



# [12] 发明专利说明书

[21] ZL 专利号 01814689.9

[45] 授权公告日 2005 年 6 月 8 日

[11] 授权公告号 CN 1205602C

[22] 申请日 2001.9.13 [21] 申请号 01814689.9  
 [30] 优先权  
 [32] 2000. 9. 20 [33] US [31] 09/665,939  
 [86] 国际申请 PCT/GB2001/004092 2001. 9. 13  
 [87] 国际公布 WO2002/025637 英 2002. 3. 28  
 [85] 进入国家阶段日期 2003. 2. 25  
 [71] 专利权人 国际商业机器公司  
 地址 美国纽约  
 [72] 发明人 詹姆斯·刘易斯 科里·奥特伽  
 审查员 吴 芸

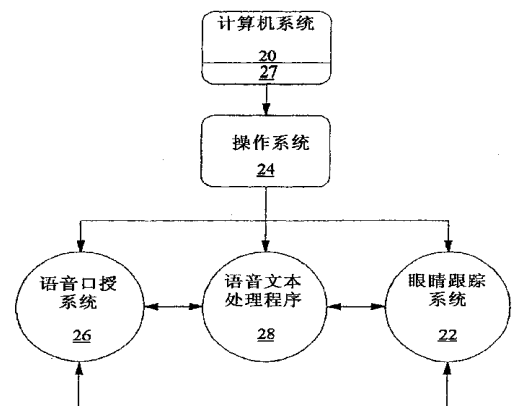
[74] 专利代理机构 中国国际贸易促进委员会专利  
 商标事务所  
 代理人 李德山

权利要求书 3 页 说明书 14 页 附图 6 页

[54] 发明名称 用于上下文语音识别的眼睛凝视

[57] 摘要

一种区分语音命令情形和语音口述情形的方法，包括：识别用户界面中的焦点；确定围绕焦点的环绕区；识别环绕区中的用户界面对象；在识别的用户界面对象中，进一步识别配置成接受语音口授文本的那些用户界面对象和未被配置成接受语音口授文本的那些用户界面对象；根据已被进一步识别为配置成接受语音口授文本的那些用户界面对象和已被进一步识别为未配置成接受语音口授文本的那些用户界面对象，计算概率；接收语音输入；根据计算的所述概率，使语音输入是语音命令还是语音口述的判断偏向一方。另外，该方法还可包括：识别位于用户界面之外的焦点；根据默认概率，使语音输入是语音命令还是语音口述的判断偏向一方。



ISSN 1008-4274

- 1、一种在电子文档中搜索匹配文本的方法，包括：
  - 识别用户界面中的焦点；
  - 确定一个围绕所述焦点的环绕区，所述环绕区包括配置成接收语音口授文本的用户界面对象内的文本主体；
  - 接收选择电子文档内的指定文本的语音命令；和
  - 在包含在所述环绕区中的所述文本主体中搜索所述指定文本的匹配文本，所述搜索局限于所述环绕区中的所述文本主体。
- 2、按照权利要求 1 所述的方法，还包括：
  - 如果在所述搜索步骤中，在所述文本主体中没有找到所述指定文本的匹配，则扩展所述环绕区，使之包括所述用户界面的一个附加区域，所述附加区域包括附加文本；和
  - 在所述附加文本中搜索所述指定文本的匹配，所述搜索局限于所述文本主体和所述附加文本。
- 3、按照权利要求 2 所述的方法，其中所述扩展步骤包括：
  - 从所述焦点把所述环绕区向外扩大预定的增量。
- 4、按照权利要求 2 所述的方法，其中所述扩展步骤包括：
  - 把所述环绕区扩大固定数量的与所述文本主体相邻的文本。
- 5、按照权利要求 2 所述的方法，其中所述扩展步骤包括：
  - 从所述焦点把所述环绕区向外扩大一个可变的增量。
- 6、一种区分语音命令情形和语音口述情形的方法，包括：
  - 识别用户界面中的焦点；
  - 确定一个围绕所述焦点的环绕区；
  - 识别所述环绕区中的用户界面对象；
  - 在所述识别的用户界面对象中，进一步识别配置成接受语音口授文本的那些用户界面对象和未被配置成接受语音口授文本的那些用户界面对象；
  - 根据已被进一步识别为配置成接受语音口授文本的那些用户界面

