



(12) 发明专利

(10) 授权公告号 CN 101313276 B

(45) 授权公告日 2013.04.24

(21) 申请号 200680040198.X

(22) 申请日 2006.10.16

(30) 优先权数据

11/262,230 2005.10.28 US

(85) PCT申请进入国家阶段日

2008.04.28

(86) PCT申请的申请数据

PCT/US2006/040537 2006.10.16

(87) PCT申请的公布数据

W02007/053294 EN 2007.05.10

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 M·V·玛哈简 A·艾斯洛 B-J·苏

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 陈斌

(51) Int. Cl.

G06F 3/16 (2006.01)

G10L 15/22 (2006.01)

(56) 对比文件

US 2004156562 A1, 2004.08.12, 参见说明书第 0019, 0029-0032 段, 附图 1, 5.

US 2004156562 A1, 2004.08.12, 参见说明书第 0019, 0029-0032 段, 附图 1, 5.

CN 1409842 A, 2003.04.09, 参见图 3, 说明书第 7-8 页.

US 2004153321 A1, 2004.08.05, 全文.

US 2002091520 A1, 2007.07.11, 参见附图 2, 3, 6, 说明书第 0031-0034 段.

审查员 胡燕

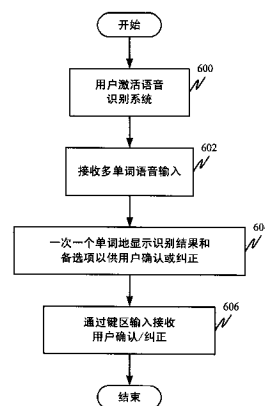
权利要求书2页 说明书12页 附图12页

(54) 发明名称

将语音和备选项输入形式组合到移动设备中

(57) 摘要

当将说出的信息输入到移动设备中时同时使用语音和备选形式的输入。该备选形式输入可用于执行对语音识别结果中的单词的顺序提交。



1. 一种用于将信息输入到移动设备中的方法,包括:  
由移动设备从用户处接收多单词语音输入;  
对所述语音输入执行语音识别以获得语音识别结果;以及  
一次一个单词地顺序地显示所述语音识别结果以供用户确认或纠正,  
其中执行语音识别包括:  
基于所述语音输入计算指示多个语音识别假设的假设点阵;  
在完成计算所述假设点阵之前,输出一初步假设点阵,以及接着继续计算所述假设点阵;

其中顺序地显示所述语音识别结果以供用户纠正或确认包括首先使用所述初步假设点阵来显示所述语音识别结果,直至完全计算出所述假设点阵,接着使用所述完全计算出的假设点阵来显示所述语音识别结果。

2. 如权利要求 1 所述的方法,其特征在于,还包括:

接收指示对给定的、所显示的单词的用户纠正或确认的用户输入。

3. 如权利要求 1 所述的方法,其特征在于,执行语音识别包括计算对应于所述语音输入的假设点阵,并且所述方法还包括:

在将下一单词添加到所述语音识别结果的显示之前,在给定的所显示的单词之后,基于指示所述给定的所显示单词的用户纠正或确认的用户输入,重新计算所述假设点阵。

4. 如权利要求 3 所述的方法,其特征在于,显示包括:

基于所述假设点阵的重新计算,将下一单词添加到所述显示。

5. 如权利要求 2 所述的方法,其特征在于,接收用户输入包括:

接收指示对所述移动设备上的输入机制的用户操纵的备选形式输入。

6. 如权利要求 5 所述的方法,其特征在于,一次一个单词地顺序显示所述语音识别结果包括:

显示最佳假设单词;以及

显示接近所述最佳假设的备选单词的列表。

7. 如权利要求 6 所述的方法,其特征在于,接收备选形式输入包括:

接收对所述备选单词之一的用户选择。

8. 如权利要求 6 所述的方法,其特征在于,接收备选形式输入包括:

接收指示用户一次一个字母地输入期望单词的用户键入输入。

9. 一种移动设备,包括:

语音识别器;以及

用户界面组件,它被配置成从所述语音识别器接收指示一多单词语音输入的识别的语音识别结果,并通过仅在所述语音识别结果中先前输出的单词被用户确认或纠正以后才输出所述语音识别结果中的下一单词来顺序地输出所述语音识别结果,

其中所述语音识别器被配置成通过基于所述语音输入计算指示多个语音识别假设的识别假设点阵来将所述语音识别结果提供给所述用户界面组件,并且其中,所述语音识别器被配置成基于用户对所显示的单词的确认或纠正来重新计算所述识别假设点阵,其中所述用户界面组件被配置成基于所述识别假设点阵的重新计算来显示下一个单词;且

其中计算识别假设点阵包括:

在完成计算所述假设点阵之前,输出一初步假设点阵,以及接着继续计算所述假设点阵;

其中顺序地输出所述语音识别结果包括首先使用所述初步假设点阵来显示所述语音识别结果,直至完全计算出所述假设点阵,接着使用所述完全计算出的假设点阵来显示所述语音识别结果。

10. 如权利要求 9 所述的移动设备,其特征在于,还包括:

显示器,所述用户界面组件被配置成通过在所述显示器上顺序地显示所述语音识别结果来输出所述语音识别结果。

11. 如权利要求 10 所述的移动设备,其特征在于,所述用户界面组件被配置成与所述语音识别结果一起显示多个备选单词作为对给定的所显示的单词的备选项。

12. 如权利要求 10 所述的移动设备,其特征在于,所述语音识别器通过输出初步的识别假设点阵来提供所述语音识别结果,并且其中所述用户界面组件被配置成最初基于所述初步的识别假设点阵来向用户显示所述语音识别结果。

13. 一种移动设备上的用户界面系统,包括:

用户可操纵的输入形式组件;

显示器;

语音识别系统,被配置成通过基于所述语音输入计算指示多个语音识别假设的假设点阵,并且其中,在完成计算所述假设点阵之前,所述语音识别系统输出一初步假设点阵,以及接着继续计算所述假设点阵;以及

用户界面组件,它被配置成显示指示一多单词语音输入的单词列表,所述列表中的每一单词在显示所述列表中的下一个单词之前被顺序地显示以供用户确认或纠正,其中首先使用初步假设点阵来显示所述指示多单词语音输入的单词列表,直至完全计算出所述假设点阵,接着使用所述完全计算出的假设点阵来显示所述指示多单词语音输入的单词列表。

14. 如权利要求 13 所述的用户界面系统,其特征在于,所述用户界面组件被配置成向语音识别系统提供指示用户对所显示的单词的纠正或确认的纠正或确认信息,用于重新计算指示所述多单词语音输入的可能单词的剩余部分。

15. 如权利要求 13 所述的用户界面系统,其特征在于,所述用户确认或纠正是通过所述用户可操纵输入形式组件来提供的。

16. 如权利要求 15 所述的用户界面系统,其特征在于,所述用户可操纵输入形式组件包括键区、软键盘、手写输入组件以及触摸屏之一。

17. 如权利要求 13 所述的用户界面系统,其特征在于,所述用户界面组件被配置成与所述列表中的每个单词一起显示一组备选单词以供用户选择。

## 将语音和备选项输入形式组合到移动设备中

[0001] 背景

[0002] 由于对用于相对较小的移动设备（诸如蜂窝电话和个人数字助理）的应用程序的使用的增加，在这种设备上的文本输入变得越来越流行。一些这样的应用程序包括电子邮件（e-mail）和短消息服务（SMS）。

[0003] 然而，移动电话、个人数字助理和其他这样的移动设备一般不具有和台式计算机上一样方便的键盘。例如，移动电话往往仅具有数字键区，其上多个字母被映射到相同的键。某些 PDA 仅具有从指示笔或类似的项目接收输入的触敏屏幕。

[0004] 由此，这种设备当前提供了允许用户使用多种不同方法中的一种通过数字键区或触摸屏幕或其他输入设备来输入文本的界面。一个这样的方法是称为多次轻叩（multi-tap）界面的确定性界面。在多次轻叩界面中，根据用户期望的相应字母，用户按下一数字键给定的次数。例如，当键区具有对应于字母“abc”的数字“2”键时，击键“2”对应于“a”，击键“22”对应于“b”，击键“222”对应于“c”，而击键“2222”对应于数字“2”。在另一示例中，击键输入 8 4444 7777 会对应于单词“this”。

[0005] 另一种已知类型的界面是预测性系统，并且被称为 Tegic 通信的 T9 界面。T9 界面允许用户轻叩对应于一期望字母的键一次，并且使用先前的击键顺序来预测期望的单词。虽然这减少了按键的次数，但是这种类型的预测性界面会遭受由共享相同的键序列的单词导致的模糊性。例如，键序列“4663”可以对应于单词“home”、“good”、“gone”、“hood”或“hone”。在这些情况下，该界面显示从该键序列生成的预测单词的列表，并且用户按下“next（下一个）”键来滚动通过各备选项。此外，由于不能预测字典外或该界面的词汇表外的单词，因此 T9 类型的界面通常与其他回退（fallback）策略，诸如多次轻叩结合，以便处理词汇表外的单词。

[0006] 一些当前的界面也提供对单词补全和单词预测的支持。例如，基于初始的键序列“466”（对应于字母“goo”），可以预测单词“good”。类似地，从初始的键序列“6676”（对应于字母“morn”），可以预测单词“morning”。类似地，基于 n 元语法语言模型预测，可以预测单词“a”作为跟在单词序列“this is”之后的下一单词。

[0007] 这些界面中没有一个真正容许任何类型的快速文本输入。实际上，这些方法的新用户通常仅能达到每分钟 5-10 个单词的文本输入速率。

[0008] 为了增加在这种通信设备上的信息输入带宽，某些设备实现了语音识别。语音具有相对高的通信带宽，估计为大约每分钟 250 个单词。然而，由于用户检查和纠正语音识别差错所花费的时间，使用常规的自动语音识别系统的文本输入的带宽在实践中要低得多，而这些语音识别差错对于当前的语音识别系统是不可避免的。

[0009] 具体地，某些当前的基于语音的文本输入方法允许用户通过在每个单词之间有轻微停顿地说话来将文本输入到蜂窝电话中。该语音识别系统接着显示一识别结果。由于直接检测通常导致差错，尤其在存在噪音时，因此用户必须选择识别结果中的错误，并接着使用备选项列表或回退输入方法来纠正它们。

[0010] 孤立的单词识别要求用户每次仅说出一个单词。该一个单词被处理并输出。用户

接着纠正该单词。虽然孤立的单词识别的确改进了识别准确度,但是孤立的单词识别界面是不自然的,并且相对于其中用户一次可说出整个词组或句子的使用连续语音识别实现的系统而言降低了数据输入速率。

