



(12) 发明专利申请

(10) 申请公布号 CN 104380284 A

(43) 申请公布日 2015.02.25

(21) 申请号 201380017905.3

G10L 13/02(2013.01)

(22) 申请日 2013.03.04

(30) 优先权数据

13/413,449 2012.03.06 US

(85) PCT国际申请进入国家阶段日

2014.09.29

(86) PCT国际申请的申请数据

PCT/US2013/028920 2013.03.04

(87) PCT国际申请的公布数据

W02013/134153 EN 2013.09.12

(71) 申请人 苹果公司

地址 美国加利福尼亚

(72) 发明人 C·B·弗雷扎齐 D·C·米尼菲

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 鲍进

(51) Int. Cl.

G06F 17/27(2006.01)

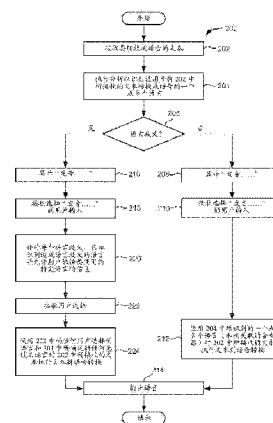
权利要求书6页 说明书17页 附图14页

(54) 发明名称

针对多种语言处理内容的语音合成

(57) 摘要

本发明公开了允许用户从多种语言中选择要用于执行文本到语音转换的语言的技术。在一些实施例中,在确定可用于执行文本的一部分的文本到语音转换的多种语言时,可将该多种语言显示给用户。用户可随后从多种语言中选择要使用的特定语言。文本的部分可随后以用户选择的语言转换成语音。



1. 一种方法,包括:

通过处理设备确定用于将文本转换成语音的一组语言,所述一组语言包括用于将所述文本的一部分转换成语音的多种候选语言;

通过所述处理设备使所述多种候选语言被显示;

接收指示从所述多种候选语言中选择第一候选语言的输入;以及

通过所述处理设备将所述文本转换成语音,所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

2. 根据权利要求1所述的方法,其中以所述第一候选语言将所述文本的所述部分转换成语音包括:使用对应于所述第一候选语言的语言合成器来产生所述文本的所述部分的语音。

3. 根据权利要求1至2中任一项所述的方法,其中所述文本的所述部分是整个所述文本或所述文本的子集。

4. 根据权利要求1所述的方法,其中将所述文本转换成语音包括:

使用第一语言合成器将所述文本中的第一字符转换成语音;以及

使用不同于所述第一语言合成器的第二语言合成器将所述文本中的所述第一字符后的第二字符转换成语音。

5. 根据权利要求4所述的方法,其中将所述文本转换成语音还包括:

读取所述文本中的所述第一字符;

确定要用于将所述第一字符转换成语音的第一语言;

将所述第一语言与所述第一字符相关联;

读取所述文本中的所述第二字符,所述第二字符在所述第一字符后;

确定要用于将所述第二字符转换成语音的第二语言,其中所述第二语言不同于所述第一语言;以及

将所述第二语言与所述第二字符相关联。

6. 根据权利要求1至5中任一项所述的方法,其中所述确定包括执行下列中的至少一者:

基于一个或多个加载的键盘,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的一个或多个标签,确定要包括在所述一组语言中的语言;

基于在分析所述文本时由语言分析工具识别的一种或多种语言,确定要包括在所述一组语言中的语言;

基于由与所述文本相关联的应用程序提供的数据,确定要包括在所述一组语言中的语言;或

基于与所述文本中的一个或多个字符相关联的编码信息,确定要包括在所述一组语言中的语言。

7. 根据权利要求1至6中任一项所述的方法,其中所述确定包括执行下列中的至少一者:

基于一个或多个用户偏好,确定要包括在所述一组语言中的语言;

基于选择用于在先文本到语音转换的语言,确定要包括在所述一组语言中的语言;

基于所述文本的来源,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的国家代码顶级域信息,确定要包括在所述一组语言中的语言;或

基于全球定位系统(GPS)信息,确定要包括在所述一组语言中的语言。

8. 根据权利要求1至7中任一项所述的方法,其中所述确定包括:

确定用于将所述文本转换成语音的第一语言列表;以及

通过从所述第一语言列表中过滤掉至少一种语言来从所述第一语言列表确定第二语言列表。

9. 一种存储用于控制处理器的多个指令的非暂态计算机可读存储介质,所述多个指令包括:

使所述处理器确定用于将文本转换成语音的一组语言的指令,所述一组语言包括用于将所述文本的一部分转换成语音的多种候选语言;

使所述处理器显示所述多种候选语言的指令;以及

使所述处理器响应于接收到指示从所述多种候选语言中选择第一候选语言的输入而将所述文本转换成语音的指令,所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

10. 根据权利要求9所述的非暂态计算机可读存储介质,其中使所述处理器以所述第一候选语言将所述文本的所述部分转换成语音的指令包括:使所述处理器使用对应于所述第一候选语言的语言合成器来产生文本的所述部分的语音的指令。

11. 根据权利要求9至10中任一项所述的非暂态计算机可读存储介质,其中所述文本的所述部分是整个所述文本或所述文本的子集。

12. 根据权利要求9所述的非暂态计算机可读存储介质,其中使所述处理器将所述文本转换成语音的所述指令包括:

使所述处理器读取所述文本中的第一字符的指令;

使所述处理器确定要用于将所述第一字符转换成语音的第一语言的指令;

使所述处理器将所述第一语言与所述第一字符相关联的指令;

使所述处理器读取所述文本中的第二字符的指令,所述第二字符在所述第一字符后;

使所述处理器确定要用于将所述第二字符转换成语音的第二语言的指令,其中所述第二语言不同于所述第一语言;

使所述处理器将所述第二语言与所述第二字符相关联的指令;

使所述处理器使用对应于所述第一语言的语言合成器将所述第一字符转换成语音的指令;以及

使所述处理器使用对应于所述第二语言的语言合成器将所述第二字符转换成语音的指令。

13. 根据权利要求9至12中任一项所述的非暂态计算机可读存储介质,其中使所述处理器确定所述一组语言的所述指令包括使所述处理器执行下列中的至少一者的指令:

基于一个或多个加载的键盘,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的一个或多个标签,确定要包括在所述一组语言中的语言;

基于在分析所述文本时由语言分析工具识别的一种或多种语言,确定要包括在所述一组语言中的语言;

基于与所述文本中的一个或多个字符相关联的字符编码信息,确定要包括在所述一组语言中的语言;

基于一个或多个用户偏好,确定要包括在所述一组语言中的语言;

基于选择用于在先文本到语音转换的语言,确定要包括在所述一组语言中的语言;

基于所述文本的来源,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的国家代码顶级域信息,确定要包括在所述一组语言中的语言;或

基于全球定位系统(GPS)信息,确定要包括在所述一组语言中的语言。

14. 根据权利要求9至13中任一项所述的非暂态计算机可读存储介质,其中使所述处理器确定所述一组语言的所述指令包括:

使所述处理器确定用于将所述文本转换成语音的第一语言列表的指令;以及

使所述处理器通过从所述第一语言列表中过滤掉至少一种语言来从所述第一语言列表确定第二语言列表的指令。

15. 一种系统,包括:

存储器,所述存储器被配置为存储文本;以及

与所述存储器耦接的一个或多个处理器,所述一个或多个处理器被配置为:

确定用于将所述文本转换成语音的一组语言,所述一组语言包括用于将所述文本的一部分转换成语音的多种候选语言;

使所述多种候选语言被显示;以及

响应于接收到指示从所述多种候选语言中选择第一候选语言的输入而将所述文本转换成语音,所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

16. 根据权利要求15所述的系统,其中所述一个或多个处理器被配置为使用对应于所述第一候选语言的语言合成器来产生文本的所述部分的语音。

17. 根据权利要求15至16中任一项所述的系统,其中所述文本的所述部分是整个所述文本或所述文本的子集。

18. 根据权利要求15所述的系统,其中所述一个或多个处理器被配置为:

读取所述文本中的第一字符;

确定要用于将所述第一字符转换成语音的第一语言;

将所述第一语言与所述第一字符相关联;

读取所述文本中的第二字符,所述第二字符在所述第一字符后;

确定要用于将所述第二字符转换成语音的第二语言,其中所述第二语言不同于所述第一语言;

将所述第二语言与所述第二字符相关联;

使用对应于所述第一语言的语言合成器将所述第一字符转换成语音;以及

使用对应于所述第二语言的语言合成器将所述第二字符转换成语音。

19. 根据权利要求15至18中任一项所述的系统,其中所述一个或多个处理器被配置为通过执行下列中的至少一者来确定所述一组语言:

基于一个或多个加载的键盘,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的一个或多个标签,确定要包括在所述一组语言中的语言;

基于在分析所述文本时由语言分析工具识别的一种或多种语言,确定要包括在所述一组语言中的语言;

基于与所述文本中的一个或多个字符相关联的字符编码信息,确定要包括在所述一组语言中的语言;

基于一个或多个用户偏好,确定要包括在所述一组语言中的语言;

基于选择用于在先文本到语音转换的语言,确定要包括在所述一组语言中的语言;

基于所述文本的来源,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的国家代码顶级域信息,确定要包括在所述一组语言中的语言;或

基于全球定位系统(GPS)信息,确定要包括在所述一组语言中的语言。

20. 根据权利要求 15 至 19 中任一项所述的系统,其中:

所述文本选自网页;并且

所述一个或多个处理器被配置为基于与托管所述网页的网站相关联的域来确定所述一组语言。

21. 一种电子设备,包括:

处理单元,所述处理单元被配置为确定用于将文本转换成语音的一组语言,所述一组语言包括用于将所述文本的一部分转换成语音的多种候选语言;

与所述处理单元耦接的显示单元,所述显示单元被配置为使所述多种候选语言被显示;

与所述处理单元耦接的输入接收单元,所述输入接收单元被配置为接收指示从所述多种候选语言中选择第一候选语言的输入;以及

所述处理单元进一步被配置为将所述文本转换成语音,所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

22. 根据权利要求 21 所述的电子设备,其中所述处理单元进一步被配置为使用对应于所述第一候选语言的语言合成单元以所述第一候选语言将所述文本的所述部分转换成语音,从而产生所述文本的所述部分的语音。

23. 根据权利要求 21 至 22 中任一项所述的电子设备,其中所述文本的所述部分是整个所述文本或所述文本的子集。

24. 根据权利要求 21 所述的电子设备,其中所述处理单元被配置为通过下列方式将所述语音转换成文本:

使用第一语言合成单元将所述文本中的第一字符转换成语音;以及

使用不同于所述第一语言合成单元的第二语言合成单元将所述文本中的所述第一字符后的第二字符转换成语音。

25. 根据权利要求 24 所述的电子设备,其中所述处理单元被配置为:

读取所述文本中的所述第一字符;

确定要用于将所述第一字符转换成语音的第一语言;

将所述第一语言与所述第一字符相关联;

读取所述文本中的所述第二字符,所述第二字符在所述第一字符后;

确定要用于将所述第二字符转换成语音的第二语言,其中所述第二语言不同于所述第

一语言 ;以及

将所述第二语言与所述第二字符相关联。

26. 根据权利要求 21 至 25 中任一项所述的电子设备,其中所述处理单元被配置为:  
基于一个或多个加载的键盘,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的一个或多个标签,确定要包括在所述一组语言中的语言;

基于在分析所述文本时由语言分析工具识别的一种或多种语言,确定要包括在所述一组语言中的语言;

基于由与所述文本相关联的应用程序提供的数据,确定要包括在所述一组语言中的语言 ;或

基于与所述文本中的一个或多个字符相关联的编码信息,确定要包括在所述一组语言中的语言。

27. 根据权利要求 21 至 26 中任一项所述的电子设备,其中所述处理单元被配置为:  
基于一个或多个用户偏好,确定要包括在所述一组语言中的语言;

基于选择用于在先文本到语音转换的语言,确定要包括在所述一组语言中的语言;

基于所述文本的来源,确定要包括在所述一组语言中的语言;

基于与所述文本相关联的国家代码顶级域信息,确定要包括在所述一组语言中的语言 ;或

基于全球定位系统 (GPS) 信息,确定要包括在所述一组语言中的语言。

28. 根据权利要求 21 至 27 中任一项所述的电子设备,其中所述处理单元被配置为:  
确定用于将所述文本转换成语音的第一语言列表 ;以及

通过从所述第一语言列表中过滤掉至少一种语言来从所述第一语言列表确定第二语言列表。

29. 一种电子设备,包括一个或多个处理器和存储由所述一个或多个处理器执行的一个或多个程序的存储器,所述一个或多个程序包括用于下列各项的指令:

确定用于将文本转换成语音的一组语言,所述一组语言包括用于将所述文本的一部分转换成语音的多种候选语言;

使所述多种候选语言被显示;

接收指示从所述多种候选语言中选择第一候选语言的输入 ;以及

将所述文本转换成语音,所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

30. 一种电子设备,包括:

用于确定用于将文本转换成语音的一组语言的装置,所述一组语言包括用于将所述文本的一部分转换成语音的多种候选语言;

用于使所述多种候选语言被显示的装置;

用于接收指示从所述多种候选语言中选择第一候选语言的输入的装置 ;以及

用于将所述文本转换成语音的装置,所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

31. 一种在电子设备中使用的信息处理装置,包括:

用于确定用于将文本转换成语音的一组语言的装置,所述一组语言包括用于将所述文

本的一部分转换成语音的多种候选语言；

用于使所述多种候选语言被显示的装置；

用于接收指示从所述多种候选语言中选择第一候选语言的输入的装置；以及

用于将所述文本转换成语音的装置，所述转换包括以所述第一候选语言将所述文本的所述部分转换成语音。

32. 一种电子设备，包括一个或多个处理器和存储由所述一个或多个处理器执行的一个或多个程序的存储器，所述一个或多个程序包括用于执行根据权利要求 1 至 8 所述的方法中任一方法的指令。

33. 一种电子设备，包括用于执行根据权利要求 1 至 8 所述的方法中的任一方法的装置。

34. 一种在电子设备中使用的信息处理装置，包括用于执行根据权利要求 1 至 8 所述的方法中任一方法的装置。

## 针对多种语言处理内容的语音合成

### 背景技术

[0001] 本发明所公开的实施例整体涉及文本到语音合成,并且更具体地涉及允许用户从多种语言中选择要用于执行文本到语音合成或转换的语言的技术。

[0002] 语言文本转换成语音的过程通常称为文本到语音合成或文本到语音转换。由于人类所说语言的多样性,各种语言可用于执行文本到语音转换。能够以多种语言执行文本到语音转换的系统通常提供多种语言合成器,每个语言合成器被配置为以特定语言将文本转换成语音。例如,英语语言合成器可提供用于将文本转换成英语语音,法语语言合成器可提供用于将文本转换成法语语音,日语语言合成器可提供用于将文本转换成日语语音等等。取决于要用于语音的特定语言,对应于特定语言的语言合成器被用于执行文本到语音转换。

[0003] 对于共用字符的语言(诸如英语、法语和德语)来说,同一段文本可使用对应于这些语言的合成器中的任何合成器来转换成语音。然而,因为用于特定语言的合成器使用特定于该语言的读音规则和发音,所以对于不同合成器来说,同一段文本的语音输出听起来将有所不同。例如,从使用英语合成器对一段文本进行文本到语音转换得到的语音听起来可能与从使用法语合成器对同一段文本进行文本到语音转换得到的语音非常不同。

