(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 108604446 B
(45)授权公告日 2023.05.16

(21)申请号 201680080197.1
(22)申请日 2016.12.29
(65)同一申请的已公布的文献号
申请公布号 CN 108604446 A
(43)申请公布日 2018.09.28
(30)优先权数据
15/009,432 2016.01.28 US
(85)PCT国际申请进入国家阶段日
2018.07.26
(86)PCT国际申请的申请数据
PCT/US2016/069182 2016.12.29
(87)PCT国际申请的公布数据
WO2017/131924 EN 2017.08.03

(73)专利权人 谷歌有限责任公司
地址 美国加利福尼亚州
(72)发明人 马修·谢里菲
雅各布·尼古劳斯·弗尔斯特
(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
专利代理师 李宝泉 周亚荣

(51)Int.Cl.
G10L 13/08(2013.01)
G06F 40/289(2020.01)
G06F 40/253(2020.01)
审查员 姚黎为

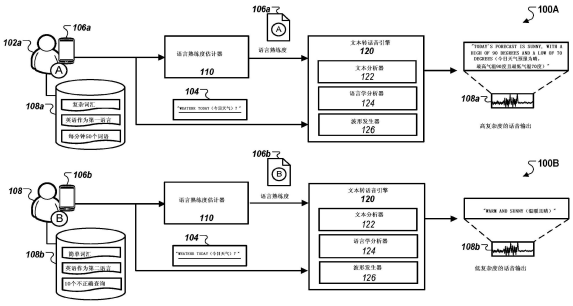
权利要求书3页 说明书15页 附图6页

(54)发明名称

自适应文本转话音输出

(57)摘要

在一些实施方式中,由一个或多个计算机确定客户端设备的用户的语言熟练度。然后,一个或多个计算机基于用户的所确定的语言熟练度来确定用于由文本转话音模块输出的文本片段。在确定用于输出的文本片段之后,一个或多个计算机生成包括文本片段的合成话语的音频数据。然后,将包括文本片段的合成话语的音频数据提供给客户端设备以用于输出。通过更好的文本转话音转换而提供改善的用户接口。



CN 108604446 B

1.一种由一个或多个计算机执行的方法,所述方法包括:

由所述一个或多个计算机确定客户端设备的用户的语言熟练度;

由所述一个或多个计算机基于所述用户的所确定的语言熟练度来确定用于由文本转话音模块输出的文本片段,其中确定所述文本片段包括:

从均具有指示不同水平的语言复杂度的语音复杂度分值的多个文本片段中选择语言复杂度分值与描述所述客户端设备的所述用户的所确定的语言熟练度的参考分值最匹配的文本片段;或者

至少基于(i)所述用户的所确定的语言熟练度和(ii)特定文本片段的复杂度分值,修改用于对所述用户的文本转话音输出的所述特定文本片段;

由所述一个或多个计算机生成包括所述文本片段的合成话语的音频数据;以及

由所述一个或多个计算机向所述客户端设备提供包括所述文本片段的所述合成话语的所述音频数据。

2.根据权利要求1所述的方法,其中,所述客户端设备显示使用文本转话音接口的移动应用。

3.根据权利要求1所述的方法,其中,确定所述用户的所述语言熟练度包括:至少基于由所述用户提交的先前查询来推断所述用户的语言熟练度。

4.根据权利要求1所述的方法,其中,确定用于由所述文本转话音模块输出的所述文本片段包括:

识别多个文本片段作为所述用户的文本转话音输出的候选,所述多个文本片段具有不同水平的语言复杂度;以及

至少基于所述客户端设备的所述用户的所确定的语言熟练度来从所述多个文本片段当中进行选择。

5.根据权利要求4所述的方法,其中,从所述多个文本片段当中进行选择包括:

针对所述多个文本片段中的每个文本片段确定语言复杂度分值;以及

选择语言复杂度分值与描述所述客户端设备的所述用户的所述语言熟练度的参考分值最匹配的文本片段。

6.根据权利要求1所述的方法,其中,确定用于由所述文本转话音模块输出的所述文本片段包括:

识别用于对所述用户的文本转话音输出的文本片段;

计算用于所述文本转话音输出的文本片段的复杂度分值;以及

至少基于所述用户的所确定的语言熟练度以及用于所述文本转话音输出的文本片段的复杂度分值,修改用于对所述用户的所述文本转话音输出的文本片段。

7.根据权利要求6所述的方法,其中,修改用于对所述用户的所述文本转话音输出的所述文本片段包括:

至少基于所述用户的所确定的语言熟练度,确定所述用户的总体复杂度分值;

确定用于对所述用户的所述文本转话音输出的所述文本片段内的个体部分的复杂度分值;

识别所述文本片段内的具有大于所述用户的所述总体复杂度分值的复杂度分值的一个或多个个体部分;以及

修改所述文本片段内的所述一个或多个个体部分，以使复杂度分值降至所述总体复杂度分值以下。

8.根据权利要求6所述的方法，其中，修改用于对所述用户的所述文本转话音输出的所述文本片段包括：

接收指示与所述用户相关联的场境的数据；

确定与所述用户相关联的所述场境的总体复杂度分值；

确定所述文本片段的所述复杂度分值超过与所述用户相关联的所述场境的总体复杂度分值；以及

修改所述文本片段，以使所述复杂度分值降至与所述用户相关联的所述场境的总体复杂度分值以下。

9.一种用于提供音频数据的系统，所述系统包括：

一个或多个计算机；以及

耦合至所述一个或多个计算机的非暂时性计算机可读介质，其上存储有指令，所述指令在由所述一个或多个计算机执行时使得所述一个或多个计算机执行操作，所述操作包括：

由所述一个或多个计算机确定客户端设备的用户的语言熟练度；

由所述一个或多个计算机基于所述用户的所确定的语言熟练度来确定用于由文本转话音模块输出的文本片段，其中确定所述文本片段包括：

从均具有指示不同水平的语言复杂度的语音复杂度分值的多个文本片段中选择语言复杂度分值与描述所述客户端设备的所述用户的所确定的语言熟练度的参考分值最匹配的文本片段；或者

至少基于(i)所述用户的所确定的语言熟练度和(ii)特定文本片段的复杂度分值，修改用于对所述用户的文本转话音输出的所述特定文本片段；

由所述一个或多个计算机生成包括所述文本片段的合成话语的音频数据；以及

由所述一个或多个计算机向所述客户端设备提供包括所述文本片段的所述合成话语的所述音频数据。

10.根据权利要求9所述的系统，其中，所述客户端设备显示使用文本转话音接口的移动应用。

11.根据权利要求9所述的系统，其中，确定所述用户的所述语言熟练度包括：至少基于由所述用户提交的先前查询来推断所述用户的语言熟练度。

12.根据权利要求9所述的系统，其中，确定用于由所述文本转话音模块输出的所述文本片段包括：

识别多个文本片段作为所述用户的文本转话音输出的候选，所述多个文本片段具有不同水平的语言复杂度；以及

至少基于所述客户端设备的所述用户的所确定的语言熟练度，从所述多个文本片段当中进行选择。

13.根据权利要求12所述的系统，其中，从所述多个文本片段当中进行选择包括：

针对所述多个文本片段中的每个文本片段确定语言复杂度分值；以及

选择语言复杂度分值与描述所述客户端设备的所述用户的所述语言熟练度的参考分

值最匹配的文本片段。

14.根据权利要求9所述的系统，其中，确定用于由所述文本转话音模块输出的所述文本片段包括：

识别用于对所述用户的文本转话音输出的文本片段；

计算用于所述文本转话音输出的文本片段的复杂度分值；以及

至少基于所述用户的所确定的语言熟练度以及用于所述文本转话音输出的文本片段的复杂度分值，修改用于对所述用户的所述文本转话音输出的文本片段。

15.根据权利要求14所述的系统，其中，修改用于对所述用户的文本转话音输出的所述文本片段包括：

至少基于所述用户的所确定的语言熟练度，确定所述用户的总体复杂度分值；

确定用于对所述用户的所述文本转话音输出的所述文本片段内的个体部分的复杂度分值；

识别所述文本片段内的具有大于所述用户的所述总体复杂度分值的复杂度分值的一个或多个个体部分；以及

修改所述文本片段内的所述一个或多个个体部分，以使复杂度分值降至所述总体复杂度分值以下。

16.一种用于提供音频数据的系统，所述系统包括一个或多个处理设备以及存储指令的一个或多个机器可读存储设备，所述指令在由所述一个或多个处理设备执行时使得所述系统执行根据权利要求1至8中的任一项所述的方法。

17.一种存储指令的一个或多个机器可读存储设备，所述指令在由一个或多个处理设备执行时使得所述一个或多个处理设备执行根据权利要求1至8中的任一项所述的方法。

自适应文本转话音输出

[0001] 相关申请的交叉引用

[0002] 本申请要求于2016年1月28日提交、名称为“ADAPTIVE TEXT-TO-SPEECH-OUTPUTS(自适应文本到话音输出)”的美国申请序列号15/009,432的优先权,其全部内容通过引用并入本文。

技术领域

[0003] 本说明书总体上涉及电子通信。

背景技术

[0004] 话音合成指代人工制作人类话音。话音合成器能够在软件或硬件组件中实现,以生成对应于文本的话音输出。譬如,文本转话音(TTS)系统通常通过串接数据库中存储的录制话音段而将正常的语言文本转换成话音。

发明内容

[0005] 随着电子计算中的更大部分已经从桌面转移到移动环境,话音合成对于用户体验而言变得愈加重要。例如,不带显示器的较小移动设备的使用率增长导致了使用文本转话音(TTS)系统来访问和使用移动设备上显示的内容日益增多。

[0006] 本说明书公开了改进的用户接口,特别是通过改进的TTS来增强计算机到用户的通信。

[0007] 现有TTS系统的一个特殊问题是,这样的系统往往无法适应不同用户的不同语言熟练度。这种灵活性欠缺往往会妨碍语言熟练度有限的用户理解复杂的文本转话音输出。譬如,使用TTS系统的非母语说话者可能由于他们的语言熟练度有限而难以理解文本转话音输出。现有TTS系统的另一个问题是,用户理解文本转话音输出的即时能力也可能随特定的用户场境(context)而变化。譬如,一些用户场境包括背景噪声,这可能会使其愈加难以理解更长或更复杂的文本转话音输出。