[0011] 然而,连续语音识别中的纠错存在问题。传统上,连续语音识别的语音识别结果是通过显示整个词组或句子的最佳假设来呈现的。为了纠错,用户接着选择误识别的单词,并从一下拉列表选择一备选项。由于差错通常成群地并且跨词界地出现,因此许多系统允许纠正整个误识别的词组。例如,话语“can you recognize speech”(你能识别语音吗)会被不正确地识别为“can you wreck a nice beach”(你能破坏一个美丽的沙滩吗)。在这种情况下,由于不正确的单词分割,不可能简单地一次一个单词地纠正识别。由此,要求用户选择短语“wreck a nice beach”并为整个短语选择一备选项。

[0012] 虽然当识别准确度较高且诸如鼠标等指点设备可用时这种方法能很好地工作,但是假定通常嘈杂的环境以及有限的处理器能力,在没有指针且其中不能呈现识别准确度的移动设备上,它就变得麻烦。在只有硬件按钮、键区或触摸屏等的设备上,难以设计出允许用户选择一定范围的单词以供纠正,而同时将击键保持在合理次数的界面。

[0013] 以上讨论仅提供了一般的背景信息,并不旨在用于协助确定所要求保护的的主题的范围。

[0014] 概述

[0015] 本发明使用语音和备选形式输入(诸如键区输入)的组合来将信息传递给移动设备。用户说出包括多个单词(诸如词组或句子)的话语。语音识别结果接着被一次一个单词地呈现给用户,以便确定或纠正。在屏幕上从第一个单词开始一次一个单词地向用户呈现最佳假设和一选择列表。如果在屏幕上呈现的最佳假设单词是正确的,那么用户可以容易地指示它。否则,如果期望单词在备选项列表中,那么在用户方面只要很少的努力(例如很少的按下按钮、击键等),用户可以快速地导航到该备选项列表,并使用各种备选输入形式之一来输入单词。

[0016] 在一个实施例中,如果没有在备选项列表中找到单词,则用户可以使用键区开始输入单词。类似地,在一个实施例中,该系统可以使用通过将来自键区输入的单词前缀的信息、语音识别结果点阵(lattice)、已经被纠正的当前单词之前的单词、语言模型等组合而获取的后验概率来重新计算最佳假设单词和备选项列表。该过程可以对输入句子中的后续单词重复。

[0017] 可以使用诸如软键盘、触摸屏输入、手写输入等其他输入形式作为键区输入的代表或附加。

[0018] 提供本概述以使用简化的形式介绍将在以下详细描述中进一步描述的一些代表性概念。本概述并不旨在确定所要求保护的的主题的关键特征或必要特征,也不旨在用于帮助确定所要求保护的的主题的范围。

[0019] 附图简述

[0020] 图 1 是其中可使用本发明的一个说明性计算环境的框图。

[0021] 图 2-4 示出了其上可采用本发明的设备的不同示例性、简化的图示实施例。

[0022] 图 5 示出了语音识别系统的一个说明性实施例。

[0023] 图 6 是依照一个实施例被配置成实现用户界面系统的设备的框图。

- [0024] 图 6A 是示出图 1-6 所示的系统的总体操作的一个实施例的流程图。
- [0025] 图 6B 是示出生成语音识别假设点阵过程中的系统操作的一个实施例的流程图。
- [0026] 图 6C 示出了一个示例性的初步假设点阵。
- [0027] 图 6D 示出了用于选择单词的示例性用户界面显示。
- [0028] 图 6E 示出了在给定用户对假设中的单词的纠正的时经修改的假设点阵。
- [0029] 图 6F 示出了用于选择语音识别假设中的单词的一个示例性用户界面显示。
- [0030] 图 7 示出了说明对假设的重新计算的一个示例性流程图。
- [0031] 图 8 示出了显示预测性单词补全的一个示例性用户界面显示。
- [0032] 详细描述

[0033] 本发明涉及将语音和备选输入形式组合,以便改进移动设备上的文本输入效率和稳健性。然而,在更详细地描述本发明之前,将描述其中可使用本发明的一个说明性环境。

[0034] 以下在图 1 中所示的计算设备 10 通常包括至少某种形式的计算机可读介质。计算机可读介质可以是可由设备 10 访问的任何可用介质。作为示例而非局限,计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于储存诸如计算机可读指令、数据结构、程序模块或其它数据等信息的任一方法或技术实现的易失性和非易失性,可移动和不可移动介质。计算机存储介质包括但不限于,RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘(DVD)或其它光盘存储、磁盒、磁带、磁盘存储或其它磁存储设备、或可以用来储存所期望的信息并是设备 10 的一部分或可由设备 10 访问的任一其它介质。通信介质通常以诸如载波或其它传输机制等已调制数据信号来体现计算机可读指令、数据结构、程序模块或其它数据,并包括任一信息传送介质。术语“已调制数据信号”指以对信号中的信息进行编码的方式设置或改变其一个或多个特征的信号。作为示例而非局限,通信介质包括有线介质,如有线网络或直接连线连接,以及无线介质,如声学、RF、红外和其它无线介质。上述任一组合也应当包括在计算机可读介质的范围之内。

[0035] 图 1 是移动设备 10 的一个实施例的框图。如图所示,移动设备 10 包括处理器 20、存储器 22、输入/输出(I/O)组件 24、台式计算机通信接口 26、收发机 27 和天线 11。在一个实施例中,移动设备 10 的这些组件被耦合,用于通过合适的总线 28 与另一移动设备通信。虽然没有在图 2 中示出,但是移动设备 10 还在 I/O 组件 24 内包括如以下有关图 5 所示和所讨论的话筒。

[0036] 存储器 22 被实现为非易失性电子存储器,诸如带有电池后备模块(未示出)的随机存取存储器(RAM),使得当移动设备 10 的总电源被关闭时,存储在存储器 22 中的信息不会丢失。存储器 22 的一部分被分配为用于程序执行的可寻址存储器,而存储器 22 的另一部分可用于存储,诸如模拟磁盘驱动器上的存储。

[0037] 存储器 22 包含操作系统 30,应用程序 16(诸如用户界面应用程序、个人信息管理器(PIM)、日程安排程序、文字处理程序、电子表格程序、因特网浏览器程序以及以下讨论的语音识别程序)、用户界面组件 17 以及对象存储 18。在操作过程中,操作系统 30 被载入到存储器 22 中,并且由处理器 20 从存储器 22 中执行。在一个实施例中,操作系统 30 是市场上可购买到的来自微软公司的 Windows CE 操作系统。操作系统 30 可以是为移动设备设计的,并实现可由 PIM、内容查看器、语音识别功能等使用的特征。这可以用任何期望的方式来完成,诸如通过所展示的应用程序编程接口或通过专用接口等。对象存储 18 中的对象可以

由 PIM、内容查看器和操作系统 30 至少部分地响应于对其的调用来维护。

[0038] 用户界面组件 17 说明性地与其他组件交互,以便将输出显示提供给用户并用于接收来自用户的输入。用户界面组件 17 在接收作为语音和键区输入的组合作用的用户输入时的操作的一个实施例在以下关于图 6A-8 来描述。

[0039] 在一个实施例中,提供 I/O 组件 24 以便于来自移动设备 10 的用户的输入和输出操作。这种组件可以包括显示器、触敏屏幕、键区、话筒、扬声器、音频发生器、振动设备、LED、按钮、滚轮或用于将信息输入到设备 10 或从设备 10 输出信息的其他机制等等。这些仅是作为示例。它们不必都存在,并且也可以提供其他或不同的机制。并且,还可以支持其他通信接口和机制,诸如有线和无线调制解调器、卫星接收器和广播调谐器等。

[0040] 可任选地提供台式计算机通信接口 26 作为任何合适的、在市场上可以购买到的通信接口。当无线收发机 27 不是用于与台式机或其他计算机 12 通信时,接口 26 被用于该目的。接口 26 可以包括例如红外线收发机或串行或并行连接。

[0041] 收发机 27 是适用于通过期望的传输来发送信号或信息的无线或其他类型的收发机。在其中收发机 27 是无线收发机的实施例中,可以使用天线 11 来发送信号或信息。收发机 27 也可以通过传输来发送其他数据。在一些实施例中,收发机 27 从台式计算机、信息源提供者或其他移动或非移动设备或电话接收信息。收发机 27 被耦合到总线 28,用于与处理器 20 通信以便存储接收到的信息并且发送要传送的信息。

[0042] 电源 35 包括用于为移动设备 10 供电的电池 37。可任选地,移动设备 10 可以从取代内置电池 37 或为其充电的外部电源 41 接收电力。例如,外部电源 41 可以包括合适的 AC 或 DC 适配器,或用于移动设备 10 的电源底座。

[0043] 应该注意,图 1 示出了图 1 中所示的其中可实现本发明的合适的操作环境的一个示例。图 1 中所示的操作环境仅是合适的操作环境的一个示例,并不旨在对本发明的使用范围或功能提出任何限制。其他可适用于本发明的公知的计算系统、环境和 / 或配置包括但不局限于蜂窝电话、个人数字助理、寻呼机、手持式或膝上型设备、多处理器系统、基于微处理器的系统、可编程消费电子设备、包括任何上述系统或设备的分布式计算环境等。

[0044] 也应该注意,本发明可以在诸如可由一个或多个计算机或其他设备执行的程序模块等计算机可执行指令的一般上下文中描述。一般地,程序模块包括例程、程序、对象、组件、数据结构等等,它们执行特定的任务或实现特定的抽象数据类型。通常程序模块的功能可以按需在各个实施例中组合或分布。

[0045] 图 2 是可以依照本发明使用的移动设备 10 的一个实施例的简化的图示说明。在该实施例中,除了天线 11 和话筒 75 之外,移动设备 10 还包括小型键盘 32、显示器 34、指示笔 36 以及扬声器 86。在图 3 所示的实施例中,显示器 34 是液晶显示器 (LCD),它们使用触敏显示屏幕以及指示笔 36。指示笔 36 用于在指定的坐标处按下或接触显示器 34 以实现某些用户输入功能。小型键盘 32 可被实现为小型字母数字键盘,其具有同样为了完成某些用户输入功能而提供的任何合适的和期望的功能键。话筒 75 被示为位于天线 11 的远端,但是它可以方便地被设置在设备 10 上的任何地方。

[0046] 图 3 是依照本发明的另一实施例的移动设备 10 的另一简化的图示说明。如图 3 中所示,移动设备 10 包括类似于关于图 2 所描述的那些项目的一些项目,并且这些项目被类似地编号。例如,如图 3 中所示,移动设备 10 也包括置于天线 11 上的话筒 75 和置于该