[0004] 在常规系统中,一般自动选择默认语言合成器来执行文本到语音转换,只要默认合成器可输出正在转换的文本的语音即可。然而,这可能产生用户不期望的语音结果。例如,如果要转换的文本是法语语言并且英语语言合成器是默认合成器,那么输出可能是以严重英语口音说出的法语。

### 发明内容

[0005] 描述了允许用户从多种语言中选择要用于执行文本到语音转换的语言的某些实施例。

[0006] 在一些实施例中,可将多种语言显示给用户。用户可随后从多种语言中选择要使用的特定语言。用户选择的语言可随后用于产生要转换成语音的文本的语音。

[0007] 在一些实施例中,在确定可适于执行文本的一部分的文本到语音转换的多种语言时,可将该多种语言显示给用户。用户可随后从多种语言中选择要使用的特定语言。当产生文本的语音时,可以用用户选择的语言将文本的该部分转换成语音。

[0008] 在一些实施例中,使用一个或多个分析准则来分析要转换成语音的文本。基于分析,可将一种或多种语言确定为可适用于将文本转换成语音。在一些情况下,分析可产生适用于文本的相同部分的多种语言。在这种情景下,可将多种语言输出给用户。用户可随后从多种语言中选择特定语言。随后使用对应于用户选择的语言的语言合成器,以用户选择的语言将文本的该部分转换成语音。

[0009] 文本的该部分(可针对文本的该部分将多种语言识别为可能的候选语言)可以是要转换成语音的整个文本或该文本的子集。文本的子集可以是例如但不限于文本内的一个或多个字符、文本内的一个或多个词、文本内的一个或多个句子等等。



[0010] 在一些实施例中,用以识别要用于将文本转换成语音的一种或多种语言的对文本的分析可基于各种准则。准则的例子可包括但不限于:一个或多个加载的键盘、与文本相关的一个或多个标签、在分析要转换成语音的文本时由语言分析工具识别的一种或多种语言、专用数据(例如,托管网页的网站的域信息、即时消息应用程序的发送方电话号码)、与文本中的一个或多个字符相关联的字符编码信息(例如Unicode信息)、一个或多个用户偏好(例如,语言或场所偏好)、用户历史信息(例如,用于文本到语音转换的先前用户语言选择)、文本的来源、与文本相关联的国家代码顶级域信息、全球定位系统(GPS)信息(例如,识别用户或用户设备的位置)等等。这些准则的各种组合用于各种实施例,以便识别要用于将文本转换成语音的一种或多种语言。

[0011] 在一些实施例中,对包括多种语言的文本来说,用于执行文本到语音转换的语言合成器自动地针对不同语言进行切换。在一些实施例中,可逐字符地处理文本并且确定要用于将字符转换成语音的语言。可识别要转换成语音的文本中的字符范围,每一个范围与一种语言相关联。当将文本转换成语音时,每一个范围部分是使用对应于与该范围相关联的的语言的合成器来转换成语音。这允许语言合成器在文本到语音转换期间自动地切换,以便促进文本到语音的转换。

[0012] 在一些实施例中,文本到语音转换可提供为基于web的服务或云服务,或可在软件即服务(Software as a Service, SaaS)模型下提供。

[0013] 根据一些具体实施,电子设备包括一个或多个处理器、存储器和一个或多个程序;一个或多个程序存储在存储器中并且被配置为由一个或多个处理器来执行,并且一个或多个程序包括用于执行上述任何方法和/或技术的操作的指令。根据一些具体实施,计算机可读存储介质中存储有指令,该指令在由电子设备执行时使该设备执行上述任何方法和/或技术的操作。根据一些具体实施,电子设备包括用于执行上述任何方法和/或技术的操作的装置。根据一些具体实施,适用于电子设备的的信息处理装置包括用于执行上述任何方法和/或技术的操作的装置。

[0014] 根据一些具体实施,电子设备包括处理单元,该处理单元被配置为确定用于将文本转换成语音的一组语言,该一组语言包括用于将文本的一部分转换成语音的多种候选语言;与处理单元耦接的显示单元,该显示单元被配置为使多种候选语言被显示;与处理单元耦接的输入接收单元,该输入接收单元被配置为接收指示从多种候选语言中选择第一候选语言的输入;以及该处理单元进一步被配置为将文本转换成语音,该转换包括以第一候选语言将文本的该部分转换成语音。

#### 附图说明

[0015] 图1描绘可包含实施例的系统的简化图;

[0016] 图2描绘简化流程图,该简化流程图描绘根据一些实施例的用于执行文本到语音转换的方法;

[0017] 图3描绘简化流程图,该简化流程图描绘根据一些实施例的用于确定执行文本到语音转换的一组语言的方法;

[0018] 图4描绘简化流程图,该简化流程图描绘根据一些实施例的用于产生文本的语音的方法;

[0019] 图 5 至图 11 描绘根据一些实施例的包括关于文本到语音转换的用户界面的实例，该用户界面可被显示给用户；

[0020] 图 12 是计算机系统的简化框图，该计算机系统可包含根据一些实施例的用于执行文本到语音转换的系统的部件；以及

[0021] 图 13 描绘根据一些实施例的用于执行文本到语音转换的分布式系统的简化图。

[0022] 图 14 例示根据一些实施例的电子设备的功能性框图。

### 具体实施方式

[0023] 在以下说明中，出于解释的目的，陈述具体细节以便提供对本发明的实施例的透彻理解。但是将显而易见的是，可在没有这些具体细节的情况下实践各种实施例。

[0024] 描述了允许用户从多种语言中选择要用于执行文本到语音转换的语言的某些实施例。在一些实施例中，可将多种语言显示给用户。用户可随后从多种语言中选择要使用的特定语言。随后可以用用户选择的语言将一段文本转换成语音。

[0025] 在一些实施例中，在确定可适于对要转换成语音的文本的一部分执行文本到语音转换的多种语言时，可将该多种语言输出给用户。用户可随后从多种语言中选择要使用的特定语言。随后可使用对应于用户选择的语言的合成器，以用户选择的语言将文本的该部分转换成语音。文本的该部分可以是要转换成语音的整个文本或该文本的子集，该子集诸如文本内的一个或多个字符、文本内的一个或多个词、文本内的一个或多个句子等等。

[0026] 图 1 描绘可包含实施例的系统 100 的简化图。在图 1 中所描绘的实施例中，系统 100 包括多个子系统，这些子系统包括文本到语音子系统 102、启发式子系统 104、音频输出子系统 106、输入 / 输出 (I/O) 子系统 108 和存储器子系统 122。可提供一个或多个通信路径，通信路径允许子系统中的一个或多个子系统与彼此通信并且交换数据。图 1 中所描绘的各种子系统可在软件中、硬件中或两者的组合中实现。在一些实施例中，软件可存储在暂态或非暂态计算机可读介质上并且由一个或多个处理器执行。

[0027] 应当理解，图 1 中所描绘的系统 100 可具有除图 1 中所描绘的那些部件之外的部件。另外，图 1 所示的实施例仅仅是可包含本发明的实施例的系统的例子。在一些其他实施例中，系统 100 可具有比图 1 所示的部件更多或更少的部件，可组合两个或更多个部件，或可具有不同的部件配置或布置。在一些实施例中，系统 100 可以是诸如移动电话、智能电话或多功能设备的便携式通信设备的一部分。便携式设备的示例性实施例包括但不限于来自 Apple Inc. (Cupertino, California) 的 iPhone<sup>®</sup>、iPod Touch<sup>®</sup> 和 iPad<sup>®</sup> 设备。在一些其他实施例中，系统 100 还可并入到诸如台式计算机、信息亭等等的其他设备中。

[0028] I/O 子系统 108 提供允许用户与系统 100 进行交互的界面。I/O 子系统 100 可允许系统 100 向用户输出信息。例如，I/O 子系统 108 可包括诸如监视器或屏幕的显示设备。I/O 子系统 108 还可允许用户向系统 100 提供输入。I/O 子系统 108 可包括允许用户向系统 100 提供输入的一个或多个输入设备。输入设备可包括但不限于鼠标、指针、键盘或其他输入设备。在一些实施例中，I/O 子系统 108 可包括不但向用户显示信息而且从用户接收输入的触敏界面（有时也称为触摸屏）。

[0029] 存储器子系统 122 可被配置为存储由本发明的一些实施例使用的数据和指令。在一些实施例中，存储器 122 可包括易失性存储器，诸如随机存取存储器或 RAM（有时称为系

统存储器)。由系统 100 的一个或多个处理器执行的指令或代码或程序可存储在 RAM 中。存储器 122 还可包括非易失性存储器,诸如一个或多个存储磁盘或存储设备、闪存存储器或其他非易失性存储器设备。

[0030] 在一些实施例中,协同工作的启发式子系统 104 和文本到语音子系统 102 负责执行与文本到语音转换相关的处理。启发式子系统 104 被配置为接收要转换成语音的文本 128。可通过 I/O 子系统 108 或从其他来源接收文本 128。例如,系统 100 的用户可使用 I/O 子系统 108 来选择文本,然后开始所选择文本的文本到语音转换过程。

[0031] 在一些实施例中,启发式子系统 104 分析文本 128 以确定适用于将文本 128 转换成语音的一种或多种语言。启发式子系统 104 可从可供系统 100 用来执行文本到语音转换的一组语言中确定一种或多种语言。可供系统 100 使用的一组语言可取决于可用的一种或多种语言合成器 112,其中每个合成器被配置用于特定语言。在一个实施例中,可将识别各种语言合成器 112 和它们的对应语言的信息提供给启发式子系统 104,并且启发式子系统 104 可使用该信息来确定用于将一段特定文本转换成语音的一种或多种语言。

[0032] 启发式子系统 104 可使用各条信息 110 来确定适用于文本 128 的一种或多种语言。在一些实施例中,可基于各种准则来执行基于启发法的分析,以便识别用于将文本 128 转换成语音的一种或多种语言。以下提供与可由启发式子系统 104 使用的各种准则和可使用某些准则的方式相关的另外细节。

[0033] 除了确定适用于将文本 128 转换成语音的一种或多种语言之外,在一些实施例中,启发式子系统 104 还被配置为确定是否存在任何语言歧义。出于本公开的目的,如果确定文本 128 的一部分可使用可用于文本到语音转换的多种语言来转换成语音,那么存在语言歧义。因为每种语言对应于特定语言合成器,所以当确定多个语言合成器可用于将文本的同一部分转换成语音时,存在语言歧义。可能造成语言歧义情况的文本 128 的该部分可以是整个文本 128 或文本 128 的子集,该子集诸如文本 128 内的一个或多个字符、文本 128 内的一个或多个词、文本 128 内的一个或多个句子等等。

[0034] 例如,在一个实施例中,启发式子系统 104 可基于由启发式子系统 104 执行的分析来确定:对于文本 128 内的一组词,可使用英语语言合成器或法语语言合成器来将这些词转换成语音。因为不止一种语言是用于将文本的同一部分转换成语音的候选语言,所以启发式子系统 104 可确定:这些词组存在语言歧义。

[0035] 又如,启发式子系统 104 可检测一组中文语言字符在要转换成语音的文本 128 内的出现。在处理这个字符组时,启发式子系统 104 可确定有四个单独的语言合成器(这些语言合成器能够读取中文字符并将它们转换成语音)可用,即:普通话语言合成器、广东话语言合成器、上海话语言合成器和台湾话语言合成器。换句话说讲,启发式子系统 104 确定有四种语言可用于将中文字符组转换成语音。同样在这种情况下,启发式子系统 104 可确定:对于中文字符组存在语言歧义。

[0036] 再如,启发式子系统 104 可基于由启发式子系统 104 执行的启发式分析来确定:对于整个文本 128,有两个单独的语言合成器(这些语言合成器能够读取文本 128 并将文本转换成语音)可用,即:法国法语语言合成器和加拿大法语语言合成器。换句话说讲,启发式子系统 104 确定:有两种语言可用于将文本 128 转换成语音。同样在这种情况下,启发式子系统 104 可确定:对于整个文本存在语言歧义。

[0037] 另外,一段特定文本可包括多个部分,每个部分都引起语言歧义。因此,在一些实例中,可确定要转换成语音的文本 128 具有多个语言歧义。例如,要转换成语音的文本可包括第一部分,该第一部分引起英语和法语之间的语言歧义。同一文本可包括不同于第一部分的第二部分,该第二部分引起普通话、广东话、上海话和台湾话之间的语言歧义。这表示英语和法语之间的第一语言歧义以及普通话、广东话、上海话和台湾话之间的第二语言歧义。因此,要转换成语音的文本中可能存在零个或更多个语言歧义。

[0038] 在执行语言歧义处理时,启发式子系统 104 随后可基于是否检测到存在语言歧义来使信息被显示给用户。在启发式子系统 104 未检测到任何语言歧义情况的情况下,启发式子系统 104 可使信息通过 I/O 子系统 108 被显示给用户,从而允许用户开始文本 128 的文本到语音转换。例如,在一个实施例中,可将“发音 (Speak)”选项显示给用户。当启发式子系统 104 接收到对用户已选择“发音”选项的指示时,可向文本到语音子系统 102 发送信号以开始文本到语音转换。启发式子系统 104 还可向文本到语音子系统 102 发送信息,该信息指示启发式子系统 104 已确定的要用于文本到语音转换的一种或多种语言。文本到语音子系统 102 可随后使用对应于由启发式子系统 104 识别的一种或多种语言的语言合成器来执行文本 128 的文本到语音转换。

[0039] 在检测到语言歧义的情况下,启发式子系统 104 可使信息被显示给用户以解决歧义。在一些实施例中,启发式子系统 104 可使“发音…… (Speak……)”选项被显示给用户。在用户选择“发音……”选项时,启发式子系统 104 可针对每个语言歧义使信息通过 I/O 子系统 108 被显示给用户,该信息识别造成歧义的多种语言。可提示用户从造成语言歧义的多种语言中选择用于执行文本到语音转换的特定语言。例如,如果语言歧义是在英语与法语这两种语言之间,那么在一个实施例中可将菜单显示给用户,该菜单将英语和法语识别为候选语言并且允许用户选择这两种语言之一来执行文本到语音转换。在用户选择特定语言时,启发式子系统 104 可向文本到语音子系统 102 发送信号,以便开始使用用户选择的语言进行的文本到语音转换。

[0040] 如以上所指出,启发式子系统 104 可针对文本到语音转换向文本到语音子系统 102 发送识别要用于文本到语音转换的一种或多种语言的信息。在检测到语言歧义的情况下,一种或多种语言可包括用户选择的语言。在未检测到语言歧义的情况下,可向文本到语音子系统 102 发送识别一个或多个“无歧义”语言(即,不会造成语言歧义的语言)的信息。在一些实例中,传送到文本到语音子系统 102 的信息可包括用户选择的语言与无歧义语言的组合。文本到语音转换可随后使用由启发式子系统 104 识别的一种或多种语言将文本 128 转换成语音。作为这种文本到语音转换的一部分,对于文本 128 中造成语言歧义的一部分,文本到语音子系统 102 使用对应于用户针对该部分选择的语言的语言合成器将该部分转换成语音。

[0041] 文本到语音子系统 102 被配置为使用由启发式子系统 104 识别的一种或多种语言将文本 128 转换成语音。在图 1 中描绘的实施例中,文本到语音转换包括一组语言合成器 112 和合成器控制器 120。图 1 中描绘的文本到语音子系统 102 的各种部件可在软件中、硬件中或两者的组合中实现。在一些实施例中,文本到语音子系统 102 可具有比图 1 所描绘的那些部件更多或更少的部件。

