置信度分值并且如果该热词置信度分值满足阈值,则计算设备对热词之后的语音执行语音识别。该计算设备可以在没有向用户显示语音识别的任何指示的情况下执行语音识别。这可以是所期望的,因为这样做给用户留下该计算设备没有活动的印象而同时还允许该计算设备与该计算设备已经等待直至该计算设备确认其计算出了最高的热词分值的情况相比更快地向用户显示基于语音识别的结果。作为示例,计算设备106计算出热词置信度分值0.85并且开始对热词之后的语音执行语音识别。计算设备106接收到置信度分值数据分组132和134并且确定热词置信度分值0.85为最高。计算设备106继续对热词之后的语音执行语音识别并且将结果呈现给用户。对于计算设备108而言,热词器126计算出热词置信度分值0.6,并且计算设备108在没有向用户显示数据的情况下开始对热词之后的语音执行语音识别。一旦计算设备108接收到包括热词置信度0.85的置信度分值数据分组130,则该计算设备停止执行语音识别。没有向用户显示数据,并且用户很可能留下了计算设备108保持着“休眠”状态的印象。

[0031] 在一些实施方式中,为了避免讲出热词之后的任何延时,分值能够在热词结束之前从热词器被报告,例如对于部分热词进行报告。例如,在用户讲出“OK Computer”时,计算设备能够在用户已经完成讲出“OK comp”时就计算部分热词置信度分值。之后,计算设备能够将该部分热词置信度分值与其它计算设备共享。具有最高的部分热词置信度分值的计算设备能够继续处理用户的语音。

[0032] 在一些实施方式中,计算设备可以在该计算设备确定热词置信度分值满足阈值时发出可听或不可听的声音,例如特定频率或频率模式的声音。该声音将用信号向其它计算设备通知该计算设备将继续处理热词之后的音频数据。其它计算设备将接收到该声音并且停止处理音频数据。例如,用户讲出“OK Computer”。计算设备中的一个计算出大于或等于阈值的热词置信度分值。一旦该计算设备确定该热词置信度分值大于或等于阈值,则该计算设备发射18千赫的声音。处于用户的邻近区域中的其它计算设备也可以计算热词置信度分值,并且在其它计算设备接收到该声音时可能正在计算热词置信度分值当中。当其它计算设备接收到该声音时,其它计算设备停止对用户语音的处理。在一些实施方式中,计算设备可以将热词置信度分值编码在可听或不可听的声音之中。例如,如果热词置信度分值为0.5,则计算设备可以生成包括对分值0.5进行编码的频率模式的可听或不可听的声音。

[0033] 在一些实施方式中,计算设备可以使用不同的音频量度来选择用于继续处理用户语音的计算设备。例如,计算设备可以使用响度来确定哪个计算设备将继续处理用户的语音。检测到最响语音的计算设备可以继续处理用户的语音。作为另一个示例,当前正在使用或者具有活动显示的计算设备可以在检测到热词后向其它计算设备通知其将继续处理用户的语音。

[0034] 在一些实施方式中,在用户正在讲话的同时处于用户的邻近区域中的每个计算设备接收音频数据并且将该音频数据发送至服务器以改进语音识别。每个计算设备能够接收到对应于用户语音的音频数据。虽然仅有一个计算设备将在用户看来是正在处理该用户的语音,但是每个计算设备都能够向服务器传送音频数据。然后,因为服务器能够对与相同话语相对应的不同音频样本进行比较,所以服务器能够使用从每个计算设备所接收的音频数据来改进语音识别。例如,用户说“Ok computer, remind me to buy milk(好的计算机,提醒我买牛奶)”。一旦用户讲完“Ok computer”,则附近的计算设备将很可能已经确定哪个计算设备具有最高的热词置信度分值并且该计算设备将在用户讲出“remind me to buy milk”时对那些词语进行处理和响应。其它计算设备也将会接收到“remind me to buy milk”。虽然所述其它计算设备将不对“remind me to buy milk”的话语作出响应,但是其它计算设备能够将与“remind me to buy milk”相对应的音频数据发送至服务器。对“remind me to buy milk”作出响应的计算设备也能够将其音频数据发送至服务器。因为服务器具有来自不同计算设备的与相同“remind me to buy milk”话语相对应的不同音频样本,所以服务器能够处理音频数据以改进语音识别。

