

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
G06F 3/01 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200610110949.2

[43] 公开日 2007年2月14日

[11] 公开号 CN 1912803A

[22] 申请日 2006.8.11

[21] 申请号 200610110949.2

[30] 优先权

[32] 2005.8.12 [33] JP [31] 2005-234032

[71] 申请人 佳能株式会社

地址 日本东京

[72] 发明人 广田诚 深田俊明 小森康弘

[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 康健峰

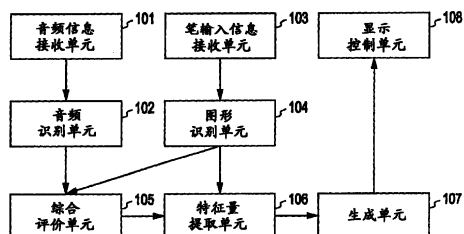
权利要求书5页 说明书18页 附图16页

## [54] 发明名称

信息处理方法和信息处理设备

## [57] 摘要

在根据本发明的用于通过组合使用语音输入来识别手写图形或文字的信息处理方法中，为了提高识别率，使给定目标经受图形识别，并且获得第一候选图形列表。在语音上识别输入语音信息，并且获得第二候选图形列表。基于通过图形识别获得的候选图形以及通过语音识别获得的候选图形，选择最可能的图形。



1. 一种信息处理设备, 包括:

图形识别装置, 被适配成通过对给定目标执行的图形识别来获得候选图形;

语音识别装置, 被适配成通过对输入语音信息执行的语音识别来获得候选图形; 以及

选择装置, 被适配成基于由图形识别装置获得的候选图形以及由语音识别装置获得的候选图形, 选择图形。

2. 根据权利要求1所述的信息处理设备, 还包括:

提取装置, 被适配成基于由选择装置选择的图形而从给定目标提取特征量;

生成装置, 被适配成基于由提取装置提取的特征量而生成由选择装置选择的图形; 以及

显示控制装置, 被适配成执行控制, 以显示由生成装置生成的图形。

3. 根据权利要求1所述的信息处理设备, 其中:

图形识别装置获得置信得分, 其表示候选图形和各个候选者是目标图形的确信度;

语音识别装置获得置信得分, 其表示候选图形和各个候选者是目标图形的确信度; 以及

选择装置相加由图形识别装置获得的候选图形和由语音识别装置获得的候选图形的对应置信度, 并且基于结果置信得分而选择图形。

4. 根据权利要求3所述的信息处理设备, 还包括:

检测装置, 被适配成检测语音信息中的噪声度;

其中选择装置根据噪声度而对由语音识别装置获得的候选图形的置信得分设置权重, 并且相应地增加置信得分。

5. 根据权利要求3所述的信息处理设备, 还包括:

接收装置, 被适配成接收来自定点设备的输入; 以及

检测装置，被适配成检测来自定点设备的输入的速度；其中  
给定目标包括由接收部件接收的来自定点设备的输入；并且  
选择装置根据由检测装置检测的来自定点设备的输入的速度，对由  
图形识别装置获得的候选图形的置信得分设置权重，并且相应地增加置  
信得分。

6. 一种信息处理设备，包括：

文字识别装置，被适配成通过对给定目标执行的文字识别来获得候  
选文字；

语音识别装置，被适配成通过对输入语音信息执行的语音识别来获  
得候选文字读音；以及

选择装置，被适配成基于由文字识别装置获得的候选文字以及由语  
音识别装置获得的候选文字，选择文字。

7. 根据权利要求6所述的信息处理设备，其中：

文字识别装置获得置信得分、以及每个候选者的至少一个读音信  
息，其中置信得分表示候选图形和各个候选者是目标图形的确信度；

语音识别装置获得置信得分，其表示候选图形和各个候选者是目标  
图形的确信度；以及

选择装置包括：第一计算装置，被适配成针对包括在文字识别结果  
中的每个候选文字，计算对候选文字给出的读音和文字识别结果中的相  
应读音之间的相似度；以及第二计算装置，被适配成基于候选文字的原  
始置信得分、相似度、以及由语音识别装置获得的读音的置信得分而计  
算校正识别置信得分，并且确定具有最大校正识别置信得分的文字作为  
选择结果。