设备的外壳上的扬声器 86。当然,话筒 75 和扬声器 86 也可以位于其他位置。同样,移动设备 10 包括触敏显示器 34,它可以结合指示笔 36 使用以便实现某些用户输入功能。应该注意,图 2 和 3 中所示的移动设备的显示器 34 可以是相同尺寸或不同尺寸的,但是通常会比用于台式计算机的常规显示器要小得多。例如,图 2 和 3 中所示的显示器 34 可以由仅有 240×329 个坐标或 160×160 个坐标或任何其他合适的大小的矩阵来定义。

[0047] 图 3 中所示的移动设备 10 也包括多个用户输入键或按钮(诸如滚动按钮 38 和/或键盘 32),它们允许用户输入数据或滚动通过菜单选项或在显示器 34 上显示的其他显示选项,而无需接触显示屏 34。此外,图 3 中所示的移动设备 10 也包括电源按钮 40,它可用于开启和关闭移动设备 10 的总电源。

[0048] 也应该注意,在图 3 所示的实施例中,移动设备 10 可以包括手写区域 42。手写区域 42 可以结合指示笔 36 使用,使得用户能够书写存储在存储器 22 中的消息,以供稍后由移动设备 10 使用。在一个实施例中,手写消息仅以手写形式存储并且可以由用户检索并显示在显示器 34 上,使得用户可以审阅输入到移动设备 10 中的手写消息。在另一实施例中,移动设备 10 配备有字符识别模块,使得用户可以通过使用指示笔 36 在区域 42 上书写字母数字信息来将该字母数字消息输入到移动设备 10 中。在该情况中,移动设备 10 中的字符识别模块识别该字母数字字符并将该字母转换成计算机可识别的、可由移动设备 10 中的应用程序 16 使用的字母数字字符。

[0049] 图 4 示出了依照本发明的一个实施例的移动设备 10 的另一实施例的图示说明。移动设备 10 具有显示区域 34、电源按钮 504、多个控制按钮 506、一串附加的控制按钮 508、话筒 509 和键区 510。键区 510 说明性地包括多个不同的字母数字按钮(其中的一些由标号 512 示出),并且也可以包括键盘按钮 514。用户可以通过使用指示笔或手指或其他机制按下按钮 512 来将字母数字信息输入到图 4 所示的设备 10 中。可以使用各种字母输入技术中的任一种,诸如确定性多次轻叩方法、预测性技术等,来通过按钮 512 输入字母数字信息。类似地,在一个实施例中,如果用户期望切换到其它的键入方法,用户只需操纵键盘按钮 514。在该情况下,设备 10 显示常规键盘的精简描绘而非字母数字按钮 512。接着用户可以通过使用指示笔轻叩所显示的键盘上的那些字符等,一次一个字母地输入文本信息。此外,也可以在各个实施例中使用其他备选的输入形式,诸如手写输入和其他触摸屏或其他输入。

[0050] 在一个实施例中,设备 10 还包括语音识别系统(这将在下文中关于图 5 更详细地描述),使得用户可以通过话筒 509 来将语音信息输入到设备 10 中。类似地,设备 10 说明性地包括由界面组件 17(图 1 中)运行的界面,它允许用户组合语音和键区输入以便将信息输入到设备 10 中。尤其是在不具有传统的键盘的移动设备上,这改进了文本输入效率和稳健性。这在下文中关于图 6A-8 更详细地描述。

[0051] 图 5 是依照一个实施例可以在上文的图 2-4 中所示的移动设备中的任一种上使用的语音识别系统 200 的一个说明性实施例的框图。

[0052] 在图 5 中,说话者 201(训练者或用户)向话筒 17 说话。由话筒 17 检测到的音频信号被转换成电信号,该电信号被提供给模数(A-D)转换器 206。

[0053] A-D 转换器 206 将来自话筒 17 的模拟信号转换成一系列数字值。在若干个实施例中,A-D 转换器 206 以 16kHz 和每个样本 16 位来对该模拟信号采样,由此创建每秒 32 千字



节的语音数据。该数字值被提供给帧构造器 207, 在一个实施例中, 帧构造器 207 将这些值分组成其开头有 10 毫秒相隔的 25 毫秒的帧。

[0054] 由帧构造器 207 创建的数据帧被提供给特征提取器 208, 特征提取器 208 从每个帧中提取特征。特征提取模块的示例包括用于执行线性预测编码 (LPC)、LPC 导出倒谱、感知线性预测 (PLP)、听觉模型特征提取以及梅尔频率 (Mel-Frequency) 倒谱系数 (MFCC) 特征提取的模块。注意, 本发明不限于这些特征提取模块, 可以在本发明的上下文中使用其他模块。

[0055] 特征提取模块产生一特征向量流, 每个特征向量与一语音信号帧相关联。

[0056] 也可以使用降噪, 使得来自提取器 208 的输出是一系列“干净的”特征向量。如果输入信号是训练信号, 则该系统“干净的”特征向量被提供给训练器 224, 该训练器如下文中更详细描述的那样使用“干净的”特征向量和训练文本 226 来训练声学模型 218 或其他模型。

[0057] 如果输入信号是测试信号, 则该系统“干净的”特征向量被提供给解码器 212, 该解码器基于特征向量流、词典 204、语言模型 216 和声学模型 218 来标识最可能的单词序列。解码所使用的具体方法对本发明而言不重要, 并且可以使用用于解码的若干个已知方法中的任一种。

[0058] 最可能的假设单词序列作为语音识别点阵被提供给置信度测量模块 220。置信度测量模块 220 部分地基于一辅助声学模型 (未示出) 来标识哪些单词最可能被语音识别器不正确地标识。置信度测量模块 220 接着将点阵中假设单词序列连同指示哪些单词可能被不正确地标识的标识符一起提供给输出模块 222。本领域的技术人员将认识到, 置信度测量模块 222 对于本发明的实现不是必需的。

[0059] 图 6 是用户界面系统 550 的框图, 其中设备 10 被配置成实现依照本发明的一个实施例的界面。图 6A 是示出依照一个实施例的移动设备 10 及其界面的操作, 并且将彼此结合地描述图 6 和 6A。虽然该界面可被部署在上文中关于图 2-4 讨论的任一种移动设备上, 但是仅是为了示例起见, 以下讨论将关于图 4 中所示的设备 10 来进行。

[0060] 依照本发明的一个实施例, 该界面允许用户将语音输入 552 和备选形式输入 554 组合以便将信息输入到设备 10 中。备选形式输入可以是使用上述形式 (软键盘、触摸屏输入、手写识别等) 中的任一种的输入。然而, 仅是为了示例起见, 此处将按照键区输入 554 来描述备选形式输入。由此, 依照一个实施例, 用户 556 首先诸如通过按住一功能按钮或操纵该用户界面上任何其他期望的按钮提供激活输入 558, 来激活设备 10 上的语音识别系统 200。这由图 6A 中的框 600 指示。接着, 用户向设备 10 上的话筒 75 说出多单词语音输入 552 (诸如词组或句子), 并且设备 10 中的语音识别系统 200 接收该多单词语音输入 522。这由图 5A 中的框 602 指示。语音识别系统 200 生成假设点阵 560 形式的语音识别结果并将点阵 560 提供给用户界面组件 17。接着, 用户界面组件 17 顺序地一次一个单词地显示 (在用户界面显示器 34 上) 语音识别结果, 以供用户使用键区 510 来确认或纠正。这由图 5A 中的框 604 来指示。用户在顺序地显示每个单词时说明性地使用键区 510 上的键来纠正它或者确认该单词是正确的。随着顺序提交的继续, 显示经纠正或确认的单词, 并且将下一顺序单词添加到显示器以供纠正或确认。这继续直至显示全部的语音识别结果。

[0061] 初看起来, 连续语音识别与逐个单词纠正机制 (此处有时称为顺序提交机制) 的

组合可能好像是不是最由的,并且无疑是反直觉的。然而,对于移动设备上当前的自动语音识别系统,相信这有助于更好的总体用户体验。

[0062] 例如,自动语音识别差错通常涉及分段差错。如在背景一节中所讨论的,假设短语“recognized speech”可能被自动语音识别系统误识别为“wreck a nice beach”。在这种情况下,显示完整的自动语音识别结果导致纠正界面的难以选择。这导致的一些问题是:用户应该选择哪些单词来纠正?当用户试图纠正单词“wreck”时,它是否应该造成该语音识别结果中的词组的剩余部分改变?如果作为用户对于不同单词的纠正的副作用其他单词开始改变,它会如何影响用户的置信度水平?

[0063] 在设计界面时所有这些问题都必须被解决,并且很难获得对所有这些问题的最优解决方案。类似地,在解决这些问题中的每一个并且向用户提供解决它们的用户界面选项时,通常导致需要相对大量的击键以供用户纠正被误识别的句子或词组。

[0064] 与此形成对比,通过顺序地从左到右呈现一个单词接一个单词的结果以供用户确认或纠正,基本上避免了所有这些问题。在“wreck a nice beach”的假设的误识别中,本发明将首先在设备 10 的显示部分 34 上仅呈现单词“wreck”以供用户确认或纠正。连同单词“wreck”一起,本系统也将说明性地显示备选项。由此,识别结果将可能包括“recognize”作为单词“wreck”的第二备选项。一旦用户将“wreck”纠正为“recognize”,那么本系统将说明性地重新计算与各个语音识别假设相关联的概率,并且接着输出下一个单词“speech”作为第一个假设,而非“a”(给定用户作出先前纠正(“wreck”到“recognize”)的上下文)。

[0065] 在一个说明性实施例中,如图 4 中所示,显示部分 34 显示语音识别结果以及显示用于当前正被确认或纠正的单词的各个备选项的下拉菜单 503。如果在语音识别结果中显示的单词是正确的,用户可以只需操纵“OK”按钮,或者用户可以滚动通过下拉菜单 503 中显示的各个备选项,并且为所显示的单词选择正确的备选项。在所示的实施例中,“OK”按钮可以位于功能按钮串 508 中,或者它可以位于键区 510 上,或者它可以是按钮 506 中的一个等。当然,也可以使用其他实施例。

[0066] 更具体地,在图 4 中对设备 10 所示的示例中,用户输入了语音输入 552 “this is speech recognition”(这是语音识别)。系统接着向用户显示第一个单词“this”,并且它已由用户确认。图 4 中的系统被示为已向用户显示了识别假设中的第二个单词,并且用户已从下拉菜单 503 中的备选项列表中选择了“is”。系统接着重新计算与各个识别假设相关联的概率,以找出最可能的单词以便作为假设中的第三个单词来显示,这还未在图 4 中示出。