[0035] 图2是用于热词检测的示例过程200的示图。过程200可以由诸如图1的计算设备108的计算设备执行。处理200计算与话语包括热词的可能性相对应的值,并且将该值与其它计算设备所计算的其它值进行比较以确定是否对该热词之后的话语的部分执行语音识别。

[0036] 计算设备接收对应于话语的音频数据(210)。用户讲出话语并且计算设备的麦克风接收到该话语的音频数据。计算设备通过对该音频数据进行缓冲、滤波、端点指示、和数字化来对该音频数据进行处理。作为示例,用户可以说出“Ok, computer”并且计算设备的麦克风将接收到对应于“Ok, computer”的音频数据。计算设备的音频子系统将对该音频数据进行采样、滤波、和端点指示以用于由计算设备进一步进行处理。

[0037] 该计算设备确与该话语包括热词的可能性相对应的第一值(220)。该计算设备通过将该话语的音频数据与包括热词的音频样本群组进行比较或者通过分析该话语的音频数据的音频特征来确定该第一值,所述第一值可以被称作热词置信度分值。该第一值可以被标准化到0至1的范围,其中1指示该话语包括热词的最高可能性。在一些实施方式中,该计算设备识别第二计算设备并且确定该第二计算设备被配置为对包括热词的话语进行响应以及被用户配置为对热词作出响应。该用户可以登录到该计算设备和第二计算设备这两者。该计算设备和第二计算设备两者都可以被配置为对用户的语音作出响应。该计算设备和第二计算设备可以连接至相同的局域网。该计算设备和第二计算设备两者都可以位于如GPS或信号强度所确定的特定距离之内,诸如十米之内。例如,计算设备可以通过短距离无线电进行通信。该计算设备可以检测到第二计算设备所传送的信号的强度为5dBm并且转换为对应的距离,诸如五米。

[0038] 计算设备接收与话语包括热词的可能性相对应的第二值,该第二值由第二计算设备确定(230)。该第二计算设备通过该第二计算设备的麦克风来接收该话语。该第二计算设备对所接收的对应于该话语的音频数据进行处理并且确定第二值或第二热词置信度分值。该第二热词置信度分值反映出如第二计算设备所计算的该话语包括热词的可能性。在一些实施方式中,该计算设备使用以下技术中的一种或多种来将该第一值传送至该第二计算设

备。该计算设备可以通过可通过互联网访问的服务器、通过位于局域网上的服务器、或者直接通过局域网或短距离无线电来将该第一值传送至该第二计算设备。该计算设备可以仅将该第一值传送至第二计算设备,或者该计算设备可以广播该第一值以使得其它计算设备也可以接收到该第一值。该计算设备可以使用与该计算设备传送第一数相同或不同的技术来从第二计算设备接收第二值。

[0039] 在一些实施方式中,计算设备可以计算话语的响度分值或话语的信噪比。该计算设备可以将该响度分值、信噪比、和热词置信度分值进行组合来确定新的值以用于与来自其它计算设备的类似值进行比较。例如,计算设备可以计算热词置信度分值和信噪比。然后,该计算设备可以将那两个分值进行组合并且与来自其它计算设备的以相似方式所计算的分值进行比较。在一些实施方式中,计算设备可以计算不同分值并且将每个分值传送至其它计算设备以用于比较。例如,计算设备可以计算话语的响度分值以及热词置信度分值。然后,该计算设备可以将那些分值传送至其它计算设备以用于比较。

[0040] 在一些实施方式中,该计算设备可以将第一标识符与第一值一起传送。该标识符可以基于计算设备的地址、用户所给出的该计算设备的名称、或者该计算设备的位置中的一个或多个。例如,标识符可以是“69.123.132.43”或“电话”。类似地,第二计算设备可以将第二标识符与第二值一起传送。在一些实施方式中,计算设备可以将第一标识符传送至特定计算设备,该计算设备先前已经将所述特定计算设备识别为被配置来对热词作出响应。例如,计算设备先前可能已经将第二计算设备识别为被配置来对热词作出响应,因为除了能够对热词作出响应之外,该计算设备的相同的用户登录到了第二计算设备。

