



(12) 发明专利

(10) 授权公告号 CN 114127665 B

(45) 授权公告日 2024. 10. 08

(21) 申请号 202080049275.8

(22) 申请日 2020.07.10

(65) 同一申请的已公布的文献号
申请公布号 CN 114127665 A

(43) 申请公布日 2022.03.01

(30) 优先权数据
62/873,775 2019.07.12 US
16/685,946 2019.11.15 US

(85) PCT国际申请进入国家阶段日
2022.01.05

(86) PCT国际申请的申请数据
PCT/US2020/041499 2020.07.10

(87) PCT国际申请的公布数据
W02021/011331 EN 2021.01.21

(73) 专利权人 高通股份有限公司
地址 美国加利福尼亚州

(72) 发明人 R.乔达里 L-H.金 S.文 Y.郭
F.萨基 E.维泽

(74) 专利代理机构 北京市柳沈律师事务所
11105
专利代理师 安之斐

(51) Int.Cl.
G06F 3/01 (2006.01)
G06F 3/0484 (2022.01)
G06F 3/04883 (2022.01)
G06F 3/16 (2006.01)
G06F 9/451 (2018.01)
G10L 15/20 (2006.01)
G10L 15/22 (2006.01)
G10L 15/26 (2006.01)

(56) 对比文件
US 2018329677 A1, 2018.11.15

审查员 牟雪

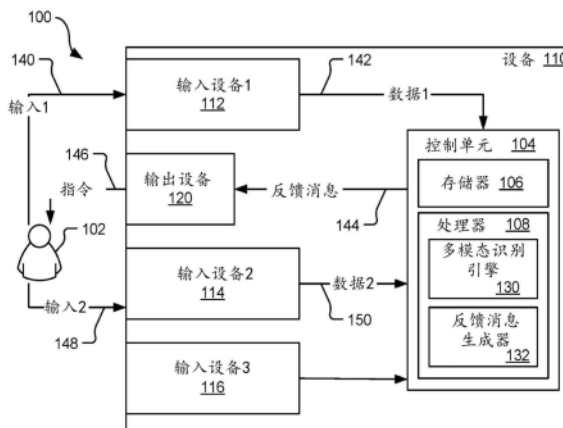
权利要求书3页 说明书19页 附图11页

(54) 发明名称

多模态用户界面

(57) 摘要

一种用于多模态用户输入的设备,包括处理器,该处理器被配置为处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式的第一输入。第一输入对应于命令。处理器被配置为基于对第一数据的处理向输出设备发送反馈消息。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。处理器被配置为:从第二输入设备接收第二数据,第二数据表示第二输入;并且更新映射,以将第一输入与由第二输入标识的命令相关联。



1. 一种用于多模态用户输入的设备,所述设备包括:
一个或多个处理器,所述一个或多个处理器被配置为:
处理从第一输入设备接收到的第一数据,所述第一数据表示来自用户的基于第一输入模式的第一输入;
基于对所述第一数据的处理向输出设备发送反馈消息,其中,所述反馈消息指示所述用户基于不同于所述第一输入模式的第二输入模式来提供标识与所述第一输入相关联的命令的第二输入;
从第二输入设备接收第二数据,所述第二数据表示所述第二输入;
基于所述第二数据标识所述命令;以及
更新将所述第一输入与所述命令相关联的映射,以响应于后续经由所述第一输入设备接收所述第一数据来使能所述命令的激活。
2. 根据权利要求1所述的设备,其中,所述第一输入模式是语音模式、手势模式或视频模式中的一种,并且其中,所述第二输入模式是所述语音模式、所述手势模式或所述视频模式中不同的一种。
3. 根据权利要求1所述的设备,其中,所述反馈消息指示所述用户提供所述第二输入,以消除所述第一输入的歧义。
4. 根据权利要求3所述的设备,其中,所述一个或多个处理器还被配置为响应于与所述第一输入的处理未能满足置信度阈值相关联的置信度来发送所述反馈消息。
5. 根据权利要求1所述的设备,其中,所更新的映射将所述第一输入和所述第二输入的组合与所述命令相关联。
6. 根据权利要求1所述的设备,其中,所述一个或多个处理器包括多模态识别引擎,所述多模态识别引擎包括:
融合嵌入网络,所述融合嵌入网络被配置为组合与所述第一输入模式相关联的第一嵌入网络的输出和与所述第二输入模式相关联的第二嵌入网络的输出,以生成组合嵌入向量;以及
分类器,所述分类器被配置为将所述组合嵌入向量映射到特定命令。
7. 根据权利要求6所述的设备,还包括存储器,所述存储器被配置为存储:
对应于所述第一嵌入网络数据和第一权重数据;以及
对应于第二用户的第二嵌入网络数据和第二权重数据,基于所述第一用户与所述第二用户之间的输入命令差异,所述第一嵌入网络数据不同于所述第二嵌入网络数据,并且基于所述第一用户与所述第二用户之间的输入模式可靠性差异,所述第一权重数据不同于所述第二权重数据。
8. 根据权利要求1所述的设备,其中,所述第一输入模式对应于视频模式,并且其中,所述一个或多个处理器被配置为响应于具有低于照明阈值的环境光度量来发送所述反馈消息。
9. 根据权利要求1所述的设备,其中,所述第一输入模式对应于语音模式,并且其中,所述一个或多个处理器被配置为响应于具有超过噪声阈值的噪声度量来发送所述反馈消息。
10. 根据权利要求1所述的设备,还包括被配置为表示图形用户界面的显示器。
11. 根据权利要求1所述的设备,还包括被配置为捕获包括一个或多个关键词或语音命

令的音频输入的一个或多个麦克风。

12. 根据权利要求1所述的设备,还包括被配置为捕获包括一个或多个手势或视觉命令的视频输入的一个或多个相机。

13. 根据权利要求1所述的设备,还包括被配置为接收表示手势输入的数据的一个或多个天线。

14. 根据权利要求1所述的设备,还包括被配置为向用户呈现或引导反馈消息的一个或多个扬声器。

15. 根据权利要求1所述的设备,其中,所述用户包括机器人或其他电子设备。

16. 根据权利要求1所述的设备,其中,所述第一输入设备和所述输出设备被结合到虚拟现实耳机或增强现实耳机中。

17. 根据权利要求1所述的设备,其中,所述第一输入设备和所述输出设备被结合到车辆中。

18. 一种用于多模态用户输入的方法,所述方法包括:

在设备的一个或多个处理器处处理从第一输入设备接收到的第一数据,所述第一数据表示来自用户的基于第一输入模式的第一输入;

基于对所述第一数据的处理,从所述一个或多个处理器向输出设备发送反馈消息,其中,所述反馈消息指示所述用户基于不同于所述第一输入模式的第二输入模式来提供标识与所述第一输入相关联的命令的第二输入;

在所述一个或多个处理器处,从第二输入设备接收第二数据,所述第二数据表示第二输入;

基于所述第二数据标识所述命令;以及

在所述一个或多个处理器处更新将所述第一输入与所述命令相关联的映射,以响应于后续经由所述第一输入设备接收所述第一数据来使能所述命令的激活。

19. 根据权利要求18所述的方法,其中,所述第一输入模式是语音模式、手势模式或视频模式中的一种,并且其中,所述第二输入模式是所述语音模式、所述手势模式或所述视频模式中不同的一种。

20. 根据权利要求18所述的方法,其中,所述反馈消息指示所述用户提供所述第二输入,以消除所述第一输入的歧义。

21. 根据权利要求20所述的方法,其中,所述反馈消息是响应于与所述第一输入的识别处理未能满足置信度阈值相关联的置信度而被发送的。

22. 根据权利要求18所述的方法,其中,所更新的映射将所述第一输入和所述第二输入的组合与所述命令相关联。

23. 根据权利要求18所述的方法,其中,更新所述映射包括以下至少一项:

更新与所述用户相关联的嵌入网络数据;或者

更新与所述用户相关联的权重数据。

24. 根据权利要求18所述的方法,其中,所述第一输入模式对应于视频模式,并且其中,所述反馈消息是响应于具有低于照明阈值的环境光度量而被发送的。

25. 根据权利要求18所述的方法,其中,所述第一输入模式对应于语音模式,并且其中,所述反馈消息是响应于具有超过噪声阈值的噪声度量而被发送的。

26. 一种用于多模态用户输入的装置,所述装置包括:

用于处理从第一输入设备接收到的第一数据的部件,所述第一数据表示来自用户的基于第一输入模式的第一输入;

用于基于对所述第一数据的处理向输出设备发送反馈消息的部件,其中,所述反馈消息指示所述用户基于不同于所述第一输入模式的第二输入模式来提供标识与所述第一输入相关联的命令的第二输入;

用于从第二输入设备接收第二数据的部件,所述第二数据表示所述第二输入;

基于所述第二数据标识所述命令;以及

用于更新将所述第一输入与所述命令相关联的映射,以响应于后续经由所述第一输入设备接收所述第一数据来使能所述命令的激活的部件。

27. 根据权利要求26所述的装置,其中,所更新的映射将所述第一输入和所述第二输入的组合与所述命令相关联。

28. 一种包括指令的非暂时性计算机可读介质,当所述指令由设备的一个或多个处理器执行时,所述指令使所述一个或多个处理器:

处理从第一输入设备接收到的第一数据,所述第一数据表示来自用户的基于第一输入模式的第一输入;

基于对所述第一数据的处理向输出设备发送反馈消息,其中,所述反馈消息指示所述用户基于不同于所述第一输入模式的第二输入模式来提供标识与所述第一输入相关联的命令的第二输入;

从第二输入设备接收第二数据,所述第二数据表示所述第二输入;

基于所述第二数据标识所述命令;以及

更新将所述第一输入与所述命令相关联的映射,以响应于后续经由所述第一输入设备接收所述第一数据来使能所述命令的激活。

29. 根据权利要求28所述的非暂时性计算机可读介质,其中,所述第一输入模式对应于视频模式,并且其中,所述反馈消息是响应于具有低于照明阈值的环境光度量而被发送的。

30. 根据权利要求28所述的非暂时性计算机可读介质,其中,所述第一输入模式对应于语音模式,并且其中,所述反馈消息是响应于具有超过噪声阈值的噪声度量而被发送的。

多模态用户界面

[0001] 优先权

[0002] 本申请要求于2019年7月12日提交的共同拥有的美国临时专利申请第62/873,775号和于2019年11月15日提交的美国非临时专利申请第16/685,946号的优先权,这些专利申请中的每一个专利申请的内容都通过引用的方式明确地整体并入本文。

技术领域

[0003] 本公开总体上涉及用户界面(user interface),更具体地,涉及支持多种用户输入模态的用户界面。

背景技术

[0004] 许多用户界面基于自动语音识别(ASR)和自然语言处理(NLP),并且经过许多不同命令、口音和语言的训练,在大量客户群中非常有用。训练此类用户界面以便在各种用户中广泛应用需要大量的资源,并且使用户界面普遍适用于大客户群而进行的许多训练因基于每个用户而造成“浪费”,因为每个单独的用户通常只使用单一的语言、口音和支持的命令子集。

发明内容

[0005] 根据本公开的一种实现方式,用于多模态用户输入的设备包括一个或多个处理器,该一个或多个处理器被配置为处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式的第一输入,并且第一输入对应于命令。一个或多个处理器被配置为基于对第一数据的处理向输出设备发送反馈消息。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。一个或多个处理器被配置为:从第二输入设备接收第二数据,第二数据表示第二输入;并且更新映射,以将第一输入与由第二输入标识的命令相关联。

[0006] 根据本公开的另一实现方式,一种用于多模态用户输入的方法包括在设备的一个或多个处理器处处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式的第一输入,并且第一输入对应于命令。该方法包括基于对第一数据的处理,从一个或多个处理器向输出设备发送反馈消息。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。该方法包括在一个或多个处理器处从第二输入设备接收第二数据。第二数据表示第二输入。该方法还包括在一个或多个处理器处更新映射,以将第一输入与由第二输入标识的命令相关联。

[0007] 根据本公开的另一实现方式,一种用于多模态用户输入的装置包括用于处理从第一输入设备接收到的第一数据的部件。第一数据表示来自用户的基于第一输入模式的第一输入,并且第一输入对应于命令。该装置包括用于基于对第一数据的处理向输出设备发送反馈消息的部件。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。该装置包括用于从第二输入设备接收第二数据的

部件,第二数据表示第二输入。该装置包括用于更新映射,以将第一输入与由第二输入标识的命令相关联的部件。

[0008] 根据本公开的另一种实现方式,非暂时性计算机可读介质包括指令,当该指令由设备的一个或多个处理器执行时,该指令使一个或多个处理器处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式的第一输入,第一输入对应于命令。当指令由一个或多个处理器执行时,该指令使一个或多个处理器基于对第一数据的处理向输出设备发送反馈消息。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。当该指令由一个或多个处理器执行时,该指令使一个或多个处理器从第二输入设备接收第二数据,第二数据表示第二输入。当指令由一个或多个处理器执行时,该指令还使一个或多个处理器更新映射,以将第一输入与由第二输入标识的命令相关联。

附图说明

[0009] 图1是根据本公开的一些示例的包括可用于处理多模态用户输入的设备的系统的特定说明性实现方式的图。

[0010] 图2是根据本公开的一些示例的图1的设备的组件的特定实现方式的图。

[0011] 图3是根据本公开的一些示例的包括可用于处理多模态用户输入的设备的系统的另一种特定实现方式的图。

[0012] 图4是根据本公开的一些示例的包括可用于处理多模态用户输入的设备的系统的另一种特定实现方式的示例的图。

[0013] 图5是根据本公开的一些示例的可用于处理多模态用户输入的设备的另一种实现方式的图。

[0014] 图6是根据本公开的一些示例的可以由图1的设备执行的处理多模态用户输入的方法的实现方式的图。

[0015] 图7是根据本公开的一些示例的可以由图1的设备执行的处理多模态用户输入的方法的另一种实现方式的图。

[0016] 图8是根据本公开的一些示例的可以由图1的设备执行的处理多模态用户输入的方法的另一种实现方式的图。

[0017] 图9是根据本公开的一些示例的可以由图1的设备执行的处理多模态用户输入的方法的另一种实现方式的图。

[0018] 图10是根据本公开的一些示例的可以由图1的设备执行的处理多模态用户输入的方法的另一种实现方式的图。

[0019] 图11是根据本公开的一些示例的可用于处理多模态用户输入的车辆图。

[0020] 图12A是根据本公开的一些示例的可用于处理多模态用户输入的虚拟现实或增强现实耳机的图。

[0021] 图12B是根据本公开的一些示例的可用于处理多模态用户输入的可穿戴电子设备的图。

[0022] 图13是根据本公开的一些示例的可用于处理多模态用户输入的设备的特定说明性示例的框图。

具体实施方式

[0023] 描述了使用多种输入模态实现用户交互的设备和方法。许多用户界面基于自动语音识别 (ASR) 和自然语言处理 (NLP), 并且经过许多不同命令、口音和语言的训练, 在大量客户群中非常有用。训练此类用户界面以便在各种用户中广泛应用需要大量的资源, 并且使用户界面普遍适用于大客户群而进行的许多训练因基于每个用户而造成“浪费”, 因为每个单独的用户通常只使用单一的语言、口音和支持的命令子集。