[0067] 在一个说明性实施例中,备选项下拉菜单 503 是出现在假设的语音识别结果中的当前插入点之下的浮动列表框,并且显示符合识别结果的当前所选的预测。显示被格式化以突出显示指定的前缀。此外,在一个说明性实施例中,框 503 中的列表的高度可以被设置成任意期望的数字,并且认为大约四个可见项目的列表限制了预测列表引入的注意力分散的量。类似地,可以将下拉菜单 503 的宽度调整为列表中最长的单词。此外,如果识别结果中的插入点太靠近文档窗口的边界,使得下拉菜单 503 中的列表框延伸到边界之外,则识别结果中的插入点和下拉菜单 503 中的预测列表可以换到下一行。

[0068] 通过键区 510 接收用户确认或纠正作为键区输入 554 由图 6A 中的框 606 示出。当

然,应该注意用户可以用各种不同的方式来提供键区输入 554,而非仅是从备选项下拉菜单 503 中选择一备选项。

[0069] 对于理想的语音识别,用户只需对每个正确识别的单词按下“OK”,从而得出非常高效的文本输入。然而,对于不理想的语音识别,用户对每个正确识别的单词按下“OK”,并且对于误识别的单词,可以或者向下滚动到期望的备选项并从备选项菜单 503 中选择它,或者开始一次一个字母地拼写正确的单词直至在下拉菜单 503 的预测列表中出现了期望的单词。

[0070] 刚才所描述的基于各个用户输入首先向用户显示的推荐的单词和备选项是说明性地从假设点阵 560 取得的,该假设点阵 560 由语音识别系统 200 响应于语音输入 552 而生成。然而,可能发生语音识别系统 200 误识别语音输入以至于要显示的正确单词实际上没有出现在假设点阵 560 中。为了处理未出现在点阵 560 中的单词,可以将来自假设点阵 560 的预测与来自语言模型(诸如 n 元语法语言模型)的预测合并,并且按照概率来排序。这样,通常可以输入语音识别系统 200 的词汇表中的单词而无需拼写出整个单词,在某些情况下无需键入一个字母,即使它们不出现在识别结果的假设点阵 560 中。这很大程度上减少了输入在最初的识别点阵 560 中没有找到的单词所需的击键。

[0071] 也可能发生用户输入的单词不仅没有出现在识别点阵 560 中,也没有出现在语音识别系统 200 的词典或词汇表中。在这种情况下,在一个实施例中,用户界面组件 17 被配置成切换到允许用户通过键区 510 拼写词汇表之外的单词的确定性的、逐个字母的输入配置。这种逐个字母的配置可以包括例如多次轻叩输入配置或键盘输入配置。

[0072] 对于键盘配置,图 4 中所示的设备 10 包括软键盘键 514。当用户操纵该键时,示出键盘的显示并且用户可以使用指示笔方便地使用“看一键打一字”来逐个字母地输入不在语音识别系统 200 的原始词汇表中的单词。接着可以按需将这些单词添加到词汇表中。

[0073] 类似地,可以在下拉菜单 503 的备选项显示的结尾处(或者在下拉菜单 503 中的任何其他位置)提供键盘选项,而非具有不变地显示的键盘按钮 514。当用户从备选项列表中操纵该选项时,再次显示键盘并且用户可以一次一个地输入字母。一旦提交了词汇表之外的单词,键盘说明性地消失,并且设备 10 中的用户界面组件 17 如上所述地转变回其先前的操作模式(可能是一次一个单词地显示以供用户确认或纠正)。

[0074] 语音识别等待时间也会成为移动设备上的问题。然而,由于本系统是顺序递交系统(因为它从句子的左侧开始一次一个单词地向用户提供语音识别结果以供确认或纠正,并向右侧前进),因此本系统可以利用从不完全的假设点阵生成的中间自动语音识别假设。换言之,在自动语音识别系统 200 完全完成了用于语音识别结果的整个假设点阵 560 的处理之前,本系统可以开始一次一个地向用户呈现单词假设。由此,在初始超时期限(诸如 500 毫秒或任何其他期望的超时期限)后,而在语音识别系统 200 生成对于用户输入的完整文本片断或句子的假设之前,本系统可以向用户显示语音识别结果中的第一个假设的单词。这使用户能以很短的等待时间开始纠正或确认语音识别结果,即使当前的移动设备具有相对有限的计算资源。

[0075] 图 6B 是更好地说明该系统如何减小语音识别等待时间的流程图。首先,用户向移动设备 10 说出完整的多单词语音输入 552。这由图 6B 中的框 650 示出。接着解码器开始计算假设点阵 560。这由框 652 示出。

[0076] 解码器接着判定点阵计算是否完成。这由框 654 示出。如果否,那么判定是否已经过了预定超时期限。这由框 656 示出。换言之,即使还没有计算出完整的假设点阵,本系统也将在预先指定的超时期限后输出中间点阵 560。由此,如果在框 656 处已经过了超时期限,那么系统输出初步的假设点阵 560,并且界面组件 17 向用户显示在语音识别结果中的第一个假设的单词(来自初步假设点阵 560)。这由图 6B 中的框 658 指示。

[0077] 在该时间期间,解码器在框 652 处继续计算完整的假设点阵。然而,界面组件 17 说明性地使用初步的假设点阵保持向用户呈现单词以供确认或纠正,直至完成了完整的假设点阵。一旦完成了完整的假设点阵 560,就输出完整的点阵以供界面组件 17 用于顺序地向用户呈现单词(一次一个单词)以供确认或纠正。这由图 6B 中的框 660 指示。

[0078] 在一个替换实施例中,当解码器计算假设点阵,并且之后计算出初步的点阵之后,用户对语音识别结果的任何确认或纠正被反馈给解码器,使得它能够考虑该用户确认或纠正信息来完成假设点阵的处理。这由框 655 指示。通过向识别器 200 提供所提交的单词序列,这提供了可由识别器用于缩小搜索空间的信息。实际上,有了该信息,识别器 200 可以剪除与所提交的单词序列不一致的所有搜索路径,以便显著加速搜索过程。当然,搜索路径剪除不仅加速了搜索,还通过允许引擎搜索与已经提交的单词序列一致的更多路径(否则的话可能被剪除)来提高准确度

[0079] 此时一个示例可以增强理解。假设用户 556 激活了设备 10 上的语音识别系统 200。还假设用户通过其话筒 75 将多单词语音输入 552 “this is speech recognition”输入到设备 10 中。设备 10 中的自动语音识别系统 200 开始处理该语音输入,以便创建指示假设的语音识别结果和备选项的假设点阵 560。然而,在自动语音识别系统 200 计算出整个假设点阵之前,可以说明性地计算初步点阵 560。图 6D 示出了从示例性语音输入生成的一个示例性的部分(或初步)假设点阵。该假设点阵一般地由图 6D 中的标号 662 示出。依照一个实施例,向用户界面组件 17 提供点阵 662,使得用户界面组件 17 可以开始向用户 556 一次一个单词地呈现来自假设的语音识别结果的单词,以供确认或纠正。

[0080] 假设在点阵 662 中,单词“this”是该点阵中表示最佳假设单词的第一个单词。由此会将单词“this”呈现给用户以供确认或纠正。图 6E 示出了设备 10 的显示器 34 的一部分,示出将单词“this”连同下拉菜单 503 中以概率得分顺序呈现的来自假设点阵 662 的备选项一起呈现给用户。在菜单 503 中列出的备选项是“Miss”和“Mrs.”。可以从点阵 662 看出点阵中有对于“this”的其他可能的备选项。用户然后可以或者通过方便地操纵“OK”按钮来接受所显示的结果“this”,或者用户可以选择如上所述的备选项之一。

[0081] 在用户作出确认或纠正所显示的语音识别结果的选择期间,解码器继续处理语音输入,以便完成语音识别点阵的计算。这可能只需花费几秒钟。由此,可能甚至在用户已纠正或确认了一个或两个单词的假设单词之前,解码器就已经完整地计算出整个假设点阵。

[0082] 图 6E 示出了解码器对示例性语音输入“this is speech recognition”计算出的完整的假设点阵 664。由于用户选择了“this”作为语音识别结果中的第一个单词,因此在点阵 664 上划去其他两个备选项“Mrs.”和“Miss”以显示在假设语音识别结果中的其他单词时不再考虑它们。实际上,由于用户已确认了单词“this”,解码器肯定地知道“this”是识别结果中正确的第一个单词。接着,解码器可以为点阵 664 中的所有其他的单词重新计算概率,并且用户界面组件 17 可以基于该重新计算将最高得分的单词呈现给用户,作为识

别结果中的下一个假设单词。

[0083] 图 6F 示出了现在向用户显示单词“is”作为语音识别结果中“this”之后的最高得分的假设单词的界面。图 6F 示出了下拉菜单 502 也显示了对单词“is”的多个备选项, 以及如果用户通过选择备选单词之一而选择纠正语音识别结果时可用于在各个备选项之间滚动的滚动条。

[0084] 也应该注意, 在图 6F 所示的实施例中, 下拉菜单 503 包括“键盘”选项, 用户可以操纵该选项来显示键盘, 使得用户能够使用指示笔或其他合适的输入机制一次一个字母地输入单词。

[0085] 图 7 是依照本发明的一个实施例更好地示出在使用初步的和完整的点阵 662 和 664 时, 图 6 中所示的系统的操作的流程图。如以上关于图 6B 所讨论的, 用户界面组件 17 首先接收初步假设点阵 (诸如图 6D 中所示的点阵 662)。这由图 6 中的框 680 示出。

[0086] 用户界面组件 17 接着输出语音识别结果中当前单词的最佳单词假设。例如, 如果这是向用户显示以供纠正或确认的第一个单词, 那么用户界面组件 17 为语音识别结果中的第一个单词位置从初步假设点阵选择最佳得分单词, 并将其显示给用户。这由图 6 中的框 682 示出, 并且其示例在图 6E 中示出。

[0087] 用户界面组件 17 接着接收关于当前所显示的单词的用户纠正或确认输入 554。这由图 6 中的框 684 示出。再一次, 这可以通过用户从备选项列表中选择一备选项、通过用户开始键入不在假设点阵中但是仍可以在自动语音识别系统 200 使用的字典或词汇表中找到的单词、或通过用户输入先前没有出现在自动语音识别系统 200 的词汇表或词典中的新单词来完成。