[0041] 计算设备将第一值与第二值进行比较(240)。之后,该计算设备基于所述比较的结果来发起对音频数据的语音识别处理(250)。在一些实施方式中,例如,计算设备在第一值大于或等于第二值时发起语音识别。如果用户讲出“ok computer, call Carol(好的计算机,给Carol打电话)”,则该计算设备由于第一值大于或等于第二值而将开始通过对“call Carol”执行语音识别来对“call Carol”进行处理。在一些实施方式中,计算设备设置激活状态。在第一值大于或等于第二值的实例中,则计算设备将激活状态设置为活动或“唤醒”。在“唤醒”状态下,计算设备显示来自语音识别的结果。

[0042] 在一些实施方式中,计算设备将第一值与第二值进行比较并且确定第一值小于第二值。基于确定第一值小于第二值,该计算设备将激活状态设置为非活动或“休眠”。在“休眠”状态下,计算设备并不对用户表现为活动或者正对音频数据进行处理。

[0043] 在一些实施方式中,当计算设备确定第一值大于或等于第二值时,该计算设备可以在将激活状态设置为活动之前等待特定时间量。该计算设备可以等待特定时间量以增加该计算设备将不从另一个计算设备接收到更高值的概率。该特定时间量可以是固定的或者可以依赖于该计算设备传送和接收值的技术而变化。在一些实施方式中,当计算设备确定第一值大于或等于第二值时,该计算设备可以在特定时间量内继续传送该第一值。通过在特定时间量内传送第一值,该计算设备增加了第一值被其它计算设备接收到的概率。在计算设备确定第一值小于第二值的实例中,该计算设备可以停止传送第一值。

[0044] 在一些实施方式中,计算设备可以在确定是否执行热词之后的命令时考虑附加信息。附加信息的一个示例可以是热词之后的话语部分。通常,热词之后的音频数据对应于针对计算设备的命令,诸如“call Sally(给Sally打电话)”、“play Halloween Movie(播放好

莱坞电影)”、或“set heat to 70degrees(将温度设置为70度)”。该计算设备可以识别处理该类型的请求或者能够处理该请求的典型设备。基于预编程的典型使用或者基于设备用户的使用模式,通常将由电话来处理给他人打电话的请求。如果用户通常在平板计算机上观看电影,则平板计算机可以处理播放电影的请求。如果恒温器能够调节温度,则恒温器可以处理温度调节。

[0045] 为了计算设备能考虑热词之后的话语的部分,该计算设备将必须在其很可能识别出热词时发起对音频数据的语音识别。该计算设备可以对话语的命令部分进行归类并且计算该类别中的命令的频率。该计算设备可以将该频率连同热词置信度分值一起传送至其它计算设备。每个计算设备可以使用该频率和热词置信度分值来确定是否执行热词之后的命令。

[0046] 例如,如果用户说出“OK computer,play Michael Jackson(好的计算机,播放 Michael Jackson)”,则如果计算设备是用户以其20%的时间来听音乐的电话,则该计算设备可以将该信息连同热词置信度分值一起传送。诸如用户使用5%的时间用来听音乐的平板计算机的计算设备可以将信息连同热词置信度分值一起传送至其它计算设备。该计算设备可以使用热词置信度分值与播放音乐的时间百分比的组合来确定是否执行该命令。

[0047] 图3示出了能够被用来实现这里所描述的技术的计算设备300和移动计算设备350。计算设备300意图表示各种形式的数字计算机,诸如膝上型电脑、桌面计算机、工作站、个人数字助理、服务器、刀片服务器、大型主机、或其它适当的计算机。移动计算设备350意图表示各种形式的移动设备,诸如个人数字助理、蜂窝电话、智能电话、和其它类似计算设备。这里所示的组件、其连接和关系、以及其功能仅意图作为示例性,而并不意味着作为限制。

[0048] 计算设备300包括处理器302、存储器304、存储设备306、连接到存储器304和高速扩展端口310的高速接口308、以及连接到低速扩展端口314和存储设备306的低速接口312。处理器302、存储器304、存储设备306、高速接口308、高速扩展端口310、和低速接口312中的每一个都使用各种总线互连,并且可以安装在公用主板上或者以其它适当方式进行安装。处理器302能够对用于在计算设备300内执行的指令进行处理,所述指令包括存储在存储器304中或存储设备306上以显示用于外部输入/输出设备上的GUI的图形信息的指令,所述外部输入/输出设备诸如耦合到高速接口308的显示器316。在其它实施方式中,可以视情况使用多个处理器和/或多个总线,连同多个存储器以及多种类型的存储器。而且,可以连接多个计算设备,其中每个设备提供必要操作的一部分(例如作为服务器组、刀片服务器群、或者多处理器系统)。