[0024] 通过实现多模态用户交互以及个性化解释用户命令的能力, 本文描述的技术能够训练多模态用户界面供特定用户使用, 从而减少或消除了对传统用户界面的广泛适用性的广泛训练。在一些实现方式中, 不同的嵌入网络用于不同的输入模态 (例如, 用于语音的嵌入网络、用于视觉输入的嵌入网络、用于手势输入的嵌入网络等), 并且被配置为区分使用相应模态接收到的不同命令。为了说明, “嵌入网络”可以包括一个或多个神经网络层, 该一个或多个神经网络层被配置 (例如, 训练) 为处理输入数据, 诸如语音数据 (例如, 时域语音数据或频域语音数据), 以生成嵌入向量。“嵌入向量”是与输入数据相比相对低维的向量 (例如, 多个值的集合), 表示输入数据, 并且可以用于区分输入数据的不同实例。不同的嵌入网络输出被转换到公共嵌入空间, 并且被融合成组合嵌入向量。例如, 语音输入的 n 维语音嵌入向量可以被转换成 k 维第一嵌入向量, 并且手势输入的 m 维手势嵌入向量可以被转换成 k 维第二嵌入向量 (其中, m 、 n 和 k 可以彼此相等或不同)。 k 维第一嵌入向量表示语音输入, 并且 k 维第二嵌入向量表示 k 维向量空间 (例如, 公共嵌入空间) 中的手势输入。 k 维第一嵌入向量和 k 维第二嵌入向量可以被组合, 诸如通过向量加法进行组合, 以生成组合嵌入向量。分类器解释组合嵌入向量以生成输出。

[0025] 嵌入网络和分类器中的每一个都可以由单独的用户更新 (例如, 训练), 以改进对经由各种模态接收到的用户命令的识别。例如, 如果接收到不能以高置信度解释的口头用户命令, 则用户界面可以向用户询问该口头命令的含义, 并且用户可以使用不同的模态输入该含义, 诸如通过执行由用户界面识别的手势输入。

[0026] 在一些实现方式中, 用户界面可以请求用户改变输入模态。例如, 如果用户的用于增加回放音量的口头命令“向上”不能可靠地与另一个命令 (例如“关”) 区分开来, 则用户界面可以生成请求用户添加另一种模态以更好地区分该命令的反馈消息 (例如, 口头或显示的)。例如, 用户可以添加视觉输入, 诸如指向上方作为“增加音量”命令。用户界面可以被更新, 以将口头输入“向上”与指向上方的视觉输入的组合识别为多模态命令, 从而增加回放音量。因此, 个性化更新 (例如, 从单模态到多模态) 可以用于改进命令识别的准确性。

[0027] 在一些实现方式中, 用户界面请求用户改变输入模态, 以更容易地消除用户输入的歧义。例如, 在音频噪声损害对用户的语音的解释的实现方式中 (例如, 在移动的车辆中), 用户界面可以生成请求用户改变模态的反馈消息, 诸如改变为视觉或手势模态。作为另一个示例, 在弱光水平损害对用户的视觉输入的解释的实现方式中, 用户界面可以生成请求用户改变模态的反馈消息, 诸如改变为使用可穿戴电子设备 (例如“智能手表”) 的运动检测器来检测手部的移动和方向的语音模态或手势模态。因此, 指示用户改变输入模态可以用于改进命令识别的准确性。

[0028] 在一些实现方式中, 作为多因素认证过程的一部分, 用户界面请求用户改变输入模态。例如, 在接收口头用户输入以执行语音认证之后, 用户界面接下来可以请求用户提供

视觉或手势输入。使用另一种输入模态提供附加用户输入的请求可以由先前用户输入的异常触发,诸如检测到语音输入具有表示用户的录制语音的回放的特征。可替代地,或者此外,请求可以被随机生成,或者作为多因素认证过程的已建立的一系列认证输入的一部分。因此,指示用户改变输入模态可以用于实现更高的准确性、更鲁棒的用户认证。如本文所使用的,对应于多因素认证过程的认证输入的用户输入不同于对应于用户命令的用户输入。为了说明,对应于命令的用户输入被用户界面解释为执行与命令(例如,“开灯”)相关联的动作或“技能”的指令,而对应于认证输入的用户输入被用于与存储的用户简档的数据进行比较,以确定用户输入源自与存储的用户简档相关联的同一用户的可能性(例如,经由生物测定数据或其他用户标识数据的比较)。

[0029] 除非上下文明确限定,否则术语“产生”用于表示其任何普通含义,诸如计算、生成和/或提供。除非上下文明确限定,否则术语“提供”用于表示其任何普通含义,诸如计算、生成和/或产生。除非上下文明确限定,否则术语“耦合”用于表示直接或间接的电连接或物理连接。如果连接是间接的,则在被“耦合”的结构之间可能存在其他块或组件。例如,扬声器可以经由中间介质(例如空气)声学耦合到附近的墙壁,该中间介质使得波(例如声音)能够从扬声器传播到墙壁(反之亦然)。

[0030] 术语“配置”可以用于指方法、装置、设备、系统或其任何组合,如其特定上下文所示。本说明书和权利要求中使用的术语“包括”不排除其他元件或操作。术语“基于”(例如“A基于B”)用于表示其任何普通含义,包括(i)“至少基于”(例如“A至少基于B”)和(ii)“等于”(例如“A等于B”)的情况。在(i)的情况,“A基于B”包括“至少基于”,这可以包括A耦合到B的配置。类似地,术语“响应于”用于表示其任何普通含义,包括“至少响应于”。术语“至少一个”用于表示其任何普通含义,包括“一个或多个”。术语“至少两个”用于表示其任何普通含义,包括“两个或更多个”。

[0031] 除非特定上下文另有说明,否则术语“装置”和“设备”一般地并且可互换地使用。除非另有说明,否则具有特定特征的装置的操作的任何公开也明确旨在公开具有类似特征的方法(反之亦然),并且根据特定配置的装置的操作的任何公开也明确旨在公开根据类似配置的方法(反之亦然)。除非特定上下文另有说明,否则术语“方法”、“处理”、“过程”和“技术”一般地并且可互换地使用。术语“元件”和“模块”可以用于表示更大配置的一部分。术语“分组”可以对应于包括报头部分和有效载荷部分的数据单元。通过引用并入文件的一部分也应理解为并入该部分中引用的术语或变量的定义,其中,这些定义出现在文件的其他地方,以及并入部分中引用的任何附图。

[0032] 如本文所使用的,术语“通信设备”指的是可以用于通过无线网络进行语音和/或数据通信的电子设备。通信设备的示例包括智能扬声器、扬声器条、蜂窝电话、个人数字助理(PDA)、手持设备、耳机、可穿戴设备、无线调制解调器、膝上型计算机、个人计算机等。

[0033] 图1描绘了系统100,其中,用户102与设备110交互以进行多模态用户输入。设备110包括第一输入设备112、第二输入设备114、可选的一个或多个附加输入设备(诸如第三输入设备116)、输出设备120和控制单元104。在一些实现方式中,作为说明性的非限制性示例,设备110可以包括便携式通信设备(例如,“智能电话”)、可穿戴设备(例如,“智能手表”)、车辆系统(例如,与汽车娱乐系统、导航系统或自动驾驶控制系统一起使用的可移动

或可移除显示器)或虚拟现实或增强现实耳机。

[0034] 第一输入设备112被配置为基于第一输入模式检测第一用户输入。在一个示例中,第一输入设备112包括麦克风,并且第一输入模式包括语音模式(例如,用于ASR/NLP)。为了说明,第一输入设备112可以包括被配置为捕获包括一个或多个关键词或语音命令的音频输入的一个或多个麦克风。

[0035] 第二输入设备114被配置为基于第二输入模式检测第二用户输入。在一个示例中,第二输入设备114包括相机,并且第二输入模式包括视频模式(例如,以检测用户102的视觉方面,诸如竖起大拇指或大拇指朝下的手部位置、面部表情等)。为了说明,第二输入设备114可以包括被配置为捕获包括一个或多个手势或视觉命令的视频输入的一个或多个相机。

[0036] 第三输入设备116被配置为基于第三输入模式检测第三用户输入。在一个示例中,第三输入设备116包括手势跟踪器,并且第三输入模式包括手势模式。第三输入设备116可以包括被配置为接收表示手势输入的数据(例如,运动数据)的一个或多个天线。为了说明,用户102可以佩戴包括运动传感器(例如,加速度计、陀螺仪等)的手镯或手表,该运动传感器跟踪用户的手部的运动并且向第三输入设备116发送运动数据。在其他实现方式中,运动跟踪电子设备可以与用户102集成,诸如人类用户102中的控制论植入件(cybernetic implant),或者在用户102是机器人的实现方式中可以是用户102的组件。

[0037] 输出设备120被配置为向用户102输出信息,诸如经由使用扬声器生成听觉输出、经由使用显示器生成视觉输出、经由一个或多个其他输出模态(例如,触觉)或其任何组合。例如,如下面进一步描述的,输出设备120可以从控制单元104接收消息数据(例如,反馈消息144),并且可以向用户102生成输出(例如,指令146)。在特定示例中,输出设备120包括被配置为表示图形用户界面的显示器、被配置为向用户102呈现或引导反馈消息144的一个或多个扬声器,或其组合。

[0038] 控制单元104被配置为从输入设备112至116接收对应于用户输入的数据,并且生成反馈消息,该消息经由输出设备120被提供给用户102。控制单元104包括耦合到一个或多个处理器(称为处理器108)的存储器106。如参考图2进一步描述的,存储器106可以包括表示一个或多个嵌入网络的数据、表示嵌入向量到组合嵌入空间的一个或多个转换的数据、以及表示一个或多个分类器的数据,这些数据可以由处理器108访问以供使用。存储器106还可以包括可由处理器108执行的指令,以实现多模态识别引擎130、反馈消息生成器132或两者。

[0039] 处理器108包括多模态识别引擎130和反馈消息生成器132。在一些实现方式中,处理器108包括一个或多个处理核心,该一个或多个处理核心被配置为执行指令以实现多模态识别引擎130和反馈消息生成器132。在一些实现方式中,处理器108包括专用电路,该专用电路被配置为实现多模态识别引擎130和反馈消息生成器132中的一个或两个。在一个示例中,处理器108被实现为集成电路(IC)。

[0040] 多模态识别引擎130被配置为从一个或多个输入设备112至116接收数据,并且处理接收到的数据以生成输出。例如,输出可以包括与接收到的输入最匹配的命令和与该命令相关联的置信度(或可能性)指示符。在一些实现方式中,多模态识别引擎130被配置为生成数据,以针对每个输入模态将特定的训练用户输入与其他训练用户输入区分开,诸如通

过为每个输入模态生成嵌入向量。多模态识别引擎130可以被配置为组合与不同输入模态相关联的嵌入向量,以生成组合嵌入向量,该组合嵌入向量表示经由输入设备112至116中的每一个检测到哪个识别的用户输入(如果有的话)作为单模态或多模态用户输入的一部分。组合嵌入向量被处理以确定输出,诸如通过使用被训练成将组合嵌入向量映射到命令的分类器处理。参考图2描述了可以在多模态识别引擎130中实现的组件的说明性示例。

[0041] 反馈消息生成器132被配置为生成反馈消息数据,该反馈消息数据经由输出设备120被输出给用户102。例如,反馈消息生成器132可以向输出设备120发送反馈消息144,以指示用户102重复未被充分识别(诸如被预测为置信度低于阈值的特定命令)的用户输入。作为其他示例,反馈消息生成器132可以向输出设备120发送反馈消息144,以指示用户102改变输入模态,或者用使用不同输入模态进行的另一个输入来增强使用一个输入模态进行的输入。其他示例包括生成反馈消息数据,该反馈消息数据提供用户102要仿真的用户的输入的录制样本、用户102要标识的用户输入的录制样本,或者帮助用户102使用设备110的其他信息。说明性示例包括跨模态样本检索,诸如响应于从用户102接收到查询显示示出对应于“向上”的运动的运动视频、生成与向上手势的运动相关联的最相似话语的音频回放,或者生成已经与用户定义的运动紧密相关联的相关话语的音频回放。在一些示例中,如下面更详细描述,反馈消息生成器132被配置为生成反馈消息144,以指示用户102根据多因素认证过程提供下一认证输入。

[0042] 在操作期间,用户102基于由第一输入设备112检测到的第一输入模式(例如,口头命令)提供第一输入140。第一输入设备112生成表示第一输入140的第一数据142,并且向控制单元104提供第一数据142。

[0043] 处理器108(例如,多模态识别引擎130)处理表示来自用户102的基于第一输入模式(例如,语音)的第一输入140的第一数据142。处理器108(例如,反馈消息生成器132)基于对第一数据142的处理向输出设备120发送反馈消息144。例如,反馈消息144指示用户102经由播放口头指令146、使用不同的输入模式提供第二输入148。第二输入148基于不同于第一输入模式的第二输入模式(例如,视频),并且可以用于更新多模态识别引擎130响应第一输入140的方式。如本文所使用的,使用不同的输入模式是指使用不同类型的输入,而不是使用相同类型的输入。每种不同类型的输入都使用一系列不同的传感器。例如,语音输入模式可以使用一个或多个麦克风。手势输入模式可以使用运动检测。视频输入模式可以使用相机和帧序列。通常,每种输入模式提供不同类型的传感器,这些传感器可以用于提供输入。

[0044] 在一些实现方式中,第一输入140是命令,并且反馈消息144指示用户102提供第二输入148来消除第一输入140的歧义。多模态识别引擎130可以响应于与第一输入140的识别处理未能满足置信度阈值相关联的置信度来发送反馈消息144,该置信度表示输出的不确定性(例如,口头输入表示“向上”还是“关”的不确定性)。如图2中进一步详细描述,用户102可以提供第二输入148(例如,指向上方),并且基于表示第二输入148的第二数据150,多模态识别引擎130可以更新第一输入140(例如,语音“向上”)到与第二输入148相关联的动作(例如,增加音乐音量)的映射。