[0042] 在一些实施例中,语言合成器 112 包括可用于执行文本到语音转换的一个或多个

合成器。每个合成器可对应于可用于文本到语音转换的特定语言（例如，用于第一语言 114 的 TTS 合成器、用于第二语言 116 的 TTS 合成器、用于第三语言 118 的 TTS 合成器等）。因此，语言合成器 112 中的合成器组识别可用于使用系统 100 进行的文本到语音转换的各种语言。语言的例子包括但不限于英语、法语（在一些实施例中，法国法语和加拿大法语可能作为两个单独的语言是可用的）、德语、日语、普通话、广东话、台湾话、上海话等等。

[0043] 在一些实施例中，合成器控制器 120 被配置为读取文本 128 并且使用一种或多种语言合成器 112 将所读取文本转换成语音。在图 4 中描绘并且在以下描述文本到语音子系统 102 可将文本 128 转换成语音的方式的实例。文本到语音子系统 102 所生成的语音可以呈音频信号的形式，该音频信号可随后使用可包括扬声器的音频输出子系统 106 作为语音 124 输出。在一些实施例中，可将音频信号作为音频文件 126（例如 .wav 文件）存储在系统 100 的存储器子系统 122 中。音频文件 126 可随后使用音频输出子系统 106 作为语音输出。

[0044] 图 1 中描绘的系统 100 可以各种配置来提供。在一些实施例中，系统 100 可被配置为分布式系统，其中系统 100 的一个或多个部件可通过一个或多个网络分布和耦接。图 13 描绘根据一些实施例的用于执行文本到语音转换的分布式系统 1300 的简化图。在图 13 中描绘的实施例中，启发式子系统 104 和文本到语音子系统 102 设置在服务器 1302 上，该服务器通过网络 1306 与远程客户端设备 1304 可通信地耦接。

[0045] 网络 1306 可包括一个或多个通信网络，通信网络可以是互联网、局域网 (LAN)、广域网 (WAN)、无线或有线网络、内联网、专用网络、公用网络、交换网络或任何其他合适的通信网络。网络 1306 可包括许多互连系统和通信链路，通信链路包括但不限于硬连线链路、光学链路、卫星或其他无线通信链路、波传播链路，或用于传送信息的任何其他方式。各种通信协议可用于促进通过网络 1306 传送信息，通信协议包括但不限于 TCP/IP 协议、HTTP 协议、可扩展标记语言 (XML)、无线应用协议 (WAP)、行业标准组织正在开发的协议、供应商专用协议、自定义协议和其他协议。

[0046] 在图 13 中描绘的配置中，客户端设备 1304 的用户可使用设备 1304 来选择要转换成语音的文本。在选择文本时，可通过网络 1306 将文本传送到服务器 1302。位于服务器 1302 上的启发式子系统 104 随后可执行分析，以便确定要用于将文本转换成语音的一种或多种语言。在一些实施例中，还可将启发式子系统 104 用来执行分析的信息从客户端设备 1304 传送到服务器 1302。在一些实施例中，服务器 1302 还可存储用于分析的信息的一部分，或可从可能通过网络 1306 连接到服务器 1302 的其他来源收集信息。作为分析的一部分，启发式子系统 104 还可确定是否存在任何语言歧义。启发式子系统 104 可随后使客户端设备 1304 显示适当的用户界面。例如，如果确定没有语言歧义，那么“发音”选项可通过客户端设备 1304 的输出界面显示给用户。如果检测到一种或多种语言歧义，那么启发式子系统 104 可使“发音……”通过客户端设备 1304 显示给用户。在用户选择“发音……”时，语言列表可被显示给用户。用户随后可使用客户端设备 1304 来选择特定语言。随后可将信息传送到服务器 1302 上的文本到语音子系统 102，该信息识别任何用户选择的语言和由启发式子系统 104 识别的任何其他无歧义语言。文本到语音子系统 102 随后可产生文本的语音，并且该语音可被传送到客户端设备 1304 以便使用客户端设备 1304 的扬声器输出给用户。

[0047] 在图 13 中描绘的配置中,文本到语音子系统 102 和启发式子系统 104 定位成远离客户端设备 1304。包括文本到语音子系统 102 和启发式子系统 104 的服务器 1302 为客户端 1304 提供文本到语音转换服务。在一些实施例中,服务器 1302 可为多个客户端提供文本到语音转换服务。可同时或以序列化方式为多个客户端服务。在一些实施例中,服务器 1302 所提供的服务可提供为基于 web 的服务或云服务,或在软件即服务 (SaaS) 模型下提供。

[0048] 应当理解,不同于图 13 中描绘的分布式系统 1300 的各种不同分布式系统配置是可能的。例如,在一些实施例中,文本到语音子系统 102 和启发式子系统 104 也可定位成远离彼此。图 13 所示的实施例因此仅仅是可包含一些实施例的系统的例子并且并非旨在进行限制。

[0049] 图 2 描绘简化流程图 200,该简化流程图描绘根据一些实施例的用于执行文本到语音转换的方法。图 2 中所描绘的处理可在由一个或多个处理器的软件(例如,代码、指令、程序)中、硬件中或软件和硬件的组合中实施。软件可存储在非暂态计算机可读存储介质上。图 2 中所描绘的特定系列的处理步骤并非旨在进行限制。

[0050] 如图 2 中所描绘,方法可在 202 处在接收要转换成语音的文本时开始。存在识别要转换成语音的文本的各种方式。在一些实施例中,仅用户使用例如 I/O 子系统 108 对文本进行的选择即可触发图 2 中所描绘的流程图 200。例如,用户可使用诸如 iPhone® 或 iPad® 的设备上的触摸屏来选择文本,并且这可使得图 2 中的处理得以开始。

[0051] 在 204 处,执行分析以识别适用于并且可用于将在 202 中接收的文本转换成语音的一组一种或多种语言。在一些实施例中,在 204 中执行的分析可包括基于系统 100 可用的或向该系统提供的信息进行的基于启发法的分析。以下关于图 3 中所描绘的流程图 300 来描述在一些实施例中与在 204 中执行的分析有关的细节。

[0052] 在 206 处,确定是否存在语言歧义。出于本公开的目的,如果在 202 中接收的文本的一部分可使用可用于文本到语音转换的多种语言来转换成语音,那么存在语言歧义。因为每种语言对应于特定语言合成器,所以当多个语言合成器可用于将在 202 中接收的文本的同一部分转换成语音时,存在语言歧义。

[0053] 文本 128 中引起语言歧义的一部分可以为整个文本或文本的子集,该子集诸如在 202 中接收的文本内的一个或多个字符、在 202 中接收的文本内的一个或多个词、在 202 中接收的文本内的一个或多个句子等等。例如,如先前所描述,如果确定文本 128 内的一个或多个词可以英语(使用英语语言合成器)或以法语(使用法语语言合成器)来转换成语音,那么可能存在语言歧义。又如,如果在 206 中确定存在用于将一组汉语语言字符转换成语音的多个语言候选者(例如,普通话、广东话、上海话和台湾话)和对应语言合成器,那么可能存在语言歧义。另外,如先前所描述,一段特定文本可能引起多个语言歧义。这可能在(例如)要转换成语音的文本包括各自引起语言歧义的多个部分时发生。

[0054] 因此,要转换成语音的文本可致使零个或更多个语言歧义出现。在 206 中执行用以检测任何语言歧义的处理。在一些实施例中,关于语言歧义和在 204 中识别的一种或多种语言,可能出现以下情景:

[0055] (1) 识别用于将整个文本转换成语音的单个语言,在这种情况下不存在语言歧义。

[0056] (2) 识别多种语言,但不存在语言歧义。这可能在要转换成语音的文本包括多个

部分并且针对每个部分确定无歧义语言时发生。例如,针对在 202 中接收的文本的第一部分可确定第一语言,并且针对在 202 中接收的文本的第二部分可确定第二语言,其中第一部分和第二部分不重叠。因为针对文本的每个部分仅确定一种语言,所以不存在语言歧义。应注意的是,多种语言的确定自身并不引起语言歧义。出于解释的目的,不是一组歧义语言的一部分的语言称为无歧义语言。

[0057] (3) 识别包括一种或多种无歧义语言和引起一种或多种语言歧义的语言的多种语言。例如,针对在 202 中接收的文本的第一部分可确定第一语言(无语言歧义),并且针对在 202 中接收的文本的第二部分可确定多种语言,从而导致第二部分的语言歧义。

[0058] (4) 识别不包括任何无歧义语言而包括引起一种或多种语言歧义的语言的多种语言。例如,针对在 202 中接收的文本的第一部分可确定第一语言和第二语言,并且针对在 202 中接收的文本的第二部分可确定多种语言,从而导致引起第一部分和第二部分的语言歧义的两组语言。

[0059] (5) 无语言被识别。这指示系统无法找到可读取在 202 中接收的文本并且将该文本转换成语音的语言合成器。在这种情况下(图 2 中未示出),处理结束并且不能对所选择文本执行文本到语音转换的消息可被显示给用户。

[0060] 在一些实施例中,如果存在任何语言歧义,那么处理继续进行 216,否则处理继续进行 208。

[0061] 如果在 206 中确定不存在语言歧义,那么可在 208 处将允许用户开始在 202 中接收的文本的文本到语音转换的信息显示给用户。例如,在一些实施例中,可将“发音”选项显示给用户。在于 210 处接收指示选择“发音”选项的用户输入时,可在 212 处使用在 204 中确定的一种或多种语言和该一种或多种语言的对应语言合成器来执行在 202 中接收的文本的文本到语音转换。在 214 处,可将在 212 中生成的语音输出给用户。例如,可使用诸如扬声器的音频输出设备来输出语音。

[0062] 回到 206,如果相反地确定存在语言歧义,那么可将允许用户选择特定语言并且解决歧义的信息显示给用户例如,可在 216 处将“发音……”选项显示给用户。“发音……”选项指示用户在可开始文本到语音转换之前必须提供一些输入。

[0063] 在 218 处,可接收指示用户已经选择“发音……”选项的信息。在 220 处,对于每个语言歧义,可将识别引起语言歧义的多种语言并且允许用户从多种语言中选择特定语言来执行文本到语音转换的信息显示给用户。因此,在 220 中,可将允许用户提供输入以解决所有语言歧义的信息显示给用户。例如,如果语言歧义是在英语和法语这两种语言之间,那么可将识别英语和法语以及允许用户选择这两种语言之一来执行文本到语音转换的信息显示给用户。

[0064] 如果存在多个语言歧义,那么可针对每个语言歧义显示信息,使得用户可选择特定语言来解决该语言歧义。例如,除了英语和法语之间的语言歧义之外,如果由于普通话、广东话、上海话和台湾话而存在另一个语言歧义,那么可将识别英语和法语以及允许用户选择这两种语言之一来执行文本到语音转换的信息显示给用户,并且另外,可将识别普通话、广东话、上海话和台湾话以及允许用户选择这四种语言之一来执行文本到语音转换的信息显示给用户。

[0065] 在 222 处,可从用户接收识别由用户选择来执行文本到语音转换的一种或多种语

言的输入。例如,可在 222 中从用户接收指示用户已经选择英语(在法语和英语之间)和普通话(在普通话、广东话、上海话和台湾话之间)的信息。

[0066] 在 224 处,使用在 222 中由用户选择的任何语言和在 204 中识别的任何其他无歧义语言将在 202 中接收的文本转换成语音。对于每种语言,使用对应语言合成器来以该语言将文本的一部分转换成语音。例如,对于文本中导致英语和法语之间的语言歧义的一部分,如果用户在 222 中选择法语,那么使用法语语言合成器将文本的该部分转换成语音。生成的语音可随后在 214 处通过诸如扬声器的音频输出设备输出给用户。

[0067] 如上所述,在一些实施例中,当检测到语言歧义时,允许用户通过从引起语言歧义的多种语言中选择特定语言来解决语言歧义。以这种方式,允许用户根据用户选择选取用于文本到语音转换的语言。

[0068] 如以所述,在 204 中,执行分析以确定可用于将文本转换成语音的一种或多种语言。图 3 描绘简化流程图 300,该简化流程图描绘根据一些实施例的用于确定执行文本到语音转换的一种或多种语言的方法。图 3 中所描绘的处理可在由一个或多个处理器执行的软件(例如,代码、指令、程序)中、硬件中或软件和硬件的组合中实施。软件可存储在非暂态计算机可读存储介质上。图 3 中所描绘的特定系列的处理步骤并非旨在进行限制。在一个实施例中,处理可由图 1 中所描绘的启发式子系统 104 执行。

[0069] 在 302 处,各种准则可用于确定要用于将文本转换成语音的一组语言。在一些实施例中,以下准则可用来(但不限于)在 302 中识别用于将文本转换成语音的语言候选项:

[0070] (1) 加载的键盘——包含系统 100 的设备(诸如 iPhone®)可具有对应于加载的一种或多种语言的一个或多个键盘。例如, iPhone®可具有加载的法语键盘和英语键盘。此信息可用于识别要用于执行文本到语音转换的语言。这里假定,如果用户已经加载特定语言键盘,那么用户可能想要使文本到语音转换使用该特定语言来执行的可能性很高。

[0071] (2) 语言标签——在一些实施例中,识别用于要转换成语音的文本的一种或多种语言的标签可与该文本相关联。多个标签可用于识别多种语言。在一些实施例中,诸如 XML 或 HTML 的标签语言可用于提供标签。例如,在 HTML 中,“lang”属性标签可用于指示文本的语言。在 XML 中,“xml:language”属性标签可用于指示文本的语言。例如,标签常常由经常将标签插入网页中以传达与网页相关的附加信息的网页设计师使用。这些标签中的一个或多个可用于识别与在网页上提供的内容相关联的语言。因此,如果由用户选择来转换成语音的文本是来自网页,那么可分析网页以确定是否存在与所选择文本相关联、识别用于文本或文本的部分的任何语言的任何标签。标签信息如果可用,那么可随后用于识别要用于文本到语音转换的一种或多种语言。在一些实施例中,根据 RFC 5646 提供的标签也可用于识别要使用的语言。

[0072] (3) 语言分析工具(或语言标记器)——可分析一段文本并且确定一个或多个相关联的语言的语言分析工具是可用的。由此类工具识别的语言可随后用于识别用于执行文本或文本的部分的文本到语音转换的一种或多种语言。一般来讲,来自语言分析工具的结果的可靠性取决于正被分析文本的长度——文本越长,结果一般越可靠。因此,给予从语言分析工具获得的结果的权重可取决于正被分析文本的长度。一般来讲,如果文本短于 300 个字,那么结果不被视为非常可靠。



[0073] (4) 专用数据——一些应用程序可提供或具有与这些应用程序的内容相关联的信息,该信息可用于确定用于由那个应用程序产生的文本的语言。因此,由与文本相关联的应用程序提供的数据可用于确定要用于文本的一种或多种语言。例如,对于通信应用程序中的文本消息,可分析发送方电话号码以确定所使用的国家代码和对应国家,该国家代码和对应国家继而可用于确定要用于文本消息的语言。又如,对于电子邮件消息,发送方国家代码或发起电子邮件所在的国家的域名可被确定,并且用于确定要用于电子邮件或电子邮件的一部分的文本到语音转换的语言。再如,对于选自网页的文本,与托管网页的网站相关联的域可用于确定要用于将文本转换成语音的语言。

[0074] (5) 基于 Unicode 或其他字符的编码——Unicode 标准是设计来支持撰写的各种不同语言的文本的全球交换、处理和显示的字符编码系统。Unicode 标准针对每一字符提供唯一数字,而不考虑语言。因此,对于已经根据 Unicode 标准编码的文本,与文本中的每个字符相关联的 Unicode 数字可用于确定要与字符相关联的语言。可对要转换成语音的文本中的字符执行逐字符 Unicode 分析,以便确定要用于文本到语音转换的语言。在一些实施例中,与文本中的字符相关联的其他字符相关的编码信息可用于识别与字符相关联的语言。