[0049] 存储器304存储计算设备300内的信息。在一些实施方式中,存储器304是一个或多个易失性存储单元。在一些实施方式中,存储器304是一个或多个非易失性存储单元。存储器304还可以是另一形式的计算机可读介质,诸如磁盘或光盘。

[0050] 存储设备306能够提供用于计算设备300的大型存储。在一些实施方式中,存储设备306可以是计算机可读介质或包含计算机可读介质,诸如软盘设备、硬盘设备、光盘设备或磁带设备、闪存或其它类似的固态存储器设备、或者设备阵列,包括在存储区域网络或其它配置中的设备。指令能够被存储在信息载体中。该指令当被一个或多个处理设备(例如处理器302)执行时执行诸如以上所描述的那些的一种或多种方法。该指令还能够由诸如计算

机或机器可读介质的一个或多个存储设备(例如存储器304、存储设备306、或者处理器302上的存储器)来存储。

[0051] 高速接口308对计算设备300的带宽密集型操作进行管理,而低速控制器312对较低带宽密集型操作进行管理。这样的功能分配仅是示例性的。在一些实施方式中,高速接口308耦合到存储器304、显示器316(例如通过图形处理器或加速器来耦合),并且耦合到可以接受各种扩展卡(未示出)的高速扩展端口310。在所述实施方式中,低速接口312耦合到存储设备306和低速扩展端口314。可以包括各种通信端口(例如USB、蓝牙、以太网、无线以太网)的低速扩展端口314可以耦合到诸如键盘、指示设备、扫描仪的一个或多个输入/输出设备,或者例如通过网络适配器耦合到诸如交换机或路由器的联网设备。

[0052] 如图所示,可以以多种不同形式来实现计算设备300。例如,计算设备300可以被实现为标准服务器320,或者多次被实现在这样的服务器的群组中。此外,计算设备300可以在诸如膝上型计算机322的个人计算机中实现。计算设备300也可以被实现为机架式服务器系统324的一部分。替选地,来自计算设备300的组件可以与诸如移动计算设备350的移动设备(未示出)中的其它组件进行组合。这样的设备中的每一个可以包含计算设备300和移动计算设备350中的一个或多个,并且整个系统可以由相互通信的多个计算设备组成。

[0053] 除其它组件之外,移动计算设备350包括处理器352、存储器364、诸如显示器354的输入/输出设备、通信接口366、以及收发器368。设备350还可以被提供有诸如微型驱动器或其它设备的存储设备,以提供额外的存储。处理器352、存储器364、显示器354、通信接口366、和收发器368中的每一个使用各种总线互连,并且组件中的若干可以被安装在共用主板上或者以其它适当方式安装。

[0054] 处理器352能够执行计算设备350内的指令,包括存储在存储器364中的指令。处理器可以被实现为包括单独的和多个模拟和数字处理器的芯片的芯片组。例如,处理器352可以提供对设备350的其它组件的协调,诸如对用户接口、移动计算设备350所运行的应用、以及移动计算设备350所进行的无线通信的控制。

[0055] 处理器352可以通过耦合到显示器354的控制接口358和显示接口356与用户进行通信。例如,显示器354可以是TFT(薄膜晶体管液晶显示器)显示器或OLED(有机发光二极管)显示器或者其它适当的显示技术。显示接口356可以包括用于驱动显示器354以向用户呈现图形和其它信息的适当电路。控制接口358可以接收来自用户的命令并且对该命令进行转换以用于向处理器352提交。此外,可以提供与处理器352通信的外部接口362,以使得设备350能够与其它设备进行近区域通信。例如,以太接口363在一些实施方式中可以提供有线通信,或者在其它实施方式中提供无线通信,并且还可以使用多个接口。

[0056] 存储器364存储移动计算设备350内的信息。存储器364可以被实现为一个或多个计算机可读介质或媒体、一个或多个易失性存储单元、或者一个或多个非易失性存储单元。也可以提供扩展存储器374并通过扩展接口372连接至设备350,例如所述扩展接口可以包括SIMM(单列直插内存模块)卡接口。扩展存储器374可以为移动计算设备350提供额外的存储空间,或者还可以存储用于移动计算设备350的应用和其它信息。具体地,扩展存储器374可以包括指令以执行或补充以上所描述的过程,并且还可以包括安全信息。因此,例如,扩展存储器374可以作为移动计算设备350的安全模块来提供,并且可以被编程有许可对移动计算设备350进行安全使用的指令。此外,安全应用连同附加信息可以经由SIMM卡来被提