[0045] 在另一种实现方式中,多模态识别引擎130更新与第二输入148组合的第一输入140到与第二输入148相关联的动作的映射。例如,当噪声条件阻止可靠地识别口头“向上”命令时,多模态识别引擎130被更新以将用户的“向上”视频输入(例如,向上指)与用户的口

头“向上”命令一起识别为调高音量的单个命令。

[0046] 因此,在一些实现方式中,用户102可以个性化设备110,以将特定输入识别为经由反馈消息144的反馈机制来执行特定动作的命令和第二输入148。为了说明,用户102可以说出当前未被多模态识别引擎130识别的命令(第一输入140),并且响应于反馈消息144,用户102可以通过输入识别的命令(第二输入148)来标识要映射到未识别的命令的动作。类似地,当用户的选择的模式变得不可靠时,设备110可以指示用户102改变输入模式。例如,当设备110在车辆(例如,汽车导航和/或娱乐系统)中实现时,在夜间驾驶期间,用户102可以被指示使用语音输入或手势输入来代替视频(由于弱光条件),并且当在窗户打开的情况下驾驶时,用户102可以被指示使用手势输入或视频输入来代替语音(由于高风噪声)。当在诸如虚拟现实或增强现实耳机的耳机中实现设备110时,可以执行类似的操作来指示用户102改变输入模式。

[0047] 在其他实现方式中,设备110用于执行多因素认证。例如,第一输入140可以对应于用户102的第一认证动作(例如,用于扬声器验证的口头验证码),并且反馈消息144指示用户102提供第二输入148(例如,显示用户102先前选择的特定手部配置)作为第二认证动作,作为多因素认证过程的一部分。设备110可以随机地或根据算法选择认证输入模式的数量和类型,以用于指示用户102执行认证动作。例如,设备110可以响应于语音输入(例如,第一输入140)可以是正在播放的录制语音的指示来生成指令146,并且可以请求“活泼”确认,诸如通过指示用户102向相机(例如,第二输入设备114)眨眼。

[0048] 尽管以上示例描述了第二输入148使用与第一输入140的模式不同的模式,但是在其他实现方式中,第二输入148可以使用与第一输入140的模式相同的模式。例如,尽管第一输入140的口头命令可能难以解释(例如,在存在环境噪声的情况下,“向上”对“关闭”),但是另一个口头命令(例如,“更大声”)可能与其他映射的命令足够不同,以选择正确的动作(例如,增加音量)。作为另一个示例,在训练过程中,用户102可以发出“更大声”作为未训练的语音命令,并且设备110可以经由指令146指示用户102标识与“更大声”的话语相关联的动作。用户102可以提供被设备110识别为增加音量的命令的第二口头话语“向上”,并且多模态识别引擎130可以更新用户输入的映射,以用“增加音量”动作来映射“更大声”。

[0049] 图2示出了根据特定实现方式的控制单元104的组件的示例,这些组件包括存储器106、多模态识别引擎130以及可由处理器108执行的一个或多个应用程序240。多模态识别引擎130包括第一嵌入网络202,第一嵌入网络202被配置为将第一用户输入(例如,语音输入)转换成第一嵌入向量(例如,第一嵌入向量“E1”)。第二嵌入网络204配置为将第二用户输入(例如,手势输入)转换成第二嵌入向量(例如,第二嵌入向量“E2”)。多模态识别引擎130可以包括一个或多个附加嵌入网络,包括被配置为将第N个用户输入(例如,视频输入)转换成第N个嵌入向量(例如,第N个嵌入向量“En”)的第N个嵌入网络206。根据本公开的一些实施例,多模态识别引擎130可以包括任何数量的嵌入网络。

[0050] 融合嵌入网络220被配置为组合嵌入网络202至206的输出,并且生成组合嵌入向量,诸如组合嵌入向量“C”228。例如,第一转换212可以将语音嵌入向量转换成“公共”嵌入空间,以生成第一公共嵌入向量222。第二转换214可以将手势嵌入向量转换到公共嵌入空间以生成第二公共嵌入向量224,第N转换216可以将视频嵌入向量转换到公共嵌入空间以生成第N公共嵌入向量226。公共嵌入向量222至226中的每一个可以分别用对应的权重W1、

W2和W3加权,并且在融合嵌入网络220处组合。映射230被配置为选择对应于组合嵌入向量228的输出232和置信度234。例如,映射230可以包括分类器231,该分类器231被配置为将组合嵌入向量映射到特定动作。为了说明,单个分类器231被用于确定由多个嵌入网络202至206的组合输入产生的输出232,而不是对每个模态输入使用单独的分类器。

[0051] 表示多模态识别引擎130使用的一个或多个参数的数据被存储在存储器106中。第一用户简档250与第一用户(例如,用户102)相关联,并且包括第一嵌入网络数据252、第一权重数据254、第一临时调节数据256和第一历史数据258。第一嵌入网络数据252包括用于配置嵌入网络(包括第一嵌入网络202、第二嵌入网络204、第N嵌入网络206和融合嵌入网络220)的数据(例如,权重或其他参数或值),以对应于第一用户。第一权重数据254包括权重值,以将权重(例如,W1、W2、W3)配置为对应于第一用户。如下面进一步描述的,第一临时调节数据256包括基于临时条件调节多模态识别引擎130的配置的值(例如,在噪声环境中减少权重W1并且增加权重W2和W3)。第一历史数据258包括与第一用户相关联的历史数据,并且使得处理器108能够基于对应于由多模态识别引擎130处理的第一用户的多模态输入的历史趋势来更新第一嵌入网络数据252、第一权重数据254或两者。

[0052] 类似地,存储器106包括与第二用户相关联的第二用户简档260,并且包括第二用户的第二嵌入网络数据262、第二权重数据264、第二临时调节数据266和第二历史数据258。基于第一用户与第二用户之间的输入命令差异,第一嵌入网络数据252不同于第二嵌入网络数据262。例如,当执行视频输入时,第一用户和第二用户可以具有不同的口音、不同的手势风格、不同的身体力学或其任何组合。第一嵌入网络数据252可以表示训练嵌入网络202至206和220以从第一用户的默认嵌入网络数据集识别用户特定变化的结果,并且第二嵌入网络数据262可以表示训练嵌入网络202至206和220以从第二用户的默认嵌入网络数据集识别用户特定变化的结果。尽管仅两个用户简档250、260被示出,但是任何数量的用户简档可以被包括来为设备110的多个用户定制多模态识别引擎130的操作。

[0053] 除了针对单独的用户变化(诸如针对不同的口音、手势风格和身体力学)进行调节之外,第一嵌入网络数据252还可以表示训练嵌入网络202至206和220以识别由第一用户确定的用户输入的第一定制集合的结果,并且第二嵌入网络数据262还可以表示训练嵌入网络202至206和220以识别由第二用户确定的用户输入的第二定制集合的结果。例如,第一用户可以定制(例如,训练)多模态识别引擎130,以将语音命令“向上”识别为在音频回放操作正在进行时增加音量的命令。相反,第二用户可以定制(例如,训练)多模态识别引擎130,以将语音命令“向上”识别为在音频回放操作正在进行时选择播放列表上的前一音频轨道的命令。

[0054] 基于第一用户与第二用户之间的输入模式可靠性差异,第一权重数据254可以不同于第二权重数据264。例如,处理器108可以诸如基于第一历史数据258确定与来自第一用户的手势输入相比,来自第一用户的语音输入不太可靠地被解释。结果,权重W1可以从默认W1值减小,并且权重W2可以从第一权重数据254中的默认W2值增加,以减少对语音输入的依赖,并且增加对来自第一用户的手势输入的依赖。相反,处理器108可以诸如基于第二历史数据268确定与来自第二用户的手势输入相比,来自第二用户的语音输入更可靠。结果,在第二权重数据264中,权重W1可以从默认W1值增加,并且权重W2可以从默认W2值减少,以减少对手势输入的依赖,并且增加对来自第二用户的语音输入的依赖。

[0055] 应用程序240包括临时调节器290和数据调节器292。临时调节器290被配置为基于临时条件来确定嵌入网络202、204、206或220中的一个或多个的调节、权重W1至W3中的一个或多个的调节,或其组合。例如,临时调节器290可以基于检测到的条件来调节权重W1至W3中的一个或多个,以强调一个或多个输入模态、去强调一个或多个输入模态,或其组合。作为说明性的非限制性示例,如下面进一步详细描述,检测到的条件可以由环境噪声数据272、环境光数据274、位置数据276或用户偏好278中的一个或多个来表示。

[0056] 数据调节器292被配置为确定嵌入网络202、204、206或220中的一个或多个的调节、权重W1至W3中的一个或多个的调节,或其组合,以更新嵌入网络数据和权重数据来表示被确定为不是基于临时条件的变化。在一些实现方式中,数据调节器292被配置为对嵌入式网络202、204、206或220中的一个或多个执行更新训练,以表示用户输入到特定命令的更新的映射,诸如响应于从用户接收到帮助多模态识别引擎130更准确地识别用户输入(例如,以适应用户的口头命令的发音与默认语音识别模型之间的差异),或响应表示输入到特定命令的定制映射的用户输入(例如,用户用双手输入“竖起大拇指”手势作为先前未知的视频输入,并且表示该视频输入应该导致设备110关闭警报)。

[0057] 在图1的系统100中实现的多模态识别引擎130的操作的说明性示例中,用户102被标识为对设备110的多模态输入的来源,诸如经由面部识别、语音识别或一些其他形式的用户识别。基于第一嵌入网络数据252更新嵌入网络202至206,基于第一权重数据254更新权重W1、W2和W3,并且基于第一临时调节数据256应用任何临时调节,以配置(例如,定制)多模态识别引擎130来识别来自用户102的输入。

[0058] 用户102提供第一输入140作为命令。第一输入140没有被识别为具有足够可靠性的任何特定命令,并且反馈消息144指示用户102提供第二输入148来消除第一输入140的歧义。例如,反馈消息144可以响应于与第一输入140的识别处理未能满足置信度阈值294相关联的置信度234而被发送,该置信度234表示输出232的不确定性(例如,口头输入表示“向上”还是“关闭”的不确定性)。在其他实现方式中,反馈消息144响应于检测到一个或多个环境条件而被发送。

[0059] 例如,在第一输入140经由视频模式被接收的实现方式中,反馈消息144响应于具有低于照明阈值286的的环境光度量284而被发送。例如,环境光数据274可以经由设备110的一个或多个传感器而被接收,并且被处理以生成环境光度量284。环境光度量284可以与照明阈值286进行比较,以确定环境照明对于可靠的视频模式输入来说是否太昏暗。反馈消息144可以通知用户昏暗的照明导致视频输入模式不可靠,并且可以指示用户使用另一种模态(例如,语音)重复输入。

[0060] 作为另一个示例,在第一输入140经由语音模式被接收的实现方式中,反馈消息144响应于具有高于噪声阈值282的值的噪声度量280(例如,信噪比或环境噪声测量)而被发送。例如,环境噪声数据272可以经由设备110的一个或多个传感器被接收(或在麦克风输入信号的语音活动检测处理期间被测量),并且被处理以生成噪声度量280。噪声度量280可以与噪声阈值282进行比较,以确定环境噪声对于可靠的语音模式输入来说是否太大。反馈消息144可以通知用户噪声环境导致语音输入模式不可靠,并且可以指示用户使用另一种模态(例如,视频)重复输入。

[0061] 用户102可以提供第二输入148(例如,指向上方),并且基于表示第二输入148的第

二数据150,多模态识别引擎130可以更新第一输入140(例如,语音“向上”)到与第二输入148相关联的动作(例如,增加音乐音量)的映射。为了说明,第一嵌入网络202、第一转换212、权重W1、融合嵌入网络220或映射230中的一个或多个可以由数据调节器292调节,以使多模态识别引擎130更准确地将用户的口头“向上”识别为增加音乐音量的命令。

[0062] 在操作的说明性示例中,如果一个输入模态被确定为具有低准确性条件,则多模态识别引擎130(例如,响应于由临时调节器290生成的临时调节数据)调节一个或多个设置,以减少或消除使用该模态生成组合嵌入向量228的输入的影响。在语音模态被确定、预测或估计为不可靠(例如,由于噪声度量280超过噪声阈值282)的示例中,当手势和视频模态被确定为对于输入识别足够可靠时,临时调节器290可以将应用于与语音输入相关联的公共嵌入向量222的权重W1设置为“0”值。应用于与手势输入相关联的公共嵌入向量224的权重W2和应用于与视频输入相关联的公共嵌入向量226的权重W3被设置为非零值(例如,在手势和视频输入被视为同等可靠的实现方式中, $W2=W3=0.5$)。当语音输入保持不可靠时,将权重W1设置为“0”值防止语音输入对所得的组合嵌入向量228产生影响。

[0063] 根据以上示例,权重的初始设置可以分配 $W1=W2=W3=1/3$,表示每个模态对于输入识别具有同等的重要性或可靠性。随后语音模态可能被确定或预测为不可靠,诸如由于检测到大量环境噪声(例如,经由检测到噪声度量280超过噪声阈值282或检测到车辆在运动时车窗是打开的)或由于在预定时间段内语音输入的阈值数量未能被准确识别。响应于语音模态被确定或预测为不可靠,临时调节器290将权重W1、W2和W3分别调节为0、1/2和1/2,以消除语音输入对输入识别的影响。响应于后续确定语音输入模态不再不可靠(例如,风噪声低于噪声阈值、窗户被关闭或车辆已经停止移动),权重W1、W2和W3可以各自返回到它们的初始值1/3。

[0064] 作为另一个示例,视频模态可以改为被确定或预测为不可靠,诸如由于检测到少量环境光(例如,环境光度量284低于照明阈值286)或者由于在预定时间段内视频输入的阈值数量未能被准确识别。响应于视频模态被确定或预测为不可靠,临时调节器290将权重W1、W2和W3分别调节为1/2、1/2和0,以消除视频输入对输入识别的影响。响应于后续确定视频输入模态不再不可靠(例如,环境光被确定为超过照明阈值),权重W1、W2和W3可以各自返回到它们的初始值1/3。