[0008] 在一些实施方式中,系统基于用户的语言熟练度来调整用于文本转话音输出的文本,以提高用户能够领会文本转话音输出的可能性。譬如,能够从在先用户活动推断用户的语言熟练度,并且将其用于将文本转话音输出调整成与用户的语言熟练度相称的适当复杂度。在一些示例中,系统获得多个候选文本片段,它们对应于不同水平的语言熟练度。然后,该系统选择与用户的语言熟练度最匹配且最密切对应的候选文本片段,并且提供所选文本片段的合成话语以供输出给用户。在其他示例中,系统在生成文本转话音输出之前将文本片段中的文本更改成更好地对应于用户的语言熟练度。能够调整文本片段的各个方面,包括词汇表、语句结构、长度等。然后,该系统提供更改后的文本片段的合成话语以供输出给用户。

[0009] 针对本文讨论的系统收集关于用户的个人信息或者可以利用个人信息的情形,可以向用户提供机会来控制程序或特征是否收集个人信息(例如,关于用户社交网络、社交动

作或活动、职业、用户偏好、或用户当前位置的信息)，或者控制是否和/或如何从内容服务器接收可能与用户更相关的内容。此外，在存储或使用某些数据之前可以通过一种或多种方式将其匿名化，使得移除个人可识别信息。例如，用户的身份可以被匿名化，使得无法确定用户的个人可识别信息，或者在获得位置信息的情况下可以将用户的地理位置泛化(诸如到城市、邮政编码或州层级)，使得无法确定用户的特定位置。因此，用户可以控制如何收集关于他或她的信息以及如何供内容服务器使用该信息。

[0010] 在一方面，一种计算机实现的方法能够包括：由一个或多个计算机确定客户端设备的用户的语言熟练度；由所述一个或多个计算机基于所述用户的所确定的语言熟练度来确定供文本转话音模块输出的文本片段；由所述一个或多个计算机生成包括所述文本片段的合成话语的音频数据；以及由所述一个或多个计算机向所述客户端设备提供包括所述文本片段的所述合成话语的所述音频数据。

[0011] 其他版本包括对应的系统和计算机程序，其被配置成执行计算机存储设备上编码的方法的动作。

[0012] 一个或多个实施方式能够包括以下可选特征。例如，在一些实施方式中，所述客户端设备显示使用文本转话音接口的移动应用。

[0013] 在一些实施方式中，确定所述用户的所述语言熟练度包括：至少基于由所述用户提交的先前查询来推断所述用户的所述语言熟练度。

[0014] 在一些实施方式中，确定供所述文本转话音模块输出的所述文本片段包括：识别多个文本片段作为所述用户的文本转话音输出的候选，所述多个文本片段具有不同水平的语言复杂度；以及至少基于所述客户端设备的所述用户的所确定的语言熟练度，从所述多个文本片段当中进行选择。

[0015] 在一些实施方式中，从所述多个文本片段当中进行选择包括：针对所述多个文本片段中的每个文本片段确定语言复杂度分值；以及选择语言复杂度分值与描述所述客户端设备的所述用户的所述语言熟练度的参考分值最匹配的文本片段。

[0016] 在一些实施方式中，确定供所述文本转话音模块输出的所述文本片段包括：识别用于对所述用户的文本转话音输出的文本片段；计算用于所述文本转话音输出的文本片段的复杂度分值；以及至少基于所述用户的所确定的语言熟练度以及用于所述文本转话音输出的文本片段的复杂度分值，修改用于对所述用户的所述文本转话音输出的文本片段。

[0017] 在一些实施方式中，修改用于对所述用户的文本转话音输出的所述文本片段包括：至少基于所述用户的所确定的语言熟练度，确定所述用户的总体复杂度分值；确定对所述用户的所述文本转话音输出的所述文本片段内的个体部分的复杂度分值；识别所述文本片段内的具有大于所述用户的所述总体复杂度分值的复杂度分值的一个或多个个体部分；以及修改所述文本片段内的所述一个或多个个体部分，以使复杂度分值降至所述总体复杂度分值以下。

[0018] 在一些实施方式中，修改用于对所述用户的所述文本转话音输出的所述文本片段包括：接收指示与所述用户相关联的场境的数据；确定与所述用户相关联的所述场境的总体复杂度分值；确定所述文本片段的所述复杂度分值超过与所述用户相关联的所述场境的总体复杂度分值；以及修改所述文本片段，以使所述复杂度分值降至与所述用户相关联的所述场境的总体复杂度分值以下。

[0019] 在另一方面，一种计算机程序包括机器可读指令，所述机器可读指令在由计算装置执行时使得其执行上述方法中的任一个。

[0020] 在另一总体方面，一种计算机实现的方法包括：接收指示与用户相关联的场境的数据；确定与所述用户相关联的所述场境的总体复杂度分值；识别用于对所述用户的文本转话音输出的文本片段；确定所述文本片段的复杂度分值超过与所述用户相关联的所述场境的所述总体复杂度分值；以及修改所述文本片段，以使所述复杂度分值降至与所述用户相关联的所述场境的所述总体复杂度分值以下。

[0021] 在一些实施方式中，确定与所述用户相关联的所述场境的所述总体复杂度分值包括：当所述用户被确定处于所述场境中时，识别由所述用户先前提交的查询内所包括的词项；以及至少基于所识别的词项来确定与所述用户相关联的所述场境的总体复杂度分值。

[0022] 在一些实施方式中，指示与所述用户相关联的所述场境的所述数据包括由所述用户先前提交的查询。

[0023] 在一些实施方式中，指示与所述用户相关联的所述场境的所述数据包括指示与所述用户相关联的当前位置的GPS信号。

[0024] 在附图和以下描述中阐述了一个或多个实施方式的细节。参阅说明书、附图、和权利要求书，其他可能特征和优势将显而易见。

[0025] 这些方面的其他实施方式包括对应的系统、装置和计算机程序，它们被配置成执行计算机存储设备上编码的方法的动作。

附图说明

[0026] 图1是图示出用于基于语言熟练度来生成文本转话音输出的过程的示例的示图。

[0027] 图2是图示出用于基于用户场境来生成自适应文本转话音输出的系统的示例的示图。

[0028] 图3是图示出用于修改文本转话音输出内的语句结构的系统的示例的示图。

[0029] 图4是图示出用于基于使用聚类技术来生成自适应文本转话音输出的系统的示例的框图。

[0030] 图5是图示出用于生成自适应文本转话音输出的过程的示例的流程图。

[0031] 图6是能够实现本文所述过程的计算设备或其部分的框图。

[0032] 在附图中，相似的附图标记在各附图中表示对应的部分。

具体实施方式

[0033] 图1是图示出用于基于语言熟练度来生成文本转话音输出的过程100A和100B的示例的示图。过程100A和100B被使用于分别针对文本查询104为具有高语言熟练度的用户102a和具有低语言熟练度的用户102b生成不同的文本转话音输出。如图所示，在用户设备106a和106b上接收到查询104之后，过程100A为用户102a生成高复杂度的文本转话音输出108a，而过程100B为用户102b生成低复杂度的输出108b。此外，执行过程100A和100B的TTS系统能够包括语言熟练度估计器110、文本转话音引擎120。此外，文本转话音引擎120能够进一步包括文本分析器122、语言学分析器124和波形发生器126。

[0034] 一般而言，根据用户的语言熟练度，能够确定用于生成文本转话音输出的文本的

内容。作为补充或者作为替选，用于生成文本转话音输出的文本能够基于用户场境来确定，用户场境例如，用户的位置或活动、存在的背景噪声、用户的当前任务等。另外，通过使用其他信息——诸如用户未能完成任务或正重复动作的指示，可以调整或确定要转换成可听形式的文本。

[0035] 在本示例中，两个用户——用户102a和用户102b——分别在用户设备106a和106b上提供相同的查询104，作为对应用、web页面或其他搜索功能的输入。譬如，查询104能够是发送到用户设备106a和106b以确定当天的天气预报的语音查询。然后，查询104被传送到文本转话音引擎120，以响应于查询104而生成文本转话音输出。

[0036] 语言熟练度估计器110能够是TTS系统内的软件模块，其基于用户数据108a来确定与特定用户(例如，用户102a或用户102b)相关联的语言熟练度分值。语言熟练度分值能够是对用户理解特定语言的通信——特别是理解特定语言的话音的能力的估计。语言熟练度的一个量度是用户成功完成语音控制任务的能力。诸如设置预约日程表约会、查找路线指引等许多类型的任务遵循用户与设备交换口语交流的一系列交互。用户通过语音接口成功完成这些任务工作流的比率是用户语言熟练度的显著指示符。例如，完成用户发起的十个语音任务中的九个语音任务的用户可能具有高语言熟练度。另一方面，未能完成用户发起的大多数语音任务的用户可能被推断为具有低语言熟练度，因为该用户可能未完全理解来自设备的通信或者可能无法提供适当的口语响应。如下文进一步讨论，当用户未完成包括标准TTS输出的工作流时，导致低语言熟练度分值，TTS可以使用适应的简化输出，其可以提高用户理解和完成各种任务的能力。

[0037] 如图所示，用户数据108a能够包括在用户提交的在先文本查询内所用的词语、英语或者TTS系统利用的任何其他语言是否是用户的母语的指示、以及反映用户语言理解技能的一组活动和/或行为。例如，如图1所示，用户的键入速度能够被使用于确定用户对语言的语言流利性。此外，基于将预定的复杂度与用户在先前文本查询中使用的词语相关联，能够将语言词汇复杂度分值或语言熟练度分值指派给用户。在另一个示例中，在先查询中的误辨识词语的数目也能够被使用于确定语言熟练度分值。譬如，大量误辨识词语能够被使用于指示低语言熟练度。在一些实施方式中，通过查找与用户相关联的存储分值来确定语言熟练度分值，该存储分值是在提交查询104之前针对用户所确定的。

[0038] 虽然图1将语言熟练度估计器110描绘为与TTS引擎120分开的组件，但在一些实施方式中，如图2所示，语言熟练度估计器110能够是TTS引擎120内的集成软件模块。在这样的情况下，涉及语言熟练度估计的操作能够由TTS引擎120直接调控。

[0039] 在一些实施方式中，指派给用户的语言熟练度分值可以是基于针对用户估计的特定用户场境。譬如，如更具体地参照图2所述，用户场境确定能够被使用于确定能够使得用户暂时具有受限语言理解能力的场境特定的语言熟练度。例如，如果用户场境指示显著的背景噪声或者如果用户参与诸如驾驶的任务，则语言熟练度分值能够被使用于指示用户的目前语言理解能力相对于其他用户场境暂时下降。

[0040] 在一些实施方式中，替代基于先前用户活动来推断语言熟练度，能够在不使用语言熟练度估计器110的情况下将语言熟练度分值直接提供给TTS引擎120。譬如，在指定用户的语言熟练度水平的登记过程期间，能够基于用户输入来对用户标定语言熟练度分值。例如，在登记期间，用户能够提供指定用户技能水平的选择，然后能够使用该选择来计算用户