供,诸如以不可非法侵入的方式在SIMM卡上设置识别信息。

[0057] 如下所讨论的,例如,存储器可以包括闪存和/或NVRAM存储器(非易失性随机存取存储器)。在一些实施方式中,指令被存储在信息载体中。该指令当被一个或多个处理设备(例如处理器352)执行时,执行诸如以上所描述的那些方法中的一种或多种方法。该指令还能够被诸如一个或多个计算机可读或机器可读介质的一个或多个存储设备(例如存储器364、扩展存储器374、或者处理器352上的存储器)所存储。在一些实施方式中,例如,该指令能够通过收发器368或外部接口362来以传播信号的方式被接收。

[0058] 在必要的情况下,移动计算设备350可以通过通信接口366来无线地通信,所述通信接口366可以包括数字信号处理电路。通信接口366可以提供在各种模式或协议下的通信,除其它之外,诸如GSM语音通话(全球移动通信系统)、SMS(短消息服务)、EMS(增强型消息服务)、或MMS(多媒体消息服务)消息发送、CDMA(码分多址)、TDMA(时分多址)、PDC(个人数字蜂窝)、WCDMA(宽带码分多址)、CDMA2000、或GPRS(通用分组无线电服务)。例如,这样的通信可以通过收发器368使用射频来发生。此外,诸如使用蓝牙、WiFi、或其它这样的收发器(未示出)可以发生短距离通信。此外,GPS(全球定位系统)接收器模块370可以向移动计算设备350提供附加的导航相关和位置相关的无线数据,所述无线数据可由在移动计算设备350上运行的应用程序视情况使用。

[0059] 移动计算设备350还可以使用音频编解码器360可听地通信,所述音频编解码器360可以接收来自用户的语音信息并将其转换为可使用的数字信息。同样,音频编解码器360可一诸如通过例如移动计算设备350的送受话器中的扬声器为用户生成可听的声音。这样的声音可以包括来自语音电话呼叫的声音,可以包括已记录的声音(例如,语音消息、音乐文件等),并且还可以包括在移动计算设备350上操作的应用所生成的声音。

[0060] 如图所示,可以以多种不同形式来实现移动计算设备350。例如,移动计算设备350可以被实现为蜂窝电话380。移动计算设备350还可以被实现为智能电话382、个人数字助理、或其它类似移动设备的一部分

[0061] 这里所描述的系统和技术各种实施方式可以以数字电子电路、集成电路、专门设计的ASIC(专用集成电路)、计算机硬件、固件、软件、和/或以上的组合来实现。这些各种实施方式可以包括一个或多个计算机程序中的实施方式,所述计算机程序可以在包括至少一个可编程处理器、至少一个输入设备和至少一个输出设备的可编程系统上执行和/或解释,所述可编程处理器是专用或通用的,被耦合以从存储系统接收数据和指令并且向所述存储系统传送数据和指令。

[0062] 这些计算机程序(也被称作程序、软件、软件应用程序、或代码)包括用于可编程处理器的机器指令,并且可以以高级程序语言和/或面向对象的编程语言、和/或汇编/机器语言来实现。如本文所使用的,术语机器可读介质和计算机可读介质是指被用来向可编程处理器提供机器指令和/或数据的任何计算机程序产品、装置和/或设备(例如,磁盘、光盘、存储器、可编程逻辑设备(PLD)),包括接收作为机器可读信号的机器指令的机器可读介质。术语机器可读信号指代用于向可编程处理器提供机器指令和/或数据的任何信号。

[0063] 为了提供与用户的交互,这里所描述的系统和技术可在具有下述的计算机上实现::用于向用户显示信息的显示设备(例如,CRT(阴极射线管)或LCD(液晶显示器)监视器),以及用户通过其可以向计算机提供输入的键盘和指示设备(例如,鼠标或轨迹球)。还

能够使用其它类型的设备来提供与用户的交互；例如，提供给用户的反馈可以为任何形式的感官反馈（例如，视觉反馈、听觉反馈、或触觉反馈）；并且可以以任何形式来接收来自用户的输入，所述形式包括以声音、语音、或触觉输入。