[0065] 在一些实现方式中,多个权重被调节以减少或消除多个输入模态对输入识别的影响。例如,在确定仅语音模态要使用的实现方式中,W1被设置为“1”,并且W2和W3被设置为“0”。为了说明,设备110可以检测低环境光条件,并且还可以确定没有活动手势检测设备被检测到(例如,用户的智能手表不存在或者没有发送运动数据)。作为另一个示例,用户102可以诸如经由输入用户偏好278来指示设备110将输入识别限制为仅处理语音输入。作为另一个示例,位置数据276可以用于确定是否限制一个或多个输入模态。例如,响应于表示用户正在操作车辆的位置数据276,临时调节器290可以限制用户输入模式以防止手势输入和视频输入的识别,诸如防止用户分心并且鼓励车辆的安全操作。响应于表示用户不再操作车辆并且在用户的家中的位置数据276,临时调节器290可以恢复用户输入模式以使得能够识别手势输入和视频输入。

[0066] 尽管在以上示例中描述了权重值的示例,但是这些示例性权重值是说明性的而非限制性的。为了说明,不是将权重设置为“0”,而是可以将权重设置为减小的值,以减小但不

消除相关输入模态对整体输入识别的影响。作为另一个示例,“可靠的”输入模态可以具有不相等的权重,这可以表示输入模态的相对可靠性。为了说明,如果手势输入被认为足够可靠,视频输入被认为比手势输入更可靠,并且语音被确定为不可靠,则权重可以被设置为诸如 $W1=0.1$ 、 $W2=0.4$ 和 $W3=0.5$ 的值。尽管在以上示例中,权重 $W1$ 、 $W2$ 和 $W3$ 的总和等于1,但是在其他实现方式中,权重 $W1$ 、 $W2$ 和 $W3$ 的总和不限于任何特定值。

[0067] 除了调节一个或多个权重以减少或消除被确定为不可靠的输入模态的影响之外或者作为替代,在一些实现方式中,多模态识别引擎130可以将相关联的嵌入网络的输出强制为来自可用技能中的“无”输出、将转换的输出强制为“无”类别输入到具有“0”值的嵌入向量的公共嵌入空间,或其组合。

[0068] 在一些实现方式中,包括多模态识别引擎130的设备110执行多个输入模态的环境感知融合。例如,响应于确定用户102正在驾驶汽车,与手势输入相关联的权重 $W2$ 可以被设置为“0”,表示用户的手部动作可能对应于汽车操作而不是手势输入,以阻止驾驶汽车时不安全的手部动作。作为另一个示例,响应于确定用户102在暗室中,与视频输入相关联的权重 $W3$ 可以被设置为“0”。作为另一个示例,响应于确定用户102处于噪声环境中,与语音输入相关联的权重 $W1$ 可以被设置为“0”。环境条件的确定可以基于集成在设备110中的一个或多个传感器(例如,环境光传感器、环境噪声传感器)、设备110外部的一个或多个传感器(例如,经由设备110与家庭自动化系统、物联网系统或另一个系统的一个或多个组件之间的通信),或其任何组合。

[0069] 图3描绘了用于多模态用户输入的系统300的示例,其中,佩戴耳机302的用户与诸如智能电话、车辆系统(诸如汽车)或结合了无线数字助理应用(例如“智能扬声器”)的扬声器系统的另一个设备进行通信。耳机302可以对应于图1的设备110,并且可以包括显示器和换能器,诸如耳塞308或其他可穿戴的噪声生成设备,以向穿戴者提供增强现实(“AR”)、虚拟现实(“VR”)或混合现实(“MR”)音频和视频输出。

[0070] 耳机302可以包括多个传感器,诸如一个或多个麦克风、一个或多个相机等,以检测用户输入。例如,经由一个或多个麦克风接收到的音频输入可以用于在集成在耳机302中或耦合到耳机302的处理器处执行一个或多个操作310。例如,对应于音频输入的音频信号可以被处理,诸如使用机器学习,以实现声音环境分类、确定耳机302的佩戴者何时在说话的自语音语音活动检测(VAD)、声学事件检测和模式控制(例如,基于序列的用户界面)。

[0071] 一个或多个操作310的结果可以用于生成一个或多个动作312。例如,动作312可以包括调谐主动噪声消除(ANC)滤波、实现一个或多个辅助收听特征、调节多麦克风声音捕获的场(例如,“AudioZoom”),或者执行增强现实渲染、虚拟现实渲染或混合现实渲染(统称为“XR”渲染)。例如,结果可以以空间透明模式在耳机302处呈现。

[0072] 在耳机302处检测到的用户输入(例如,经由一个或多个麦克风、运动检测器、手势检测器、相机等)可以用于启动一个或多个基于语音的操作304的执行,诸如自动语音识别和自然语言处理、搜索或查询响应,或者两者。一个或多个基于语音的操作304可以使用机器学习来执行,诸如在智能电话或与耳机302通信的其他便携式通信设备上执行。数据通信305(例如,无线网络通信、有线通信或两者)可以包括向外部处理资源306(例如,基于云的ASR/NLP和结合机器学习的搜索服务器)发送音频语音数据。搜索和查询结果可以经由耳机302传送回用户。

[0073] 图4描绘了调节多麦克风声音捕获的场(例如,“AudioZoom”)的示例400,诸如可以由图3的耳机302在说明性的非限制性示例中执行。多个麦克风,诸如代表性的麦克风412、414和416围绕用户布置。用户被示出为在极坐标系中居中,并且面向0度角方向。麦克风412、414和416可以包括定向麦克风、非定向麦克风,或两者,并且捕获用户周围的音频环境。在第一配置402中,来自麦克风412至416的音频的声音处理产生对用户的听觉输出(例如,经由耳机或耳塞),其表示没有用户导向的调节的音频环境。

[0074] 在第二配置404中,响应于经由多模态接口的用户输入(例如,作为说明性示例的用户手势、话语、视频输入或其组合),来自麦克风412至416的音频的声音处理被调节为强调(例如,放大)源自或到达特定空间区域420(例如,90度角方向的区域或用户的左侧)的声音,同时衰减源自空间区域420之外的区域的声音。作为说明性的非限制性示例,导致转换到第二配置404的用户输入的示例可以包括基于语音模态的“向左缩放”语音序列、基于手势模态的“向左指手”或“向左指手指”手势序列,或者基于音频(非语音)模态的“发出啪嗒声”音频序列。

[0075] 在一些实现方式中,如上参考图1至图4描述的多模态界面响应于上下文,诸如用户附近的位置或活动(例如,在客厅看电视,或在厨房洗碗)。例如,使用基于手表或臂章的加速度计捕获的手势可以基于检测到的上下文来解释。例如,挥动一只手可以被解释为目标命令“开灯”,向左翻转一只手可以被解释为“下一首歌”或“下一个频道”,向右翻转一只手可以被解释为“上一首歌”、“上一个频道”或“开门”。作为说明性的非限制性示例,其中形成闭合拳头的“抓取”手势可以被解释为“选择呼叫”或“选择频道”,长抓取可以被解释为“停止歌曲”、“取消警报”或“关门”,并且手指伸出的手部的逆时针旋转可以被解释为“发现家庭设备”。可以经由与检测到的声学事件/环境序列相关联来确定上下文。例如,各种声学事件可以被检测来推断声学环境(例如,用户在哪里)或者监控适当的反馈定时。这种可检测的声学事件的示例包括:吹风机、真空吸尘器、音乐、厨房通风橱、烹饪、吃饭、洗碗、室内空调、微波炉、洗衣机、烘干机、淋浴和看电视。

[0076] 用于手势识别的数据集可以包括:表示手部运动的三维(3-D)加速度计和陀螺仪传感器数据(例如,沿着x、y和z轴)。传感器信号的每个分量(例如,来自加速度计和陀螺仪的信号)可以在固定宽度的窗口中,诸如3秒的窗口(例如,150个读数/窗口)。作为说明性的非限制性示例,多个手势类别可以实现,诸如:下一个、上一个、向上/增加、向下/减少、开、关和未知。一种或多种数据增强技术可以被实现,诸如置换、时间扭曲、缩放、幅度扭曲、抖动和裁剪。

[0077] 特征提取可以被执行,诸如基于用于手势识别的数据集的统计特征。为了说明,作为说明性的非限制性示例,提取的特征可以对应于最小值、最大值、Var、平均值、标准偏差、MSE(最小均方误差)、ACF(自相关)、ACV(自协方差)、偏斜度、峰度、平均交叉率、抖动或3分位数。

[0078] 一个或多个模型可以用于手势识别,诸如支持向量机(SVM)、梯度增强、分类器、堆叠的长短期记忆递归神经网络(LSTM-RNN)、利用attention的序列到序列编码器-解码器模型、一个或多个其他模型,或其任何组合。

[0079] 在一些方面,多模态识别引擎130可以学习或训练以生成直接映射到目标动作的序列嵌入向量。输入序列的示例包括加速度计或陀螺仪时间序列(例如,用于手势输入)、语

音命令时间序列或音频时间序列。利用attention的编码器-解码器LSTM-RNN可以用于学习生成嵌入向量,该嵌入向量将可变长度时间序列信号表示为固定长度和区别向量,诸如生成到softmax层的输出,以表示与输入序列相关联的目标动作类别。

[0080] 在一些方面,多模态识别引擎130可以使用嵌入向量来登记和设计不同的动作类别。例如,许多不同的输入序列可以被登记,并且一个或多个分类器可以使用映射到每个目标动作的嵌入向量来设计。例如,SVM、K均值、K近邻(KNN)、余弦(cos)距离或其他设计可以被实现来将嵌入映射到目标动作。用户序列可以被测试,以验证更新后的系统的准确性。

[0081] 在一些方面,在登记和SVM/K均值/KNN设计之后,与分类器评估相关联的度量显示类别之间的辨别太模糊,并且序列成形可以被执行。在这种情况下,反馈消息生成器132可以生成反馈,以向用户指示导致与其他类别混淆的某些有问题的序列。例如,混淆类别的运动、音频或语音序列可以诸如经由输出设备120向用户回放。用户可以意识到哪些序列会引起混淆,并且可以说出/用手势表示新的序列,以改善目标类别之间的辨别并且提供歧义消除。可替代地,用户可以经由多模态用户界面将混淆的输入序列匹配到相同的动作/类别,使得输入序列之间的模糊性变得没有意义。在接收到用户反馈之后,多模态识别引擎130可以重新登记和修改SVM/K均值/KNN设计,并且反馈消息生成器132可以在混淆的情况下重新提示用户重复输入序列,直到序列映射变得彼此足够不同。例如,“混淆矩阵”可以表示不同序列之间的模糊量,并且训练可以重复,直到混淆矩阵是准对角的。

[0082] 在一些方面,具有多模态输入的序列成形可以响应于检测到导致与其他动作类别混淆的“有问题的”输入序列而被执行。设备110可以向用户请求用户是否希望对每个有问题的输入序列使用多模态输入。例如,如果多模态识别引擎130难以区分用户对于“关”和“调低音量”的特定手势,则输出设备120可以向用户输出查询:“您想要对“关”类别使用口头命令“关”吗?”作为另一个示例,输出设备120可以输出查询:“您想要使用口头命令“降低音量”用于“调低音量”类别吗?”响应于用户选择使用多模态输入(例如,由于手势混淆而添加口头命令),多模态提示可以被激活,并且设备110可以使用多模态输入序列来调节混淆矩阵以包括混淆概率。

[0083] 在一些方面,交互式顺序验证可以用于消除输入序列的歧义。例如,用户可能忘记哪些类别注册了哪些多模态输入。基于对话的交互可以在用户与设备110之间进行。例如,如果手势输入被检测为“关”类别,则输出设备120可以询问用户:“您是指“关”类别还是“下一个”类别吗?”用户可以回答“关”,并且多模态识别引擎130可以激活“关”命令。

[0084] 图5描绘了设备502的实现方式500,设备502包括集成在分立组件(诸如半导体芯片或封装,如参考图13进一步描绘的)中的多模态识别引擎130和反馈消息生成器132。为了说明,设备502可以包括一个或多个处理器(例如,处理器108),该一个或多个处理器被配置为执行存储的指令以执行关于多模态识别引擎130和反馈消息生成器132描述的操作。设备502包括传感器数据输入端510,诸如第一总线接口,以使得传感器数据504(诸如来自图1的一个或多个输入设备112诸116的数据)能够从设备502外部的一个或多个传感器接收。设备502还包括输出端512,诸如第二总线接口,以使得能够发送反馈消息144(例如,到输出设备120)。设备502能够实现多模态用户界面处理,作为包括多个传感器和输出设备的系统中的组件,诸如图11描绘的车辆、图12A描绘的虚拟现实或增强现实耳机、图12B描绘的可穿戴电子设备,或图13描绘的无线通信设备。

[0085] 参考图6,描绘了处理多模态用户输入的方法600的特定实现方式,作为说明性的非限制性示例,该方法可以由图1的设备110或控制单元104、图5的设备502,或两者来执行。

[0086] 方法600包括在设备的处理器处处理从第一输入设备接收到的第一数据(602处)。第一数据表示来自用户的基于第一输入模式的第一输入。例如,参考图1,处理器108处理从第一输入设备112接收到的第一数据142。第一数据142表示来自用户102的基于第一输入模式的第一输入140。

[0087] 方法600还包括基于对第一数据的处理,从设备的处理器向输出设备发送反馈消息(604处)。反馈消息指示用户提供基于不同于第一输入模式的第二输入模式的第二输入。例如,参考图1,控制单元104基于对第一数据142的处理向输出设备120发送反馈消息144。反馈消息144指示用户102提供基于第二输入模式的第二输入148。

[0088] 方法600可以由现场可编程门阵列(FPGA)设备、专用集成电路(ASIC)、诸如中央处理单元(CPU)的处理单元、数字信号处理器(DSP)、控制器、另一种硬件设备、固件设备,或其任何组合来实现。作为示例,如本文所描述,方法600可以由执行指令的处理器来执行。

[0089] 参考图7,描绘了处理多模态用户输入的方法700的特定实现方式,作为说明性的非限制性示例,该方法可以由图1的控制单元104、图5的设备502,或两者来执行。

[0090] 方法700包括处理从第一输入设备接收到的第一数据(702处)。第一数据表示对应于来自用户的基于第一输入模式的命令的第一输入。例如,参考图1,处理器108处理从第一输入设备112接收到的第一数据142。第一数据142表示对应于来自用户102的基于第一输入模式的命令的第一输入140。

[0091] 方法700还包括基于对第一数据的处理向输出设备发送反馈消息(704处)。反馈消息指示用户提供基于不同于第一输入模式的第二输入模式的第二输入,以消除第一输入的歧义。例如,参考图1,控制单元104基于对第一数据142的处理向输出设备120发送反馈消息144。反馈消息144指示用户102提供基于不同于第一输入模式的第二输入模式的第二输入148,以消除第一输入140的歧义。