的适当语言熟练度。在其他示例中，用户能够提供其他类型的信息，诸如群体特征信息、教育水平、居住地等，所述信息能够被使用于指定用户的语言熟练度水平。

[0041] 在上述示例中，语言熟练度分值能够是基于最近生成的用户活动数据来周期性调整的离散值集合，或者是在登记过程期间初步标定的连续分值。在第一种情况下，语言熟练度分值的值能够基于指示用户的目前语言理解力和熟练度可能变弱的一个或多个因素(例如，指示显著背景噪声的用户场境)来偏置。在第二种情况下，语言熟练度分值的值能够在初步计算之后预设并且仅在指示用户的语言熟练度已提高的特定标志性事件(milestone events)(例如，给定语言的键入速率升高或者纠正比率降低)之后调整。在其他情况下，这两种技术的组合能够被使用于基于特定文本输入来可变地调整文本转话音输出。在这样的情况下，各自表示用户语言技能的特定方面的多个语言熟练度分值能够被使用于确定如何针对用户最佳地调整文本转话音输出。例如，一个语言熟练度分值能够表示用户词汇的复杂度，而另一个语言熟练度分值能够被使用于表示用户的语法技能。

[0042] TTS引擎120能够使用语言熟练度分值来生成适应于用户的语言熟练度分值所指示的语言熟练度的文本转话音输出。在一些情况下，TTS引擎120基于从针对文本查询104的候选TTS字符串集合中选择特定TTS字符串来适配文本转话音输出。在这样的情况下，TTS引擎120基于使用用户的语言熟练度分值来预测用户会准确领会候选TTS字符串中的每个TTS字符串的可能性，选择特定TTS字符串。参照图2提供了有关这些技术的更加具体的描述。替选地，在其他情况下，TTS引擎120能够选择基线TTS字符串并且基于用户的语言熟练度分值所指示的语言熟练度来调整TTS字符串的结构。在这样的情况下，TTS引擎120能够调整基线TTS字符串的语法、提供词语替换和/或降低语句复杂度以生成用户更有可能理解的适应TTS字符串。参照图3提供了有关这些技术的更加具体的描述。

[0043] 仍参照图1，TTS引擎120可以针对用户102a和102b生成不同的文本转话音输出，因为所述用户的语言熟练度分值不同。例如，在过程100A中，语言熟练度分值106a指示高英语语言熟练度，其是从指示以下的用户数据108a推断得出的：用户102a具有复杂词汇，以英语作为第一语言，以及在先用户查询中每分钟词语相对高。基于语言熟练度分值106a的值，TTS引擎120生成包括复杂语法结构的高复杂度的文本转话音输出108a。如图所示，文本转话音输出108a包括描述今日天气预报为晴的独立分句，此外还包括关于当日高温和低温的附加信息的从属分句。

[0044] 在过程100B的示例中，语言熟练度分值106b指示低英语语言熟练度，其是从指示以下的用户活动数据108b推断得出的：用户102b具有简单词汇，以英语作为第二语言，以及先前曾提供十个不正确的查询。在该示例中，TTS引擎120生成低复杂度的文本转话音输出108b，其包括相对于文本转话音输出108a更简单的语法结构。譬如，替代单个语句内包括多个分句，文本转话音输出108b包括单个独立分句，其传达与文本转话音输出108a相同的主要信息(例如，今日天气预报为晴)，但不包括有关当日高温和低温的附加信息。

[0045] 能够通过各种不同设备和软件模块来执行用于TTS输出的文本适配。例如，服务器系统的TTS引擎可以包括这样的功能：基于语言熟练度分值来调整文本并然后输出包括经调整的文本的合成话语的音频。另举一例，服务器系统的预处理模块可以调整文本并且将经调整的文本传递到TTS引擎以供进行话音合成。另举一例，用户设备可以包括TTS引擎或者TTS引擎和文本预处理器，以能够生成适当的TTS输出。

[0046] 在一些实施方式中，TTS系统能够包括配置成与客户端设备的第三方移动应用或web页面交换通信的软件模块。譬如，该系统的TTS功能能够通过应用包接口(API)来对第三方移动应用可用。API能够包括已定义的一组协议，应用或web站点能够使用这些协议从运行TTS引擎120的服务器系统请求TTS音频。在一些实施方式中，API能够使在用户设备上本地运行的TTS功能可用。例如，API可以通过进程间通信(IPC)、远程过程调用(RPC)、或者其他系统调用或函数而可用于应用或web页面。TTS引擎以及相关联的语言熟练度分析或文本预处理可以在用户设备上本地运行，以针对用户语言熟练度确定适当文本并且还生成合成话音的音频。

[0047] 例如，第三方应用或web页面能够使用API来生成基于第三方应用或web页面的语音接口的任务流而提供给用户的语音指令集。API能够指定应用或web页面应提供要转换成话音的文本。在一些情况下，能够提供其他信息，诸如用户标识符或语言熟练度分值。

[0048] 在TTS引擎120通过使用API与第三方应用交换通信的实施方式中，TTS引擎120能够被使用于确定是否应在生成该文本的文本转话音输出之前调整来自第三方应用的文本片段。例如，API能够包括计算机实现的协议，所述协议指定第三方应用内发起自适应文本转话音输出的生成的条件。

[0049] 作为一个示例，一个API可以许可应用提交多个不同的文本片段作为TTS输出的候选，其中不同的文本片段对应于不同的语言熟练度水平。例如，候选能够是具有等同含义但复杂度水平不同的文本片段(例如，高复杂度响应、中复杂度响应、和低复杂度响应)。然后，TTS引擎120可以确定理解每个候选所需的语言熟练度，确定用户的适当语言熟练度分值，并且选择与该语言熟练度分值最佳对应的候选文本。然后，TTS引擎120例如通过使用API的网络将所选文本的合成音频提供回给应用。在一些情况下，API能够在用户设备106a和106b上本地可用。在这样的情况下，API能够通过各种类型的进程间通信(IPC)或者经由系统调用而可访问。例如，用户设备106a和106b上的API的输出能够是TTS引擎120的文本转话音输出，因为API在用户设备106a和106b上本地操作。

[0050] 在另一个示例中，API能够允许第三方应用提供单个文本片段以及指示是否许可TTS引擎120修改文本片段以生成具有不同复杂度的文本片段的值。如果App或web页面指示许可更改，则TTS系统120可以对文本作出各种变化，例如，当语言熟练度分值表明原始文本比用户能够在口述响应中理解的更加复杂时，降低文本的复杂度。在其他示例中，API允许第三方应用还提供用户数据(例如，在第三方应用上提交的在先用户查询)连同文本片段，使得TTS引擎120能够确定与用户相关联的用户场境并且基于所确定的用户场境来调整生成特定的文本转话音输出。类似地，API能够允许应用提供来自用户设备的场境数据(例如，全球定位信号、加速计数据、环境噪声水平等)或者用户场境的指示，以允许TTS引擎120调整最终将通过第三方应用提供给用户的文本转话音输出。在一些情况下，第三方应用也能够向API提供能够用来确定用户的语言熟练度的数据。

[0051] 在一些实施方式中，TTS引擎120能够在不使用用户的语言熟练度或者没有确定与用户相关联的场境情况下调整针对用户查询的文本转话音输出。在这样的实施方式中，TTS引擎120能够基于接收到用户已误解输出(例如，多次重试相同的查询或任务)的信号来确定初始文本转话音输出对于用户而言过于复杂。作为响应，TTS引擎120能够降低重试的查询或相关查询的后续文本转话音响应的复杂度。因此，当用户未能成功完成动作时，TTS引

擎120可以逐步降低理解TTS输出所需的细节量或语言熟练度，直到其达到用户理解的水平。

[0052] 图2是图示出基于用户场境来自适应地生成文本转话音输出的系统200的示例的示图。简而言之，系统200能够包括TTS引擎210，其包括查询分析器211、语言熟练度估计器212、插值器213、语言学分析器214、重排名器215、和波形发生器216。系统200也包括存储一组场境简档232的场境储存库220以及存储用户历史数据234的用户历史管理器230。在一些情况下，TTS引擎210对应于如参照图1所述的TTS引擎120。

[0053] 在本示例中，用户202在用户设备208上初步提交查询204，该查询包括与用户当天的第一个约见相关的信息的请求。然后，用户设备208能够将查询204和与用户202相关联的场境数据206分别传送到查询分析器211和语言熟练度估计器212。可以使用相同的技术来适配并非对查询的响应的其他类型的TTS输出，例如，日历提醒、通知、任务工作流等。

[0054] 场境数据206能够包括关于与用户202相关联的特定场境的信息，诸如重复文本查询之间的时间间隔、指示与用户202相关联的位置、速度或运动模式的全球定位信号(GPS)数据、在特定时间段内提交到TTS引擎210的在先文本查询、或者能够指示与TTS引擎210相关的用户活动的其他类型的背景信息。在一些情况下，场境数据206能够指示提交到TTS引擎210的查询204的类型，诸如查询204是与用户动作相关联的文本片段还是传送到TTS引擎210以生成文本转话音输出的指令。

[0055] 在接收到查询204之后，查询分析器211解析查询204以识别响应于查询204的信息。例如，在查询204是语音查询的一些情况下，查询分析器211初步生成语音查询的转录，并且然后处理查询204内的各个词语或片段以确定响应于查询204的信息，例如，通过将查询提供给搜索引擎并且接收搜索结果来确定。然后，查询204的转录和所识别的信息能够被传送到语言学分析器214。

[0056] 现描述语言熟练度估计器212，在接收到场境数据206之后，通过使用参照图1所述的技术，语言熟练度估计器212基于所接收的场境数据206来计算用户202的语言熟练度。特别地，语言熟练度估计器212解析存储在储存库220上的各个场境简档232。场境简档232可以是归档库，其包括与特定用户场境相关联并且能够被包括在文本转话音输出中的相关类型的信息。场境简档232附加地指定与每种类型的信息相关联的值，其表示当用户202目前处于与场境简档232相关联的场境内时用户202可能理解每种类型的信息的程度。

[0057] 在图2所示的示例中，场境简档232指定用户202目前处于指示该用户202处于每日通勤上下班的场境中。此外，场境简档232也指定用户202有可能领会的各个词语和短语的值。譬如，数据或时间信息与“SINCE”的值“0.9”相关联，指示用户202更有可能理解与约见相关联的广义信息(例如，下一个即将发生的约见的时间)204，而非与约见相关联的详细信息(例如，参加约见的参与方或约见地点)。在该示例中，值的差异指示用户理解特定类型信息的能力的差异，因为用户理解复杂或详细信息的能力下降。