[0075] (6) 用户偏好和历史——如果与用户偏好或过去语言选择相关的信息可用,那么可用于确定要用于文本到语音转换的语言。例如,用户偏好或历史信息可指示:用户的语言偏好被设定为法语或者用户已经重复地选取法语作为选择的语言。此信息可随后用于识别要用于文本到语音转换的语言,并且可选择法语来包括在用于文本到语音转换的候选语言列表中。

[0076] (7) 文本的来源——如果与文本的来源相关的信息是可用的,那么该信息可用于识别要用于文本到语音转换的一种或多种语言。例如,如果选择来进行文本到语音转换的文本是来自用户接收的电子邮件,那么识别电子邮件的发送方的信息可用于确定要使用的语言。例如,如果选择来进行文本到语音转换的文本是电子邮件的一部分,那么电子邮件的发送方信息可用于确定要用于文本到语音转换的特定语言。

[0077] (8) 国家代码顶级域(ccTLD)——国家代码顶级域(ccTLD)是一般用于或保留给国家、主权国或属地的互联网顶级域。这些一般为两个字母长,并且可与诸如浏览器或域的应用程序相关联。ccTLD 的例子包括但不限于:用于美国的 .us、用于日本的 .jp、用于法国的 .fr 等。当可用于正被转换成语音的文本时,这些 ccTLD 可用来识别对应国家和要用于将文本转换成语音的对应语言。

[0078] (9) 全球定位系统(GPS)信息——许多设备(尤其是移动设备)现在包含能够提供关于设备的位置以及因此关于设备的用户的位置的信息的 GPS 芯片。因此,如果指示文本到语音转换正被执行所针对的用户的位置或指示用于选择用于转换成语音的文本的设备的位置的 GPS 信息是可用的,那么此信息可用于确定要用于文本到语音转换的一种或多种语言。例如,如果 GPS 信息指示用户在墨西哥,那么西班牙语可被选择为用于执行文本到语音转换的语言。

[0079] 以上识别的准则列表并不意味着是详尽的或限制性的。在一些其他实施例中,可使用比上述更多或更少的准则。另外,使用各种准则的方式在实施例之间也可改变。例如,在一个实施例中,每个准则可独立于其他准则用于识别用于将文本转换成语音的零个或更

多个可能的语言候选项。在此类实施例中,在 302 中,可从针对每个准则执行的分析识别一组零个或更多个语言候选项。在另一个实施例中,在 302 中,两个或更多个准则可结合来识别候选语言。基于准则的处理可并行地、以串行化方式或以并行和串行化方式的组合来执行。

[0080] 在 302 中执行的处理产生零种或更多种语言的列表作为用于将文本转换成语音的潜在候选项。在一些实施例中,如果在列表中存在多种语言,或者如果在此列表中存在任何语言歧义,那么可在 304 中尝试解决语言歧义,或者如果不能解决语言歧义,那么可尝试减少语言歧义中涉及的语言的数目。例如,如果在 302 中生成的列表包括由多种语言引起的语言歧义,那么在 304 中尝试取决于基于启发法的分析来解决这个语言歧义。如果不能解决语言歧义,那么取决于基于启发法的分析来尝试减少引起歧义的语言的数目。因此,在 304 中可执行过滤在 302 中生成的候选语言列表的处理。在完成 304 中的处理之后,处理继续进行图 2 中的 206。

[0081] 各种不同的准则可用于执行 304 中的处理。在一些实施例中,用于 304 中的处理的准则可包括在 302 中使用(和上述)的准则或在 302 中没有使用的其他准则中的一个或多个。在一个实施例中,加权可用来减少候选语言列表。在一个此类实施例中,权重可被计算并且被分配给在 302 中使用的每个准则。与准则相关联的权重可随后用来保留或者过滤掉基于该准则识别的一种或多种语言。例如,如果与特定准则相关联的权重不满足某一配置阈值,那么可过滤掉使用该准则识别的语言。

[0082] 例如,在一个实施例中,在 302 中生成的语言列表可包括基于从语言分析工具获得的结果所确定的一种或多种语言。如以上所指出,来自此类工具的结果的可靠性与所分析文本的长度成正比,并且可能是不可靠的,例如,如果文本的长度小于 300 个字。因此,在 304 中,可进行关于所分析文本的长度的确定,并且基于所确定的长度,可将权重分配给这个准则,使得如果文本的长度不超过某一阈值,那么可过滤掉使用这个准则识别的语言。例如,如果确定所分析文本的长度小于 300 个字,那么可过滤掉基于来自这个工具的结果所识别的一种或多种语言。

[0083] 又如,与用户偏好和历史相关的信息可用来过滤在 302 中生成的候选语言列表。例如,用户偏好或历史信息可指示用户的语言偏好被设定为法语或者用户已经重复地选取法语作为选择的语言。因此,如果存在英语和法语之间的语言歧义,那么在 304 中,可通过选择法语而不是英语作为要用于文本到语音转换的语言来解决歧义。在这种情景下,英语将从候选语言列表中被过滤掉。

[0084] 再如,用户偏好可包括场所偏好。例如,用户的场所偏好可被设定为台湾。在此类情景下,如果存在普通话、广东话、上海话和台湾话这四种语言之间的语言歧义,那么作为用户的场所偏好的结果,台湾话将被识别为要使用的语言,并且普通话、广东话和上海话可被过滤掉。

[0085] 在一些实施例中,某些准则可在识别要用于文本到语音转换的语言中被给予比其他因素更大的权重。例如,在一些实施例中,各种因素的重要性可如下(以重要性降序排列):Unicode 或其他字符相关的信息、加载的键盘、与文本相关联的标签和来自语言分析工具的结果。

[0086] 如先前所描述,在一些实施例中,要用于文本到语音转换的一种或多种语言的列

表由启发式子系统 104 转发到文本到语音子系统 102。这个列表可包括零个或更多个无歧义语言和（如果检测到任何语言歧义）一个或多个用户选择的语言。文本到语音子系统 102 随后在图 2 中的 212 或 224 中使用这些语言来执行文本到语音转换。

[0087] 图 4 描绘简化流程图 400, 该简化流程图描绘根据一些实施例的用于产生文本的语音的方法。图 4 中所描绘的处理可在由一个或多个处理器执行的软件（例如, 代码、指令、程序）中、硬件中或软件和硬件的组合中实施。软件可存储在非暂态计算机可读存储介质上。图 4 中所描绘的特定系列的处理步骤并非旨在进行限制。在一个实施例中, 处理可由图 1 中所描绘的文本到语音子系统 102 执行。

[0088] 在 402 处, 可读取来自要转换成语音的文本的第一字符。在 404 处, 确定要用于将所读取字符转换成语音的语言。在一个实施例中, 语言选自启发式子系统 104 提供的语言列表。在 404 中选择来用于所读取字符的语言可以为无歧义语言, 或者如果字符是从文本中引起语言歧义的一部分读取, 那么该语言为由用户选择来用于该部分的语言。

[0089] 在 406 处, 将所选择语言与所读取字符相关联。各种技术可用于将所选择语言与字符相关联。在一个实施例中, 相关联信息可存储在存储器中。

[0090] 在 408 处, 确定要转换成语音的文本中的所有字符是否已经被处理。如果在 408 中确定所有字符已经被处理, 那么这指示整个文本已经被处理, 并且处理继续进行 416。如果在 408 中确定所有字符尚未都被处理, 那么这指示文本中存在需要处理的另外的字符, 并且在 410 处从文本中读取下一个字符。

[0091] 在 412 处, 确定当前所选择的语言（即, 选择用于先前所读取的字符的语言）是否可用来发出在 410 中读取的字符的读音。在一些实施例中, 目前所选择的语言可能由于各种原因而不用来发出在 410 中读取的字符的读音。在一个实例中, 对应于与先前所读取字符相关联的语言的语言合成器可能无法读取并且发出在 410 中读取的下一个字符的读音。例如, 先前字符可能已经是英语字符, 并且在 410 中读取的下一个字符是汉语字符。在另一个实例中, 先前字符可能已经与无歧义语言相关联, 并且在 410 中读取的下一个字符是来自文本的歧义部分并且将与用户选择的语言相关联, 其中用户选择的语言不同于用于先前字符的无歧义语言。

[0092] 如果在 412 中确定当前所选择的语言可用来发出在 410 中读取的字符的读音, 那么处理继续进行 406。如果在 412 中确定当前所选择的语言不能用来读出在 410 中读取的字符, 那么在 414 处, 确定要用于将在 410 中读取的字符转换成语音的新语言。新语言可以为无歧义语言或用户选择的语言。处理随后继续进行 406, 其中将所选择的语言与字符相关联。

[0093] 以上述方式, 处理继续直到文本中的所有字符已经被处理并且已将语言与所有字符相关联为止。在所有字符已经被处理之后, 在 416 处, 确定文本内具有公用相关联语言的一个或多个字符的范围, 并且针对每个范围, 将与范围内的一个或多个字符相关联的语言与该范围相关联。例如, 文本可包括具有使英语作为相关联语言的字符的第一部分和具有以日语作为相关联语言的字符的第二部分。在此实例中, 第一部分中的字符可表示以英语作为相关联语言的字符的第一范围, 并且第二部分中的字符表示以日语作为相关联语言的第二个范围。

[0094] 在 418 处, 产生文本的语音。文本到语音转换处理包括: 针对在 416 中确定的每个

范围,使用对应于与该范围相关联的语言的语言合成器来将该范围中的字符转换成语音。例如,对于具有英语作为相关联语言的字符范围,英语语言合成器用来将该字符范围转换成语音。在 418 中生成的文本的语音可随后在 420 中被输出。例如,可使用诸如扬声器的音频输出设备来输出生成的语音。

[0095] 在图 4 中所描绘和描述的实施例中,逐字符地读取要转换成语音的文本,并且将一种语言与每个字符相关联。当在文本中遇到不能由当前所选择的语言转换成语音的字符时,自动选择可读取字符并且将该字符转换成语音的新语言,并且将该新语言与该字符相关联。以这种方式,可基于逐字符分析将多种语言与要转换成语音的文本的不同部分相关联。此外,此类分析允许自动地选择并且切换语言合成器,以便促进文本到语音的转换。例如,如果英语语言合成器当前被选择并且用来执行文本到语音转换,那么当在文本中遇到汉语字符范围时,用于文本到语音转换的语言合成器自动地从英语语言合成器切换到可将汉语字符转换成语音的语言合成器,并且新选择的语言合成器用来将汉语字符转换成语音。如果稍后,在正被转换成语音的文本中遇到另一个英语字符范围,那么用于文本到语音转换的语言合成器可自动地从汉语语言合成器切换到英语语言合成器,并且用来将英语字符范围转换成语音。以这种方式,在一些实施例中,用于文本到语音转换的语言合成器自动地且适当地被切换。

[0096] 图 5 至图 11 描绘根据一些实施例的包括可被显示给用户的关于文本到语音转换的用户界面的实例。这些图中描绘的用户界面可例如显示在如 iPhone<sup>®</sup>、iPad<sup>®</sup> 等设备上。图 5 至图 11 中描绘的实例并非旨在具有限制性。

[0097] 在图 5 中描绘的实例中,用户已选择文本 502。响应于用户选择,标注菜单 504 被显示给用户。标注菜单 504 包括“发音”选项 506。“发音”选项 506 的显示指示:启发式子系统 104 对文本 502 的分析未指示任何语言歧义。用户可通过长按“发音”选项 506 来开始文本 502 的文本到语音转换。

[0098] 在图 6 中描绘的实例中,用户已选择与图 5 中的文本 502 相同的文本 602。响应于用户选择,标注菜单 604 被显示给用户。然而,与图 5 不同,图 6 中的标注菜单 604 显示“发音……”选项 606。“发音……”选项 606 的显示指示:启发式子系统 104 对文本 602 的分析指示存在语言歧义。用户可长按图 6 中的选项 606,并且这可使图 7 中描绘的标注菜单 702 被显示。如图 7 所示,标注菜单 702 识别造成语言歧义的两种语言:英语 704 和法语 706。用户随后可选择英语 704 或法语 706 之一,并且在选择时开始使用选择的语言和对应语言合成器进行文本 602 的文本到语音转换。

[0099] 应注意,虽然在图 5 和图 6 中选择相同的文本来进行文本到语音转换,但是所得的启发式分析产生两个不同结果。这可能是因为在图 5 中描绘的实例中,语言标签器可能已指示英语作为语言并且任何其他启发式标准未识别其他语言。然而,在图 6 中描绘的实例中,语言标签器可能已得出英语作为候选语言,但同时法语键盘可能已加载,因此得出法语作为另一个可能的候选语言,从而引起语言歧义。为了解决语言歧义,如图 7 中一样将两种语言显示给用户,并且用户随后可选择要使用的具体语言。例如,如果用户选择英语 704,那么使用英语语言合成器执行文本 602 的文本到语音转换。另一方面,如果用户选择法语 706,那么使用法语语言合成器执行文本 602 的文本到语音转换。

[0100] 在图 8 中描绘的实例中,用户已选择文本 802。响应于用户选择,包括“发音……”

选项 806 的标注菜单 804 被显示给用户。“发音……”选项 806 的显示指示：启发式子系统 104 对文本 802 的分析指示存在语言歧义。用户可长按图 8 中的选项 806，并且这可使图 9 中描绘的标注菜单 902 被显示。如图 9 所示，标注菜单 902 识别两个用户可选择的选项：英语 & 和日语 904 以及法语和日语 906。在这种情景下，英语和法语两种语言造成语言歧义，并且日语已识别为无歧义语言。这可能是由于，文本 802 的分析指示：该文本包括两个段落 808 和 810。对于段落 808，英语和法语均可识别为用于文本到语音转换的候选语言，从而引起语言歧义。对于段落 810，仅日语可识别为语言，并且因此，关于段落 810 不存在歧义。用户随后可选择英语 & 和日语 904 或法语 & 和日语 906 之一，并且在选择时开始使用选择的语言 and 对应语言合成器进行文本 802 的文本到语音转换。如果选择英语 & 和日语 904，那么使用英语语言合成器将段落 808 中的文本转换成语音，并且使用日语语言合成器将段落 810 中的文本（包括日语字符）转换成语音。

[0101] 在图 9 中描绘的实例中，当执行文本到语音转换时，文本到语音子系统 102 使用英语语言合成器，直到在文本中遇到日语字符，并且在遇到日语字符时自动切换为使用日语语言合成器。

[0102] 在图 10 中描绘的实例中，用户已选择包括中文字符的文本 1002。响应于用户选择，包括“发音……”选项 1006 的标注菜单 1004 被显示给用户。“发音……”选项 1006 的显示指示：启发式子系统 104 对文本 1002 的分析指示存在语言歧义。例如，启发式子系统 104 可确定文本 1002 包括可用普通话、广东话或台湾话发音的中文字符，并且这个语言歧义不能使用启发法来解决。当用户长按图 10 中的选项 1006 时，如图 11 中所描绘的标注菜单 1102 可被显示。如图 11 所示，标注菜单 1102 包括三个用户可选择的选项，这些选项对应于造成歧义的三种语言，即普通话 1104、广东话 1106 和台湾话 1108。用户可随后从标注菜单 1102 选择普通话 1104、广东话 1106 和台湾话 1108 之一，并且使用对应于用户选择的语言合成器将文本 1002 转换成语音。