[0092] 方法700可以由现场可编程门阵列(FPGA)设备、专用集成电路(ASIC)、诸如中央处理单元(CPU)的处理单元、DSP、控制器、另一种硬件设备、固件设备,或其任何组合来实现。作为示例,如本文所描述,方法700可以由执行指令的处理器来执行。

[0093] 参考图8,描绘了处理多模态用户输入的方法800的特定实现方式,作为说明性的非限制性示例,该方法可以由图1的控制单元104、图5的设备502,或两者来执行。

[0094] 方法800包括处理从第一输入设备接收到的第一数据(802处)。第一数据表示来自用户的基于第一输入模式的第一输入,并且第一数据对应于用户的第一认证动作。例如,参考图1,处理器108处理从第一输入设备112接收到的第一数据142。第一数据142表示来自用户102的基于第一输入模式的第一输入140,并且第一数据142对应于用户102的第一认证动作。

[0095] 方法800还包括基于对第一数据的处理向输出设备发送反馈消息(804处)。反馈消息指示用户基于不同于第一输入模式的第二输入模式提供第二输入,作为第二认证动作,作为多因素认证过程的一部分。例如,参考图1,控制单元104基于对第一数据142的处理向输出设备120发送反馈消息144。反馈消息144指示用户102基于不同于第一输入模式的第二输入模式来提供第二输入148,作为第二认证动作,作为多因素认证过程的一部分。

[0096] 方法800可以由现场可编程门阵列 (FPGA) 设备、专用集成电路 (ASIC)、诸如中央处理单元 (CPU) 的处理单元、DSP、控制器、另一种硬件设备、固件设备,或其任何组合来实现。作为示例,如本文所描述,方法800可以由执行指令的处理器来执行。

[0097] 参考图9,描绘了处理多模态用户输入的方法900的特定实现方式,作为说明性的非限制性示例,该方法可以由图1的控制单元104、图5的设备502,或两者来执行。

[0098] 方法900包括基于第一输入模式检测第一用户输入 (902处)。例如,参考图1,第一输入设备112基于第一输入模式检测第一用户输入140。

[0099] 方法900还包括基于第二输入模式检测第二用户输入 (904处)。例如,参考图1,第二输入设备114基于第二输入模式检测第二用户输入148。

[0100] 方法900还包括使用第一嵌入网络生成第一嵌入向量,该第一嵌入网络被配置为将第一用户输入转换成第一嵌入向量 (906处)。例如,参考图2,第一嵌入网络202通过将第一用户输入转换成第一嵌入向量来生成第一嵌入向量。

[0101] 方法900还包括使用第二嵌入网络生成第二嵌入向量,该第二嵌入网络被配置为将第二用户输入转换成第二嵌入向量 (908处)。例如,参考图2,第二嵌入网络204通过将第二用户输入转换成第二嵌入向量来生成第二嵌入向量。

[0102] 方法900还包括使用融合嵌入网络生成组合嵌入向量,该融合嵌入网络被配置为组合第一嵌入网络的输出和第二嵌入网络的输出以生成组合嵌入向量 (910处)。例如,参考图2,融合嵌入网络220组合第一嵌入网络202的输出和第二嵌入网络204的输出,以生成组合嵌入向量。

[0103] 方法900还包括使用分类器将组合嵌入向量映射到特定动作 (912处)。例如,参考图2,映射230将组合嵌入向量映射到特定动作。

[0104] 方法900可以由现场可编程门阵列 (FPGA) 设备、专用集成电路 (ASIC)、诸如中央处理单元 (CPU) 的处理单元、DSP、控制器、另一种硬件设备、固件设备,或其任何组合来实现。作为示例,如本文所描述,方法900可以由执行指令的处理器来执行。

[0105] 参考图10,描绘了处理多模态用户输入的方法1000的特定实现方式,作为说明性的非限制性示例,该方法可以由图1的控制单元104、图5的设备502,或两者来执行。

[0106] 方法1000包括处理从第一输入设备接收到的第一数据 (1002处)。第一数据表示来自用户的基于第一输入模式的第一输入,并且第一输入对应于命令。例如,参考图1,处理器108处理从第一输入设备112接收到的第一数据142。第一数据142表示来自用户102的基于第一输入模式的第一输入140。

[0107] 方法1000还包括基于对第一数据的处理向输出设备发送反馈消息 (1004处)。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。例如,参考图1,控制单元104基于对第一数据142的处理向输出设备120发送反馈消息144。反馈消息144指示用户102使用另一种输入模式重新输入与第一输入140相关联的命令。在一个示例中,第一输入模式是语音模式、手势模式或视频模式中的一种,并且第二输入模式是语音模式、手势模式或视频模式中不同的一种。

[0108] 在一些实现方式中,反馈消息指示用户提供第二输入来消除第一输入的歧义。例如,反馈消息可以响应于与第一输入的识别处理未能满足置信度阈值相关联的置信度 (诸如当置信度234小于置信度阈值294时) 而被发送。在一些示例中,第一输入模式对应于视频

模式,并且反馈消息响应于具有低于照明阈值的值的环境光度量(诸如具有小于照明阈值286的值的的环境光度量284)而被发送。在其他示例中,第一输入模式对应于语音模式,并且反馈消息响应于具有超过噪声阈值的值的噪声度量(诸如具有大于噪声阈值282的值的噪声度量280)而被发送。

[0109] 方法1000还包括从第二输入设备接收第二数据,第二数据表示第二输入(1006处)。例如,参考图1,控制单元104接收与第二输入148相关联的第二数据150,该第二数据150标识与第一输入140相关联的特定命令。

[0110] 方法1000还包括更新映射以将第一输入与第二输入标识的命令相关联(1008处)。例如,参考图1,控制单元104更新用户输入到命令的映射,以将第一输入140与特定命令相关联。在一些实现方式中,更新的映射将第一输入和第二输入的组合与命令相关联,以诸如为用户定制输入到命令的映射,使得经由同时(例如,在时间上至少部分重叠)经由第一输入模式接收第一输入和经由第二输入模式接收第二输入来更可靠地识别命令。在一些实现方式中,更新映射包括更新与用户相关联的嵌入网络数据(例如,第一嵌入网络数据252)或更新与用户相关联的权重数据(例如,第一权重数据254)中的至少一个。

[0111] 方法1000可以由现场可编程门阵列(FPGA)设备、专用集成电路(ASIC)、诸如中央处理单元(CPU)的处理单元、DSP、控制器、另一种硬件设备、固件设备,或其任何组合来实现。作为示例,如本文所描述,方法1000可以由执行指令的处理器来执行。

[0112] 图11描绘了集成到诸如汽车仪表盘设备1102的车辆仪表盘设备中的设备110的实现方式1100的示例。多个传感器1150可以包括一个或多个麦克风、照相机或其他传感器,并且可以对应于图1的输入设备112至116。尽管一个或多个传感器1150被示出为在单个位置,但是在其他实现方式中,该一个或多个传感器1150可以位于车辆的其他位置处,诸如分布在车辆的车厢内的不同位置处,诸如位于车辆中每个座位附近的一个或多个麦克风和一个或多个相机的阵列,以检测来自车辆驾驶员和来自每个乘客的多模态输入。

[0113] 诸如显示器1120的视觉接口设备可以对应于输出设备120,并且被安装在汽车仪表盘设备1102内或定位在汽车仪表盘设备1102上(例如,可拆卸地固定到车辆手持设备支架上),以便汽车的驾驶员可见。多模态识别引擎130和反馈消息生成器132用虚线边界示出,以表示多模态识别引擎130和反馈消息生成器132对车辆的乘客不可见。多模态识别引擎130和反馈消息生成器132可以在还包括显示器1120和传感器1150的设备中(诸如在图1的设备110中)实现,或者可以诸如在图5的设备502中与显示器1120和传感器1150分离并且耦合。

[0114] 图12A描绘了集成到诸如虚拟现实耳机、增强现实耳机或混合现实耳机的耳机1202中的多模态识别引擎130和反馈消息生成器132的示例。诸如显示器1220的视觉接口设备可以对应于输出设备120,并且被定位在用户的眼睛的前方,以使得能够在耳机1202被佩戴时向用户显示增强现实图像或场景或虚拟现实图像或场景。传感器1250可以包括一个或多个麦克风、照相机或其他传感器,并且可以对应于图1的输入设备112至116。尽管一个或多个传感器1250被示出为在单个位置,但是在其他实现方式中,该一个或多个传感器1250可以位于耳机1202的其他位置处,诸如分布在耳机1202周围的一个或多个麦克风和一个或多个相机的阵列,以检测多模态输入。

[0115] 图12B描绘了集成到可穿戴电子设备1204中的多模态识别引擎130和反馈消息生

成器132的示例,可穿戴电子设备1204被示出为包括显示器1220和传感器1250的“智能手表”。例如,传感器1250能够基于诸如视频、语音和手势的模式来检测用户输入。此外,尽管一个或多个传感器1250被示出为在单个位置,但是在其他实现方式中,一个或多个传感器1250可以位于可穿戴电子设备1204的其他位置处。

[0116] 图13描绘了诸如在无线通信设备实现方式(例如,智能电话)或数字助理设备实现方式中包括多模态识别引擎130的设备1300的特定说明性实现方式的框图。在各种实现方式中,设备1300可以具有比图13所示出的组件更多或更少的组件。在说明性实现方式中,设备1300可以对应于设备110。在说明性实现方式中,设备1300可以执行参考图1至图12B描述的一个或多个操作。

[0117] 在特定实现方式中,设备1300包括处理器1306(例如,对应于处理器108的中央处理单元(CPU)),该处理器1306包括多模态识别引擎130。设备1300可以包括一个或多个附加处理器1310(例如,一个或多个DSP)。处理器1310可以包括语音和音乐编解码器-解码器(编解码器)1308。语音和音乐编解码器1308可以包括声码器(“声码器”)编码器1336、声码器解码器1338,或两者。

[0118] 设备1300可以包括存储器1386和编解码器1334。存储器1386可以对应于存储器106,并且可以包括可由处理器1306(或一个或多个附加处理器1310)执行的指令1356,以实现参考多模态识别引擎130、反馈消息生成器132、一个或多个应用程序240,或其任何组合描述的功能。设备1300可以包括经由收发器1350耦合到一个或多个天线1352的无线控制器1340。在一些实现方式中,一个或多个天线1352包括被配置为接收表示手势输入的数据的一个或多个天线。

[0119] 设备1300可以包括耦合到显示控制器1326的显示器1328(例如,输出设备120)。显示器1328可以被配置为表示输出反馈消息144(例如,指令146)的图形用户界面。编解码器1334可以包括数模转换器(DAC)1302和模数转换器(ADC)1304。在特定实现方式中,编解码器1334可以从一个或多个麦克风1312(例如,配置为捕获包括一个或多个关键词或语音命令的音频输入的第一输入设备112)接收模拟信号,使用模数转换器1304将模拟信号转换成数字信号,并且将数字信号提供给语音和音乐编解码器1308。语音和音乐编解码器1308可以处理数字信号。

[0120] 在特定实现方式中,语音和音乐编解码器1308可以向编解码器1334提供表示音频回放信号的数字信号。编解码器1334可以使用数模转换器1302将数字信号转换成模拟信号,并且可以向一个或多个扬声器1314提供模拟信号,以生成可听信号。一个或多个扬声器1314可以对应于输出设备120,并且可以被配置为呈现图1的反馈消息144或者向用户引导反馈消息144。

[0121] 在特定实现方式中,设备1300包括一个或多个输入设备1330。输入设备1330可以对应于图1的输入设备112至116中的一个或多个。例如,输入设备1330可以包括被配置为捕获包括一个或多个手势或视觉命令的视频输入的一个或多个相机。

[0122] 在特定实现方式中,设备1300可以被包括在系统级封装或系统级芯片设备1322中。在特定实现方式中,存储器1386、处理器1306、处理器1310、显示控制器1326、编解码器1334和无线控制器1340被包括在系统级封装或片上系统设备1322中。在特定实现方式中,输入设备1330(例如,图1的输入设备112至116中的一个或多个)和电源1344耦合到系统级

封装或片上系统设备1322。此外,在特定实现方式中,如图13所示,显示器1328、输入设备1330、麦克风1312、天线1352和电源1344在系统级封装或系统级芯片设备1322的外部。在特定实现方式中,显示器1328、输入设备1330、麦克风1312、扬声器1314、天线1352和电源1344中的每一个可以耦合到系统级封装或片上系统设备1322的组件,诸如接口或控制器。

[0123] 作为说明性的非限制性示例,设备1300可以包括移动通信设备、智能电话、蜂窝电话、膝上型计算机、计算机、平板电脑、个人数字助理、显示设备、电视、游戏控制台、音乐播放器、收音机、数字视频播放器、数字视频光盘(DVD)或蓝光光盘播放器、调谐器、照相机、导航设备、虚拟现实耳机或增强现实耳机、可穿戴电子设备、车辆控制台设备,或其任何组合。

[0124] 结合描述的实现方式,用于多模态用户输入的设备包括多模态识别引擎,该多模态识别引擎处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式(例如,语音模式、手势模式或视频模式)的第一输入。反馈消息生成器基于对第一数据的处理向输出设备发送反馈消息,该反馈消息指示用户提供基于不同于第一输入模式的第二输入模式的第二输入。

[0125] 结合描述的实现方式,用于多模态用户输入的设备包括多模态识别引擎,该多模态识别引擎处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式的第一输入。反馈消息生成器基于对第一数据的处理向输出设备发送反馈消息。反馈消息指示用户标识与第一输入相关联的动作。多模态识别引擎接收第二输入,该第二输入标识与第一输入相关联的特定动作,并且更新用户输入到动作的映射,以将第一输入与特定动作相关联。

[0126] 结合描述的实现方式,一种用于多模态用户输入的装置包括用于处理从第一输入设备接收到的第一数据的部件。第一数据表示来自用户的基于第一输入模式的第一输入,并且第一输入对应于命令。例如,用于处理第一数据的部件可以包括处理器108、多模态识别引擎130、可由处理器(1306、1310)执行的指令1356、一个或多个其他设备、模块、电路、组件,或其组合。

[0127] 该装置包括用于基于对第一数据的处理向输出设备发送反馈消息的部件。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。例如,用于发送的部件可以包括处理器108、多模态识别引擎130、反馈消息生成器132、可由处理器(1306、1310)执行的指令1356、一个或多个其他设备、模块、电路、组件,或其组合。