8. 根据权利要求6所述的信息处理设备，还包括：

短语词典，包括短语的标记和读音的组合；

显示控制装置，被适配成控制由选择装置选择的文字，以将其显示  
在对应于给定目标的位置；

确认装置，被适配成确认由显示控制装置显示的文字；

接收装置，被适配成在确认了一个或多个文字之后，接收给定目标

和语音信息;

搜索装置, 被适配成在短语词典中搜索具有与标记部分匹配的已确认一个或多个文字的短语;

后续文字获得装置, 被适配成在由搜索装置搜索的一个或多个短语的标记中, 获得已确认一个或多个文字的后续文字;

第二文字识别装置, 被适配成相对于由后续文字获得装置获得的后续文字, 使由接收装置接收的连续坐标信息经受文字识别, 以便帮助将后续文字包括在识别结果中;

第二语音识别装置, 被适配成在语音上识别由接收装置接收的语音信息, 以便帮助将后续文字包括在识别结果中;

第二选择装置, 被适配成基于由第二文字识别装置获得的候选文字和由第二语音识别装置获得的候选文字读音而选择文字; 以及

第二显示控制装置, 被适配成控制由第二选择装置选择的文字, 以将其显示在对应于坐标信息的位置。

9. 一种信息处理设备, 包括:

识别装置, 被适配成使给定目标经受模式识别, 以获得候选模式;

语音识别装置, 被适配成在语音上识别输入语音信息, 以获得用于指定模式的信息; 以及

选择装置, 被适配成基于由识别装置获得的候选模式以及由语音识别装置获得的用于指定模式的信息, 选择模式。

10. 一种信息处理设备, 包括:

识别装置, 被适配成使给定目标经受模式识别, 以获得候选模式;

语音识别装置, 被适配成对由识别装置获得的候选模式设置权重, 并且在语音上识别输入语音信息, 以获得候选模式; 以及

选择装置, 被适配成在由语音识别装置获得的候选模式当中选择模式。

11. 一种信息处理设备, 包括:

语音识别装置, 被适配成在语音上识别输入语音信息, 以获得候选模式;

识别装置，被适配成对由语音识别装置获得的候选模式设置权重，并且使给定目标经受模式识别，以获得候选模式；以及

选择装置，被适配成在由识别装置获得的候选模式当中选择模式。

12. 一种信息处理方法，包括：

图形识别结果获得步骤，获得作为对给定目标执行图形识别的结果的候选图形；

语音识别结果获得步骤，获得作为在语音上识别语音信息的结果的候选图形；以及

选择步骤，基于在图形识别结果获得步骤中获得的候选图形以及在语音识别结果获得步骤中获得的候选图形，选择图形。

13. 一种信息处理方法，包括：

文字识别结果获得步骤，获得作为对给定目标执行文字识别的结果的候选文字；

语音识别结果获得步骤，获得作为在语音上识别语音信息的结果的候选文字读音；以及

选择步骤，基于在文字识别结果获得步骤中获得的候选文字以及在语音识别结果获得步骤中获得的候选文字读音，选择文字。

14. 一种信息处理方法，包括：

识别步骤，使给定目标经受模式识别，以获得候选模式；

语音识别步骤，在语音上识别输入语音信息，以获得用于指定模式的信息；以及

选择步骤，基于在识别步骤中获得的候选模式以及在语音识别步骤中获得的用于指定模式的信息，选择模式。

15. 一种信息处理方法，包括：

识别步骤，使给定目标经受模式识别，以获得候选模式；

语音识别步骤，对在识别步骤中获得的候选模式设置权重，并且在语音上识别输入语音信息，以获得候选模式；以及

选择步骤，从在语音识别步骤中获得的候选模式当中选择模式。

16. 一种信息处理方法，包括：

---

语音识别步骤，在语音上识别输入语音信息，以获得候选模式；  
识别步骤，对在语音识别步骤中获得的候选模式设置权重，并且使  
给定目标经受模式识别，以获得候选模式；以及  
选择步骤，从在识别步骤中获得的候选模式当中选择模式。