对象和已被进一步识别为未配置成接受语音口授文本的那些用户界面对象，计算概率；

接收语音输入；和

根据计算的所述概率，使所述语音输入是语音命令还是语音口述的判断偏向一方。

7、按照权利要求6所述的方法，还包括：

识别位于所述用户界面之外的焦点；和

根据默认概率，使所述语音输入是语音命令还是语音口述的决定偏向一方。

8、一种在电子文档中搜索匹配文本的设备，包括：

识别装置，用于识别用户界面中的焦点；

确定装置，用于确定一个围绕所述焦点的环绕区，所述环绕区包括配置成接收语音口授文本的用户界面对象内的文本主体；

接收装置，用于接收在电子文档内选择指定文本的语音命令；和

搜索装置，用于在包含在所述环绕区中的所述文本主体中搜索所述指定文本的匹配，所述搜索局限于所述环绕区中的所述文本主体。

9、按照权利要求8所述的设备，还包括：

扩展装置，如果所述搜索装置在所述文本主体中没有找到所述指定文本的匹配，则扩展装置扩展所述环绕区，使之包括所述用户界面的一个附加区域，所述附加区域包括附加文本，

其中，搜索装置在所述附加文本中搜索所述指定文本的匹配，所述搜索局限于所述文本主体和所述附加文本。

10、按照权利要求9所述的设备，其中所述扩展装置包括：

用于从所述焦点把所述环绕区向外扩大预定的增量的装置。

11、按照权利要求9所述的设备，其中所述扩展装置包括：

用于把所述环绕区扩大固定数量的与所述文本主体相邻的文本的装置。

12、按照权利要求9所述的设备，其中所述扩展装置包括：

用于从所述焦点把所述环绕区向外扩大一个可变的增量的装置。

**13、一种区分语音命令情形和语音口述情形设备，包括：**

**识别用户界面中的焦点的装置；**

**确定围绕所述焦点的环绕区的装置；**

**识别所述环绕区中的用户界面对象的装置；**

**在所述识别的用户界面对象中，进一步识别配置成接受语音口授文本的那些用户界面对象和未被配置成接受语音口授文本的那些用户界面对象的装置；**

**根据已被进一步识别为配置成接受语音口授文本的那些用户界面对象和已被进一步识别为未配置成接受语音口授文本的那些用户界面对象，计算概率的装置；**

**接收语音输入的装置；和**

**根据计算的所述概率使所述语音输入是语音命令还是语音口述的判断偏向一方的装置。**

**14、按照权利要求 13 所述的设备，还包括：**

**识别位于所述用户界面之外的焦点的装置；和**

**根据默认概率使所述语音输入的装置是语音命令还是语音口述的判断偏向一方的装置。**

## 用于上下文语音识别的眼睛凝视

### 技术领域

本发明涉及语音识别，更具体地说，涉及增强用户口授、用户语音命令和文本之间的区别的方法。

### 背景技术

语音识别是计算机把麦克风接收的声信号转换成文本的过程。识别的文本随后可在各种计算机应用软件中用于诸如文档准备、数据输入和命令及控制之类的用途。语音口授系统还向用户提供一种不需用手的计算机系统操作方法。

就电子文档准备来说，目前可用的语音识别系统提供使用户能够选择电子文档中一部分文本的用户语音命令。这种用户语音命令一般采用诸如“SELECT<文本>”之类的语法，这里用户语音命令“SELECT”表示在该命令之后的文本应被选择或突出显示。在选择一部分文本之后，用户可对选择的文本进行一系列后续操作中的任意操作。

从而，如果用户说“SELECT how are you”，则语音口授系统将在电子文档的文本主体内搜索文本短语“how are you”。一旦在文本主体中确定该短语的位置，则可选择或突出显示该短语。随后，用户可对选择的文本进行操作，例如删除操作，加粗/倾斜/下划线操作或者校正操作。就进一步举例说明来说，一旦突出显示文本“how are you”，则可用来源于后续用户话语的不同文本替换用户选择的文本部分。按照这种方式，用户可以不用手校正电子文档。

目前，“SELECT”命令，或者用于选择文本的其它类似用户语音命令的已知实现方式存在几个缺点。一个缺点是文本主体内，用户想选择的短语或单词可能多次出现。例如，在文本主体内，可能多

次出现单词 “ the”。从而，如果用户说 “ SELECT the”，则语音口授系统不能确定用户想要选择单词 “ the”的哪一次出现。

就解决该问题而论，常规的语音口授系统依赖于确定用户想要选择哪一次出现的用户所需单词或短语的一套规则。例如，语音口授系统可从活动窗口的顶部开始，选择该单词或短语的第一次出现。但是，如果用户不想选择该单词或短语的第一次出现，则常规的语音口授系统可向用户提供选择该单词的另一次出现的能力。特别地，某些常规语音口授系统提供诸如 “ NEXT” 或者 “ PREVIOUS” 之类的导航语音命令。

通过发出语音命令 “ NEXT”，用户指示语音口授系统定位并选择所需单词或短语的下一次出现。类似地，命令 “ PREVIOUS” 指示语音口授系统定位并选择所需单词或短语的上一次出现。虽然这样的常规系统使用户能够到达特定单词或短语的所需出现位置，但是用户必须研究到达所需出现位置的策略。这会导致时间的浪费和使用户受挫，尤其是在用户发觉语音口授系统不准确或者效率不高的情况下。

常规的语音口授系统内，常规的文本选择方法的另一缺点是当搜索用户指定的单词或短语时，这种语音口授系统一般搜索出现在用户屏幕上的整个文本主体。在语音口授系统语法中，出现在用户屏幕上的每个单词被激活，在语音口授系统看来是可能性相同的候选者。由于用户只需要单个单词或短语，激活并搜索出现在用户屏幕上的整个文本主体效率不高。此外，该技术会增大发生误识别的概率。

常规语音口授系统内，常规的文本选择方法的又一缺点是，对于语音口授系统来说，用户是在语音口授过程中发音一个单词还是在发出一个语音命令，例如激活下拉菜单的语音命令，常常不太明显。例如，如果用户发出单词 “ File”，根据情况，用户可能打算激活菜单条中的 File 菜单或者在电子文档中插入单词 “ File”。因此，对于常规的语音口授系统来说，用户话语是语音命令还是语音口述并不总是很明显。

因此，尽管目前可用的语音口授系统提供与计算机交互作用，从而以语音方式命令应用程序、提供电子文档中的语音口述以及选择电子文档内的文本的方法，但是仍然需要一种区别用户语音命令、用户口述、文本和它们的组合的改进方法。

### 发明内容

于是，这里公开的本发明提供一种通过结合语音口授系统利用眼睛跟踪系统，区分电子文档中文本的不同出现位置，以及区分语音命令情形和语音口述情形的方法和设备。本发明的方法和设备可包括和语音口授系统协同使用，以便在语音口授系统中确定用户凝视的焦点的眼睛跟踪系统（ETS）。特别地，ETS与语音口授系统的协同应用可提高“SELECT”用户语音命令功能，或者在语音口授系统中用于选择文本主体内一部分文本的其它任意用户语音命令的准确性。本发明中ETS的应用还可通过简化用户口述和语音命令的区分来提高系统的性能。

相应地，在第一方面，本发明提供一种在电子文档中搜索匹配文本的方法，所述方法可包括识别用户界面中的焦点，并且确定焦点周围的环绕区。特别地，环绕区可包括配置成接收语音口授文本的用户界面对象内的文本主体。另外，该方法还可包括接收用于选择电子文档内的指定文本的语音命令，搜索包含在环绕区中的文本主体以匹配所述指定文本。重要的是，搜索可局限于该环绕区中的文本主体。