[0064] 这里所描述的系统和技术能够以计算系统来实现，所述计算系统包括：后端组件（例如作为数据服务器），或者包括中间组件（例如应用服务器），或者包括前端组件（例如，具有用户通过其可以与这里所描述的系统和技术实施方式进行交互的图形用户界面或Web浏览器的客户端计算机），或者这样的后端、中间件、或前端组件的组合。所述系统的组件能够通过任何形式或介质的数字数据通信（例如，通信网络）互连。通信网络的示例包括局域网（LAN）、广域网（WAN）、和互联网。

[0065] 计算系统能够包括客户端和服务器。客户端和服务器通常彼此远离并且典型地通过通信网络进行交互。客户端和服务器的关系依靠在各自计算机上运行并且彼此具有客户端—服务器关系的计算机程序产生。

[0066] 虽然上文中已经详细描述了一些实施方式，但是其它修改也是可能的。例如，虽然客户端应用被描述为访问（一个或多个）委托，但是在其它实施方式中，（一个或多个）委托可以被一个或多个处理器所实现的其它应用所采用，所述其他应用诸如在一个或多个服务器上执行的应用。此外，在图中所描绘的逻辑流不需要所示出的特定顺序或连续顺序来实现所期望的结果。此外，可以提供其它动作，或者可以从所描述的流中消除动作，并且可以向所描述的系统添加其他组件或者从其去除其他组件。因此，其它实施方式在以下权利要求书的范围之内。