[0103] 在一些实施例中，用户对特定语言的选择可存储为用户偏好信息的一部分。例如，对于图 10 和图 11 中描绘的实例，如果用户选择普通话 1104 用于文本到语音转换，那么这个偏好可存储为用户偏好的一部分并且用于后续语言分析。例如，下一次用户选择中文文本时，可使用包括用户偏好的启发法自动解决普通话、广东话和台湾话之间的语言歧义。在这种情景下，普通话可被自动选择并且用于文本到语音转换。因此，用户无需进行另一个语言选择。以这种方式，用户过去的选择可影响用于文本到语音转换的未来语言选择。

[0104] 在一些实施例中，可清除用户进行的先前语言选择，从而允许用户进行新的可能有所改变的选择。例如，用户可能先前已选择加拿大法语而不是法国法语作为选择的语言。可提供允许用户重置这个选择的选项。在一个实施例中，可提供用户可选择的来回切换特征“发音选择”，这个特征可被打开或关闭，并且状态的变化使用户的先前选择被清除。

[0105] 图 1 中描绘的系统 100 可并入到各种系统和设备中。图 12 是计算机系统 1200 的简化框图，该计算机系统可包含根据一些实施例的系统 100 的部件。如图 12 所示，计算机系统 1200 包括通过总线子系统 1204 与多个外围子系统进行通信的处理器 1202。这些外围子系统可包括存储子系统 1206（包括存储子系统 1208 和文件存储子系统 1210）、用户界面输入设备 1212、用户界面输出设备 1214 和网络接口子系统 1216。

[0106] 总线子系统 1204 提供用于允许计算机系统 1200 的各种部件和子系统按预期彼此

通信的机构。虽然总线子系统 1204 被示意性地示出为单个总线,但是总线子系统的另选实施例可利用多个总线。

[0107] 可以实现为一个或多个集成电路(例如常规的微处理器或微控制器)的处理器 1202 可控制计算机系统 1200 的操作。在各种实施例中,处理器 1202 可响应于程序代码执行多种程序,并且可维持多个同时执行的程序或过程。在任何给定时间,要执行的程序的部分或全部可驻留在处理器 1202 中和/或存储子系统 1206 中。通过合适的编程,处理器 1202 可提供上述用于执行文本到语音转换和分析的各种功能。

[0108] 网络接口子系统 1216 提供通向其他计算机系统和网络的接口。网络接口子系统 1216 用作用于从其他系统接收数据并且从计算机系统 1200 向其他系统传输数据的接口。例如,网络接口子系统 1216 可允许计算机系统 1200 通过互联网连接到客户端设备。在一些实施例中,网络接口 1216 可包括用于访问无线语音和/或数据网络(例如使用蜂窝电话技术;如 3G、4G 或 EDGE 的高级数据网络技术;WiFi (IEEE 802.11 系列标准);或其他移动通信技术;或它们的任何组合)的射频(RF)收发器部件, GPS 接收器部件,和/或其他部件。在一些实施例中,除了或代替无线接口,网络接口 1216 可提供有线网络连接性(例如以太网)。

[0109] 用户界面输入设备 1212 可包括:键盘、诸如鼠标或轨迹球的指向设备、并入到显示器中的触摸板或触摸屏、滚轮、点击式触摸转盘(click wheel)、刻度盘、按钮、开关、小键盘、诸如语音识别系统的音频输入设备、麦克风、以及其他类型的输入设备。一般来讲,使用术语“输入设备”旨在包括用于向计算机系统 1200 输入信息的所有可能类型的设备和机构。例如,在 iPhone® 中,用户输入设备 1212 可包括由 iPhone® 提供的一个或多个按钮、触摸屏等等。用户可使用输入设备 1212 中的一或多个输入设备提供与选择哪种语言用于文本到语音转换有关的输入。用户还可使用输入设备 1212 中的一或多个输入设备选择要转换成语音的文本。

[0110] 用户界面输出设备 1214 可包括显示器子系统、指示灯,或如音频输出设备的非视觉显示器等。显示器子系统可以为阴极射线管(CRT)、如液晶显示器(LCD)的平板设备、投影设备、触摸屏等等。一般来讲,使用术语“输出设备”旨在包括用于从计算机系统 1200 输出信息的所有可能类型的设备和机构。例如,用于选择要用于文本到语音转换的语言的菜单和其他选项可通过输出设备显示给用户。另外,可通过音频输出设备输出语音。

[0111] 存储子系统 1206 提供计算机可读存储介质,用于存储提供一些实施例的功能的基本编程和数据构造。存储子系统 1206 可例如以任何组合方式使用磁盘、闪存存储器或任何其他存储介质来实现,并且可根据需要包括易失性和/或非易失性存储器。在由处理器执行时提供上述功能的软件(程序、代码模块、指令)可存储在存储子系统 1206 中。这些软件模块或指令可由处理器 1202 执行。存储子系统 1206 还可提供存储库,用于存储根据本发明所使用的数据。存储子系统 1206 可包括存储器子系统 1208 和文件/磁盘存储子系统 1210。

[0112] 存储器子系统 1208 可包括多个存储器,这些存储器包括:用于在程序执行期间存储指令和数据的主随机存取存储器(RAM) 1218,和存储有固定指令的只读存储器(ROM) 1220。文件存储子系统 1210 提供对程序和数据文件的永久性(非易失性)存储,并且可包括硬盘驱动器、软盘驱动器以及相关联的可移除介质、光盘只读存储器(CD-ROM) 驱

动器、光驱、可移除介质盒和其他类似的存储介质。

[0113] 计算机系统 1200 可以是各种类型,包括个人计算机、便携式设备(例如 iPhone<sup>®</sup>、iPad<sup>®</sup>)、工作站、网络计算机、大型机、信息亭、服务器或任何其他数据处理系统。由于计算机和网络不断变化的性质,对图 12 中描绘的计算机系统 1200 的描述旨在仅仅作为具体实例。比图 12 中描绘的系统具有更多或更少部件的许多其他配置是可能的。

[0114] 根据一些实施例,图 14 示出根据如上所述的本发明的原理来配置的电子设备 1400 的功能框图。设备的功能块可由硬件、软件或硬件和软件的组合来实现,以便实行本发明的原理。本领域的技术人员应当理解,图 14 中描绘的功能块可被组合或分成子块,以便实现如上所述的本发明的原理。因此,本文中的描述可支持对本文所述的功能块的任何可能的组合或分离或进一步限定。

[0115] 如图 14 所示,电子设备 1400 包括各自耦接到处理单元 1406 的输入接收单元 1402 和显示单元 1404。在一些具体实施中,处理单元 1406 包括语言确定单元 1408 和文本到语音转换单元 1410。在一些具体实施中,文本到语音转换单元 1410 包括一种或多种语言合成单元 1412-n。

[0116] 处理单元 1406 被配置为确定用于将文本转换成语音的一组语言(例如用语言确定单元 1408),该一组语言包括用于将文本的一部分转换成语音的多种候选语言。显示单元 1404 被配置为使多种候选语言被显示。输入接收单元 1402 被配置为接收指示从多种候选语言中选择第一候选语言的输入。处理单元 1406 被进一步配置为将文本转换成语音(例如,用文本到语音转换单元 1410 和 / 或语言合成单元 1412-n 中的一种或多种语言合成单元),该转换包括以第一候选语言将文本的该部分转换成语音。在一些具体实施中,文本的该部分是整个文本或文本的子集。

[0117] 在一些具体实施中,处理单元 1406 被进一步配置为使用对应于第一候选语言的语言合成单元(例如语言合成单元 1412-n 之一),以第一候选语言将文本的该部分转换成语音,从而产生文本的该部分的语音。

[0118] 在一些具体实施中,处理单元 1406 被配置为通过以下方式将语音转换成文本:使用第一语言合成单元(例如语言合成单元 1412-n 中的第一个)将文本中的第一字符转换成语音;以及使用不同于第一语言合成单元的第二语言合成单元(例如语言合成单元 1412-n 中的第二个)将文本中第一字符后的第二字符转换成语音。

[0119] 在一些具体实施中,处理单元 1406 被配置为(例如用语言确定单元 1408):读取文本中的第一字符;确定要用于将第一字符转换成语音的第一语言;将第一语言与第一字符相关联;读取文本中的第二字符,第二字符在第一字符后;确定要用于将第二字符转换成语音的第二语言,其中第二语言不同于第一语言;并且将第二语言与第二字符相关联。

[0120] 在一些具体实施中,处理单元 1406 被配置为(例如用语言确定单元 1408):基于一个或多个加载的键盘确定要包括在一组语言中的语言;基于与文本相关联的一个或多个标签确定要包括在一组语言中的语言;基于在分析文本时由语言分析工具识别的一种或多种语言确定要包括在一组语言中的语言;基于由与文本相关联的应用程序提供的确定要包括在一组语言中的语言;或基于与文本中的一个或多个字符相关联的编码信息确定要包括在一组语言中的语言;或以任何组合方式执行这些操作中的任一个。

[0121] 在一些具体实施中,处理单元 1406 被配置为(例如用语言确定单元 1408):基于

一个或多个用户偏好确定要包括在一组语言中的语言 ; 基于选择用于在先文本到语音转换的语言确定要包括在一组语言中的语言 ; 基于文本的来源确定要包括在一组语言中的语言 ; 基于与文本相关联的国家代码顶级域信息确定要包括在一组语言中的语言 ; 或基于全球定位系统 (GPS) 信息确定要包括在一组语言中的语言。

[0122] 在一些具体实施中, 处理单元 1406 被配置为 ( 例如用语言确定单元 1408) : 确定用于将文本转换成语音的第一语言列表 ; 并且通过从第一语言列表中过滤掉至少一种语言来从第一语言列表确定第二语言列表。

[0123] 上述各种实施例可使用专用部件和 / 或可编程处理器和 / 或其他可编程设备的任何组合来实现。各种实施例可仅在硬件中、仅在软件中、或使用它们的组合来实现。本文所述的各种过程可以任何组合方式在相同处理器或不同处理器上实现。因此, 在将部件描述为被配置为执行某些操作的情况下, 这种配置可例如通过以下方式来实现 : 设计电子电路以执行操作、将可编程电子电路 ( 如微处理器 ) 编程以执行操作、或它们的任何组合。过程可使用多种技术进行通信, 这些技术包括但不限于用于过程间通信的常规技术, 并且不同的过程对可使用不同的技术, 或相同的过程对可在不同时间使用不同的技术。尽管上述实施例可引用具体硬件和软件部件, 但本领域的技术人员将会知道, 也可以使用硬件和 / 或软件部件的不同组合, 并且被描述为在硬件中实现的特定操作也可能在软件中实现, 或反之亦然。

[0124] 各种实施例不限于在某些具体数据处理环境中操作, 而是可自由地在多个数据处理环境中操作。另外, 虽然已使用一系列特定事务来描述实施例, 但这并不旨在具有限制性。