如果在搜索步骤中，在文本主体中没有找到指定文本的匹配文本，则搜索电子文档中的匹配文本的方法最好还可包括扩展环绕区，使之包括用户界面中的附加区域。特别地，借助所述扩展包括进来的附加区域可包括附加文本。因此，可搜索所述附加文本，寻找所述指定文本的匹配文本。最后，如前所述，搜索可局限于所述文本主体和所述附加文本。

扩展步骤最好可包括把环绕区从焦点向外扩大预定的增量。或者，扩展步骤可包括把环绕区扩大固定数量的与文本主体相邻的文本。最后，扩展步骤可包括把环绕区从焦点向外扩大一个可变的增量。

相应地，在第二方面，本发明提供一种区分语音命令情形和语音口述情形的方法，可包括识别用户界面中的焦点；确定焦点周围的环绕区；识别环绕区中的用户界面对象；在所识别的用户界面对象中，进一步识别配置成接受语音口授文本的那些用户界面对象和未被配置成接受语音口授文本的那些用户界面对象；根据已被进一步识别为配置成接受语音口授文本的那些用户界面对象，和已被进一步识别为未配置成接受语音口授文本的那些用户界面对象，计算概率；接收语音输入；和根据计算的概率，使语音输入是语音命令还是语音口述的决定倾向一方。另外，该方法可包括在用户界面之外确定一个焦点；并且根据默认概率，使语音输入是语音命令还是语音口述的决定倾向一方。

相应地，在第三方面，本发明提供一种在电子文档中搜索匹配文本的设备，包括：识别装置，用于识别用户界面中的焦点；确定装置，用于确定一个围绕所述焦点的环绕区，所述环绕区包括配置成接收语音口授文本的用户界面对象内的文本主体；接收装置，用于接收在电子文档内选择指定文本的语音命令；和搜索装置，用于在包含在所述环绕区中的所述文本主体中搜索所述指定文本的匹配，所述搜索局限于所述环绕区中的所述文本主体。

相应地，在第四方面，本发明提供一种区分语音命令情形和语音口述情形设备，包括：识别用户界面中的焦点的装置；确定围绕所述焦点的环绕区的装置；识别所述环绕区中的用户界面对象的装置；在所识别的用户界面对象中，进一步识别配置成接受语音口授文本的那些用户界面对象和未被配置成接受语音口授文本的那些用户界面对象的装置；根据已被进一步识别为配置成接受语音口授文本的那些用户界面对象和已被进一步识别为未配置成接受语音口授文本的那些用户界面对象，计算概率的装置；接收语音输入的装置；和根据计算的所述概率使所述语音输入是语音命令还是语音口述的判断偏向一方的装置。

相应地，在第五方面，本发明提供一种计算机程序产品，所述计

计算机程序产品包括保存在计算机可读存储介质上的计算机程序代码，当在计算机上被执行时，所述计算机程序代码执行如上所述的方法。

### 附图说明

下面将参考附图中图解说明的本发明的优选实施例，举例说明本发明，其中：

图 1 是与这里公开的本发明交互作用的用户例证描述。

图 2 是图解说明适合于本发明之用的计算机系统的方框图。

图 3 是表示图 1 的计算机系统的典型高级架构的方框图。

图 4 是表示包括语音识别引擎的典型组件的方框图。

图 5A 和 5B 一起构成图解说明通过结合语音口授系统，利用眼睛跟踪系统，区别电子文档中文本的不同出现和区别语音命令情形和语音口述情形的方法的流程图。

### 具体实施方式

结合语音口授系统使用眼睛跟踪系统 (ETS) 可改进语音口授系统的性能。具体地说，根据本发明的方案，ETS 可帮助语音口授系统区别文本主体内文本的多次出现。另外，ETS 可帮助语音口授系统分析语音输入，从而区别语音命令和语音口述。可通过在 ETS 中检测用户凝视的焦点的屏幕位置实现这种改进。有利的是，屏幕位置（在或不在屏幕上）可被传送给语音口授系统。根据用户凝视焦点的位置，可确定焦点周围的区域（称为“环绕区”），该区域可帮助区分语音输入是语音命令还是语音口述。另外，环绕区可用于识别用户指定选择的文本的特定出现。

图 1 举例描述了与这里公开的发明交互作用的用户。在图 1 中，用户凝视视频显示终端 (VDT) 32 上的某一位置。用位于 VDT 32 的屏幕上的星号表示用户凝视的焦点。图中另外还描述了具有头戴式硬

件接口 29 的 ETS。在眼睛跟踪和测量领域中，ETS 众所周知。ETS 目前可从市场上购得，比如由 Virginia 的 Fairfax 的 LC Technologies 公司生产的 THE EYEGAZE DEVELOPMENT SYSTEM，以及均由 Massachusetts，Boston 的 SensoMotoric Instruments 公司（“SensoMotoric Instruments”是 SensoMotoric Instruments 公司的商标）生产的 EYEMOUSE（“Eyemouse”是 SensoMotoric Instruments 公司的商标）和 EYELINK（“Eyelink”是 SR Research Ltd 公司的商标）。

ETS 的配置可包括眼睛跟踪硬件接口 29 和图像处理系统 34。眼睛跟踪硬件接口 29 可以是可从 LC Technologies 公司获得的桌面安装装置。图 2 中表示了例证的桌面安装的眼睛跟踪装置。或者，眼睛跟踪硬件接口 29 可以是可从 SensoMotoric Instruments 公司获得的如图 1 中所示的头戴式装置。无论是桌面安装装置还是头戴式装置，眼睛跟踪硬件接口 29 均可把关于用户眼睛的信息传送给图像处理系统 34。

图像处理系统可以是独立的图像处理系统，或者可存在于常规的计算机之内。在图像处理系统存在于常规的计算机之内的情况下，常规计算机可利用图像处理电路和图像处理软件的组合完成图像处理系统的功能。本领域的技术人员应认识到本发明并不受选用的 ETS 的限制。相反，可采用能够把用户凝视焦点的位置传送给计算机的任何适当的 ETS。