[0058] 与各个词语和短语相关联的值能够基于来自先前用户会话的用户活动数据来确定，该先前用户会话中用户202先前处于场境数据206所指示的场境中。譬如，能够从用户历史管理器230传送历史用户数据，该用户历史管理器检索存储在查询日志234中的数据。在本示例中，基于确定与约见地点相比用户通常更频繁地访问与约见相关联的日期和时间信息，能够增大日期和时间信息的值。

[0059] 在语言熟练度估计器212选择与所接收的场境数据206相对应的特定场境简档232之后，语言熟练度估计器212将所选择的场境简档232传送到插值器213。插值器213解析所选择的场境简档232，并且提取所包括的各个词语和短语及其关联值。在一些情况下，插值器213将不同类型的信息和关联值直接传送到语言学分析器214，以供生成文本转话音输出候选的列表240a。在这样的情况下，插值器213从所选择的场境简档232中提取特定类型的信息和相关联的值并且将它们传送到语言学分析器214。在其他情况下，插值器213也能够将所选择的场境简档232传送到重排名器215。

[0060] 在一些情况下，能够向TTS引擎210提供结构化数据的集合(例如，日历事件的字段)。在这样的情况下，插值器213能够将结构化数据转换成水平与场境简档232所指示的用户熟练度相匹配的文本。例如，TTS引擎210可以访问指示一个或多个语法的数据——所述一个或多个语法指示表达结构化数据中的信息的不同细节或复杂度水平，并且基于用户的语言熟练度分值来选择适当的语法。类似地，TTS引擎210能够使用字典来选择鉴于语言熟练度分值适当的词语。

[0061] 语言学分析器214对查询204内所包括的信息执行诸如标准化的处理操作。譬如，查询分析器211能够将发音转录(phonetic transcription)指派给查询204内所包括的每个词语或摘录(snippet)，并且使用文本转话音转换将查询204划分成诸如短语、分句、和语句的超音段单元(prosodic unit)。语言学分析器214也生成列表240a，该列表240a包括识别为响应于查询204的多个文本转话音输出候选。在本示例中，列表240a包括具有不同复杂度水平的多个文本转话音输出候选。例如，响应“At 12:00PM with Mr.John near Dupont Circle(下午12点在杜邦圈附近约见John先生)”是最复杂的响应，因为它标识约见的时间、约见的地点、将与其约见的个体。相比之下，响应“In three hours(在三小时内)”最不复杂，因为它仅识别约见的时间。

[0062] 列表240a也包括基于每个文本转话音输出候选可能响应于查询204的可能性的文本转话音候选的基线排名。在本示例中，列表240a指示最复杂的文本转话音输出候选最有可能响应于查询204，因为它包括与查询204的内容相关联的最大信息量。

[0063] 在语言学分析器生成文本转话音输出候选的列表240a之后，重排名器215基于所接收的场境数据206来生成包括文本转话音输出候选的经调整的排名的列表240b。譬如，重排名器215能够基于与所选择的场境简档232中所包括的特定类型的信息相关联的分值来调整排名。

[0064] 在本示例中，重排名器215基于场境简档232来将最简单的文本转话音输出排名为最高，该场境简档指示：鉴于指示用户正通勤上下班的目前用户场境，指示用户202有可能领会文本转话音响应内的日期和时间信息，但不太可能理解文本转话音响应内的参与方名称或位置信息。就此而言，所接收的场境数据206能够被使用于调整特定文本转话音输出候选的选择，以增加用户202会理解TTS引擎210的文本转话音输出204c的内容的可能性。

[0065] 图3是图示出用于修改文本转话音输出内的语句结构的系统300的示例的示图。简而言之，TTS引擎310接收用户(例如，用户202)的查询302和语言熟练度简档304。然后，TTS引擎310执行操作312、314和316以生成响应于查询302的经调整的文本转话音输出302c。在一些情况下，TTS引擎310对应于参照图1所述的TTS引擎120或者参照图2所述的TTS引擎210。

[0066] 一般而言,TTS引擎310能够通过使用不同类型的调整技术来修改针对查询302的基线文本转话音输出306a的语句结构。举例而言,TTS引擎310能够基于确定与各个词语或短语相关联的复杂度分值大于用户的语言复杂度简档304所指示的阈值分值来替换基线文本转话音输出306a内的词语或短语。另举一例,TTS引擎310能够重新布置各个语句分句,使得基于语言复杂度简档304来将基线文本转话音输出306a的总体复杂度降低到令人满意的水平。TTS引擎310也能够重排序词语、分割或组合语句、以及作出其他变化来调整文本的复杂度。

[0067] 更详细地,在操作312期间,TTS引擎310初步生成响应于查询302的基线文本转话音输出306a。然后,TTS引擎310将基线文本转话音输出306a解析成片段312a至312c。TTS引擎310也检测指示各个片段之间的断点的标点符号(例如,逗号、句点、分号等)。TTS引擎310也计算片段312a至312c中的每个的复杂度分值。在一些情况下,复杂度分值能够基于特定词语在特定语言内的频率来计算。替选的技术能够包括基于用户使用的频率或者用户访问的历史内容(例如,新闻报道、web页面等)中的出现频率来计算复杂度分值。在这些示例中的每个中,复杂度分值能够被使用于指示有可能被用户领会的词语以及不太可能被用户领会的其他词语。

[0068] 在本示例中,基于分别包括诸如“FORECAST(天气预报)”和“CONSISTENT(持续)”的高度复杂的词项,片段312a和312b被确定为相对复杂。然而,片段312c被确定为相对简单,因为所包括的词项相对简单。这种确定通过与片段312c的复杂度分值(例如,0.41)相比具有更高复杂度分值(例如,0.83、0.75)的片段312a和312b来表示。

[0069] 如上所述,语言熟练度简档304能够被使用于计算阈值复杂度分值,其指示用户能领会的最大复杂度。在本示例中,阈值复杂度分值能够被计算为“0.7”,使得TTS 310确定片段312a和312b不太可能被用户领会。

[0070] 在识别具有大于语言熟练度简档304所指示的阈值复杂度分值的关联复杂度分值的各个片段之后,在操作314期间,TTS引擎310用预测为更可能被用户理解的替选项目来替换所识别的词语。如图3所示,“FORECAST”能够被替换为“WEATHER()天气”,并且“CONSISTENT”能够被替换为“CHANGE(变化)”。在这些示例中,片段314a和314b表示具有低于语言熟练度简档304所指示的阈值复杂度分值的更复杂度分值的更简单的替选方案。

[0071] 在一些实施方式中,TTS引擎310能够使用经训练的跳跃元语法(skip-gram)模型来处理高复杂度词语的词语替换,该跳跃元语法模型使用无监督技术来确定适当复杂的词语以替换高度复杂的词语。在一些情况下,TTS引擎310也能够使用同义词词典或同义词数据来处理对高度复杂词语的词语替换。

[0072] 现介绍操作316,基于计算与特定语句结构相关联的复杂度以及基于语言熟练度简档304所指示的语言熟练度来确定用户是否将能够理解语句结构,能够调整查询的语句分句。

[0073] 在本示例中,基于确定基线文本转话音响应306a包括三个语句分句(例如,“today's forecast is sunny(今日天气预报为晴)”、“but not consistent(而不持续)”和“and warm(并且温暖)”),TTS引擎310确定基线文本转话音响应306a具有高语句复杂度。作为响应,TTS引擎310能够生成经调整的语句部分316a和316b,它们将从属分句与独立分句组合成不包括进行划分的标点符号的单个分句。结果,经调整的文本转话音响应306b包

括更简单的词汇(例如,“WEATHER”、“CHANGE”)以及更简单的语句结构(例如,没有分句划分),这会提高用户理解经调整的文本转话音输出306b的可能性。然后,生成经调整的文本转话音输出306b,以供TTS引擎310作为输出306c输出。

[0074] 在一些实施方式中,TTS引擎310能够基于使用用户特定的重构算法来执行语句结构调整,该用户特定的重构算法包括使用加权因子来调整基线查询302a以避免识别为对用户成问题的特定语句结构。例如,用户特定的重构算法能够指定选项来降低包含从属分句的权重或者提高具有简单主谓宾顺序的语句分句的权重。

[0075] 图4是图示出基于使用聚类技术来自适应地生成文本转话音输出的系统400的示例的框图。系统400包括语言熟练度估计器410、用户相似度确定器420、复杂度优化器、和机器学习系统440。

[0076] 简而言之,语言熟练度估计器410从多个用户402接收数据。然后,语言熟练度估计器410针对多个用户402中的每个用户估计一组语言复杂度简档412,然后将其发送到用户相似度确定器420。用户相似度确定器420识别相似用户的用户集群424。然后,复杂度优化器430和机器学习系统440分析用户集群424内的每个用户的语言复杂度简档412以及从多个用户402接收的场境数据,以便生成复杂度映射442。

[0077] 一般而言,系统400能够被使用于分析用户群体的主动语言复杂度与被动语言复杂度之间的关系。主动语言复杂度指代检测到的用户所提供的语言输入(例如,文本查询、语音输入等)。被动语言复杂度指代用户理解或领会提供给用户的话音信号的能力。就此而言,系统400能够使用所确定的多个用户的主动语言复杂度与被动语言复杂度之间的关系来确定每个个体用户的适当被动语言复杂度,其中特定用户具有理解文本转话音输出的最高可能性。

[0078] 多个用户402能够是使用与TTS引擎(例如,TTS引擎120)相关联的应用的多个用户。譬如,多个用户402能够是一组使用移动应用的用户,该移动应用利用TTS引擎通过移动应用的用户接口向用户提供文本转话音特征。在这样的情况下,来自多个用户402的数据(例如,在先用户查询、用户选择等)能够被移动应用跟踪并且被聚合以供语言熟练度估计器410分析。

[0079] 语言熟练度估计器410能够使用与前文参照图1所述基本上相似的技术来初步测量多个用户402的被动语言复杂度。然后,语言熟练度估计器410能够生成语言复杂度简档412,其包括针对多个用户402中的每个用户的个体语言复杂度简档。每个个体语言复杂度简档包括指示多个用户402中的每个用户的被动语言复杂度和主动语言复杂度的数据。

[0080] 用户相似度确定器420使用该组语言熟练度简档412内所包括的语言复杂度数据来识别多个用户402内的相似用户。在一些情况下,用户相似度确定器420能够将具有相似主动语言复杂度(例如,提供相似的语言输入、话音查询等)的用户分组。在其他情况下,用户相似度确定器420能够通过比较在先用户提交的查询中所包括的词语、移动应用上的特定用户行为或用户位置来确定相似用户。然后,用户相似度确定器420聚类相似用户以生成用户集群424。

[0081] 在一些实施方式中,用户相似度确定器420基于所存储的集群数据422来生成用户集群424,该集群数据包括指定集群中的用户的聚合数据。例如,集群数据422能够通过指示与多个用户402相关联的被动语言复杂度的特定参数(例如,不正确查询响应的数目等)来