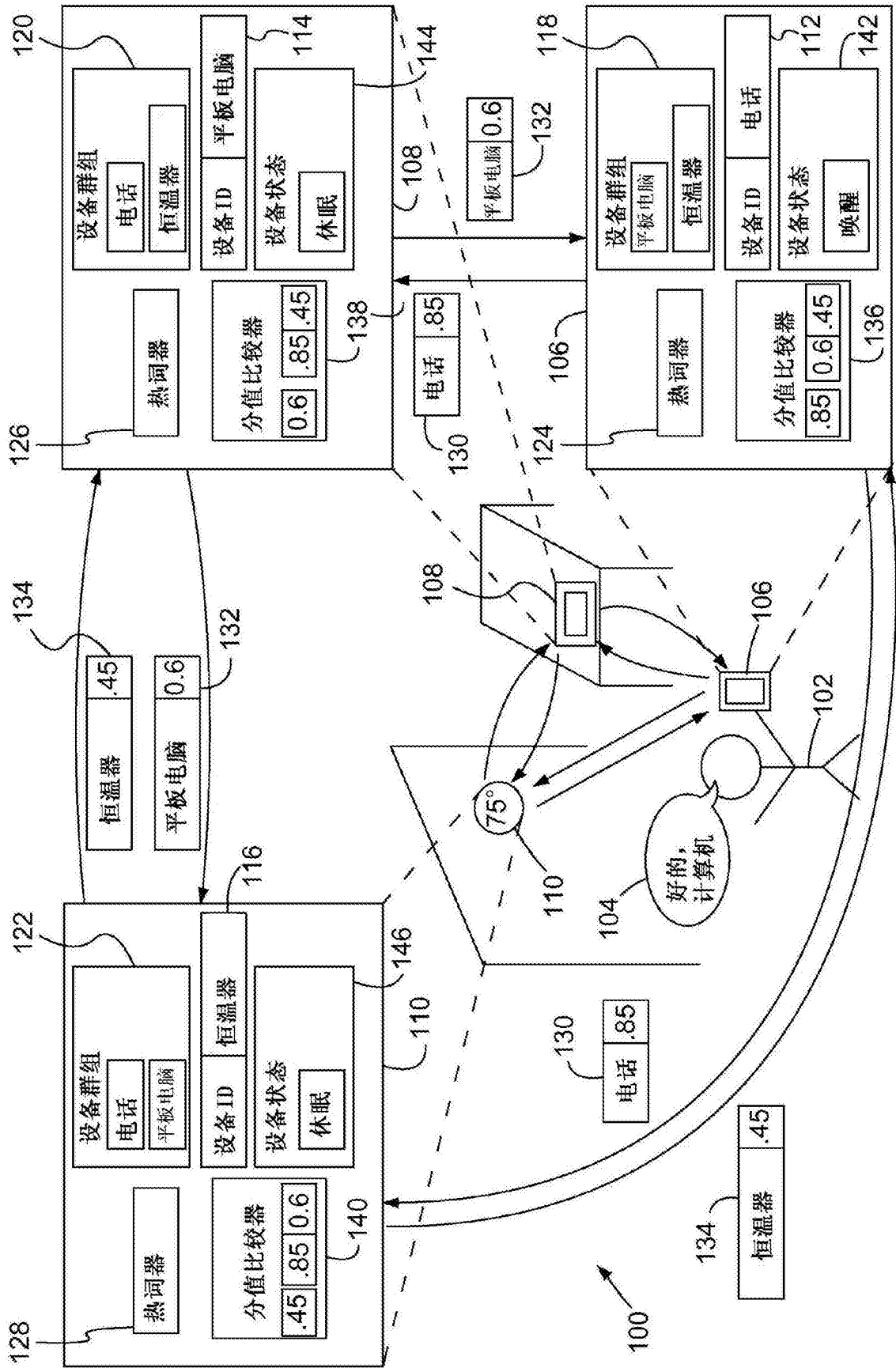


图1

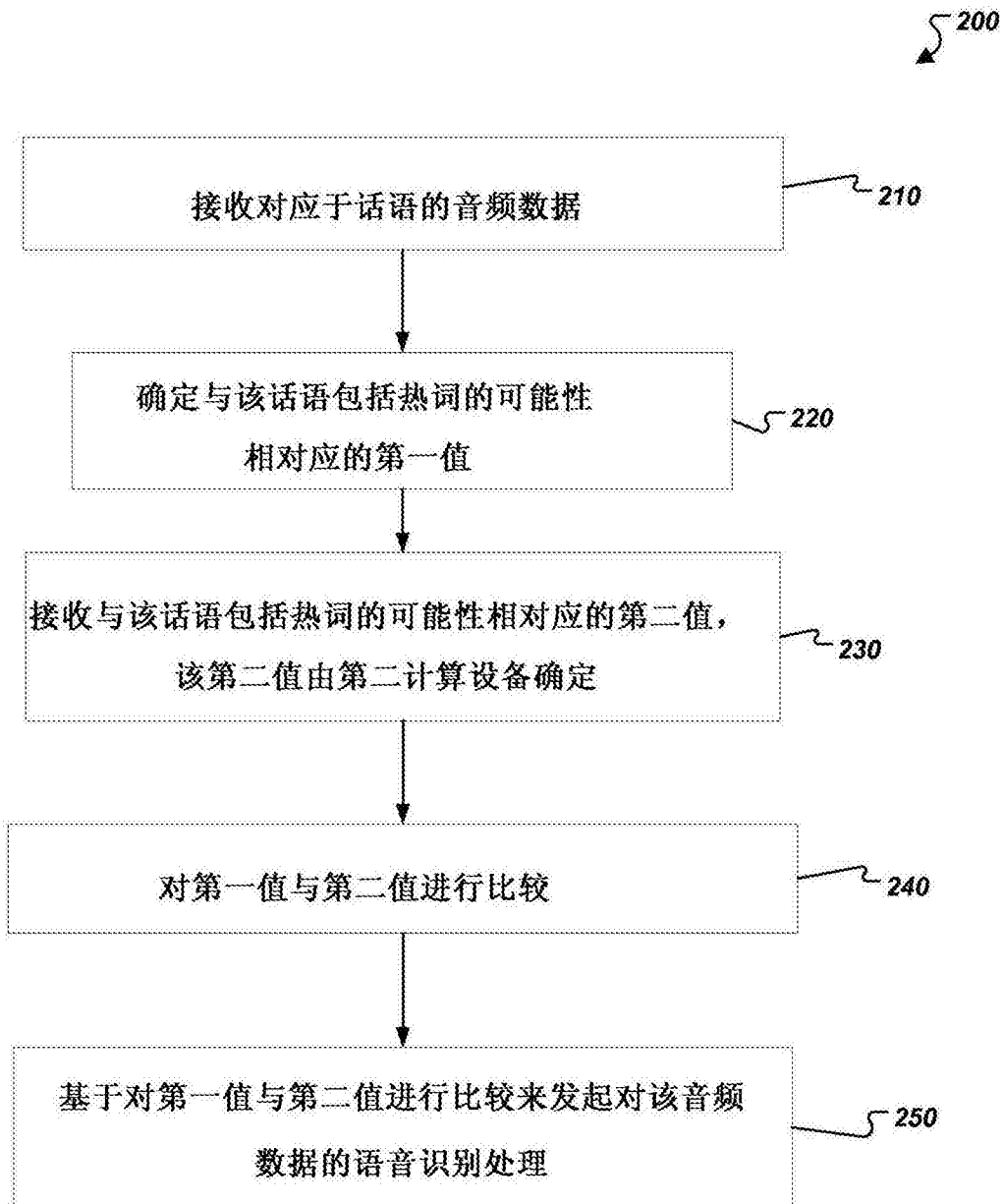