17. 一种记录介质，用于存储使计算机执行根据权利要求 12 至 16  
任一所述的信息处理方法的控制程序。

## 信息处理方法和信息处理设备

### 技术领域

本发明涉及一种对诸如文字或图形的输入信息执行处理的信息处理方法。具体地说，本发明涉及一种用于提高输入文字、图形等的识别率的技术。

### 背景技术

可以将手写文字或图形输入到具有能够接收通过笔或输入笔的直接输入的屏幕的设备。这样的设备的示例包括触摸屏、电子白板、具有触摸屏的 PDA、以及平板 PC。通过该设备，可以如同在纸张或白板上绘画手写文字或图形一样实现自然的输入。如果与手写文字识别或图形识别相关的技术与该设备集成在一起，则可以将通过手写而生成的笔划信息转换成文本等形式的数字数据，由此有可能将该数据用于各种后续处理。

同时，随着语音识别技术的发展和硬件性能的提高，变得有可能不仅向 PC 或工作站而且向各种计算机控制设备输入语音。

在这样的情形下，代替单独采用笔输入或者单独采用语音输入，提出了用于通过组合使用笔和语音经由多个模式输入来提高操作效率的技术。在该技术领域内，术语多模式有时用来表示使用两个或更多个不同模式的输入。日本专利公开 No. 06-085983 提出了这样一种系统，其中通过以手写方式用笔绘画图形并且同时念出图形类型，对手写图形进行整形并且显示。

手写输入和语音输入都具有实现直观输入的优点，但是具有“误识”的缺点。例如，将考虑这样的情况，其中用户通过手写来绘画图形，并且对该图形应用图形识别处理，以便进行自动整形。即使当用户意欲绘画“四边形”时，该图形可能由于误识而被识别为“圆形”，并且被整形为

圆形以便显示。与此相对，根据日本专利公开 No. 06-085983，当用户绘画图形同时念出图形类型时，在语音上识别该发音，以识别图形类型，从而根据所识别的图形类型对输入图形进行整形。然而，语音识别也遭受误识。即使当用户念出“四边形”时，如果该发音被识别为“三角形”，则将图形整形为三角形，并且显示，从而执行用户不意欲的错误整形。也就是，存在由于误识而不能实现系统目的的问题。

在本例中，为了描述起见，采取了使用图形识别和语音识别的特定示例，但是诸如文字识别、面部识别、以及姿势识别的模式识别也具有相同的问题。

### 发明内容

本发明是鉴于上述问题而提出的，以便提高诸如输入文字或图形的信息的识别准确度。

为了解决上述问题，根据本发明的一方面，提供了一种信息处理设备，包括：图形识别装置，被适配成通过对给定目标执行的图形识别来获得候选图形；语音识别装置，被适配成通过对输入语音信息执行的语音识别来获得候选图形；以及选择装置，被适配成基于由图形识别装置获得的候选图形以及由语音识别装置获得的候选图形，选择图形。

另外，为了解决上述问题，根据本发明的另一方面，提供了一种信息处理设备，包括：文字识别装置，被适配成通过对给定目标执行的文字识别来获得候选文字；语音识别装置，被适配成通过对输入语音信息执行的语音识别来获得候选文字读音；以及选择装置，被适配成基于由文字识别装置获得的候选文字以及由语音识别装置获得的候选文字，选择文字。

另外，为了解决上述问题，根据本发明的另一方面，提供了一种信息处理方法，包括：图形识别结果获得步骤，获得作为对给定目标执行图形识别的结果的候选图形；语音识别结果获得步骤，获得作为在语音上识别语音信息的结果的候选图形；以及选择步骤，基于在图形识别结果获得步骤中获得的候选图形以及在语音识别结果获得步骤中获得的候