分组。

[0082] 在生成用户集群424之后，复杂度优化器430更改由TTS系统输出的语言的复杂度，并且使用指示用户表现的一组参数(例如，理解率、语音动作流完成率或回答成功率)来测量用户的被动语言复杂度，所述一组参数指示用户理解由TTS系统输出的语言的能力。譬如，所述参数能够被使用于表征每个集群424内的用户理解给定文本转话音输出的程度。在这样的情况下，复杂度优化器430能够初步向用户提供低复杂度的话音信号，并且以递归方式提供一定复杂度范围内的附加话音信号。

[0083] 在一些实施方式中，复杂度优化器430也能够确定与每个用户集群424相关联的各个用户场境的最优被动语言复杂度。譬如，在使用该组参数测量用户的语言熟练度之后，复杂度优化器430然后能够通过从多个用户402接收的场境数据将所测得的数据分类，以使得能够针对每个用户场境确定最优被动语言复杂度。

[0084] 在聚集被动语言复杂度范围的表现数据之后，机器学习系统440然后在表现参数指示用户的语言理解最强的情况下确定特定被动语言复杂度。譬如，机器学习系统440聚合特定用户集群424内的所有用户的表现数据，以确定主动语言复杂度、被动语言复杂度与用户场境之间的关系。

[0085] 然后，能够将用户集群424的聚合数据与用户集群424内的每个用户的个体数据进行比较，以确定用户集群424内的每个用户的实际语言复杂度分值。譬如，如图4所示，复杂度映射442能够表示主动语言复杂度与被动语言复杂度之间的关系，以推断实际语言复杂度，其对应于映射到最优被动语言复杂度的主动语言复杂度。

[0086] 复杂度映射442表示多个用户402内的所有用户集群的主动语言复杂度、TTS复杂度与被动语言复杂度之间的关系，其然后能够被使用于预测个体用户后续查询的适当TTS复杂度。例如，如上所述，用户输入(例如，查询、文本消息、电子邮件等)能够被使用于将相似用户分组成用户集群424。针对每个集群，该系统提供需要不同水平的语言熟练度来理解的TTS输出。然后，该系统评估从用户接收的响应以及不同TTS输出的任务完成率，以确定适合于每个集群中的用户的语言复杂度水平。该系统存储集群标识符和与所标识的集群相对应的TTS复杂度分值之间的映射442。然后，该系统使用复杂度映射442来确定针对用户的TTS输出的适当复杂度水平。例如，该系统识别表示用户的主动语言熟练度的集群，在映射442中查找该集群的对应TTS复杂度分值(例如，指示被动语言理解水平)，并且生成具有由检索到的TTS复杂度分值所指示的复杂度水平的TTS输出。

[0087] 然后，通过使用参照图1至3所述的技术，能够使用针对用户所确定的实际语言复杂度来调整TTS系统。就此而言，来自一组相似用户(例如，用户集群424)的聚合语言复杂度数据能够被使用于智能地调整TTS系统针对单个用户的表现。

[0088] 图5是图示出用于自适应地生成文本转话音输出的过程500的示例的流程图。简而言之，过程500能够包括确定客户端设备的用户的语言熟练度(510)，确定文本片段以供文本转话音模块输出(520)，生成包括文本片段的合成话语的音频数据(530)，以及将音频数据提供给客户端设备(540)。

[0089] 更详细地，过程500能够包括确定客户端设备的用户的语言熟练度(510)。譬如，如参照图1所述，语言熟练度估计器110能够使用各种技术来确定用户的语言熟练度。在一些情况下，语言熟练度能够表示指示语言熟练度水平的指派分值。在其他情况下，语言熟练度

能够表示来自多个语言熟练度类别的指派类别。在其他情况下，能够基于指示用户的熟练度水平的用户输入和/或行为来确定语言熟练度。

[0090] 在一些实施方式中，能够根据不同的用户信号来推断语言熟练度。譬如，如参照图1所述，能够根据用户输入的词汇复杂度、用户的数据录入速率、来自话音输入的误标识词语的数目、不同水平的TTS复杂度下完成的语音动作的数目、或者用户查看的文本(例如，书籍、文章、web页面上的文本等)的复杂度水平来推断语言熟练度。

[0091] 过程500能够包括确定文本片段以供文本转话音模块输出(520)。譬如，TTS引擎能够基于确定用户的语言熟练度来调整基线文本片段。在一些情况下，如参照图2所述，能够基于与用户相关联的用户场境来调整用于输出的文本片段。在其他情况下，如参照图3所述，通过词语替换或语句重构以便降低文本片段的复杂度，也能够调整用于输出的文本片段。例如，调整能够是基于文本片段中所包括的个体词语的罕见程度、所用动词的类型(例如，复合动词或动词时态)、文本片段的语言学结构(例如，从属分句的数目、相关词语之间的间隔量、短语嵌套的程度等)。在其他示例中，调整也能够是基于上述语言学量度以及语言学特性的参考测量(例如，主语与动词之间的平均间隔、形容词与名词之间的间隔等)。在这样的示例中，参考测量能够表示平均值或者可能包括针对不同复杂度水平的范围或示例。

[0092] 在一些实施方式中，确定用于输出的文本片段能够包括选择具有与描述用户的语言熟练度水平的参考分值最佳匹配的分值的文本片段。在其他实施方式中，能够针对复杂度对各个词语或短语进行评分，并且然后能够替换、删除或重构最复杂的词语，使得总体复杂度符合用户的适当水平。

[0093] 过程500能够包括生成包括文本片段的合成话语的音频数据(530)。

[0094] 过程500能够包括将音频数据提供给客户端设备(540)。

[0095] 图6是能够作为客户端或者作为一个服务器或多个服务器而用于实现本文所述的系统和方法的计算设备600、650的框图。计算设备600意图表示各种形式的数字计算机，诸如膝上型计算机、桌面型计算机、工作站、个人数字助理、服务器、刀片服务器、大型主机以及其他适当的计算机。计算设备650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、以及其他类似的计算设备。附加地，计算设备600或650能够包括通用串行总线(USB)闪速驱动器。USB闪速驱动器能够存储操作系统和其他应用。USB闪速驱动器能够包括输入/输出组件，诸如能够插入另一个计算设备的USB端口的无线发射器或USB连接器。在此所示的组件、它们的连接和关系以及它们的功能意在仅为示例性，而非意在限制本文所述和/或所要求保护的本发明的实施方式。

[0096] 计算设备600包括处理器602、存储器604、存储设备606、连接至存储器604和高速扩展端口610的高速接口608以及连接至低速总线614和存储设备606的低速接口612。组件602、604、606、608、610和612中的每个使用各种总线互连，并且能够被安装在公共母板上或者酌情以其他方式安装。处理器602能够处理用于在计算设备600内执行的指令，包括存储在存储器604中或者存储设备606上的指令，以对诸如耦合至高速接口608的显示器616的外部输入/输出设备上的显示GUI的图形信息。在其他实施方式中，能够酌情连同多个存储器和存储器类型一起使用多个处理器和/或多个总线。此外，能够连接多个计算设备600，每个设备提供所需操作的多个部分，例如，作为服务器组、刀片服务器组、或者多处理器系统。

[0097] 存储器604存储计算设备600内的信息。在一种实施方式中，存储器604是一个或多个易失性存储器单元。在另一种实施方式中，存储器604是一个或多个非易失性存储器单元。存储器604也能够是另一种形式的计算机可读介质，诸如磁盘或光盘。

[0098] 存储设备606能够为计算设备600提供大容量存储。在一种实施方式中，存储设备606能够是或包含：计算机可读介质，诸如软盘设备、硬盘设备、光盘设备或者磁带设备、闪速存储器或者其他类似的固态存储器设备、或者包括存储区域网络或者其他配置中的设备的设备阵列。计算机程序产品能够有形地体现于信息载体中。计算机程序产品也能够包含指令，这些指令当被执行时执行一个或多个方法，诸如上述那些方法。信息载体是计算机或机器可读介质，诸如存储器604、存储设备606或者处理器602上的存储器。

[0099] 高速控制器608管理用于计算设备600的带宽密集(bandwidth-intensive)操作，而低速控制器612管理较低带宽密集操作。这样的功能分配仅为示例性。在一种实施方式中，高速控制器608耦合至存储器604、显示器616(例如，通过图形处理器或者加速器)以及能够接受各种扩展卡(未示出)的高速扩展端口610。在本实施方式中，低速控制器612耦合至存储设备606以及低速扩展端口614。能够包括各种通信端口(例如，USB、蓝牙、以太网、无线以太网)的低速扩展端口能够耦合至一个或多个输入/输出设备，诸如键盘、指示设备、麦克风/扬声器对、扫描仪、或者诸如通过例如网络适配器的交换机或者路由器的网络设备。如图所示，计算设备600能够通过数种不同的形式来实现。例如，它能够被实现为标准服务器620或者在这样的服务器组中多重实现。它还能够被实现为机架式服务器系统624的一部分。此外，它能够被实现在诸如膝上型计算机622的个人计算机中。替选地，来自计算设备600中的组件能够与诸如设备650的移动设备(未示出)中的其他组件组合。这样的设备中的每一个能够包含计算设备600、650中的一个或多个，并且整个系统能够由彼此通信的多个计算设备600、650组成。

[0100] 如图所示，计算设备600能够通过数种不同的形式来实现。例如，它能够被实现为标准服务器620或者在这样的服务器群组中多重实现。它还能够被实现为机架式服务器系统624的一部分。此外，它能够被实现在诸如膝上型计算机622的个人计算机中。替选地，来自计算设备600中的组件能够与诸如设备650的移动设备(未示出)中的其他组件组合。这样的设备中的每一个设备能够包含计算设备600、650中的一个或多个，并且整个系统能够由彼此通信的多个计算设备600、650组成。

[0101] 计算设备650包括处理器652、存储器664、诸如显示器654、通信接口666和收发器668的输入/输出设备以及其他组件。设备650还能够具有用于提供附加存储的存储设备，诸如微驱动器或者其他设备。组件650、652、664、654、666和668中的每个组件使用各个总线互联，并且这些组件中的若干组件能够被安装在公共母板上或者酌情以其他方式安装。

[0102] 处理器652能够执行计算设备650内的指令，包括存储在存储器664中的指令。处理器能够被实现为包括分开的多个模拟和数字处理器的芯片的芯片组。附加地，处理器能够使用多个架构中的任何一个来实现。例如，处理器610能够是CISC(复杂指令集计算机)处理器、RISC(精简指令集计算机)处理器、或者MISC(最小指令集计算机)处理器。例如，处理器能够提供与设备650的其他组件的协作，诸如用户接口的控件、设备650运行的应用以及设备650的无线通信。