[0128] 该装置包括用于从第二输入设备接收第二数据的部件,第二数据表示第二输入。例如,用于接收第二数据的部件可以包括处理器108、多模态识别引擎130、可由处理器(1306、1310)执行的指令1356、一个或多个其他设备、模块、电路、组件,或其组合。

[0129] 该装置包括用于更新映射,以将第一输入与由第二输入标识的命令相关联的部件。例如,用于更新的部件可以包括处理器108、多模态识别引擎130、数据调节器292、可由处理器(1306、1310)执行的指令1356、一个或多个其他设备、模块、电路、组件,或其组合。

[0130] 在一些实现方式中,非暂时性计算机可读介质(例如,存储器106、存储器1386,或其任何组合)包括指令(例如,指令1356),当该指令由设备的一个或多个处理器(例如,处理器108、处理器1306、处理器1310,或其任何组合)执行时,该指令使一个或多个处理器执行用于处理多模态用户输入的操作,诸如通过执行对应于图6至10的一个或多个方法的全部

或部分的操作。在一个示例中,当指令由一个或多个处理器执行时,该指令使一个或多个处理器处理从第一输入设备接收到的第一数据。第一数据表示来自用户的基于第一输入模式的第一输入,第一输入对应于命令。当指令由一个或多个处理器执行时,该指令使一个或多个处理器基于对第一数据的处理向输出设备发送反馈消息。反馈消息指示用户基于不同于第一输入模式的第二输入模式来提供标识与第一输入相关联的命令的第二输入。当该指令由一个或多个处理器执行时,该指令使一个或多个处理器从第二输入设备接收第二数据,第二数据表示第二输入。当指令由一个或多个处理器执行时,该指令还使一个或多个处理器更新映射,以将第一输入与由第二输入标识的命令相关联。

[0131] 本领域技术人员将进一步理解,结合本文公开的实现方式描述的各种说明性逻辑块、配置、模块、电路和算法步骤可以被实现为电子硬件、由处理器执行的计算机软件,或两者的组合。各种说明性的组件、块、配置、模块、电路和步骤已经在上面在它们的功能方面进行了一般性描述。这种功能是实现为硬件还是处理器可执行指令取决于特定的应用和对整个系统施加的设计约束。技术人员可以针对每个特定应用以不同的方式实现所描述的功能,并且这种实现方式决策不应被解释为导致脱离本公开的范围。

[0132] 结合本文公开的实现方式描述的方法或算法的步骤可以直接体现在硬件、由处理器执行的软件模块,或两者的组合中。软件模块可以位于随机存取存储器(RAM)、闪存、只读存储器(ROM)、可编程只读存储器(PROM)、可擦除可编程只读存储器(EPROM)、电可擦除可编程只读存储器(EEPROM)、寄存器、硬盘、可移动盘、光盘只读存储器(CD-ROM),或本领域已知的任何其他形式的非瞬态存储介质中。示例性存储介质耦合到处理器,使得处理器可以从该存储介质读取信息,以及向该存储介质写入信息。或者,该存储介质可以集成到该处理器中。该处理器和该存储介质可以位于专用集成电路(ASIC)中。ASIC可以位于计算设备或用户终端中。或者,该处理器和存储介质可以作为分立组件位于计算设备或用户终端中。

[0133] 所公开的实现方式的先前描述被提供,以使本领域技术人员能够实现或使用所公开的实现方式。对这些实现方式的各种修改对于本领域技术人员来说将是显而易见的,并且在不脱离本公开的范围的情况下,本文定义的原理可以应用于其他实现方式。因此,本公开并不旨在限于本文所示出的实现方式,而是符合与由所附权利要求限定的原理和新颖特征相一致的尽可能宽的范围。