选图形，选择图形。

另外，为了解决上述问题，根据本发明的另一方面，提供了一种信息处理方法，包括：文字识别结果获得步骤，获得作为对给定目标执行文字识别的结果的候选文字；语音识别结果获得步骤，获得作为在语音上识别语音信息的结果的候选文字读音；以及选择步骤，基于在文字识别结果获得步骤中获得的候选文字以及在语音识别结果获得步骤中获得的候选文字读音，选择文字。

根据下面参照附图对示例性实施例的描述，本发明的其它特征将会变得清楚。

#### 附图说明

图 1 是根据本发明第一实施例的信息处理设备中的输入处理的框图。

图 2 是根据第一实施例的信息处理设备的硬件配置图。

图 3 是示出根据第一实施例的图形识别单元的操作的流程图。

图 4 是示出根据第一实施例的语音识别单元的操作的流程图。

图 5 是示出根据第一实施例的综合评价单元的操作的流程图。

图 6 示出了根据第一实施例的、用户在触摸屏上通过手写来绘画图形同时进行语音输入的情况。

图 7 示出了根据第一实施例的、来自图形识别单元的识别结果的示例。

图 8 示出了根据第一实施例的、来自语音识别单元的识别结果的示例。

图 9 示出了根据第一实施例的、由综合评价单元做出的综合评价的示例。

图 10 是示出根据第一实施例的、应当包括与语音输入相对应的笔输入的时间间隔的图。

图 11 是示出根据第一实施例的综合评价单元和生成单元的操作的流程图。

图 12 示出了根据第一实施例的、由综合评价单元得到的特征点的示例。

图 13 示出了如何显示根据第一实施例基于输入而生成的图形。

图 14 是根据第二实施例的信息处理设备中的输入处理的框图。

图 15 是示出根据第二实施例的文字识别单元的操作的流程图。

图 16 是根据第四实施例的图形识别处理的流程图。

图 17 是示出根据第二实施例的综合评价单元的操作的流程图。

图 18 示出了根据第二实施例的、用户在触摸屏上通过手写来绘画文字同时进行语音输入的情况。

图 19 是示出根据第二实施例的文字识别单元的操作的流程图。

图 20 是示出根据第二实施例的语音识别单元的操作的图。

图 21 示出了根据第二实施例的、由综合评价单元做出的综合评价的示例。

图 22 示出了以任意字体显示根据第二实施例基于输入而识别的文字的情况。

图 23 示出了根据第四实施例的、用户在触摸屏上通过手写来绘画文字同时进行语音输入的情况。

图 24 是根据第五实施例的多功能设备的硬件配置图。

图 25 是根据第五实施例的图形生成处理的流程图。

## 具体实施方式

以下参照附图仅仅作为示例来描述本发明的优选实施例。

### 第一实施例

在本实施例中，将描述这样的示例，其中用笔绘画图形同时念出图形类型，以对输入图形进行整形，以显示在具有大触摸屏的信息处理设备中，其中该信息处理设备可以在诸如会议的情形中使用。

图 1 是根据第一实施例的信息处理设备中的输入处理的框图。在该图中，标号 101 表示语音信息接收单元，其用于接收语音信息。标号 102 表示语音识别单元，其用于在语音上识别所接收的语音信息。标号

103 表示笔输入信息接收单元，其用于接收笔输入信息。标号 104 表示图形识别单元，其用于基于所接收的坐标信息而识别图形。标号 105 表示综合评价单元，其用于通过语音识别结果和图形识别结果的综合评价来选择图形。标号 106 表示特征量提取单元，其用于从由笔输入信息接收单元 103 接收的笔输入信息提取特征量。标号 107 表示生成单元，其用于基于由特征量提取单元 106 提取的特征量，生成由综合评价单元选择的图形。标号 108 表示显示控制单元，其用于执行控制，以便显示由生成单元 107 生成的图形等。