图 2 图解说明了当图像处理系统 34 是基于常规计算机的图像处理系统时的情况。特别地，图像处理系统 34 可包括常规计算机 20，常规计算机 20 包括中央处理器（CPU），一个或多个存储器和相关电路。常规计算机 20 可包括计算机存储器 27，存储器 27 最好由电子随机存取存储器 27A 和诸如磁盘驱动器之类的批量数据存储介质 27B 组成。最后，计算机 20 可包括指示器 21，例如鼠标，以及与之工作连接的至少一个用户界面显示器 32，例如视频数据终端（VDT）。

特别地，计算机 20 可被配置成执行语音识别及文本-语音（TTS）转换。因此，计算机 20 还可包括音频输入装置 30，例如麦克风。另

外，计算机 20 可包括音频输出装置 23，例如扬声器。音频输入装置 30 和音频输出装置 23 都可通过适当的接口电路或“声卡（Sound board）”（未示出）与计算机 20 工作连接。这样，通过音频输入装置 30，用户语音可被接收到计算机 20 中，通过音频输出装置 23，可向用户提供合成语音及其它音频。利用可从市场上购得的多种高速多媒体个人计算机中的任意一种，可满足关于如上所述的常规计算机 20 的各种硬件要求。

根据本发明，计算机 20 还可包括通过计算机 20 的通信端口（未示出）与计算机 20 工作连接，并且通过适当的图像处理电路和软件与计算机 20 通信链接的眼睛跟踪硬件接口 29（这里图示的是桌面型）。具体地说，图像处理电路和软件可确定用户凝视焦点的位置，并且可把该信息传送给与图像处理软件通信连接的计算机应用程序。本发明中，语音口授系统可与图像处理软件通信连接，语音口授系统可从所述图像处理软件接收指示用户凝视焦点的位置的数据。

图 3 图解说明了包含 ETS 的有语音能力的计算机系统的典型架构，其中计算机系统被配置成区分电子文档中文本的不同出现，以及区分语音命令情形和语音口述情形。如图 3 所示，计算机系统 20 在存储器 27 中可包括操作系统 24、语音口授系统 26 和眼睛跟踪系统 22。在所举例子中，还提供了语音文本处理应用程序 28。但是，本发明并不局限于这方面，语音口授系统 26 可和有语音能力的任何应用程序一起使用。

图 3 中，语音口授系统 26、语音文本处理程序 28 和眼睛跟踪系统 22 被表示成单独的应用程序。但是，应注意本发明并不局限于此，这些不同的应用程序可被实现成单一的更复杂的应用程序。例如，语音口授系统 26 可与语音文本处理应用程序 28 结合，或者和与语音口授系统一起使用的其它任意应用程序结合。另外，眼睛跟踪系统 22 可作为包含在计算机 20 中的应用程序，或者可存在于能够通过数据链路与计算机 20 通信的独立 ETS 之内。系统还可包括语音导航应用程序（未示出），以便针对其它应用程序的语音操作协调语音口授系统的

操作，不过对这里所述的本发明的操作来说不是必需的。

图 4 是表示举例说明语音口授系统 26 中语音信号的语音-文本转换的典型组件的方框图。一般来说，通过如图 2 中所示的音频输入装置可接收模拟语音信号，并在音频电路中将其处理成数字语音信号。具体地说，通过以固定的频率，一般为每 10-20 毫秒对语音信号采样，可把语音信号转换成数字化数据集。随后，音频电路可把数字化的语音信号传送给语音口授系统 26。

“表达”方框 35 可接收数字化的语音信号，并且能够产生数字化语音信号的表达，所述表达可在语音识别过程的后续阶段中用于确定一部分语音信号对应于特定语音事件的概率。该过程用来强调从操作系统接收的语音信号的对知觉重要的、和扬声器无关的特征。

在“建立模型/分类”方框 36 中，算法可进一步处理语音信号，使和扬声器无关的声学模型适应当前扬声器的声学模型。最后，在搜索方框 38 中，使用搜索算法把搜索引擎引导到对应于语音信号的最可能单词。在声学模型 40、词汇模型 42、语言模型 44 和训练数据 46 的帮助下，进行搜索方框 38 中的搜索过程。

最好，本发明提供一种区分电子文档中文本的不同出现，以及区分语音命令情形和语音口述情形的方法和设备。本发明的方法和设备可包括 ETS 和语音口授系统的协同应用。特别地，这种组合可提高语音口授系统中“SELECT”用户语音命令功能，或者在文本主体内选择一部分文本的其它任意用户语音命令的精度。通过帮助语音口授系统把语音输入解释为语音口述或语音命令，这种组合还可提高语音口授系统性能。

通过根据检测到的用户凝视焦点，计算时间上最接近用户凝视的语音输入是语音口述或语音命令之一的概率，实现对语音口授系统的上述提高。计算的概率可用于使语音口授系统倾向于把语音输入解释为语音口述或语音命令之一。具体地说，语音口授系统可确定环绕所检测到的焦点的一个可调整屏幕区（“环绕区”），其结果是语音口授系统可连续捕捉并更新和位于环绕区内的文本和对象有关的信

息。

当接收语音输入时，语音口授系统可确定环绕区是否主要包含用户界面对象或者文本输入区。如果环绕区主要包含文本输入区，则语音口授系统可推断语音输入应被理解为插入文本输入区的语音口授。相反，如果环绕区主要包括用户界面对象，则语音口授系统可把语音输入理解为语音命令。最后，在语音输入被理解为在文本输入区中选择文本主体中的文本的语音命令的情况下，语音口授系统可根据环绕区中的文本，而不是文本输入区中的整个文本确定要选择的文本。按照这种方式，可更高效地把语音口授系统资源专用于较小的文本区域，而不是电子文档中的整个文本主体。

图 5A 和 5B 一起构成图解说明区分电子文档中文本的不同出现，以及区分语音命令情形和语音口述情形的方法的流程图。可结合为语音口授系统和 ETS 的使用而配置的计算机系统执行所述方法。图 5A 开始于步骤 50，在步骤 50 中，在向语音口授系统提供语音输入的同时，用户自然地凝视 VDT 32 上的不同位置（在屏幕上）或者凝视在 VDT 32 之外（脱离屏幕）。