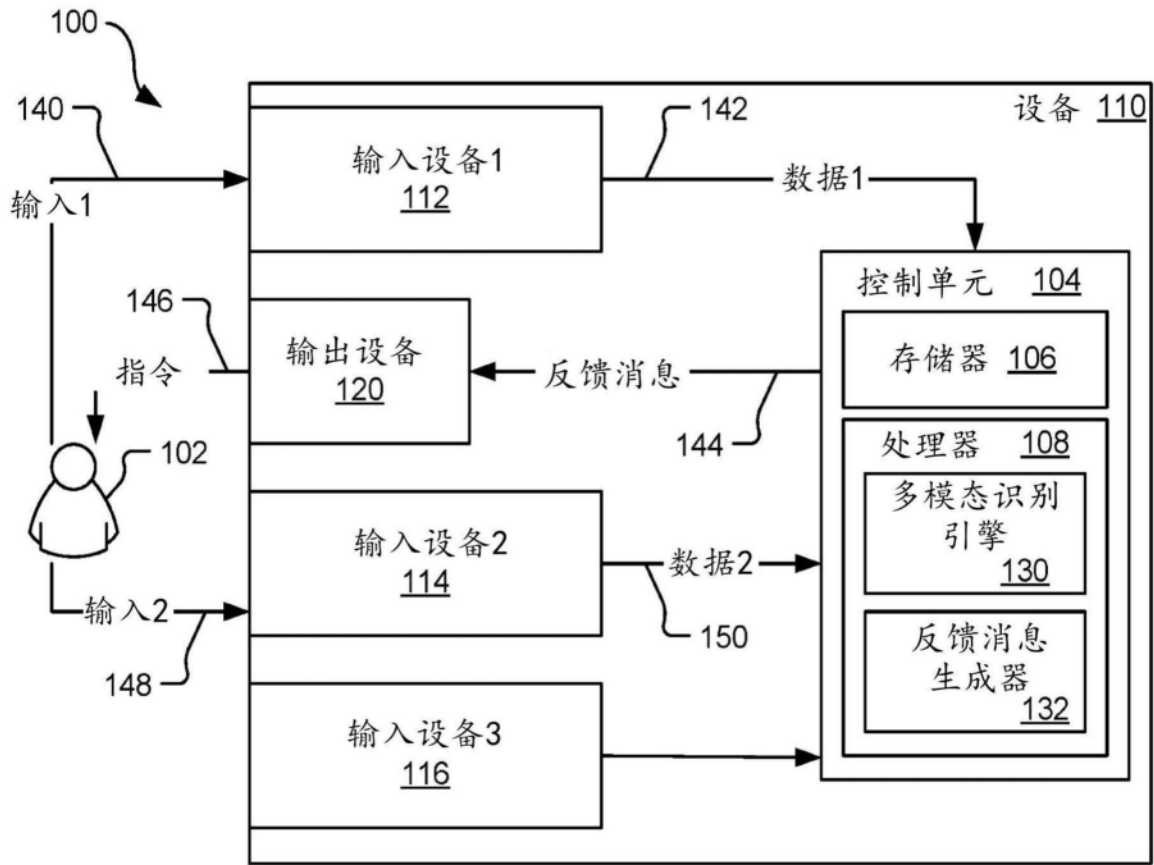


图1

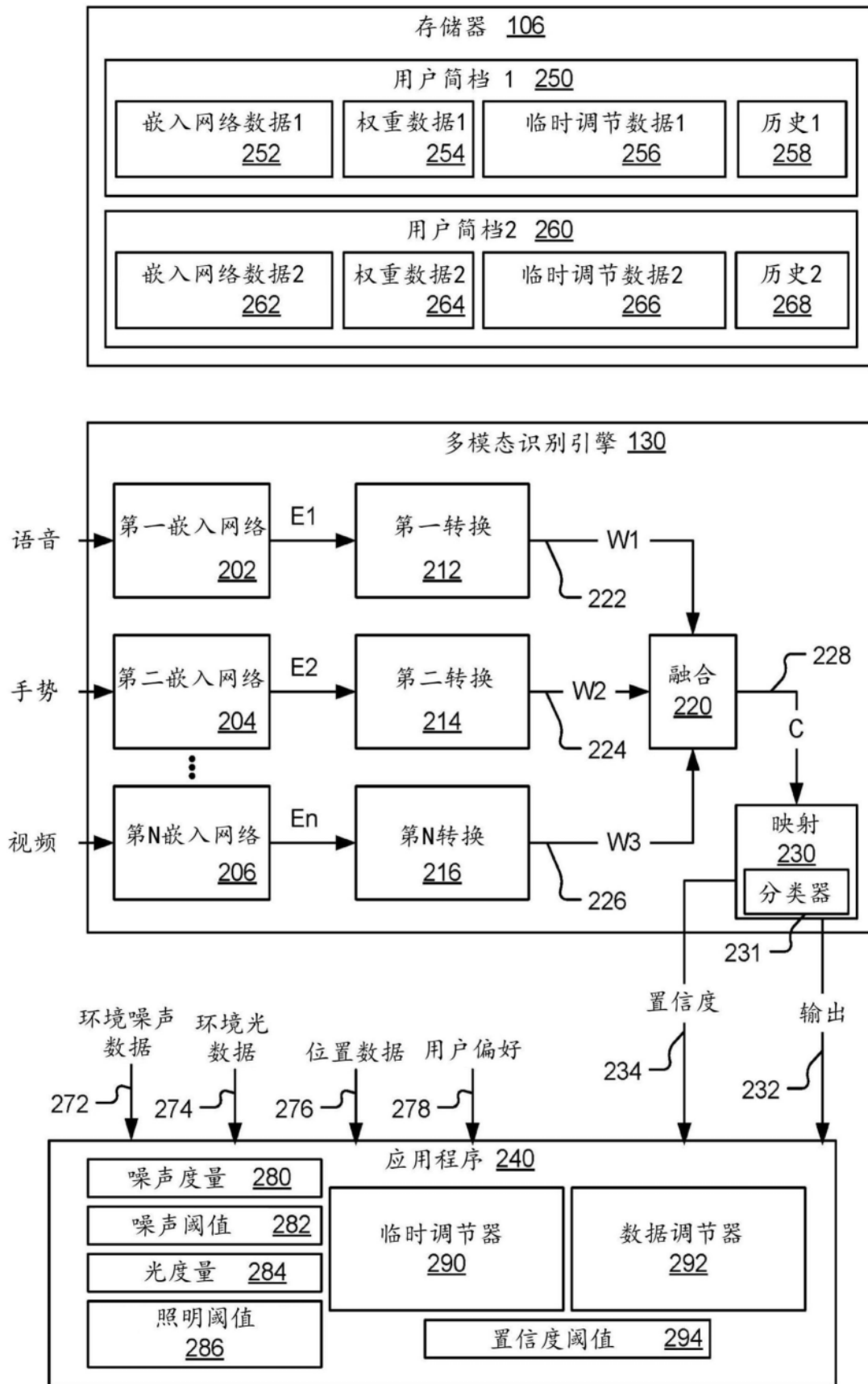


图2

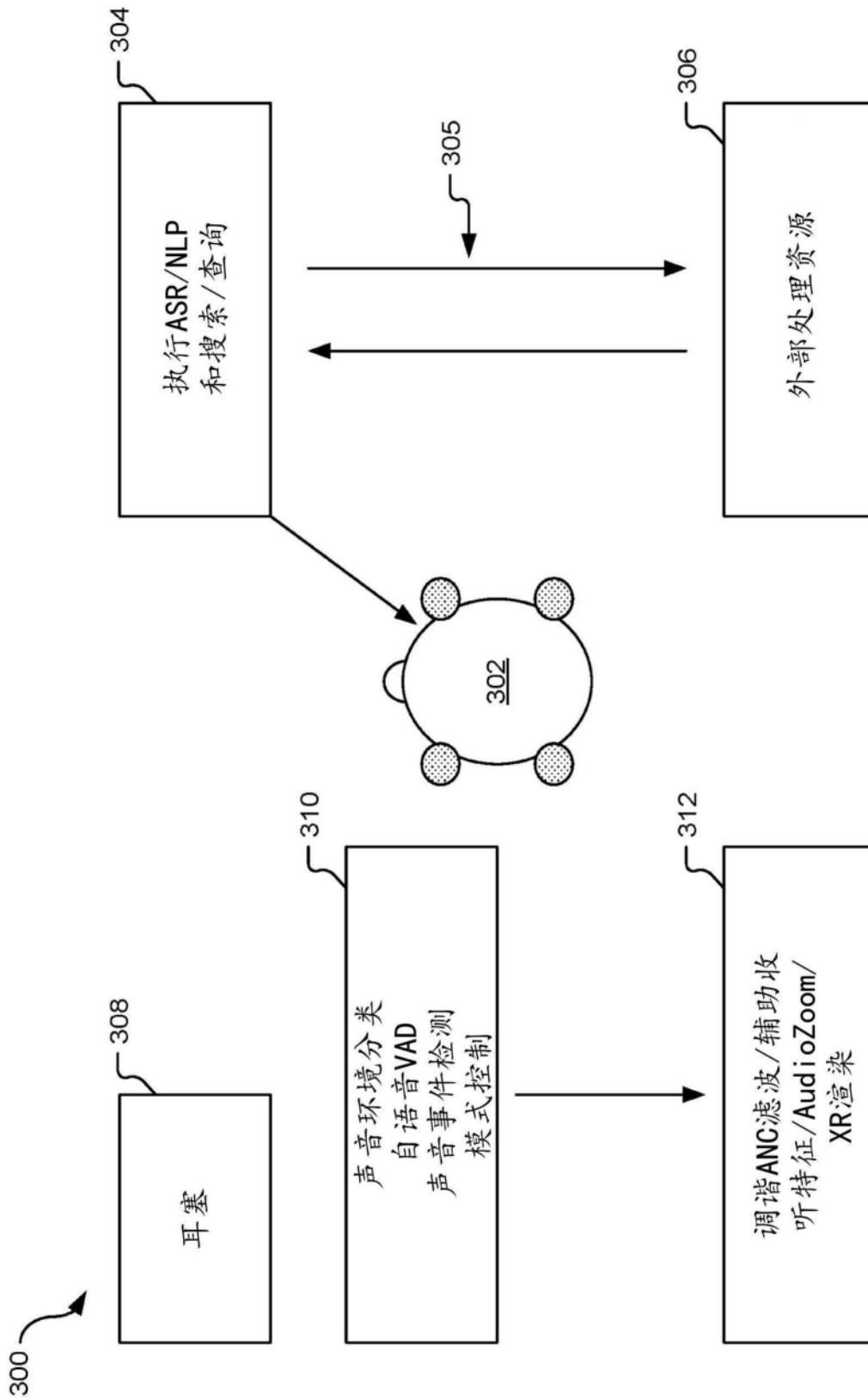


图3

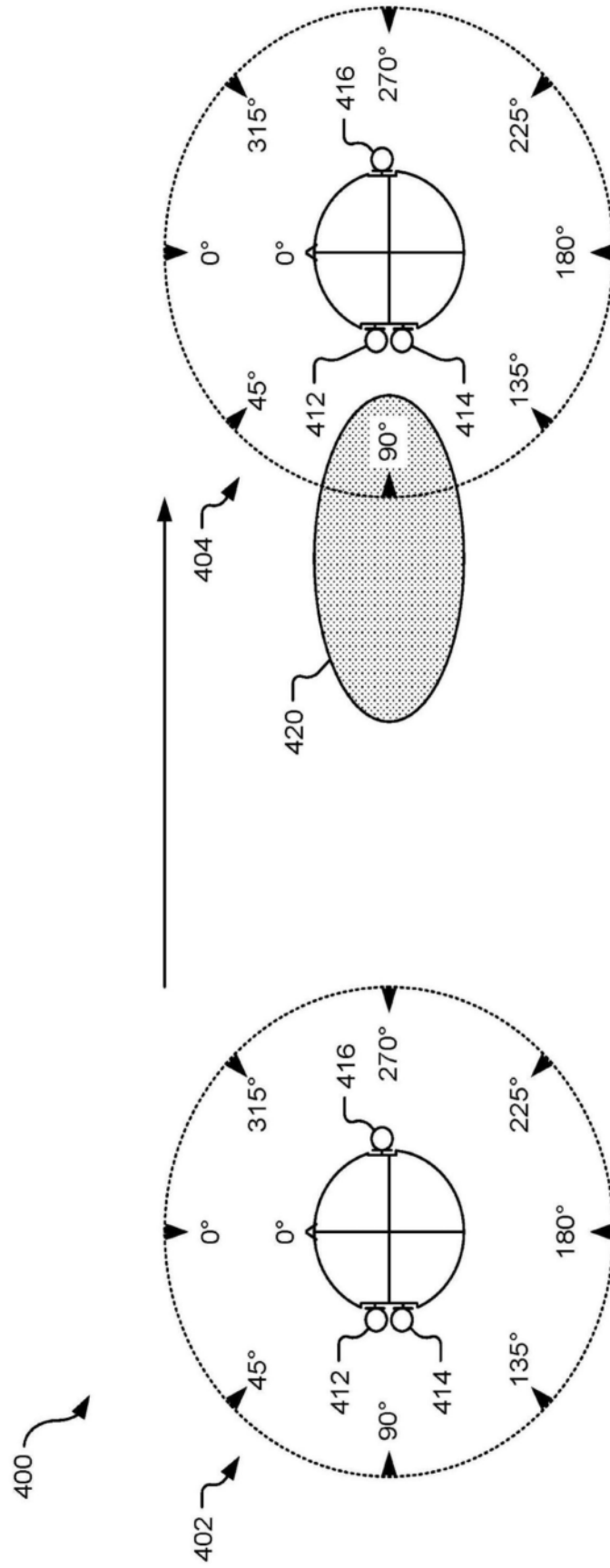


图4

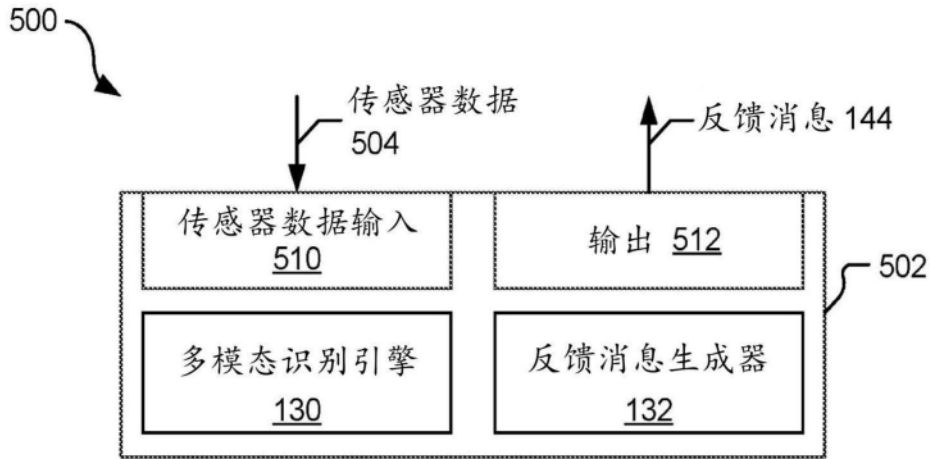


图5

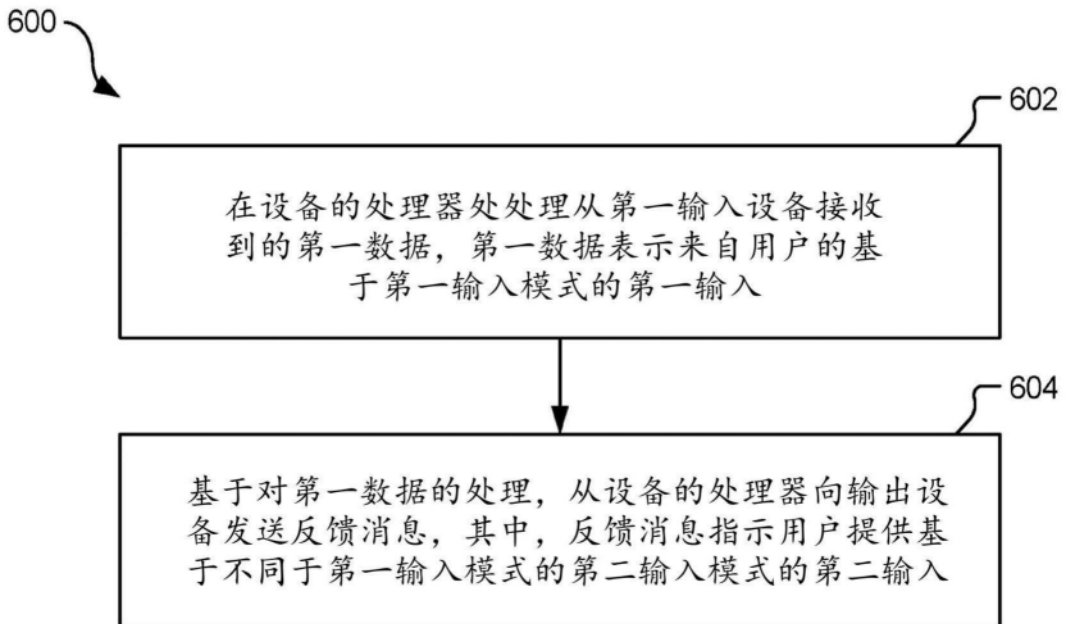


图6

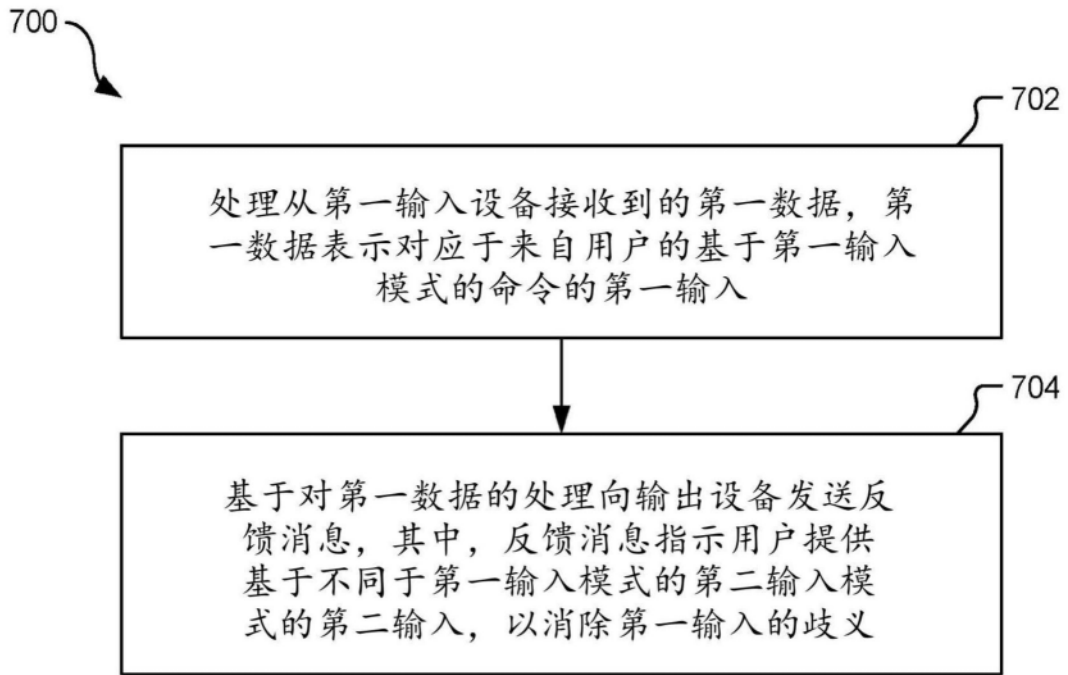


图7