图2

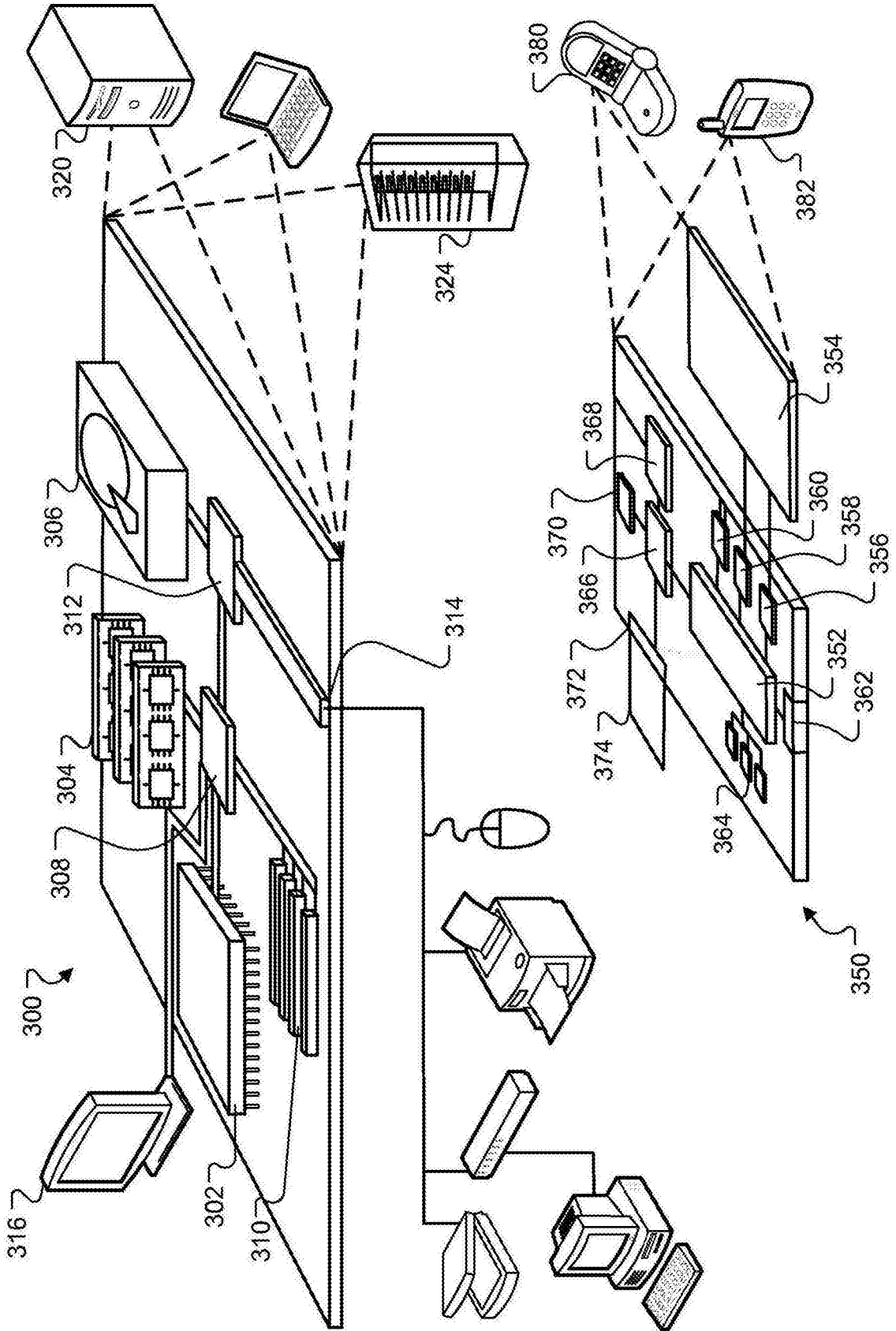


图3