[0088] 应该注意, 在第二种情况下 (其中用户开始键入不在假设点阵中但是仍可以在自动语音识别系统 200 使用的词典中找到的单词) 可以使用前缀反馈。这在图 8 中更好地示出。例如, 假设所考虑的语音识别结果的正确的单词是“demonstrates”。还假设基于语音输入 552, 单词“demonstrates”未出现在自动语音识别系统 200 生成的假设点阵中。然而, 假设单词“demonstrates”在自动语音识别系统 200 使用的词典中。在这种情况下, 用户将开始一次一个字母地键入单词 (诸如通过选择键盘选项或多次轻叩输入选项)。随着用户输入每个字母, 自动语音识别系统 200 基于已经输入的前缀字母使用预测性单词补全。在一个说明性实施例中, 系统还突出显示用户已经输入的字母, 使得用户可以方便地确定已经输入了哪些字母。从图 8 中可以看出用户已经输入了字母“demon”并且已经预测到单词“demonstrates”。

[0089] 也应该注意, 即时当单词已经出现在假设点阵中时也可以使用该选项 (一次一个字母地输入单词)。换言之, 用户可以方便地输入允许用户一次一个字母地输入单词的输入配置, 而非用户滚动通过下拉菜单 503 中的备选项列表以便找到正确的备选项。基于用户输入的每个字母, 系统 200 重新计算各个单词的概率并且基于给定前缀的情况下最高概率的单词来重新排列所显示的替换项。

[0090] 在一个实施例中, 为了重新排列单词, 在给定用户识别出先前单词的上下文以及给定用户输入的前缀字母的情况下, 不仅基于由用户已经输入的前缀字母, 而且还基于出现在语音识别假设中的该单词位置中的单词以及也基于用户已经确认或纠正的先前的单词, 并且进一步基于其他排序组件, 诸如上下文相关组件 (例如 n 元语法语言模型) 来执行

重新排列。

[0091] 无论如何,接收用户纠正或确认输入 554 由图 7 中的框 684 示出。用户界面组件 17 基于用户输入纠正或确认单词,如图 6 中框 686 所示。

[0092] 如果刚确认或纠正的单词是语音识别结果中前几个单词中的一个,则可能用户界面组件 17 正基于初步点阵向用户提供假设单词,如以上有关图 6B 所描述的。因此,判定是否已经接收到完整的点阵。这由图 6 中的框 688 示出。如果是,那么如框 690 所示为所有今后的处理使用完整的点阵。如果还没有接收到完整的点阵,那么再次使用初步点阵来处理语音识别结果中的下一个单词。

[0093] 一旦用户确认或纠正了正在处理的当前单词,用户界面组件 17 判定在假设的语音识别结果中是否有更多的单词。这由图 7 中的框 692 示出。

[0094] 如果是,那么自动语音识别解码器为可能作为语音识别结果中的下一个单词提议的每个可能的单词重新计算得分。再一次,这一对下一单词的得分的重新计算可以基于由用户已经确认或纠正的单词、基于在假设点阵中找到的单词、基于语言模型得分或基于其他期望的建模得分。

[0095] 为了从点阵生成候选单词,首先确定通过对应于该结果的已确认单词序列可以从初始点阵节点获得哪个候选单词集。来自这些候选节点的与点阵中的外出弧相关联的该单词列表形成了由点阵预测的候选单词。例如,在图 6F 所示的点阵 664 中,假设用户已确认或纠正了单词“this is”,则在给定已经确认的单词情况下,接着在语音识别结果中可能的备选项是“speech”、“beach”和“bee”。

[0096] 为了确定每个候选单词的概率,通过使用动态编程组合匹配路径的概率来计算每个候选节点的前向概率。对于每个候选单词过渡,从后验前向概率、局部过渡概率和后向得分来计算总体过渡概率。通过组合来自相应的候选单词过渡的概率来确定每个候选单词的最终概率。在一个实施例中,可以通过添加概率来准确地计算概率组合,或通过取最大值以 Viterbi 方式来估算概率组合。为了减少计算,随着用户提交每个单词递增地计算候选节点和对应的概率。当然应该注意,这仅是计算与语音识别结果中的下一个单词相关联的得分的一种方式。基于该信息重新计算得分在图 7 中由框 696 示出。

[0097] 在框 692 处,如果没有更多的单词要处理,那么语音识别结果完成,并且处理结束。

[0098] 可以看出,通过组合用于移动设备上的文本输入的键区和语音输入,当单词被误识别时,顺序提交范例优于传统的随机访问纠正。它采用较少的击键,并且话语允许系统显示具有不同分割的单词备选项,而同时以非常直接的方式呈现结果。因此,当识别点阵表示正确的识别时,用户无需因为不正确的单词分割而纠正多单词短语,由此导致更少的击键。当选择多个单词以供纠正时,这也避免了备选项的组合激增的问题。

[0099] 此外,知道先前提交的单词使得系统能够基于语言模型和与已提交单词的声学对齐依照它们的后验概率来重新排列假设。由此,感知到的准确度比传统系统高,在传统系统中在纠正之后不能改变对话语的剩余部分的假设。通过仅显示要纠正和提交的下一个单词,顺序提交系统改进了感知的准确度并且导致击键的减少。

[0100] 类似地,顺序提交是基于现有文本输入方法的用户熟悉的一次一个单词的输入界面。在语音输入不合适的情况下,用户可以方便地跳过说出期望话语的第一步骤,而开始仅

使用键区来输入文本。由此,该系统是非常灵活的。

[0101] 尽管以对结构特征和 / 或方法动作专用的语言描述了本主题,但是可以理解,所附权利要求书中定义的本主题不必限于先前所描述的具体特征或动作。相反,这些具体特征和动作是作为实现权利要求书的示例来公开的。