在步骤 55 中，ETS 确定用户凝视的焦点的位置。借助图像处理电路和软件的帮助，ETS 确定用户凝视的焦点是在屏幕上的位置还是在屏幕之外。无论如何，ETS 把该信息传送给语音口授系统。在步骤 60 中，语音口授系统从 ETS 接收用户的焦点位置。如果用户凝视焦点的位置在屏幕上，则系统进行到步骤 70。否则，系统进行到步骤 65。

如果在步骤 60 中确定焦点的位置在屏幕上，则 ETS 已确定用户凝视焦点的屏幕位置。从而，在步骤 70 中，可在焦点周围确定一个环绕区。在一个代表性的实施例中，可根据从焦点向外延伸的规定半径，用周界确定该环绕区。或者，可通过把预定的几何区域覆盖在焦点上来确定环绕区。

同样，本发明并不局限于计算环绕区的上述方法。相反，计算环绕区的任意适当方法都可满足本发明的目的。此外，本领域的技术人员要认识到，和如何确定环绕区或者环绕区的最终形状无关，外周界

内的区域的默认面积或大小可以是用户可调整的数值。例如，用户可指定默认面积，或者另一方面，用户可规定环绕区从焦点向外延伸的半径。

在步骤 75 中，在确定环绕区之后，可捕捉和环绕区内的文本和对象相关的信息，供确定语音输入应被理解为语音口述还是理解为语音命令之用，以及供识别电子文档中指定文本的特定出现之用。特别地，捕捉的信息可包括，例如专用于显示不适于接收语音口述文本的用户界面对象的象素的数目，以及专用于显示适于接收语音口述文本的用户界面对象的象素的数目。应认识到，通过确定语音口授系统可将其资源投入其中的有限区域，语音口授系统可实现更高的效率。例如，语音口授系统只需激活包含在环绕区内找到的文本的部分语音口授文法，而不是整个语音口授语法。

在步骤 80 中，可根据哪个语音口述可被理解为语音命令或者语音口述，计算概率。具体地说，可通过计算环绕区的可口授面积和环绕区的总面积的比值，计算概率。例如，如果 70% 的环绕区可接收用户口授，则概率为 70% 或者 0.70。同样，本发明并不局限于计算概率的上述特定方式。事实上，概率的其它计算可以基于例如环绕区内的文本或者口授单词的数目和环绕区内适合于用户语音命令的对象的数目的比值。但是，无论如何计算概率，都应认识到概率最好既不为零，又不为 1（指示后续用户话语绝对是用户口述或者用户语音命令）。禁止这种极端概率值使得当用户凝视屏幕之外时，希望向语音口授系统口授语音成为可能。

如果在判定步骤 60 中，确定用户凝视的焦点在屏幕之外，在步骤 65 中，系统可向概率赋予默认值。该默认值被称为默认概率，并可由用户预先配置。默认概率指示当用户凝视屏幕之外时，后续语音输入是语音口述或语音命令之一的统计可能性。因此，基于默认概率的统计分析可指示当用户注视屏幕之外时，用户打算使语音输入被理解为语音口授的可能性。

默认概率可具有从零（0.00）到 1（1.00）变化的可调数值。特别

地，本领域的技术人员应认识到向默认概率赋予较高的数值表示在语音口授过程中，用户不需注视屏幕的假定。但是，默认值最好不指示当用户注视屏幕之外时提供的语音输入应被绝对理解为语音口述或者语音命令。这样的确定概率会在语音口授系统内导致错误。

在步骤 85 中，在计算概率或者依赖默认概率之后，可接收语音输入。根据在 ETS 的帮助下得到的概率，可分析语音输入，从而确定语音输入是应被理解为语音口述还是应被理解为语音命令。随后，该方法可继续处理该语音输入，通过跳转圆圈 A 到达图 5B 的判定步骤 95。

在判定步骤 95 中，可确定在步骤 85 中接收的语音输入是“SELECT”语音命令还是在电子文档内选择文本的其它类似语音命令。如果语音输入未被理解为 SELECT 命令，则该方法进行到步骤 97，在步骤 97 中可发生两种操作之一。首先，如果语音输入，虽然不是 SELECT 语音命令，但被确定为另一语音命令，则可如同在常规的有语音功能的应用程序中那样执行该语音命令。其次，如果语音输入被确定为语音口述，则语音识别引擎可把该语音输入转换成文本。随后，可把转换文本插入配置成接收转换文本的用户界面对象中。在任一种情况下，该方法都可通过跳转圆圈 C 返回图 5A 的步骤 50，重复上述过程。

返回判定步骤 95。如果确定在步骤 85 接收的语音输入是 SELECT 语音命令或者在电子文档内选择文本的其它类似语音命令，则在步骤 100 中，可确定 SELECT 命令指定的文本是否位于包含在环绕区中的文本主体中。例如，如果语音输入已被理解为 SELECT 命令“SELECT mouse”，则可确定包含在环绕区中的文本主体是否包括单词“mouse”。如果对于指定文本，在步骤 100 中找到了匹配，则方法可前进到步骤 105。否则，方法可在步骤 110 中继续。

如果根据步骤 100 找到了指定文本的匹配，则在步骤 105 中，可选择指定文本的最恰当的匹配。更具体地说，如果在环绕区中的文本主体内只存在一个匹配，则可选择该文本的所述单一匹配文本，一般通过突出显示该文本的匹配出现来实现。相反，如果在环绕区中的文

本主体内多次出现匹配文本，则可确定环绕区中文本主体内指定文本的哪一次出现最接近焦点。从而用户凝视的焦点可用于确定应选择哪一个匹配文本。同样，本发明并不局限于此，也可采用在匹配文本的多次出现中选择一个匹配文本的其它适当方法。这样的备选方法包括在环绕区中的文本主体内选择首先出现的匹配文本。