[0103] 处理器652能够通过耦合至显示器654的控制接口658以及显示器接口656与用户

通信。例如，显示器654能够是TFT(薄膜晶体管液晶显示)显示器或OLED(有机电致发光二极管)显示器或者其他适当的显示技术。显示器接口656能够包括用于驱动显示器654以将图形信息和其他信息呈现给用户的适当电路。控制接口658能够从用户接收命令并且将它们进行变换以供提交给处理器652。此外，能够提供外部接口662与处理器652通信，以便启用设备650与其他设备的近区通信。例如，外部接口662能够在一些实施方式中提供有线通信，或者在其他实施方式中提供无线通信，并且也能够使用多个接口。

[0104] 存储器664存储计算设备650内的信息。存储器664能够被实现为一个或多个计算机可读介质、一个或多个易失性存储器单元或者一个或多个非易失性存储器单元中的一个或多个。还能够提供扩展存储器674并且通过扩展接口672将其连接至设备650，例如，该扩展接口能够包括SIMM(单列直插内存模块)卡接口。这样的扩展存储器674能够提供用于设备650的额外存储空间，或者也能够存储用于设备650的应用或其他信息。具体地，扩展存储器674能够包括用于实行或者补充上述过程的指令，并且也能够包括安全信息。因此，例如，扩展存储器674能够被提供为用于设备650的安全模块，并且能够用许可安全使用设备650的指令来进行编程。此外，能够经由SIMM卡提供安全应用连同附加信息，诸如，以不可破解(non-hackable)的方式将识别信息置于SIMM卡上。

[0105] 例如，存储器能够包括闪速存储器和/或NVRAM存储器，如下文所讨论。在一种实施方式中，计算机程序产品有形地体现于信息载体中。计算机程序产品包含指令，所述指令当被执行时执行一个或多个方法，诸如上述那些方法。信息载体是能够例如通过收发器668或者外部接口662接收的计算机或机器可读介质，诸如存储器664、扩展存储器674或者处理器652上的存储器。

[0106] 设备650能够通过通信接口666进行无线通信，该通信接口必要时能够包括数字信号处理电路。通信接口666能够提供各种模式或协议下的通信，诸如GSM语音呼叫、SMS、EMS或MMS消息收发、CDMA、TDMA、PDC、WCDMA、CDMA2000或GPRS等等。例如，这样的通信能够通过射频收发器668来进行。此外，能够发生近程通信，诸如使用蓝牙、Wi-Fi或者其他这样的收发器(未示出)。此外，GPS(全球定位系统)接收器模块670能够将额外的导航和位置相关的无线数据提供给设备650，其能够酌情供设备650上运行的应用使用。

[0107] 设备650也能够使用音频编解码器660来进行可听通信，该音频编解码器能够从用户接收口述信息并且将其变换为可用的数字信息。音频编解码器660同样能够为用户生成可听声音，诸如通过扬声器，例如，在设备650的手持设备中。这样的声音能够包括来自语音电话呼叫的声音，能够包括录制的声音，例如，语音消息、音乐文件等，并且也能够包括通过在设备650上操作的应用所生成的声音。

[0108] 如图所示，计算设备650能够通过数种不同的形式来实现。例如，它能够被实现为蜂窝电话480。它也能够被实现为智能电话682、个人数字助理或者其他类似移动设备的一部分。

[0109] 本文所述的系统和方法的各种实施方式能够在数字电子电路、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/或这些实施方式的组合中实现。这些各种实施方式能够包括可编程系统上的可执行和/或可解释的一个或多个计算机程序中的实施方式，所述可编程系统包括耦合以从存储系统接收数据和指令并且向存储系统发送数据和指令的、能够是专用或通用的至少一个可编程处理器、存储系统、至少一个输入设备以

及至少一个输出设备。

[0110] 这些计算机程序(也被称作程序、软件、软件应用或者代码)包括用于可编程处理器的机器指令,并且能够以高级程序语言和/或面向对象的编程语言和/或汇编/机器语言来实现。如本文所用,术语“机器可读介质”、“计算机可读介质”指代用于将机器指令和/或数据提供给可编程处理器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),包括接收机器指令作为机器可读信号的机器可读介质。术语“机器可读信号”指代用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0111] 为提供与用户的交互,本文所述的系统和技术能够在计算机上来实现,该计算机具有用于将信息显示给用户的显示设备(例如CRT(阴极射线管)或者LCD(液晶显示器)监视器)以及用户能够借以向计算机提供输入的键盘和指示设备(例如鼠标或跟踪球)。其他种类的设备也能够被使用于提供与用户的交互;例如,提供给用户的反馈能够是任何类型的感官反馈,例如,视觉反馈、听觉反馈或者触觉反馈;并且来自用户的输入能够以任何形式来接收,包括声音、话音或者触觉输入。

[0112] 本文所述的系统和技术能够在包括后端组件(例如作为数据服务器)、包括中间件组件(例如应用服务器)、包括前端组件(例如具有用户能够借以与本文所述的系统和技术的实施方式交互的图形用户界面或者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任何组合的计算系统中来实现。该系统的组件能够通过任何形式或者介质的数字数据通信来互联,例如通信网络。通信网络的示例包括局域网(“LAN”)、广域网(“WAN”)以及互联网。

[0113] 计算系统能够包括客户端和服务器。客户端与服务器一般彼此远离,并且通常通过通信网络相交互。客户端与服务器的关系借助在相应计算上运行并且彼此具有客户端-服务器关系的计算机程序来实现。

[0114] 本文已对数个实施例予以描述。然而,应当理解到,在不脱离本发明的精神和范围的情况下,能够作出各种修改。此外,在附图中所描绘的逻辑流程不要求所示的特定次序或顺序次序来获得期望的结果。此外,能够从所述流程中提供其他步骤或者取消一些步骤,并且能够向所述系统添加其他组件或者从所述系统移除一些组件。因此,其他实施例落入所附权利要求书的范围内。

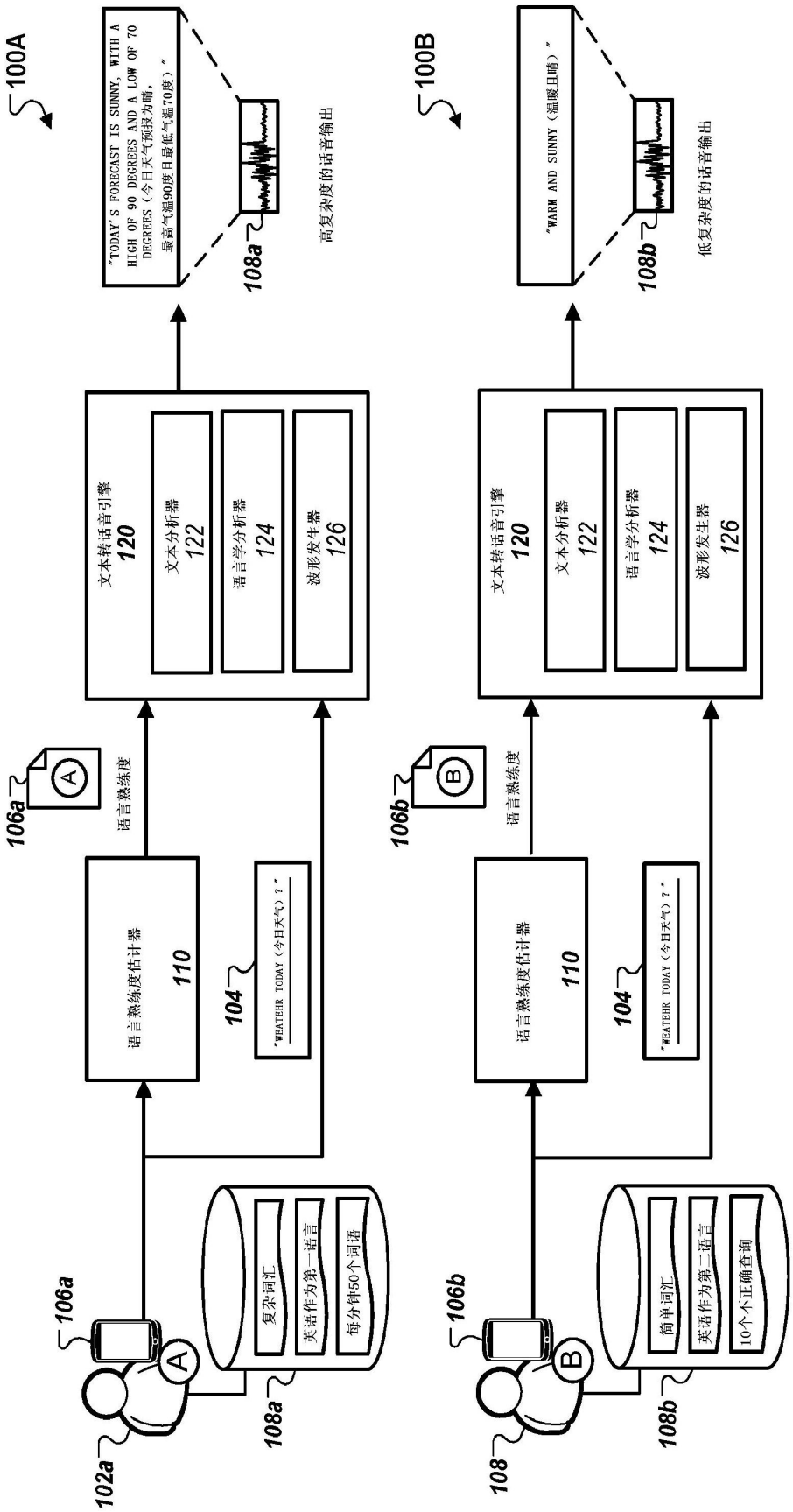

100A
"TODAY'S FORECAST IS SUNNY, WITH A HIGH OF 90 DEGREES AND A LOW OF 70 DEGREES"（今日天气预报为晴，最高气温90度且最低气温70度）"
108a
高复杂度的话音输出
文本转话音引擎 120
文本分析器 122
语言学分析器 124
波形发生器 126
106a
A
语言熟练度
语言熟练度估计器 110
104
"WEATEHR TODAY（今日天气）?"
复杂词汇
英语作为第二语言
每分钟50个词语
108a
106a
A
102a
100B
"WARM AND SUNNY（温暖且晴）"
108b
低复杂度的话音输出
文本转话音引擎 120
文本分析器 122
语言学分析器 124
波形发生器 126
106b
B
语言熟练度
语言熟练度估计器 110
104
"WEATEHR TODAY（今日天气）?"
简单词汇
英语作为第二语言
10个不正确查询
108b
106b
B
108