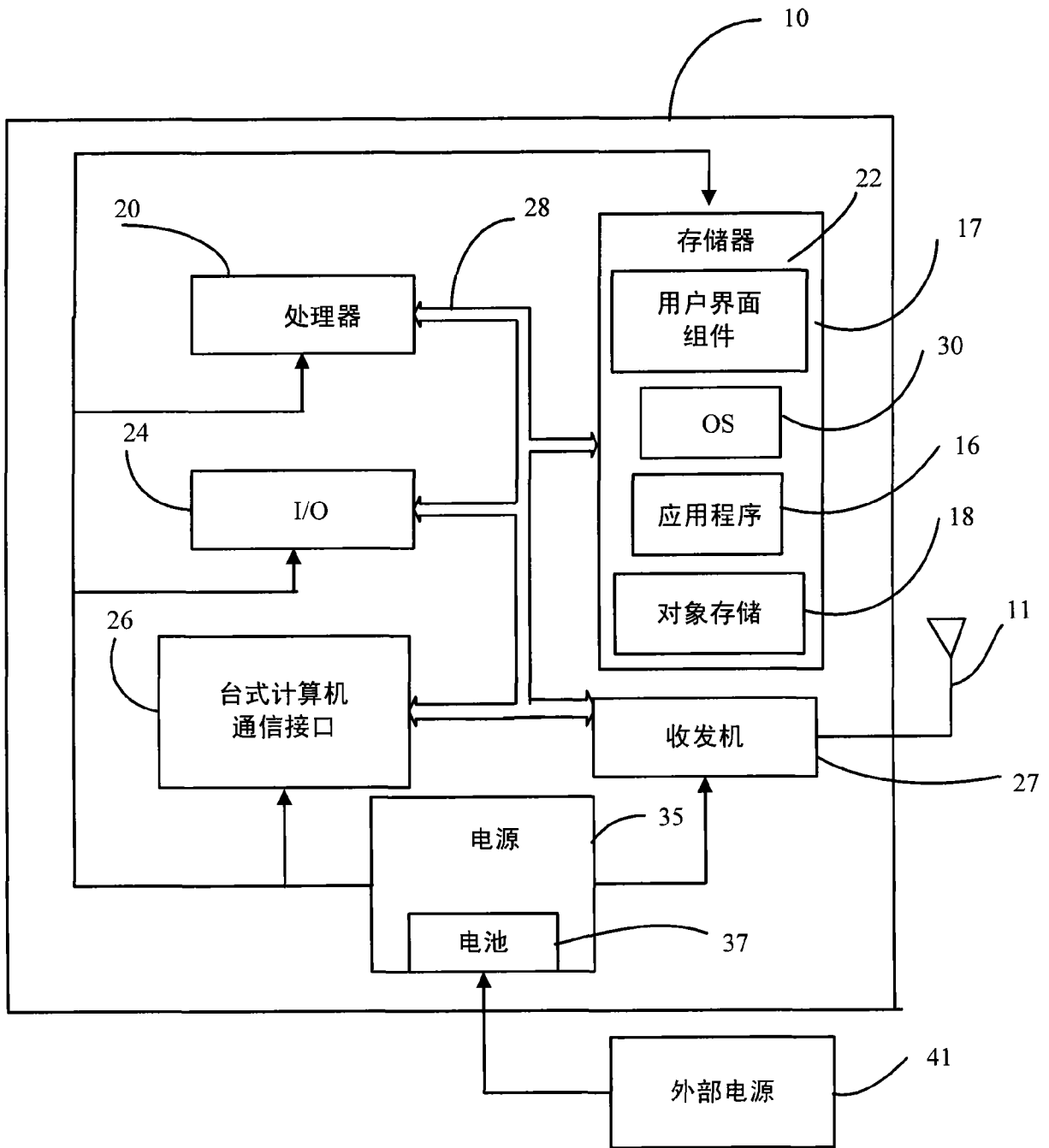


图 1



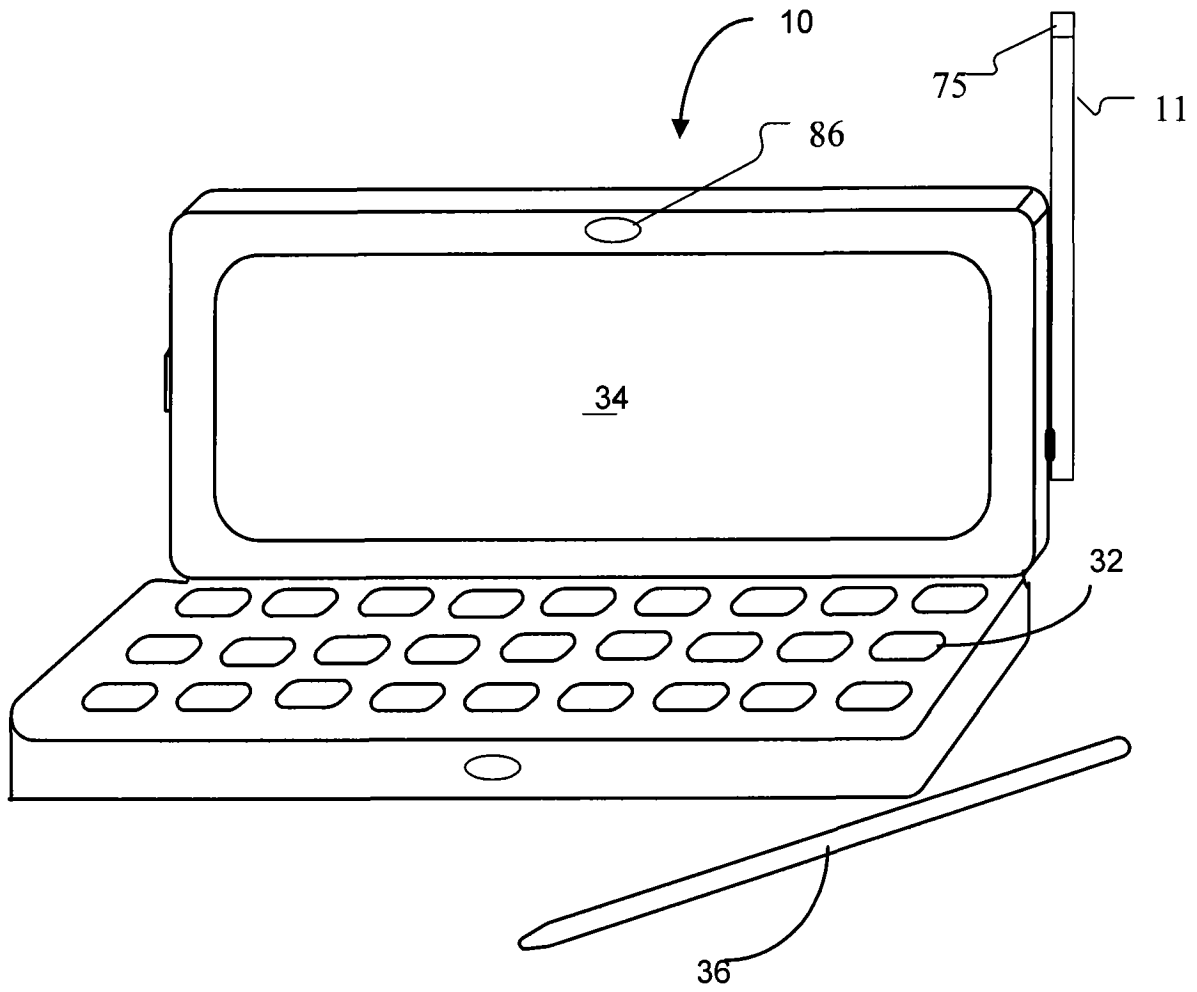


图 2

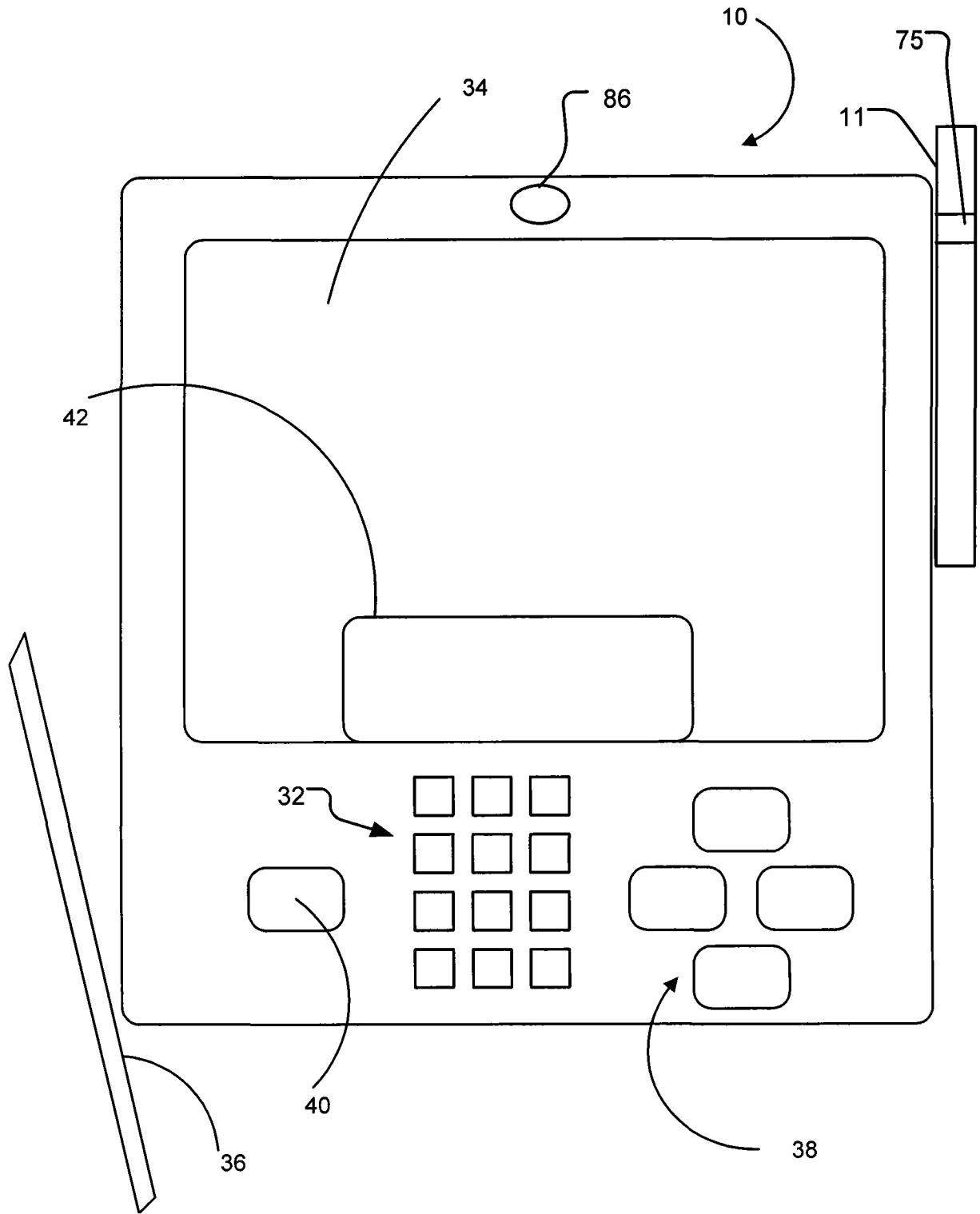


图 3

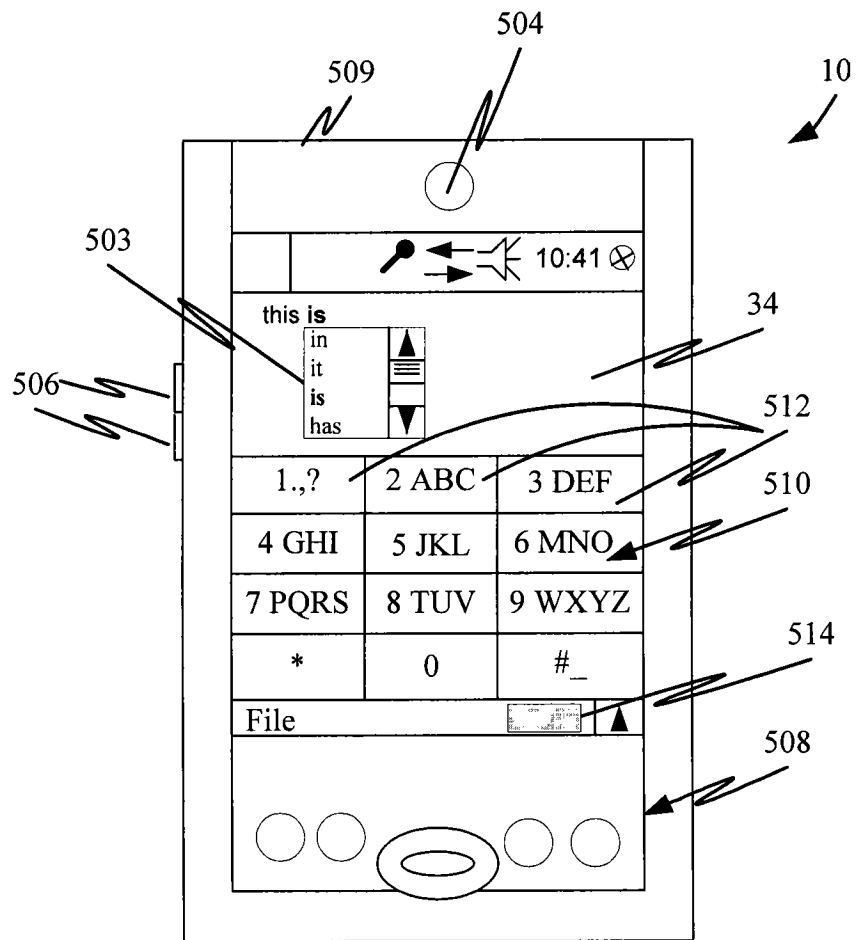


图 4

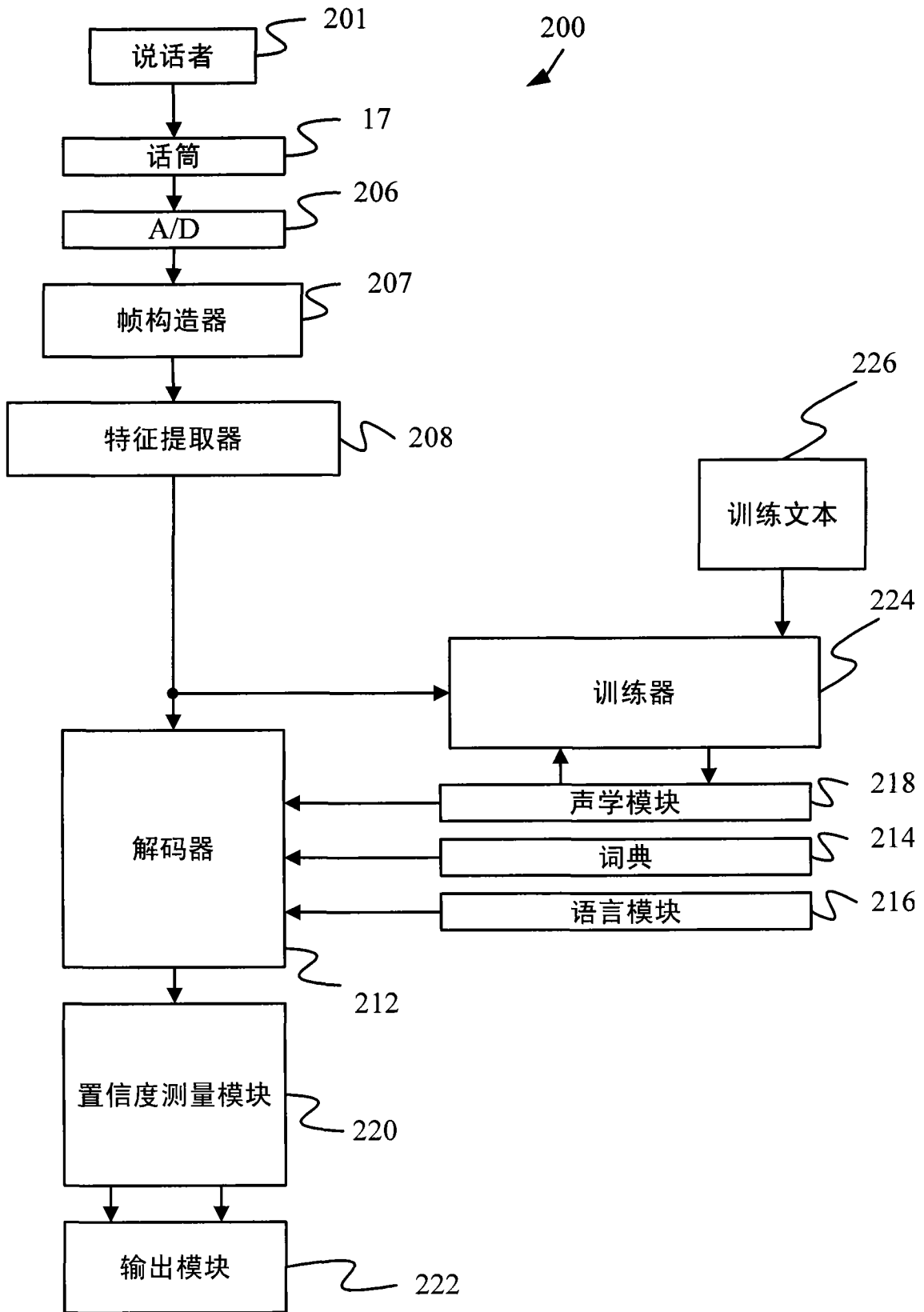


图 5