一旦确定了指定文本的适当出现，则可选定被确定的文本，一般通过在视觉上突出显示该文本来实现。应认识到在选择了指定文本的不正确或不合乎需要的出现的情况下，可使用诸如“PREVIOUS”或者“NEXT”之类常规语音命令定位到环绕区中指定文本的其它出现位置。在任何情况下，该方法可通过跳转圆圈 C 返回图 5A 的步骤 50，重新开始该过程。从而，通过重复该过程，该方法可再次计算环绕区，确定随后接收的语音输入是语音口述还是语音命令的概率。

现在返回判定步骤 110，如果在环绕区中的文本主体内没有找到任何匹配，则可确定环绕区是否包括被配置成接收语音口授的整个可视用户界面。如果是，则可确定在屏幕上的文本主体中不存在任何匹配，可如步骤 115 中那样通知用户。在未在图 5B 中描述的另一实施例中，在屏幕上不存在任何匹配的情况下，系统可向用户提供继续以及进一步扩展针对用户指定文本的搜索的附加选择。例如，可询问用户是否希望搜索当前打开的电子文档的其它部分。或者，可向用户提供更多的有针对性的选择，例如在环绕区之前或者之后使环绕区扩展预定或者用户可调数目的单词或段落。在任何情况下，随后方法可通过跳转圆圈 C 返回图 5A 的步骤 50，重新开始该过程。

相反，如果在步骤 100 中确定环绕区不包含被配置成接收语音口授的整个可视用户界面，则不能假定在屏幕上的文本主体中不存在任何匹配。从而，继续进行步骤 120，可扩展环绕区所覆盖的面积以便包括更多的文本。执行环绕区的扩展的任意适当方法都可满足需要。例如，可从用户焦点沿所有方向把环绕区的外周界向外延伸相同的预定的或者动态计算的数值。或者，可从焦点把环绕区向外延伸一个代表一面积值的预定数值。

在本发明的一个实施例中，默认预定值可用于确定扩展的程度。默认值可以是可调的，以便提供微调能力。按照这种方式，用户可规定在重复搜索过程中，环绕区应增长多大。以前一例子为例，如果在环绕区中的文本主体内没有找到用户指定的文本“mouse”，则可从焦点沿所有方向把环绕区的周界向外扩大1厘米。或者，可把环绕区扩大5平方厘米或者特定数目的象素的预定面积。

在扩大环绕区之后，在步骤125中，可计算、收集并保存和新扩大的环绕区内的对象和文本有关的信息，以供本发明的方法未来之用。另外，在语音口授系统语法内，可激活新扩大的环绕区内的新的文本主体。另外，可确定存在于新扩大的环绕区内的对象的属性。在识别新扩大的环绕区内的文本和对象之后，通过跳转圆圈B从步骤100开始，可重复文本主体中匹配文本的搜索。按照这种方式，该方法可系统地并且递增地把对文本主体内用户指定文本的搜索扩大到并超出文本主体的屏幕部分之外。

特别地，可用硬件、软件或者硬件和软件的组合实现本发明。可在一个计算机系统中以集中方式实现本发明，或者按照其中不同的部件分散在几个互连计算机系统中的分布方式实现本发明。适合于执行这里描述的方法的任意类型的计算机系统或其它设备都是适宜的。硬件和软件的典型组合可以是一个通用计算机系统加上计算机程序，当程序被加载并被执行时，其控制计算机系统，使其执行这里描述的方法。

本发明还可嵌入计算机程序产品中，所述计算机程序产品包括使这里描述的方法的实现成为可能的所有特征，并且当被载入计算机系统时，能够执行这些方法。本文中，计算机程序方法或计算机程序意指一组指令的采用任何语言、代码或符号的任何表达，所述一组指令用来使具有信息处理能力的系统直接或者在下述任一或两种操作之后执行特定功能：a) 转换为另一语言、代码或符号；b) 以不同的材料形式复制。