图 2 是根据本实施例的信息处理设备的硬件配置图。在该图中，标号 201 表示 CPU，其用于根据下面将要描述的、用于执行信息处理设备的操作过程的程序来控制输入处理。标号 202 表示 RAM，其用于提供上述程序的操作所需的存储区域。标号 203 表示 ROM，其用于保存程序等。标号 204 表示触摸屏，可以执行通过笔和手指的输入。笔输入信息接收单元 103 接收在触摸屏 204 上输入的笔输入信息。另外，显示控制单元 108 控制触摸屏 204 显示上述图形等。标号 205 表示硬盘驱动器(HDD)，其用于保存程序等。在本实施例中要执行的程序可以是保存在 ROM 203 中的程序、或者保存在 HDD 205 中的程序。另外，该程序可以被分割并且保存在 ROM 203 和 HDD 205 中，并且通过引用这些程序来执行。标号 206 表示 A/D 转换器，其用于将输入语音转换成数字信号。标号 207 表示麦克风，其用于接收语音。标号 208 表示总线。

在本实施例中，如图 6 所示，将考虑这样的情况，其中，用户使用笔 602 在触摸屏 601 上绘画手写笔划 603，同时念出“四边形”。下面将参照流程图来描述具有上述配置的信息处理设备的操作。

图 3 是示出笔输入信息接收单元 103 和图形识别单元 104 的操作的流程图。用于执行该流程图的程序保存在 ROM 203 或 HDD 205 中，并且在 CPU 201 的控制下执行。

首先，检查是否存在从用户到触摸屏 204 的笔输入(S301)。当存在笔输入时，由笔输入信息接收单元 103 接收笔输入信息，并且图形识别单元 104 基于所接收的笔输入信息而识别图形(S302)。将识别结果推入

到在图中未示出的图形识别结果栈中(S303)。图形识别结果栈存储在RAM 202或HDD 205中。图7示出了推入到图形识别结果栈中的图形识别结果的示例。以这种方式,图形识别结果包括笔输入开始和结束时间戳(701、702)、以及用置信得分表示图形类型的N个最佳候选者703(在本例中,N=3)。置信得分可以通过公知方法来得出。可以使用任何方法,只要该方法能够得出与置信得分相同的概念即可,例如不相似度。

图4是示出语音信息接收单元101和语音识别单元102的操作的流程图。用于执行该流程图的程序保存在ROM 203或HDD 205中,并且在CPU 201的控制下执行。

首先,语音信息接收单元检查是否存在来自用户的语音输入(S401)。语音输入的检测是公知技术,并且例如,基于输入语音功率的幅度等而确定检测。当存在语音输入时,接收并且在语音上识别语音输入(S402)。此后,将识别结果推入到在图中未示出的语音识别结果栈中(S403)。语音识别结果栈保存在RAM 202或HDD 205中。图8示出了推入到语音识别结果栈中的语音识别结果的示例。以这种方式,语音识别结果包括发音开始和结束时间戳(801、802)、以及用置信得分表示图形类型的N个最佳候选者803(在本例中,N=3)。语音识别置信得分可以通过公知方法来得出。

图5是示出综合评价单元105的操作的流程图。用于执行该流程图的程序保存在ROM 203或HDD 205中,并且在CPU 201的控制下执行。

综合评价单元105监视图形识别结果栈,以检查是否有任何图形识别结果保存在该栈中(S501)。当保存了图形识别结果时,检查与该图形识别结果相对应的语音识别结果是否保存在语音识别结果栈中(S502)。这里,可以一直监视或者定期检查图形识别结果栈。另外,可以监视语音识别结果栈,而非图形识别结果栈。在这种情况下,S501和S502的顺序颠倒。

首先,例如基于时间戳而确定语音识别结果是否对应于图形识别结

果。具体地说，如图 10 所示，如果图形识别结果的笔输入开始时间戳 701 或笔输入结束时间戳 702 包括在从语音开始时间之前的 T1 秒跨越到结束时间之后的 T2 秒的时间段 D 内，则相互关联语音识别结果和图形识别结果。在本例中，将在假定图 8 的语音识别结果对应于图 7 的图形识别结果时给出描述。