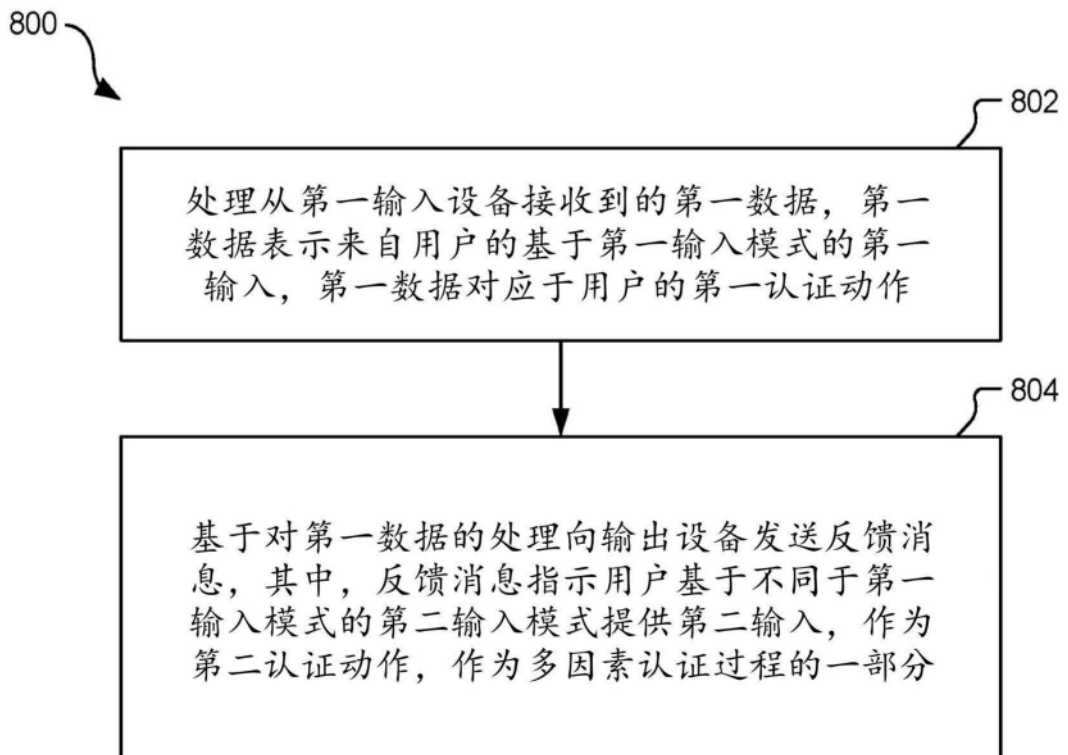


图8

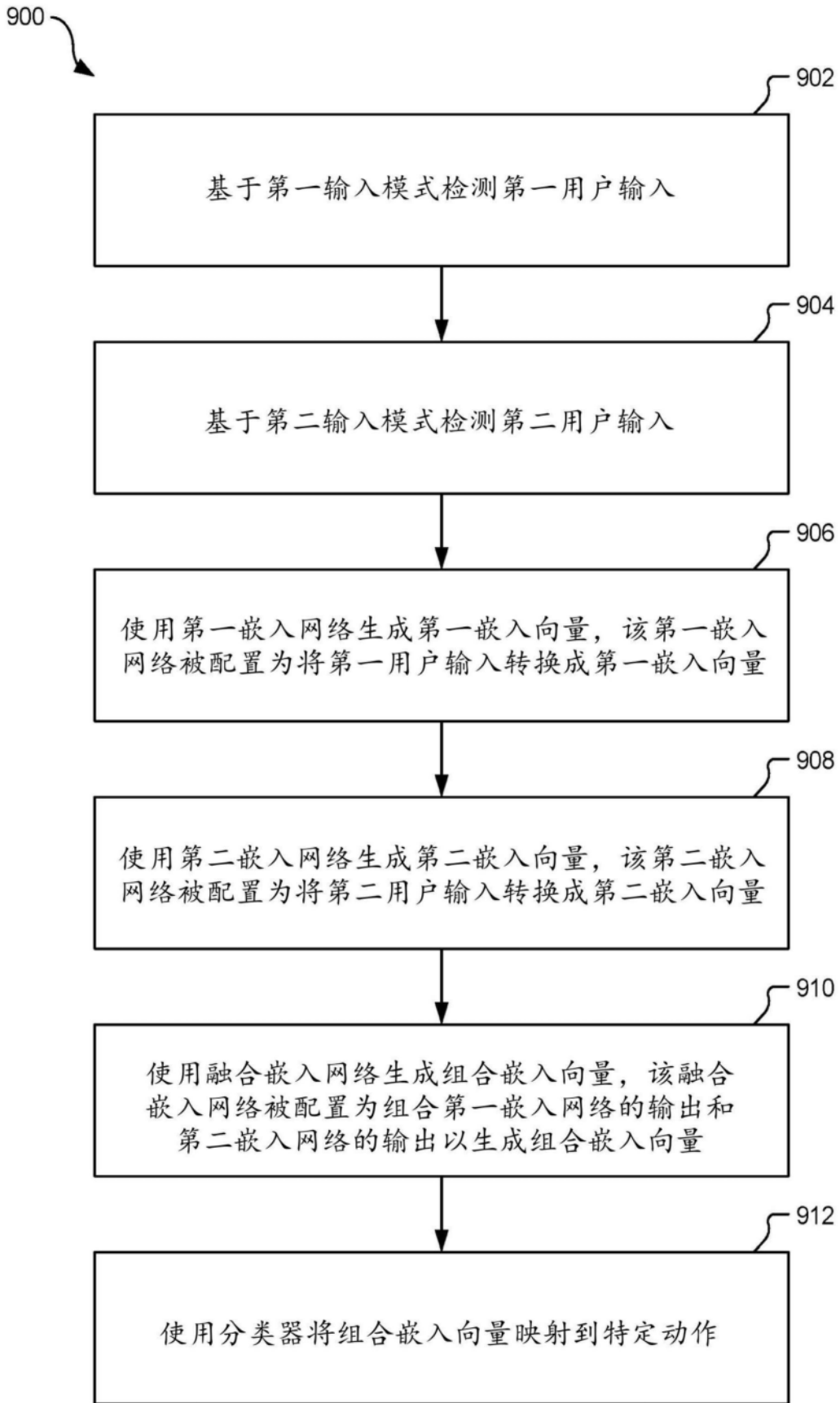


图9

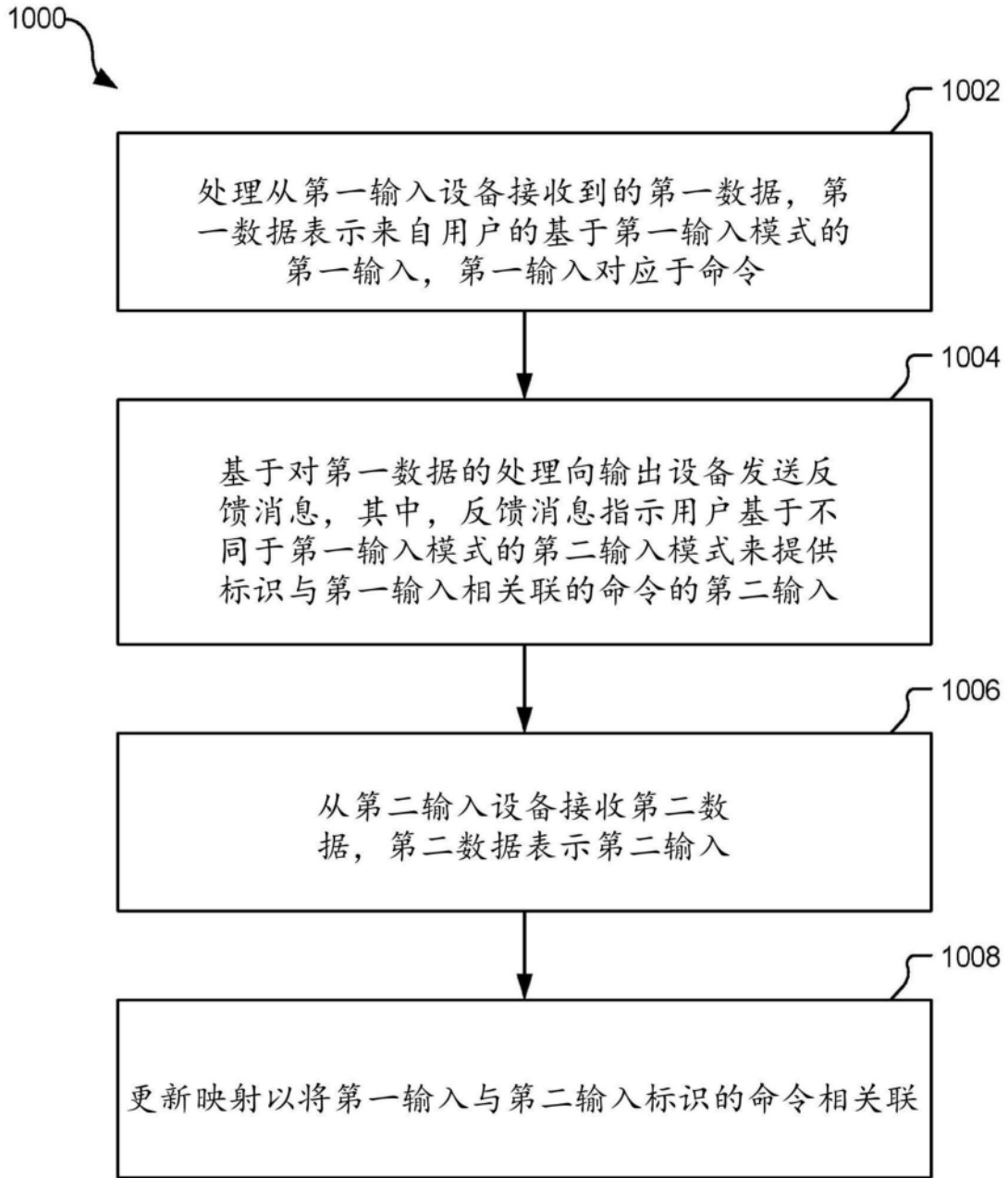


图10

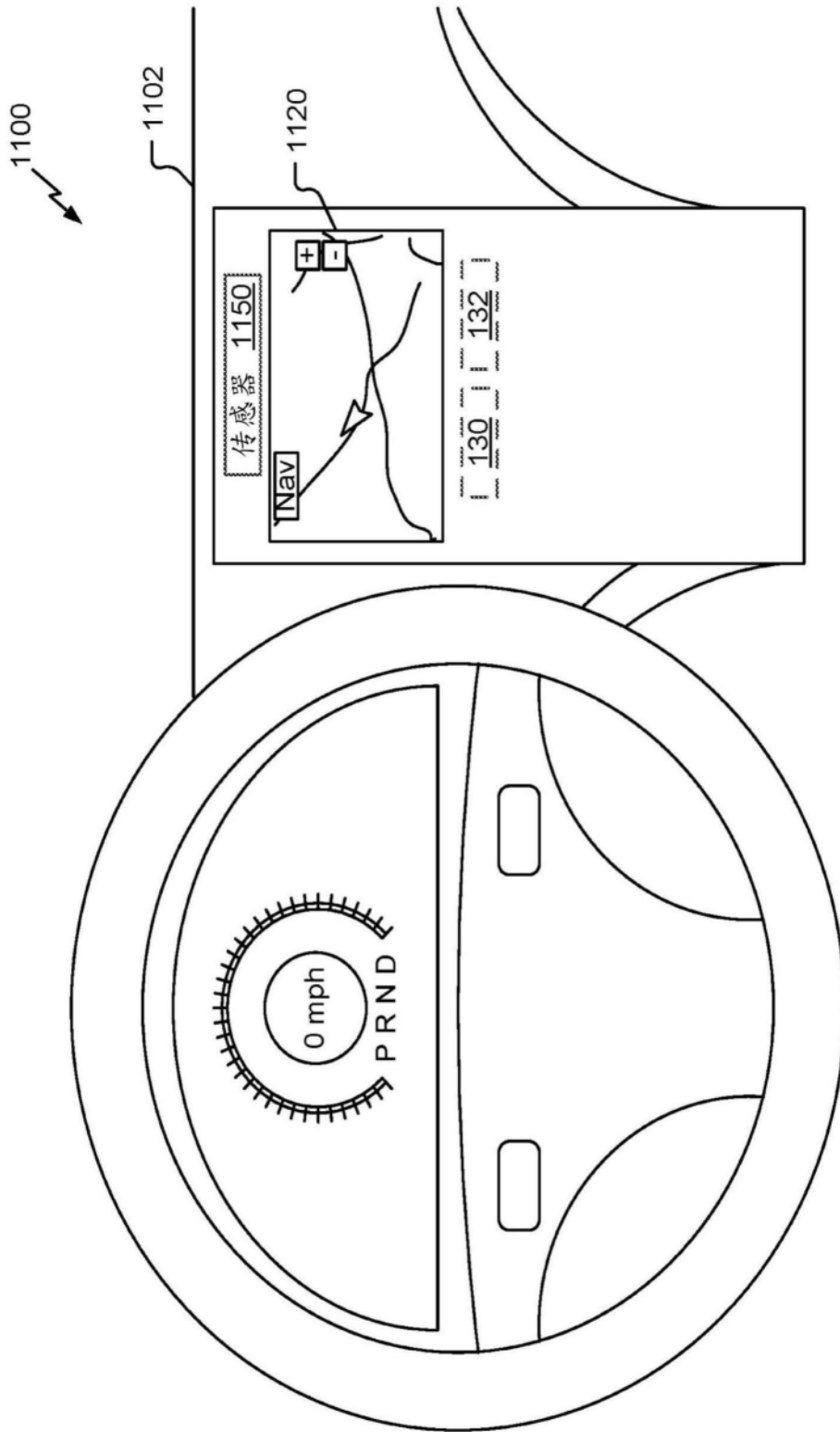


图11

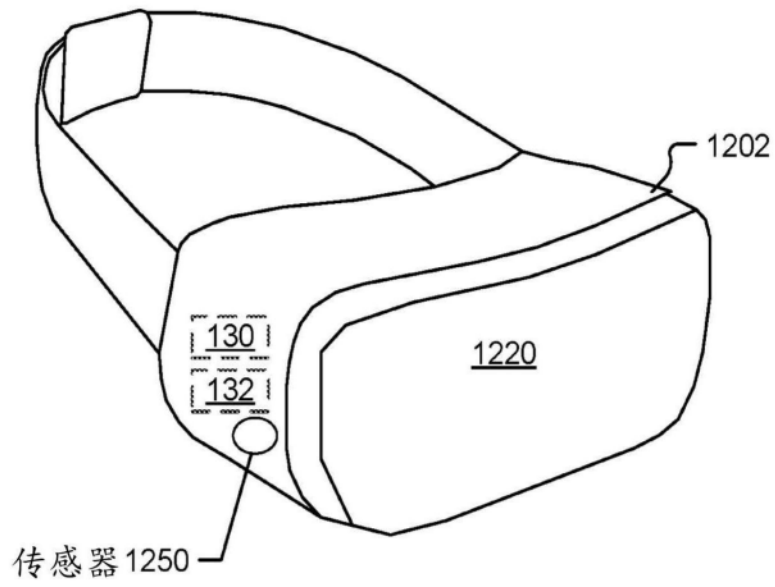


图12A

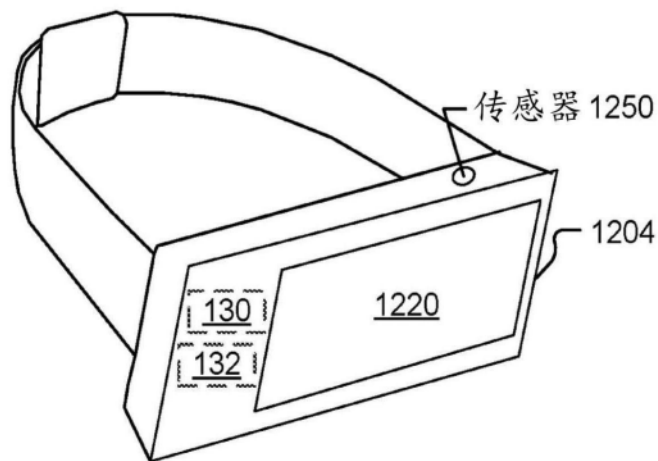


图12B

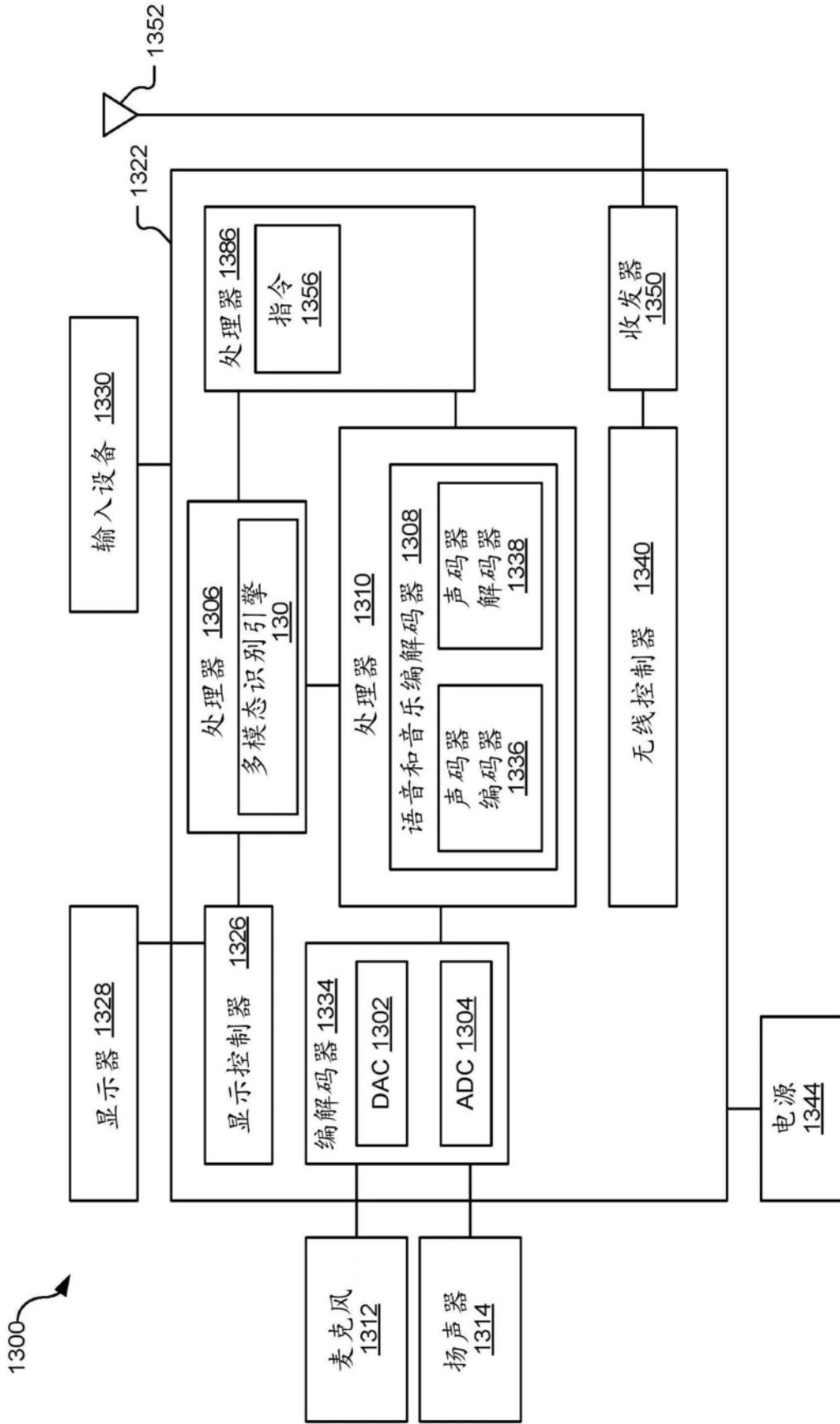


图13