虽然上面的说明书举例说明了本发明的优选实施例，但是要明白

---

**本发明并不局限于这里公开的具体结构。在不脱离本质属性的情况下，可以其它具体形式体现本发明。因此，本发明的范围只应由下述权利要求限定，而不是由前述说明书限定。**

图1

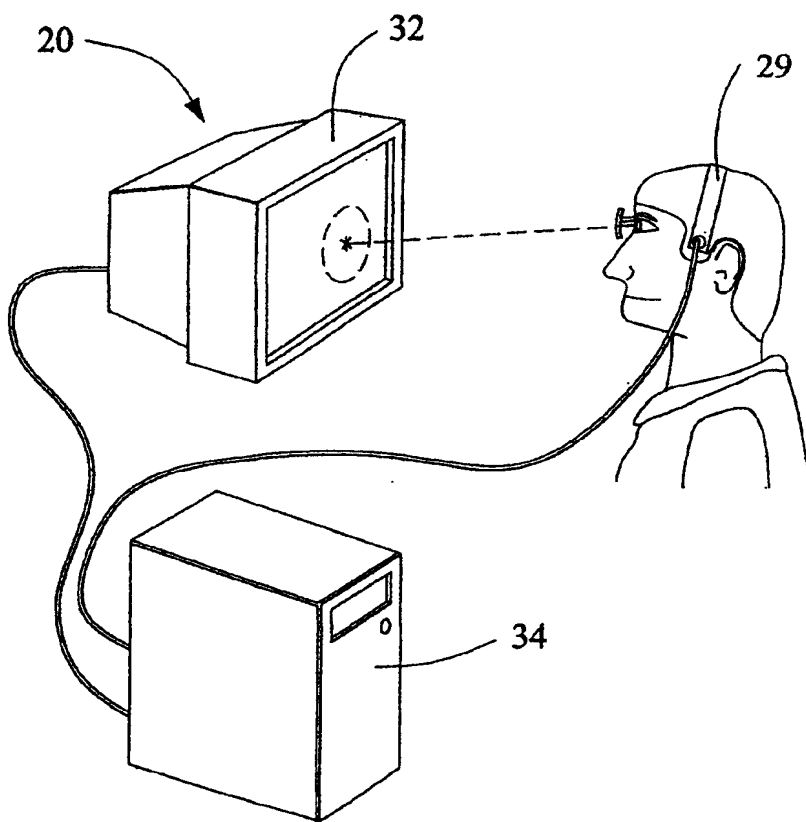


图2

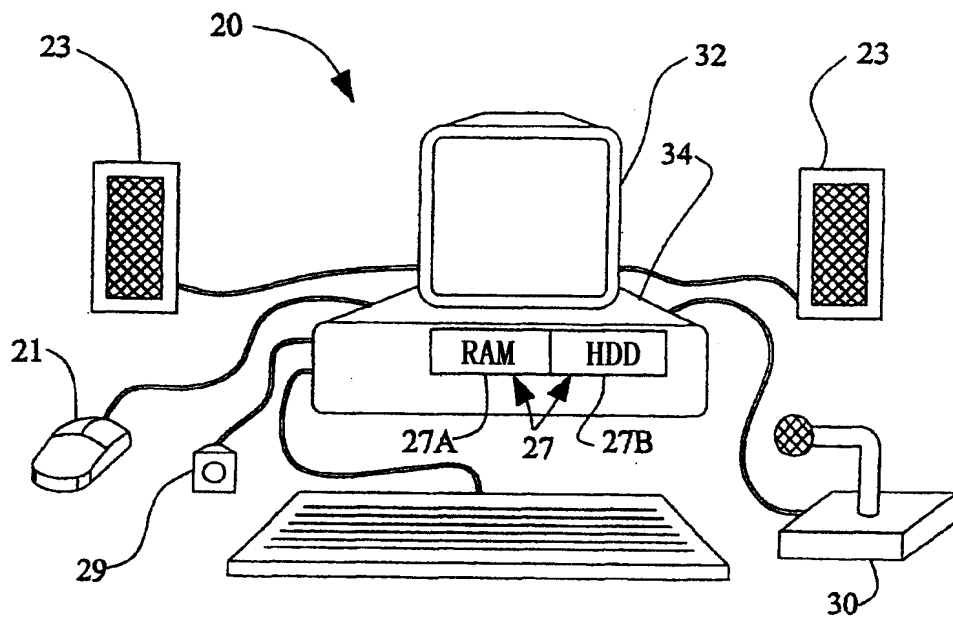
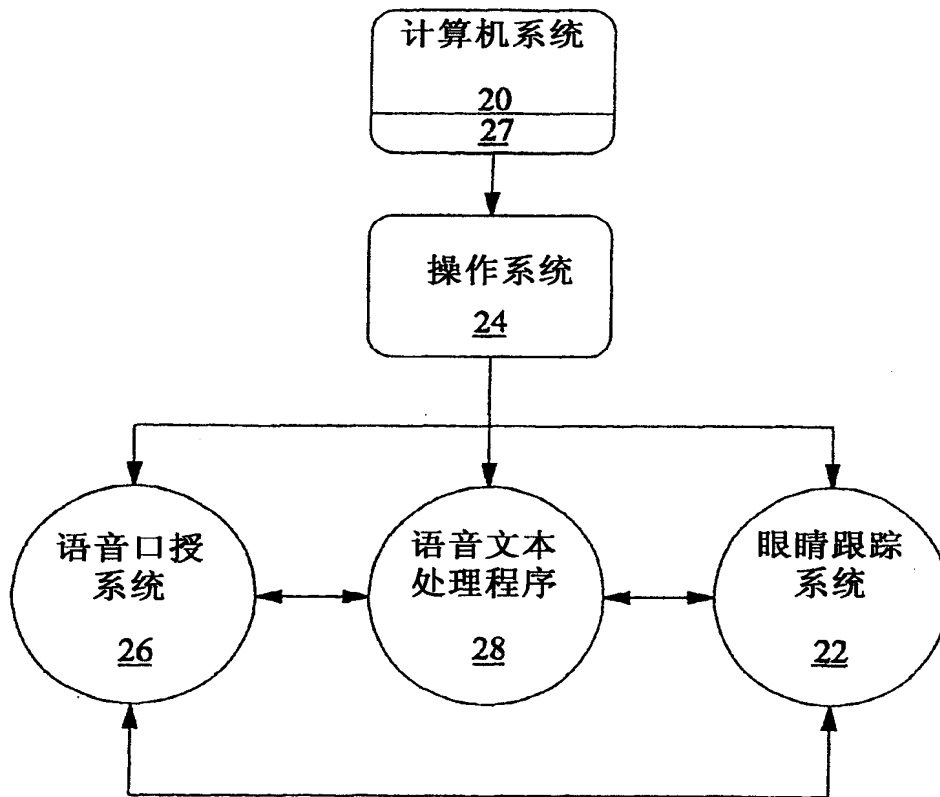