图1

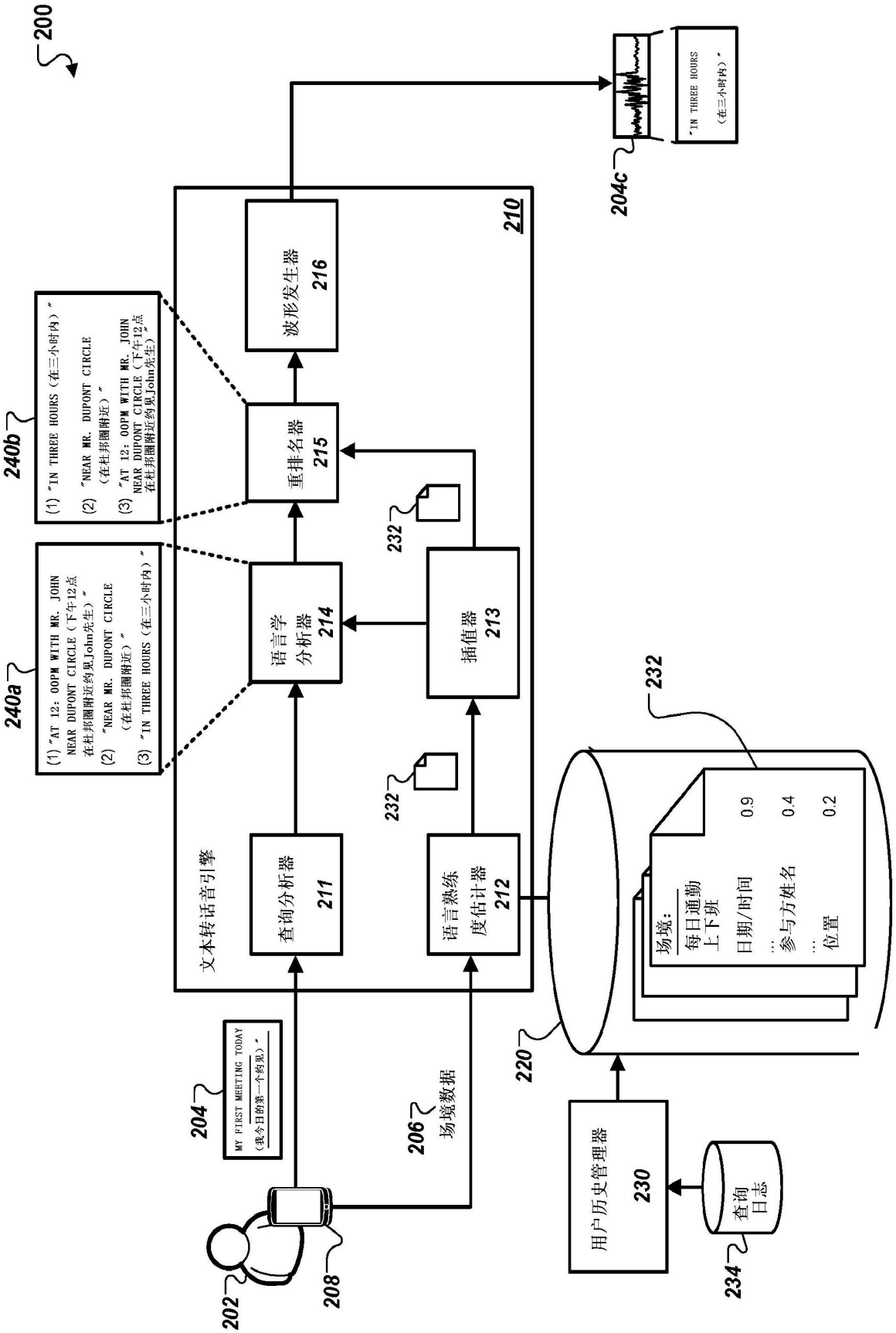

200
文本转话音引擎
210
查询分析器 211
语言熟练度估计器 212
语言学分析器 214
插值器 213
重排名器 215
波形发生器 216
240a
(1) "AT 12: 00PM WITH MR. JOHN NEAR DUPONT CIRCLE (下午12点在杜邦圈附近约见John先生)"
(2) "NEAR MR. DUPONT CIRCLE (在杜邦圈附近)"
(3) "IN THREE HOURS (在三小时内)"
240b
(1) "IN THREE HOURS (在三小时内)"
(2) "NEAR MR. DUPONT CIRCLE (在杜邦圈附近)"
(3) "AT 12: 00PM WITH MR. JOHN NEAR DUPONT CIRCLE (下午12点在杜邦圈附近约见John先生)"
204
MY FIRST MEETING TODAY (我今日的第一个约见)
204c
"IN THREE HOURS (在三小时内)"
206
场境数据
208
202
220
232
场境:
每日通勤上下班
日期/时间 0.9
参与方姓名 0.4
位置 0.2
用户历史管理器 230
查询日志
234

图2

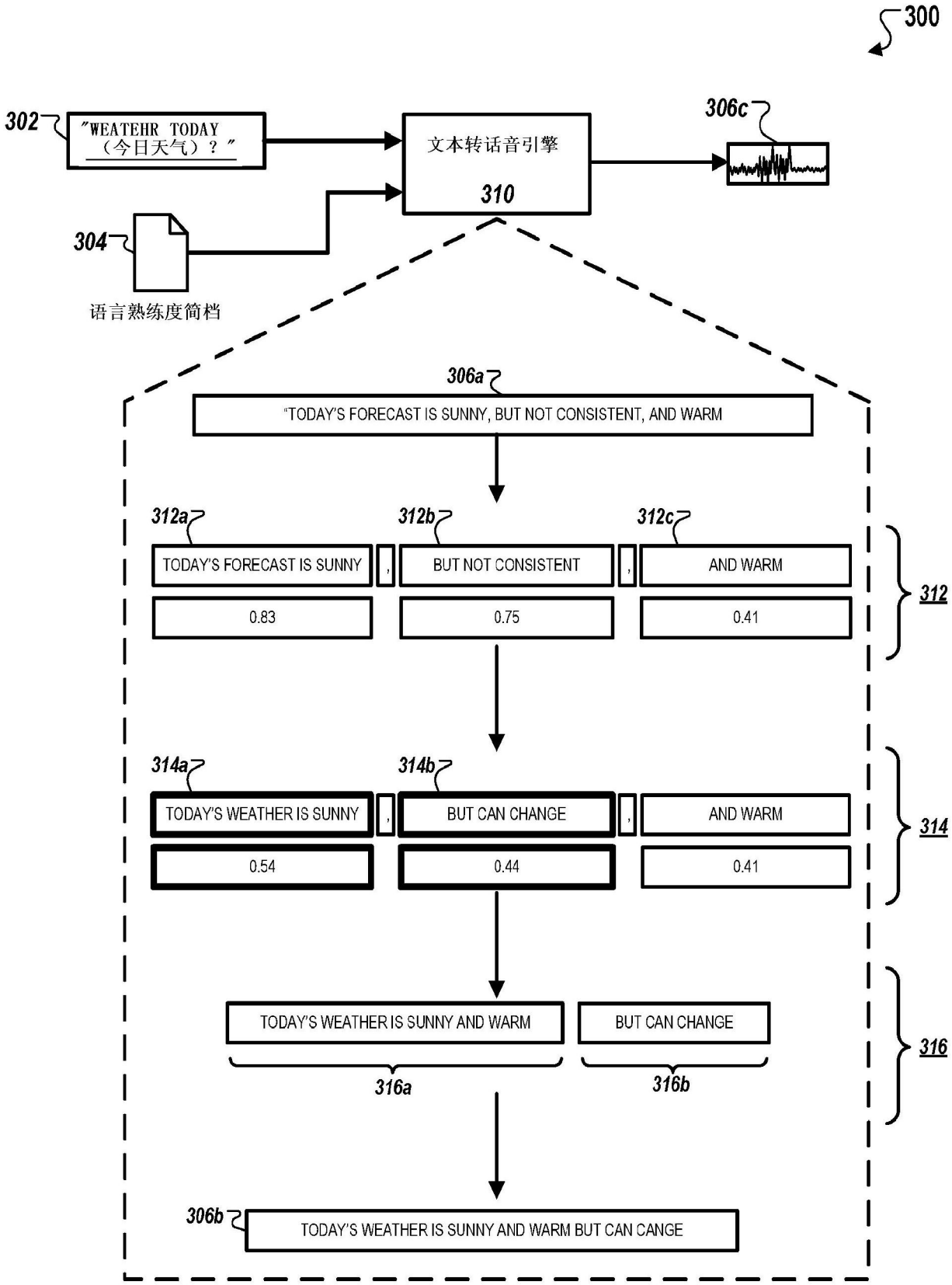

300
302
"WEATEHR TODAY
（今日天气）？"
文本转话音引擎
310
306c
304
语言熟练度简档
306a
"TODAY'S FORECAST IS SUNNY, BUT NOT CONSISTENT, AND WARM
312a
312b
312c
TODAY'S FORECAST IS SUNNY
,
BUT NOT CONSISTENT
,
AND WARM
0.83
0.75
0.41
312
314a
314b
TODAY'S WEATHER IS SUNNY
,
BUT CAN CHANGE
,
AND WARM
0.54
0.44
0.41
314
TODAY'S WEATHER IS SUNNY AND WARM
BUT CAN CHANGE
316a
316b
316
306b
TODAY'S WEATHER IS SUNNY AND WARM BUT CAN CANGE