[0125] 因此, 虽然已关于具体实施例描述本发明, 但这些实施例并不旨在具有限制性。各种修改和等同物在以下权利要求书的范围内。



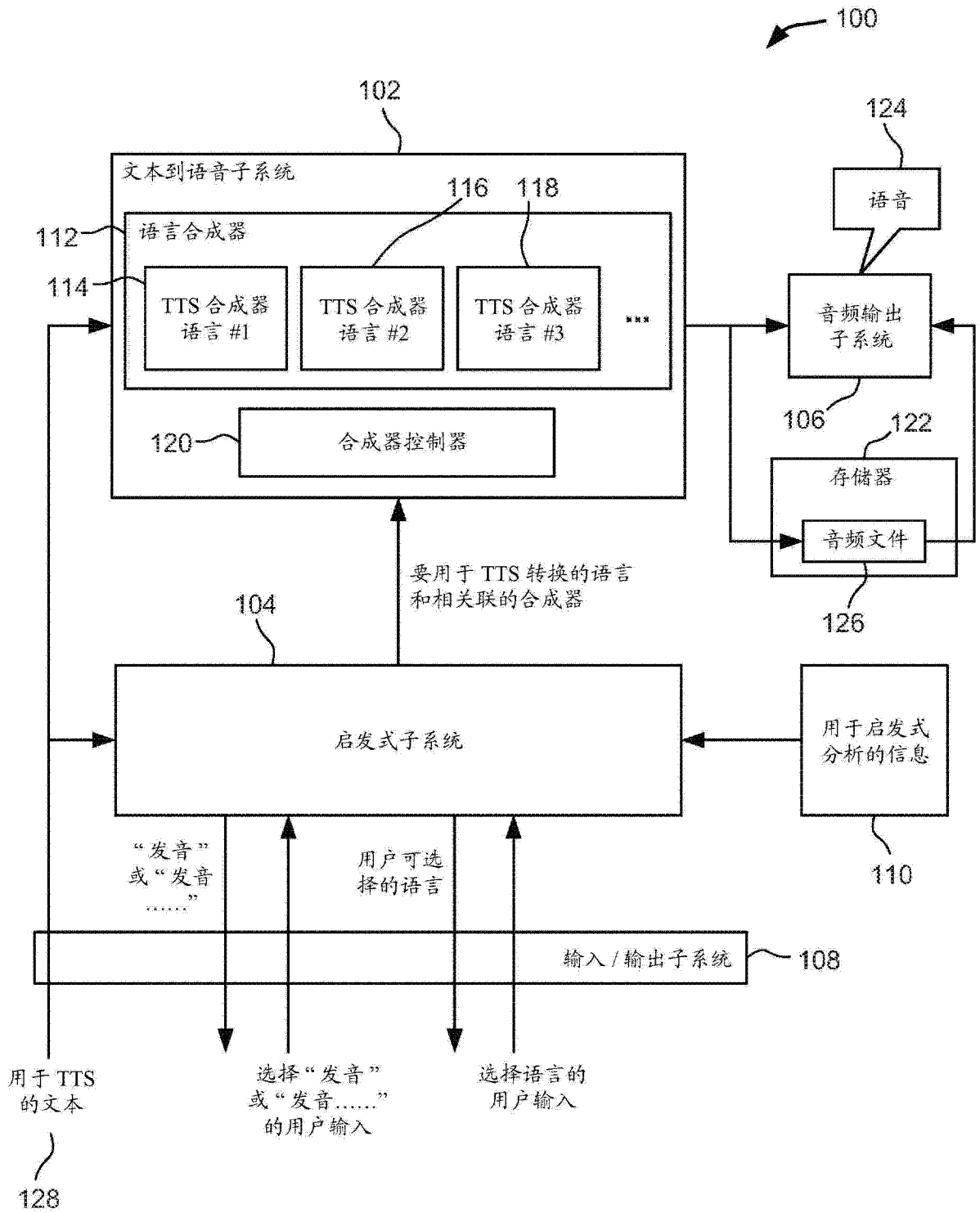


图 1

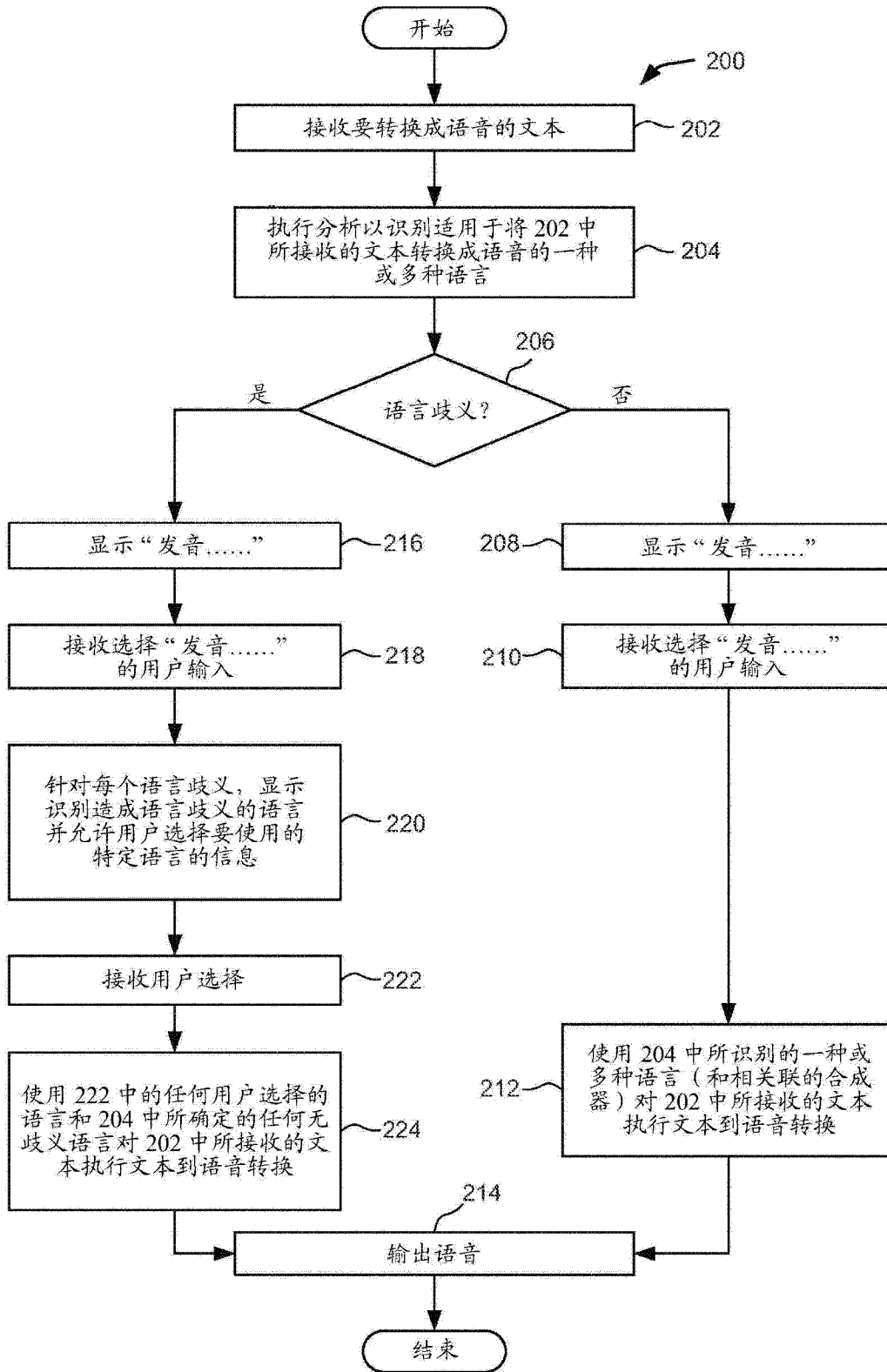


图 2

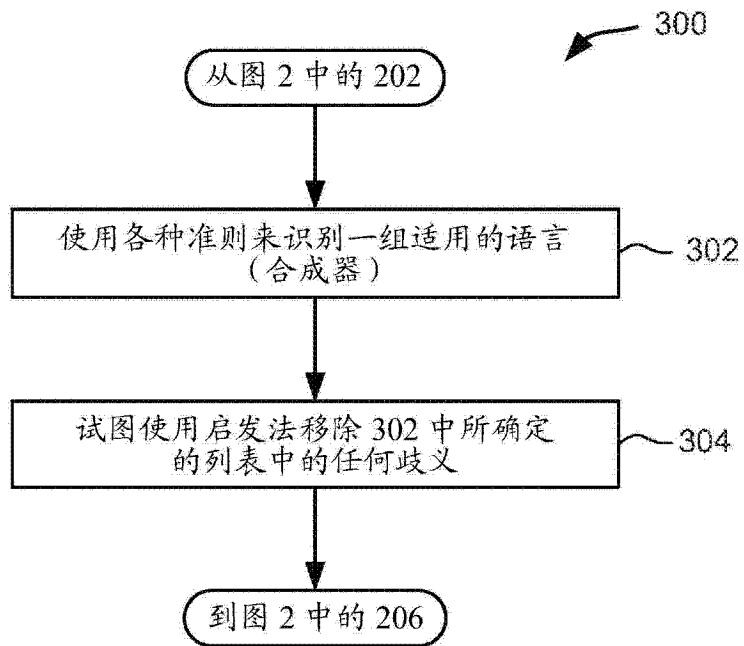


图 3

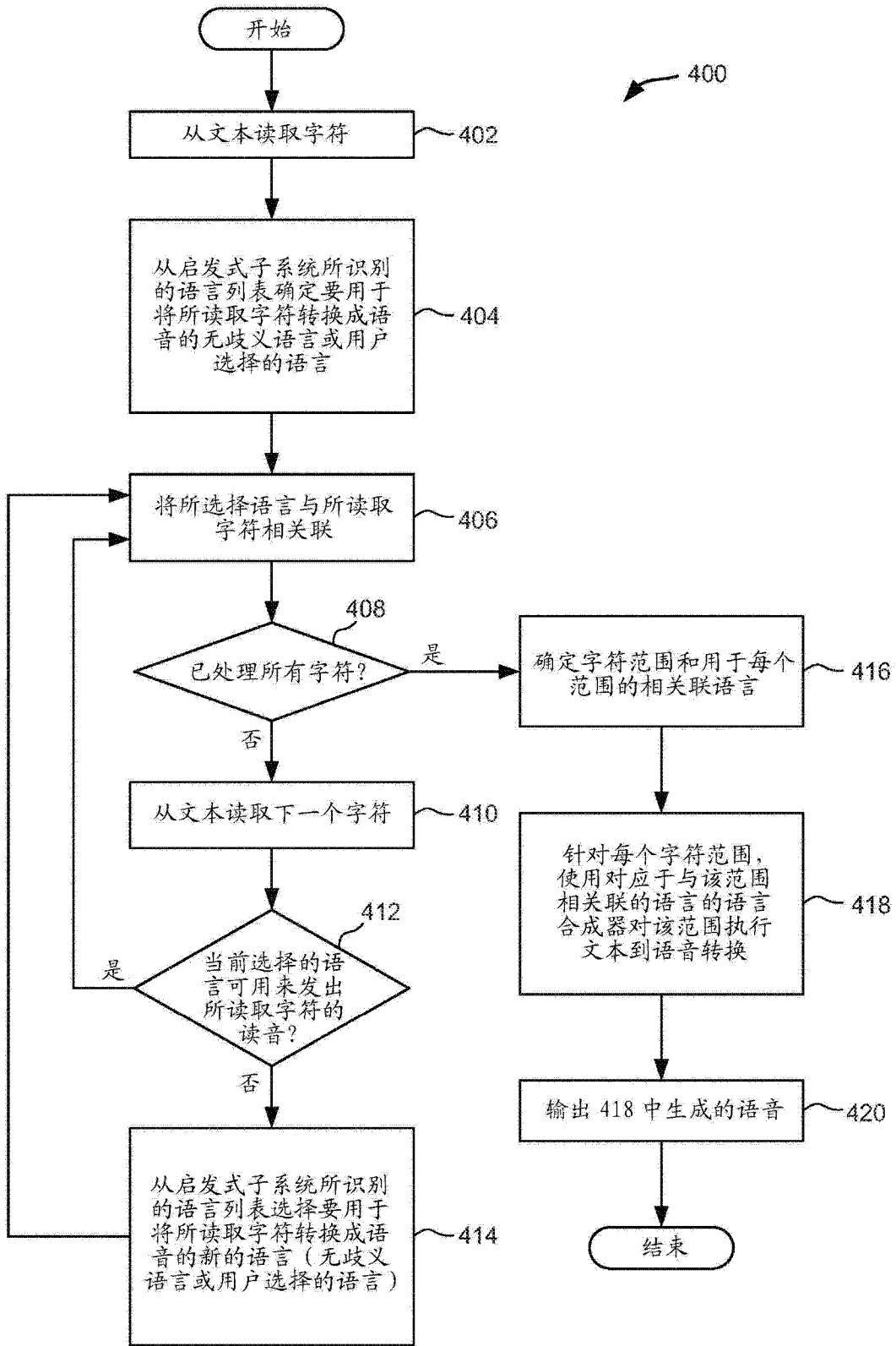