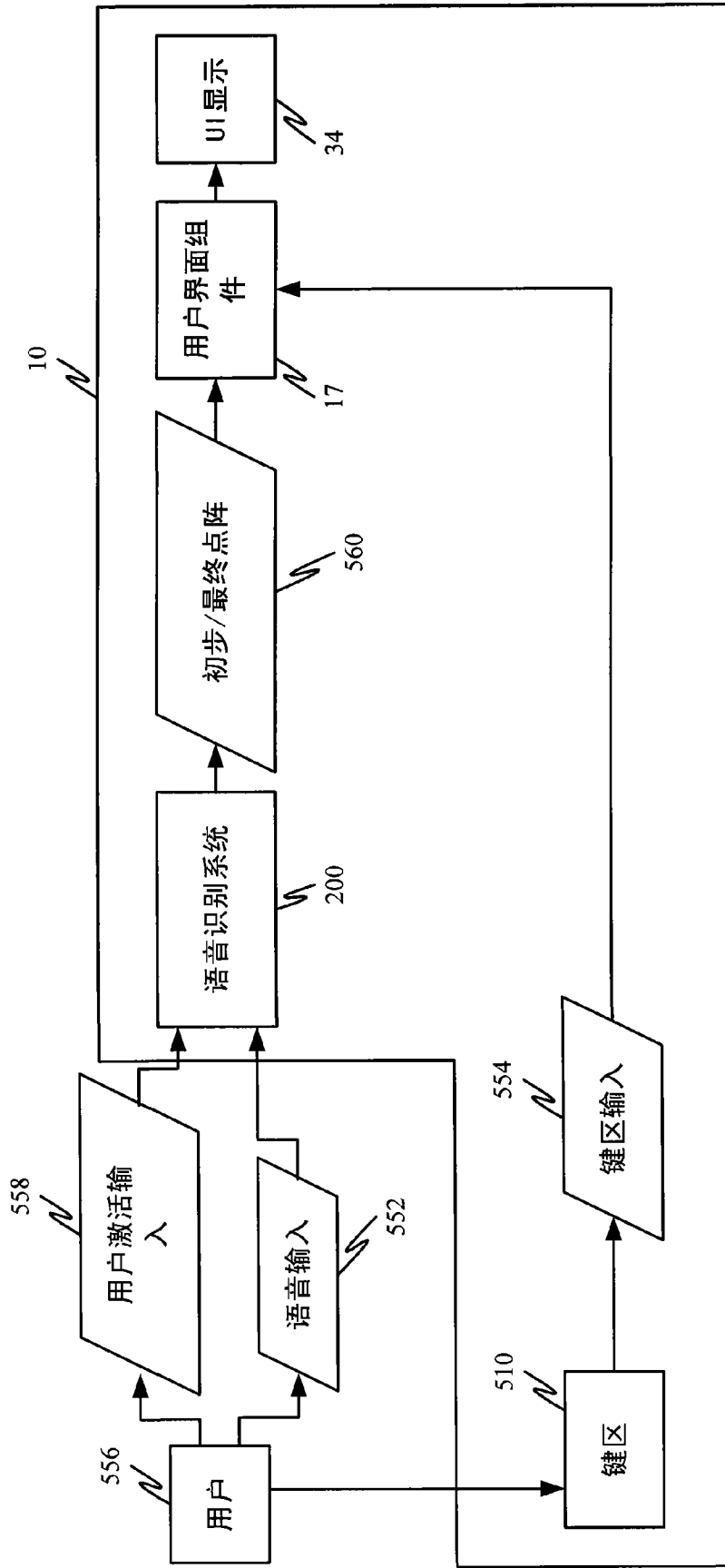


图 6

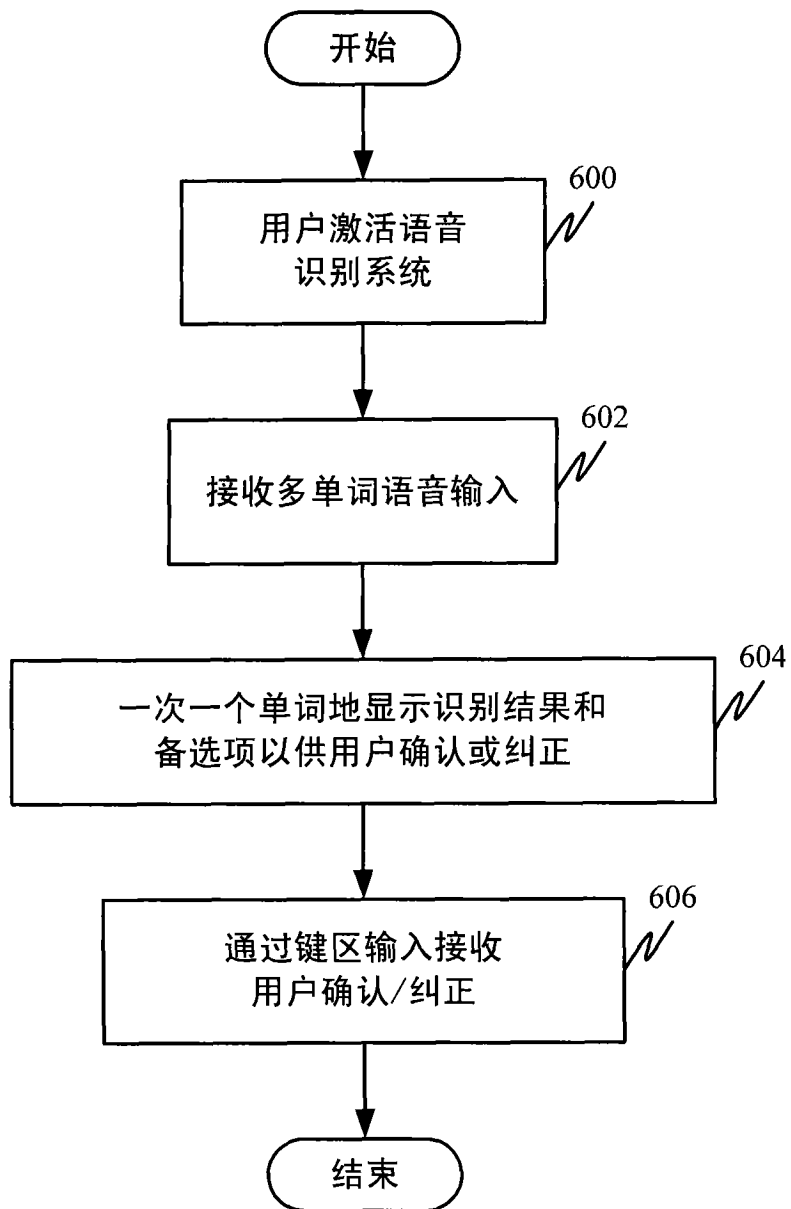


图 6A

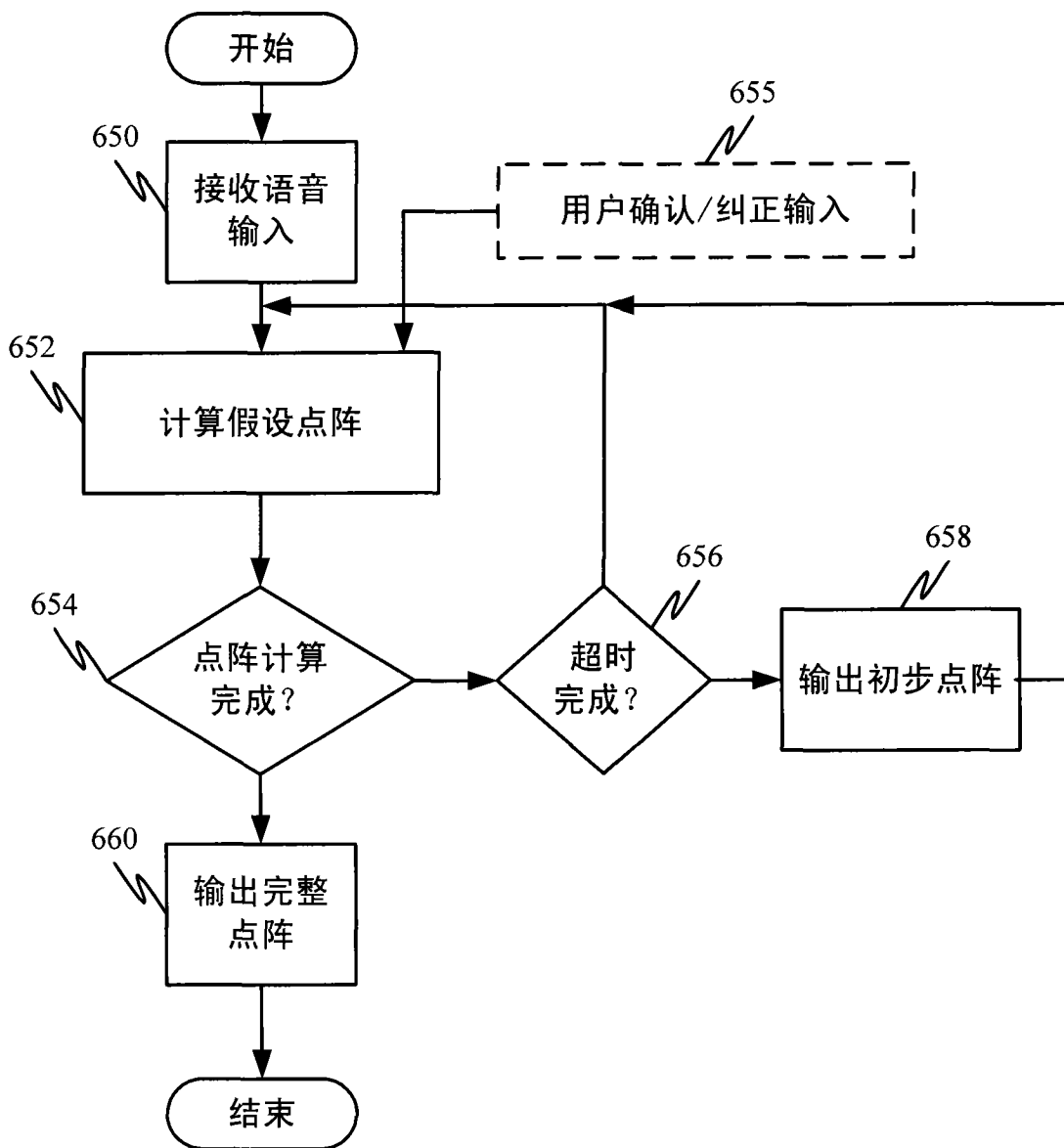


图 6B

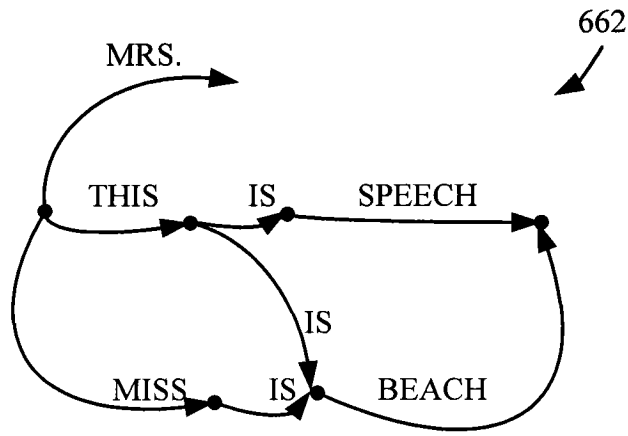


图 6C

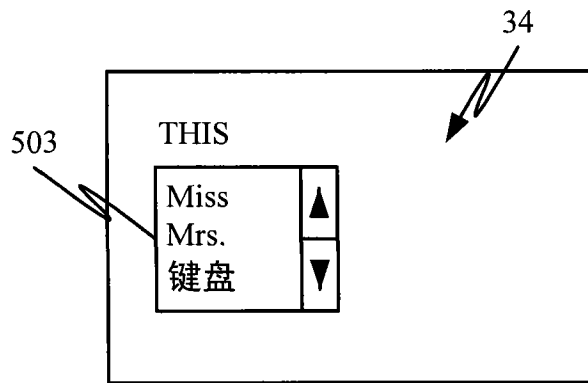