图3



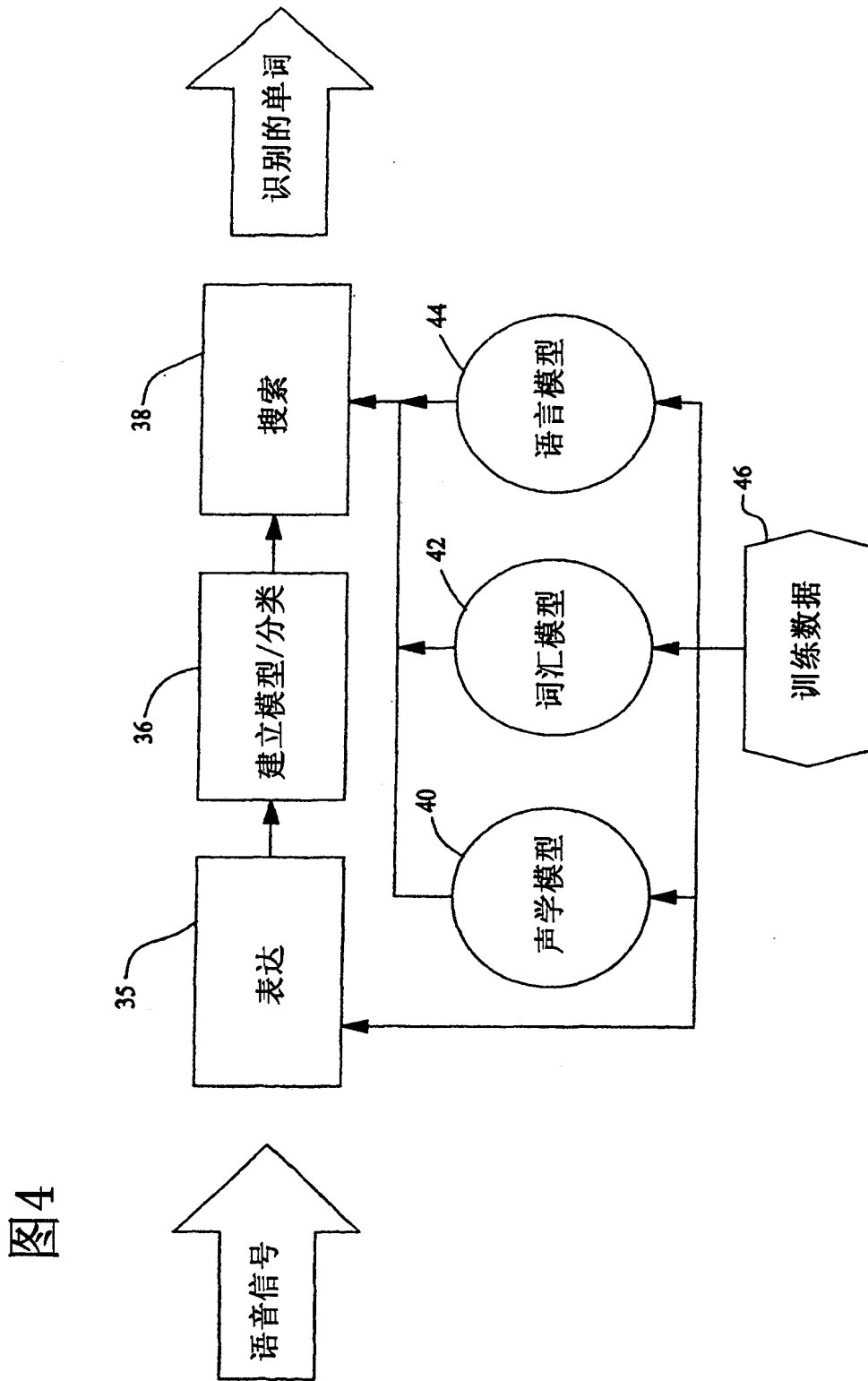


图4

图5A

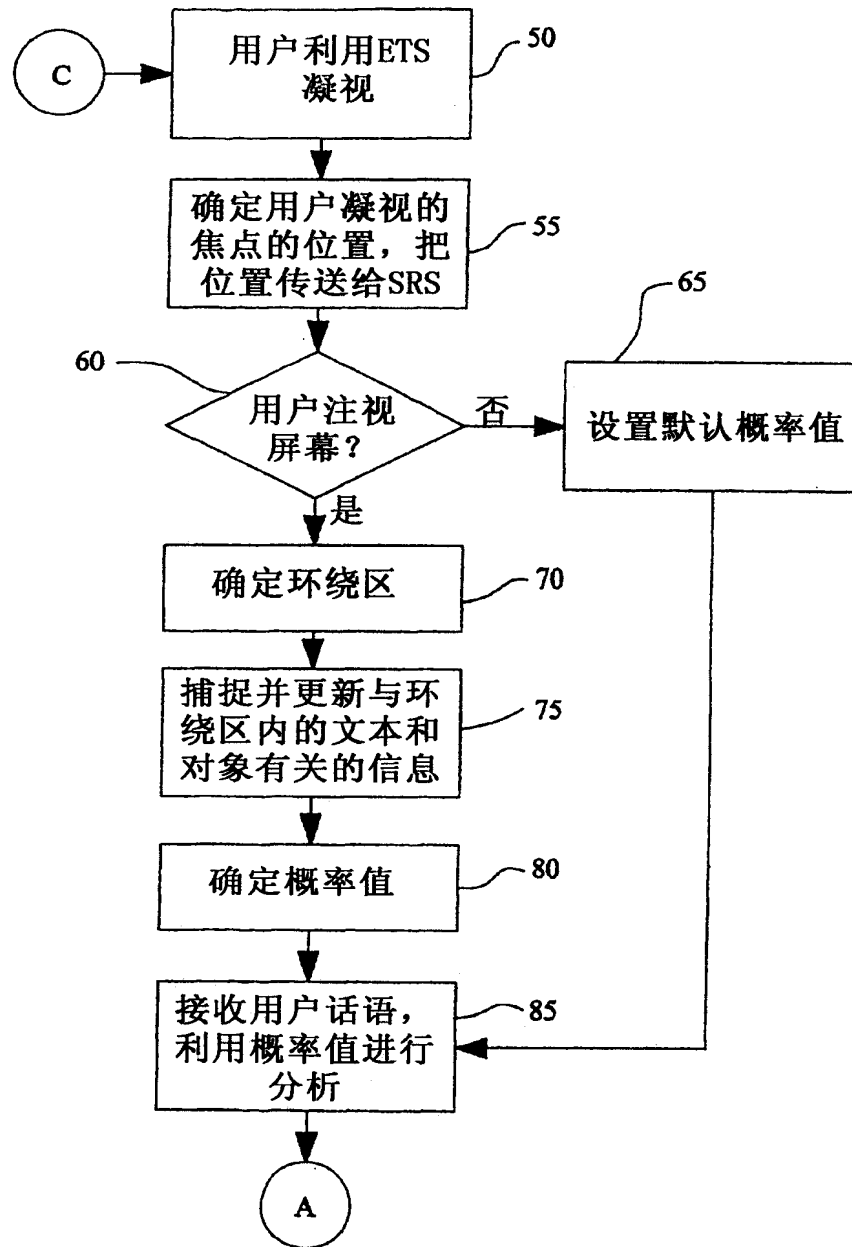


图5B