图 4

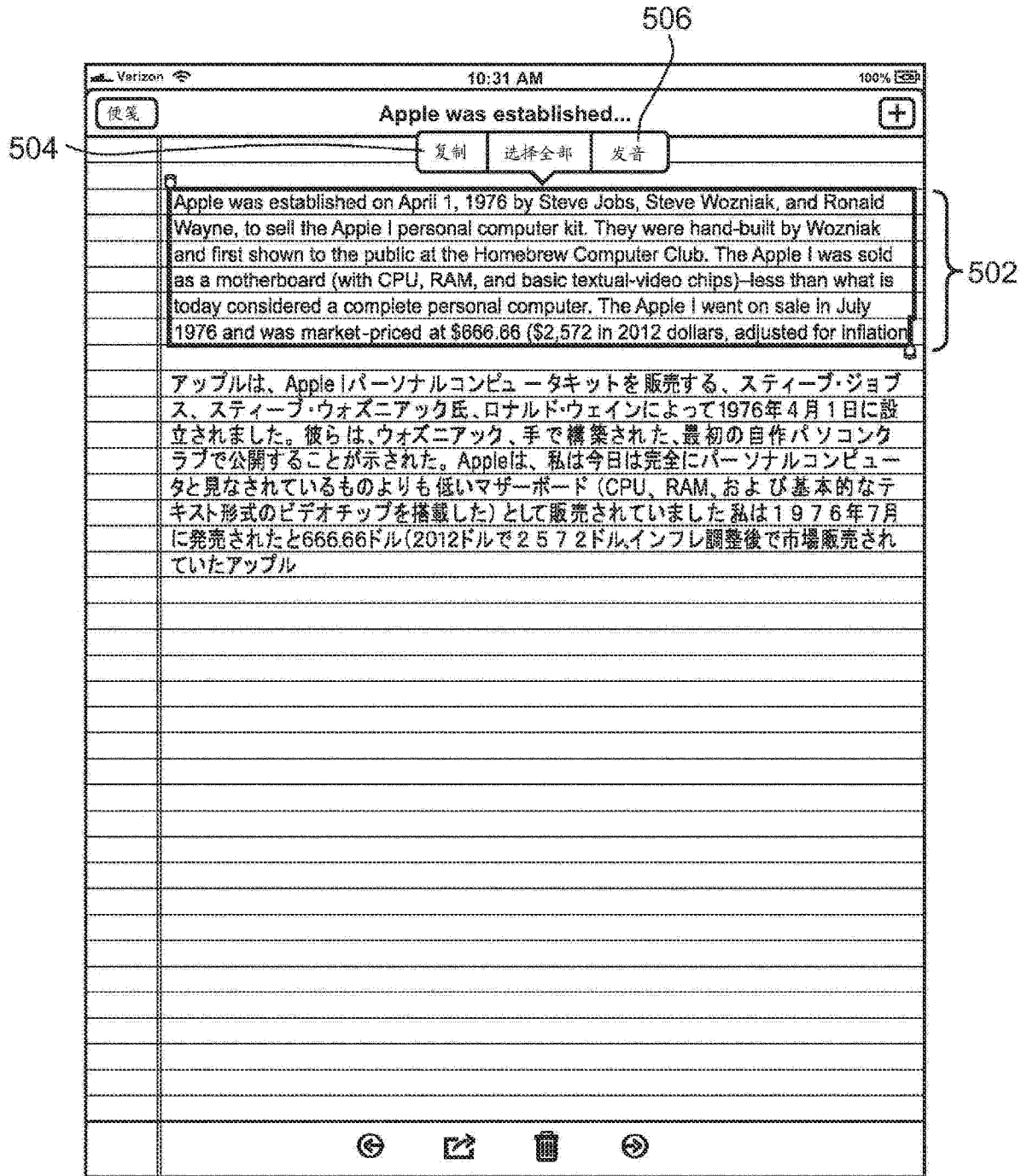


图 5

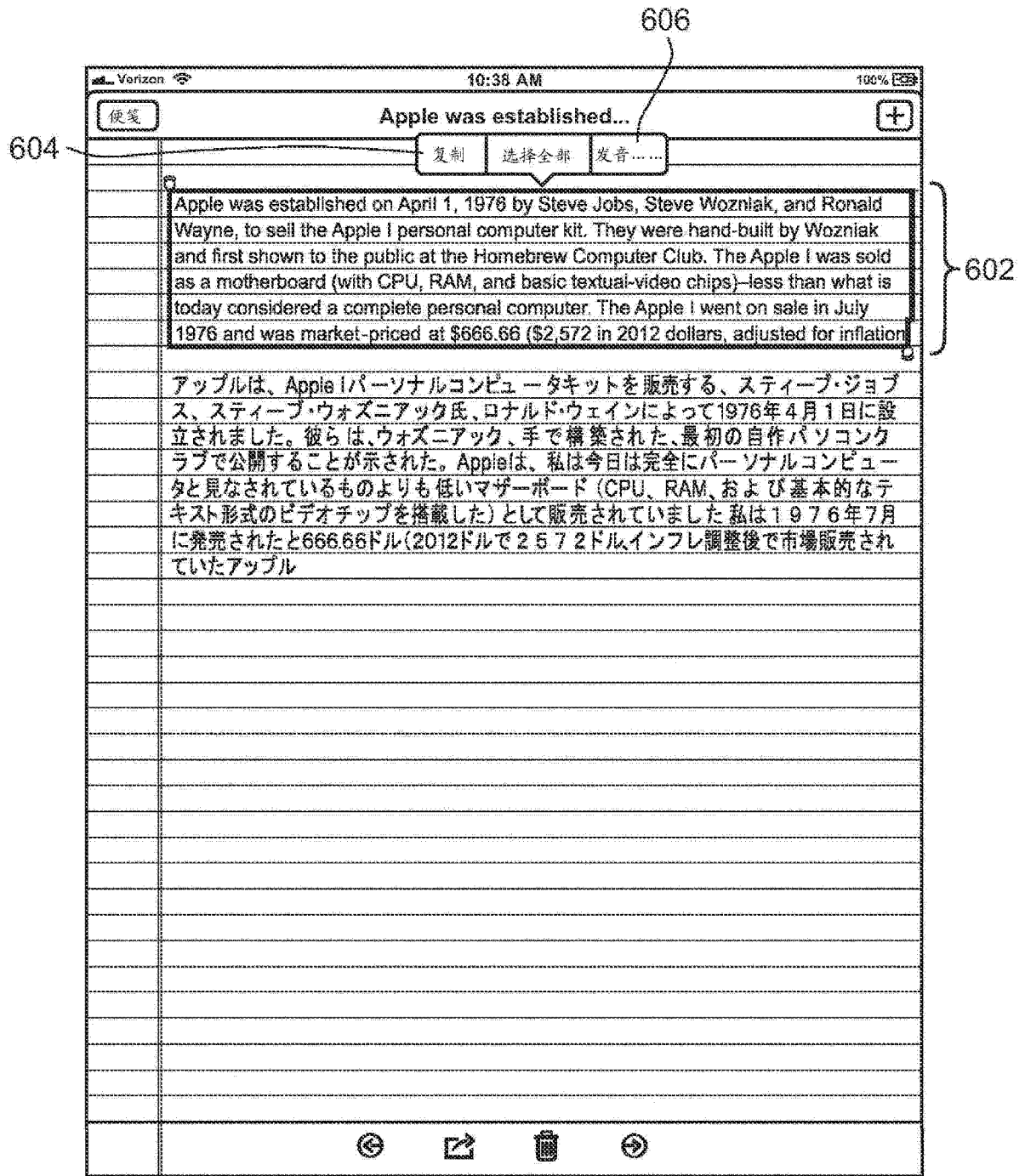


图 6

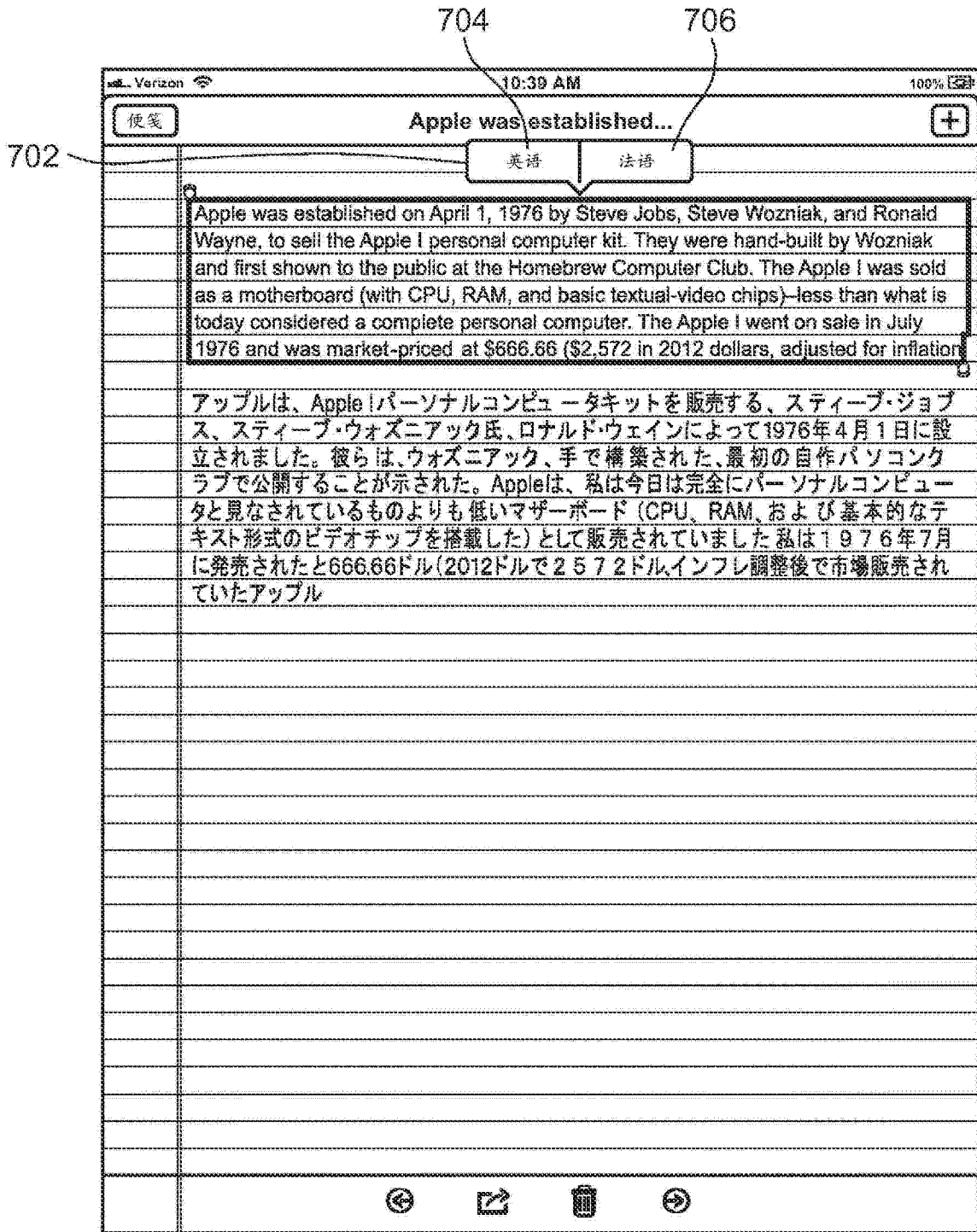


图 7

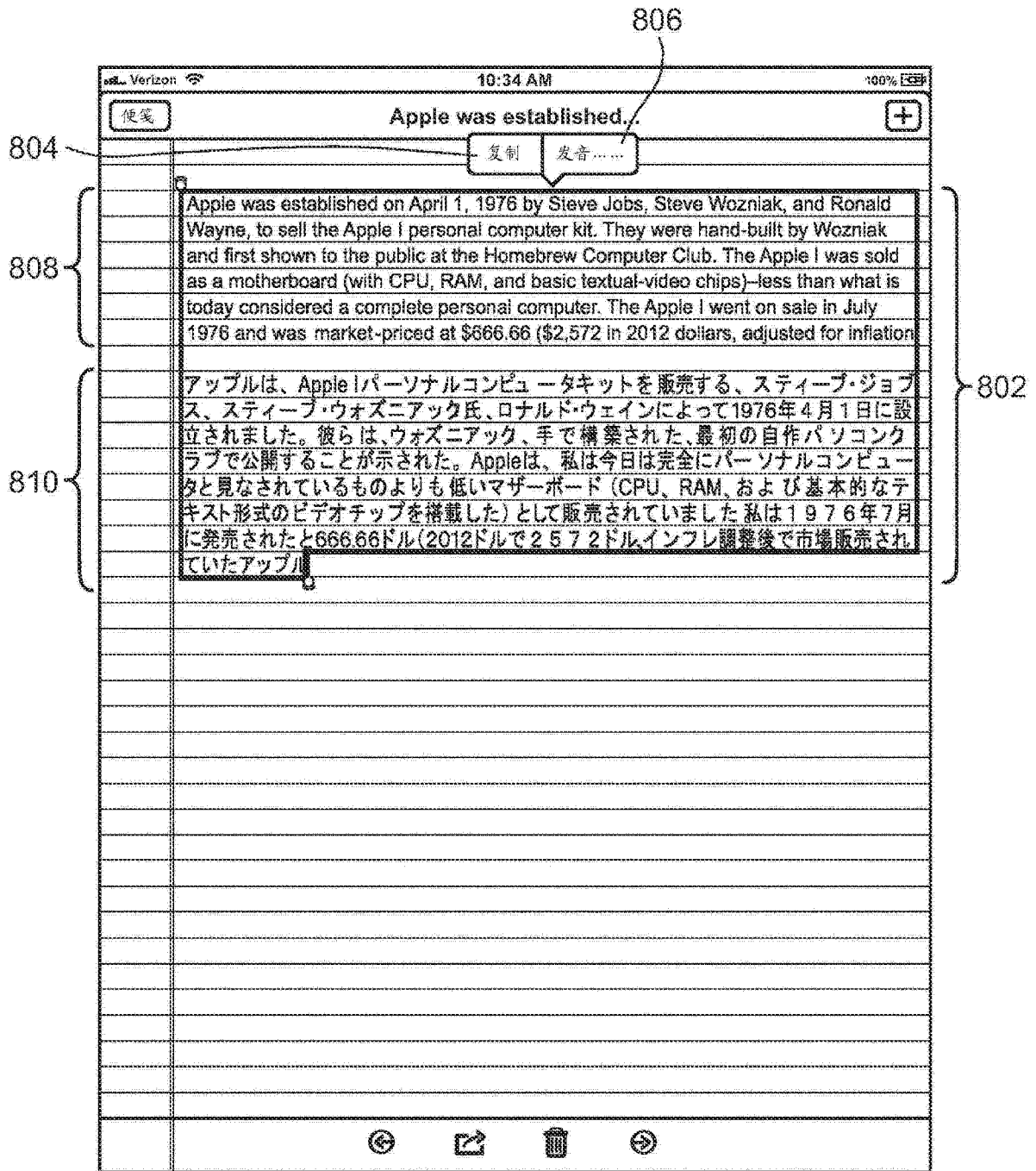


图 8



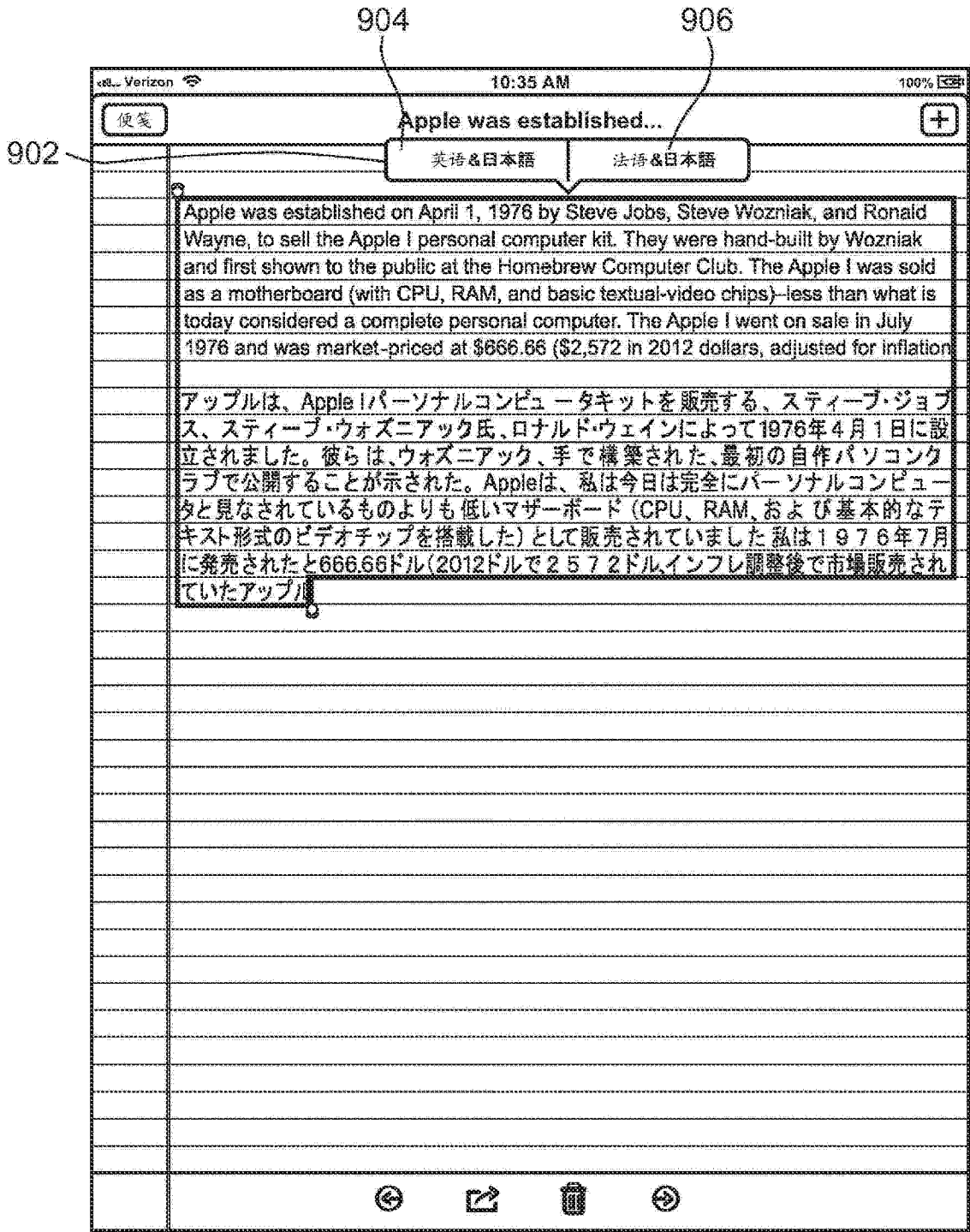