图 6D

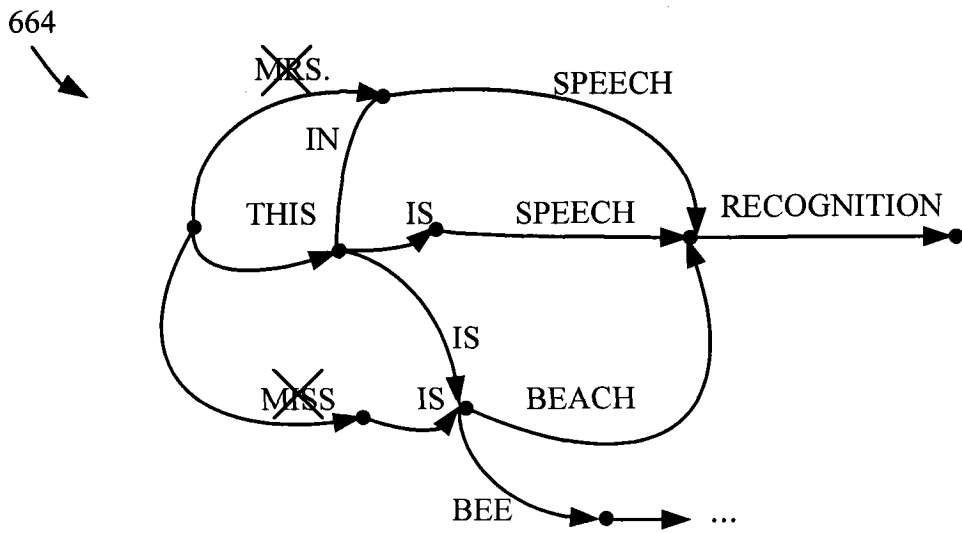


图 6E



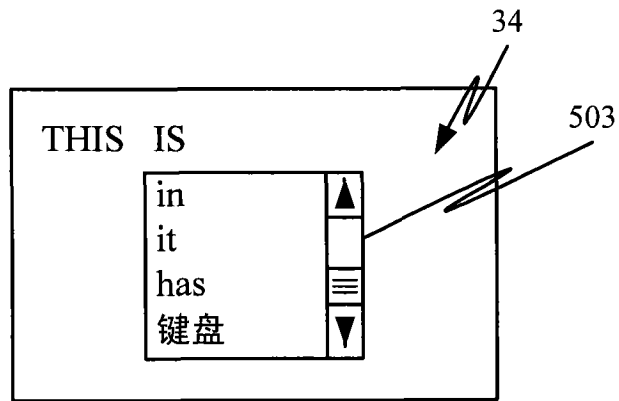


图 6F

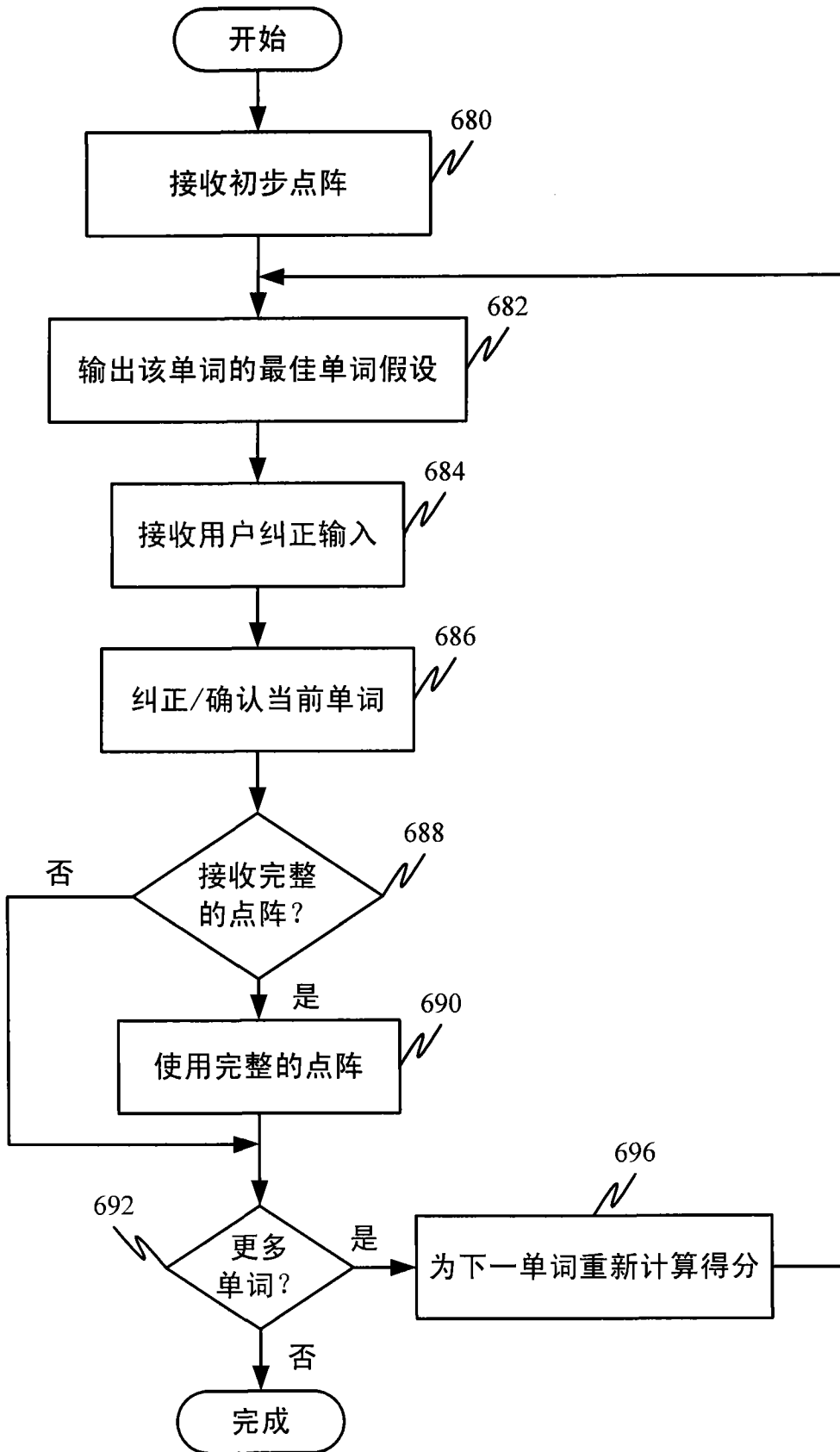


图 7

**demonstrates**

demonstration	▲
<b>demonstrates</b>	≡
demonstrate	□
demonstrated	▼

图 8