图3

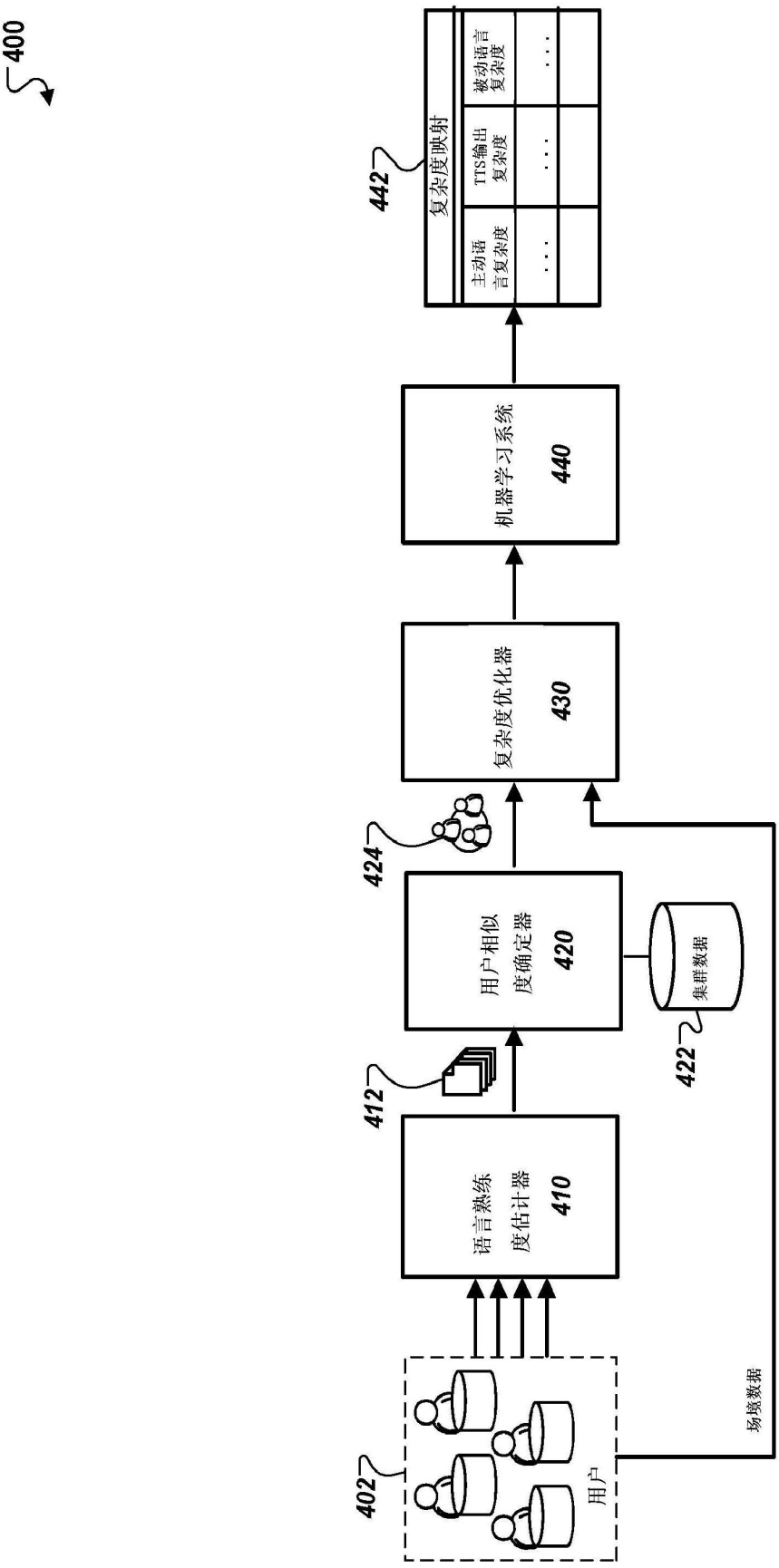

400
402
用户
场境数据
412
语言熟练
度估计器
410
422
集群数据
424
用户相似
度确定器
420
复杂度优化器
430
机器学习系统
440
442
复杂度映射
主动语
言复杂度
TTS输出
复杂度
被动语言
复杂度

图4

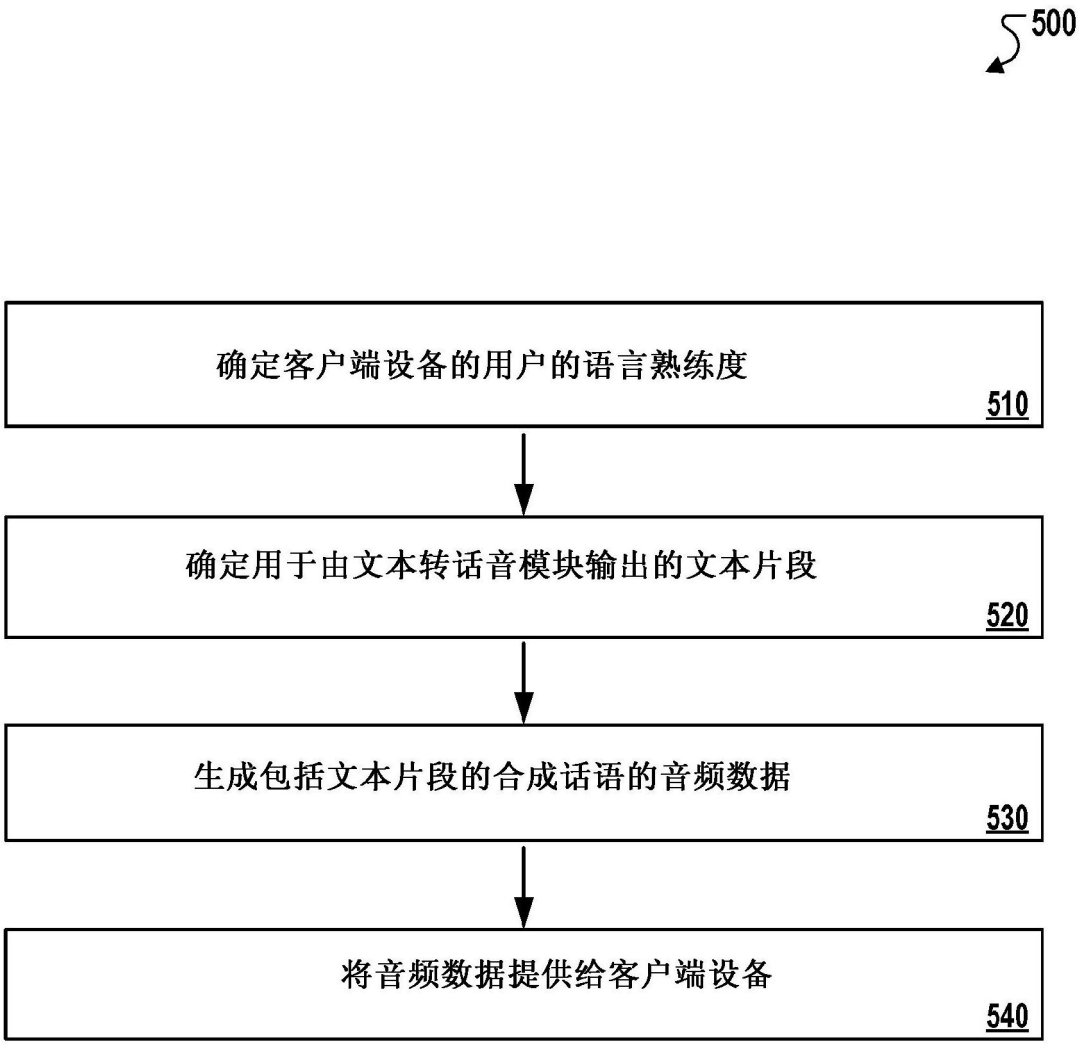

500
确定客户端设备的用户的语言熟练度
510
确定用于由文本转话音模块输出的文本片段
520
生成包括文本片段的合成话语的音频数据
530
将音频数据提供给客户端设备
540

图5

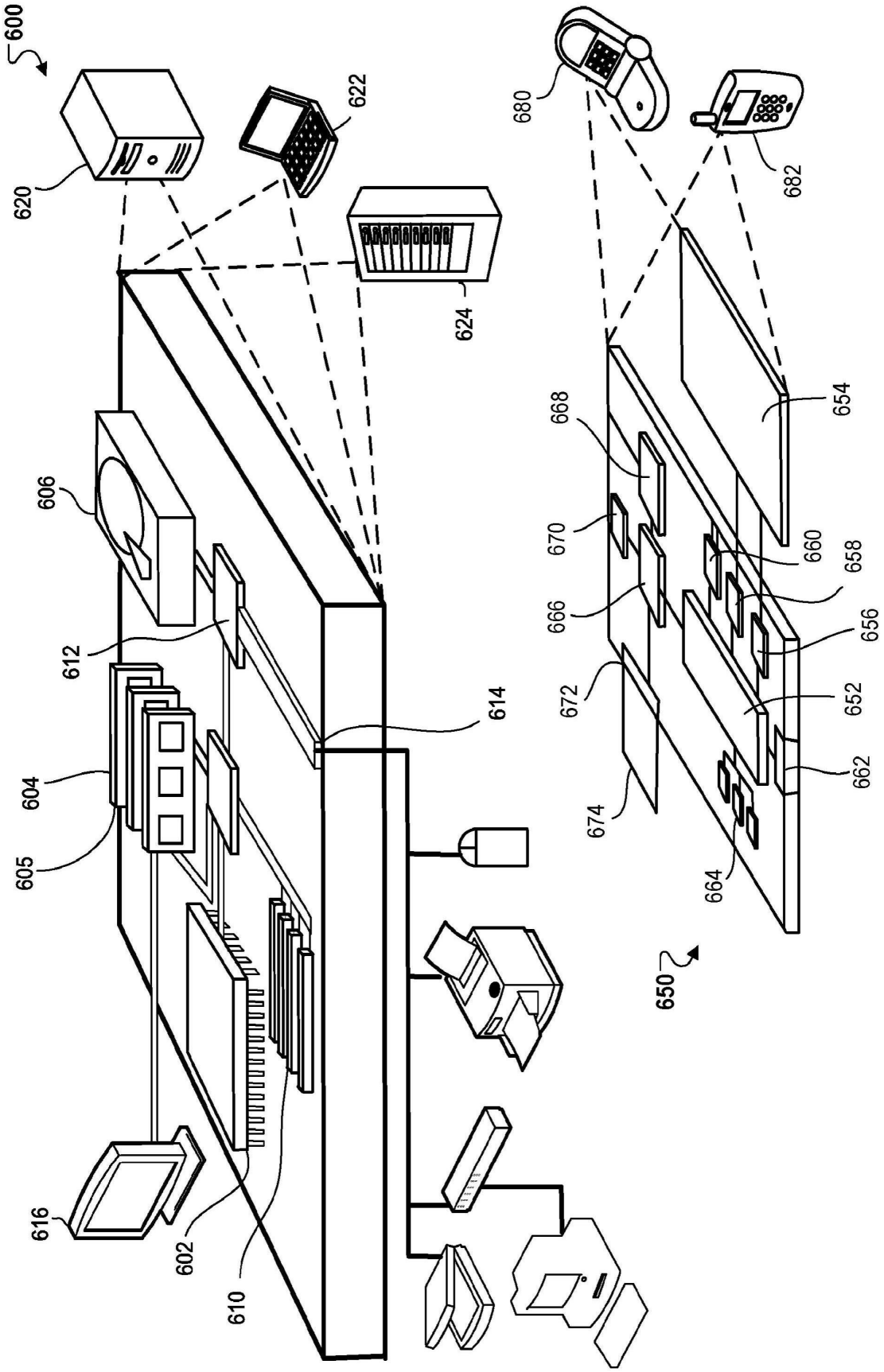

600
620
622
624
606
612
614
604
605
616
602
610
680
682
668
654
670
660
658
666
656
672
652
662
674
664
650

图6