图 9

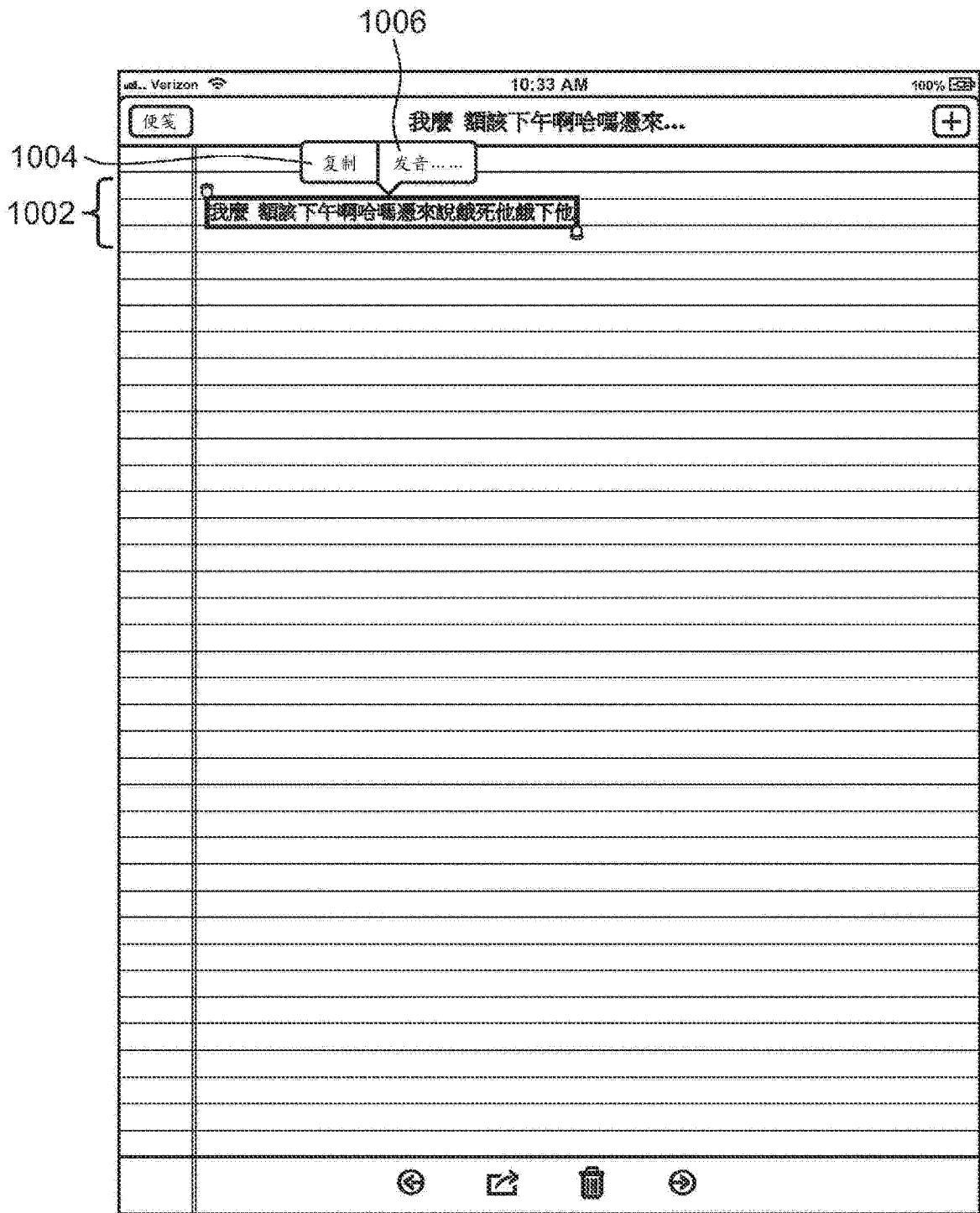


图 10

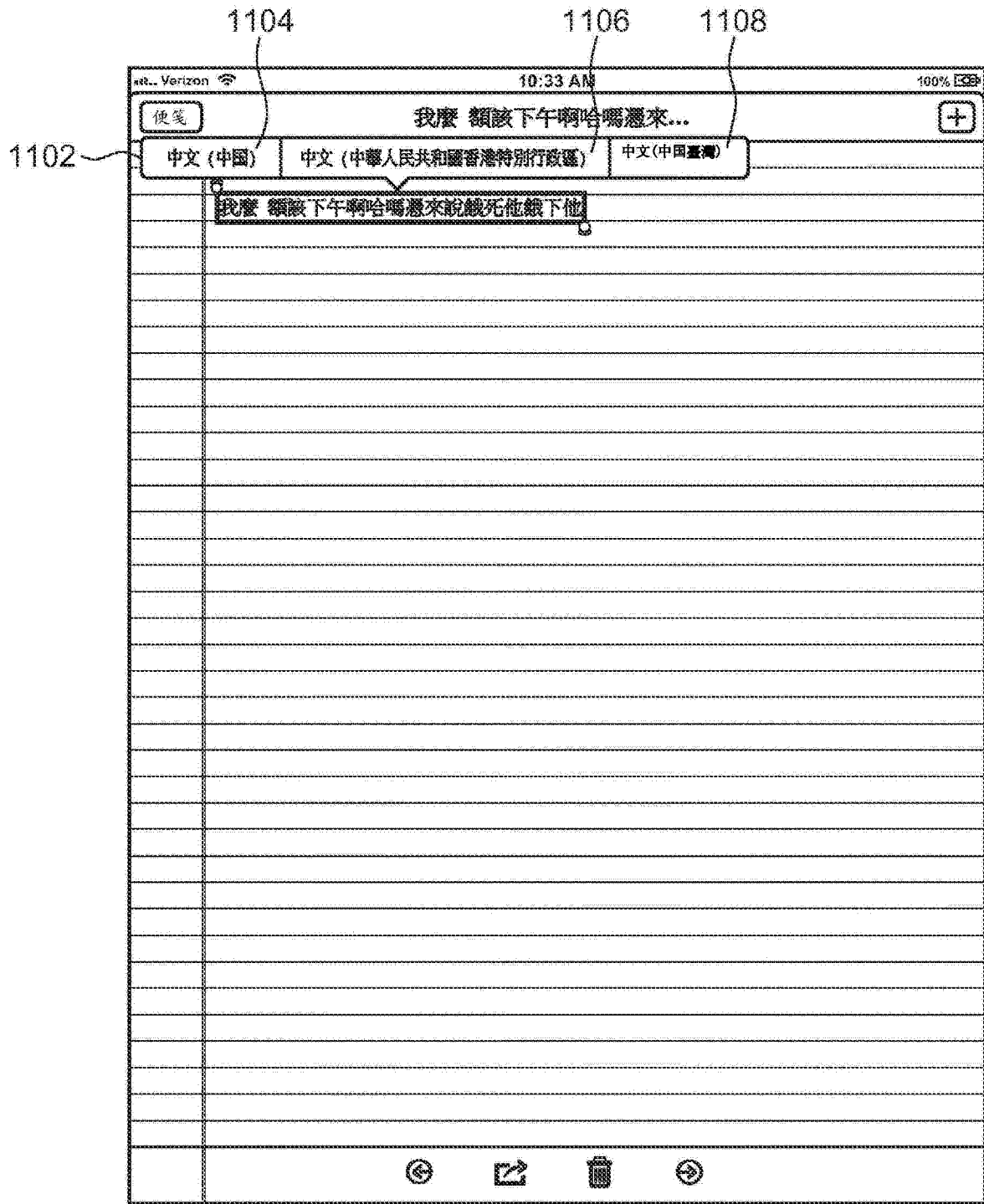


图 11

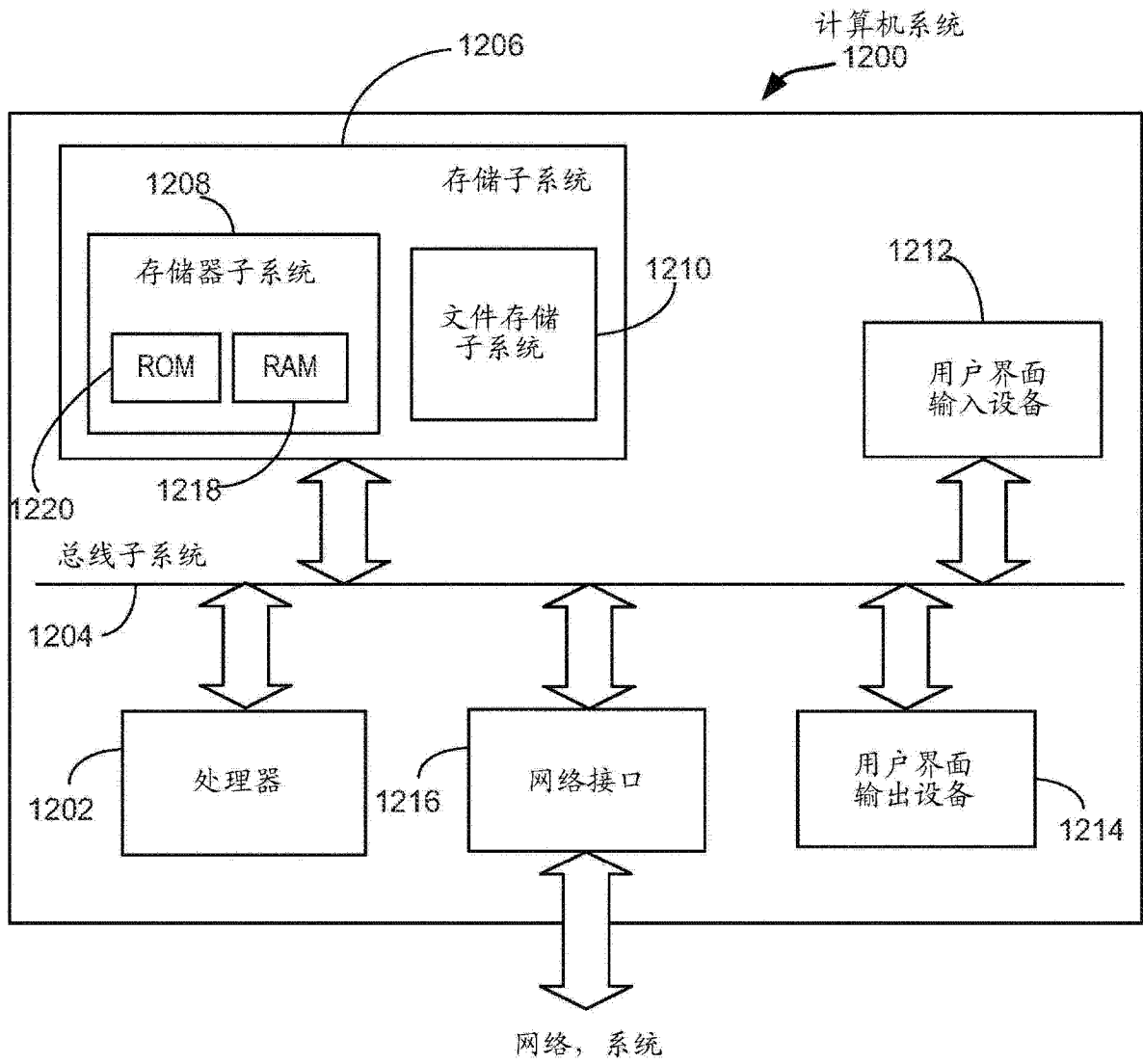


图 12

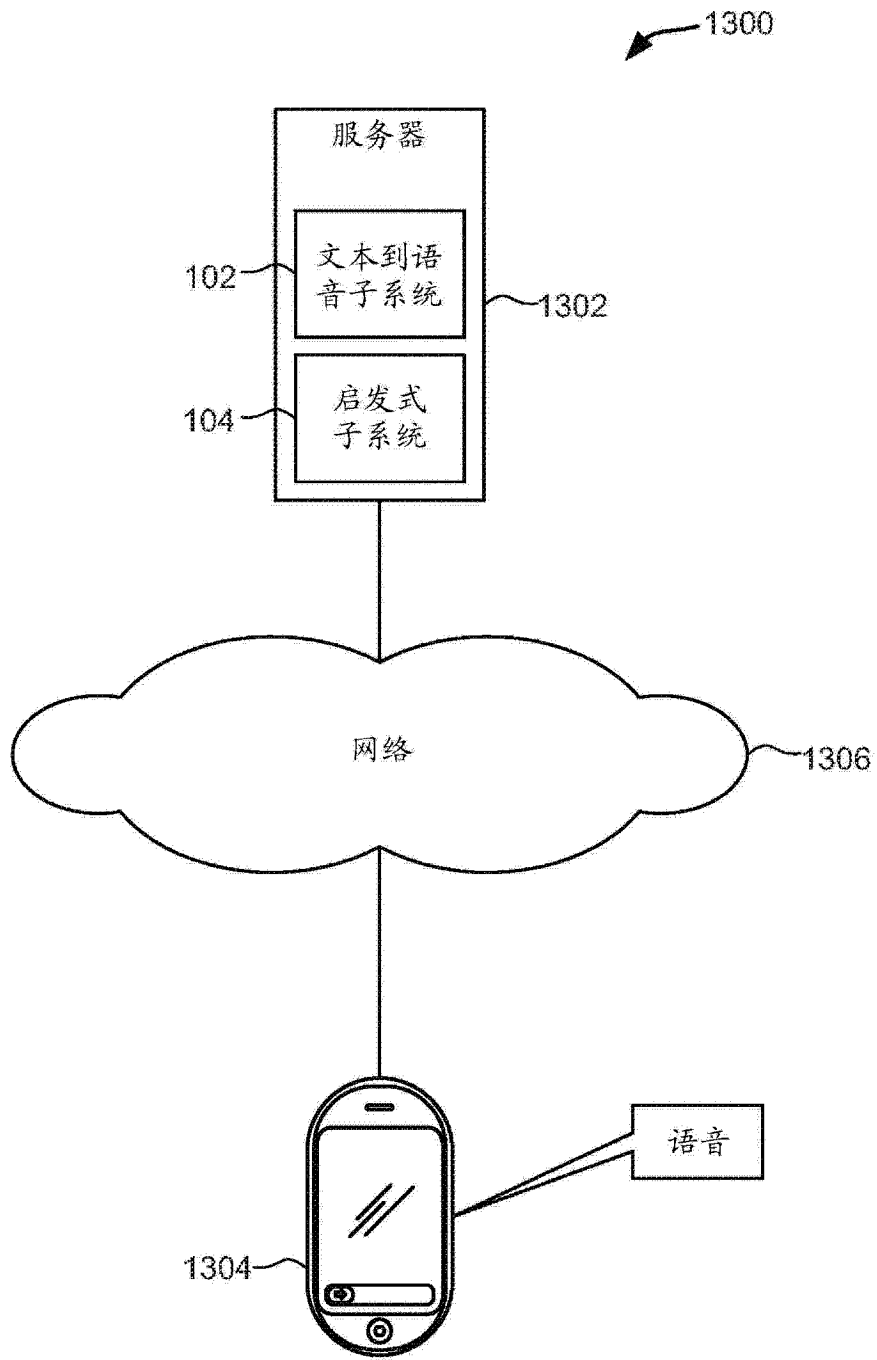


图 13

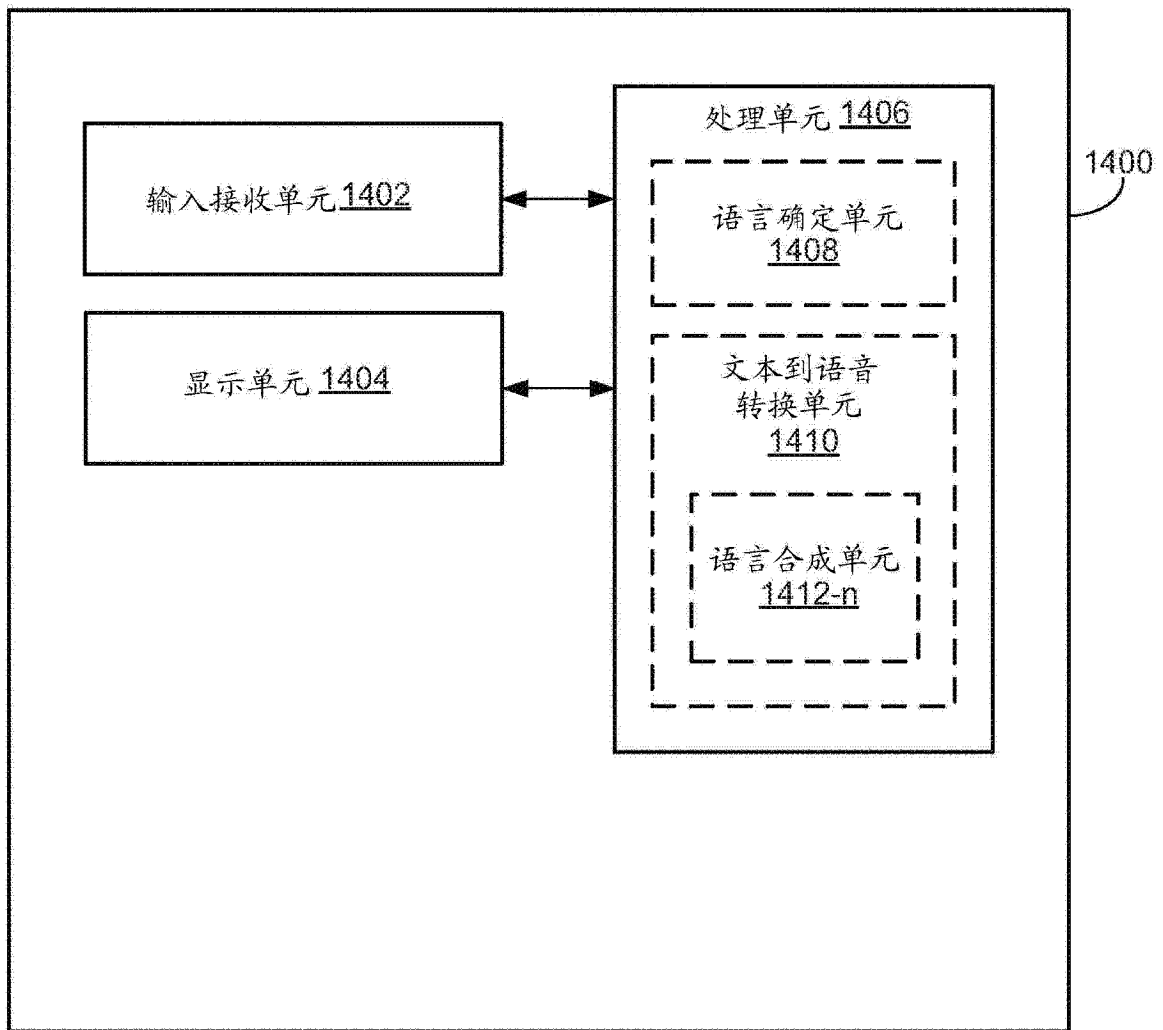


图 14