当在图形识别结果中的一个或多个图形类型和语音识别结果中的一个或多个图形类型当中找到对应于图形识别结果的语音识别结果时，对匹配图形类型的置信度进行求和。在图 7 和 8 中，对作为图形识别结果的椭圆形的置信得分 60 和作为语音识别结果的椭圆形的置信得分 10 进行求和，并且获得 70。三角形和四边形的置信度分别是 90 和 110。然后，选择具有求和置信得分的最大值的图形类型(S503)。在本例中，选择具有最大求和置信得分的四边形。应当注意，在 S502 中，如果不存在对应的语音识别结果，则在图形识别结果中的一个或多个图形类型当中选择具有最大置信得分的图形(S504)。如同以上述方式一样，完成用于在图形识别结果中的一个或多个图形类型当中选择一个图形的处理，并且全部清除图形识别结果栈和语音识别结果栈(S505)。作为上述处理的结果，如图 9 所示基于综合评价结果而将图形类型确定为“四边形”。

图 11 是示出特征量提取单元 106、生成单元 107 和显示控制单元 108 的操作的流程图。用于执行该流程图的程序保存在 ROM 203 或 HDD 205 中，并且在 CPU 201 的控制下执行。

根据由综合评价单元 105 确定的图形类型，在本例中提取输入笔划的特征点。在四边形的情况下，找出图 12 的四个顶点 1201 到 1204(S1101)。用于找出特征点的算法是公知技术。根据特征点，如图 13 的标号 1301 所示生成图形(S1102)，并且显示该图形(S1103)。在显示的时候，关闭原始笔划显示。整形后的图形存储在 RAM 202 和/或 HDD 205 中。

在上述示例中，根据从笔输入笔划的图形识别，识别结果的第一候选者是“椭圆形”，其不同于用户的意图。另一方面，语音识别结果的第一候选者是“三角形”，其也不同于用户的意图。然而，这两个识别结果

都经受综合评价，从而有可能作为“四边形”整形和显示该图形，而符合用户的意图。以这种方式，图形识别结果和语音识别结果经受综合评价，从而获得提高按照用户意图的识别准确度的效果。

应当注意，在本实施例中，已经描述了直到图形生成步骤的过程，但是仅仅直到图 5 的图形类型选择处理的过程就满足本发明的要求。

## 第二实施例

在第二实施例中，将描述这样的情况，其中在由信息处理设备提供的触摸屏上使用笔绘画文字，同时念出文字的读音，并且整形和显示输入文字。图 14 是根据本实施例的信息处理设备中的输入处理的框图。在该图中，标号 1401 表示语音信息接收单元。标号 1402 表示语音识别单元。标号 1403 表示笔输入信息接收单元。标号 1404 表示文字识别单元。标号 1405 表示综合评价单元。标号 1406 表示显示控制单元，其用于执行控制，以便显示由综合评价单元 1405 输出的文字等。综合评价单元 1407 是单汉字字典。标号 1408 表示语音识别语法。

根据第二实施例的信息处理设备的硬件配置与图 2 的相同。在本实施例中，将考虑这样的情况，其中用户使用笔 1802 在触摸屏 1801 上通过手写来绘画文字笔划 1803，同时念出“ai”，如图 18 所示。应当注意，在图 18 中，用户以书写汉字“蓝”的意图绘画文字笔划。汉字“蓝”具有“ai”的读音。

图 15 是示出笔输入信息接收单元 1403 和文字识别单元 1404 的操作的流程图。用于执行该流程图的程序保存在 ROM 203 或 HDD 205 中，并且在 CPU 201 的控制下执行。

首先，检查是否存在从用户到触摸屏 204 的笔输入(S1501)。当存在笔输入时，笔输入信息接收单元接收笔输入信息，然后文字识别单元 104 将笔输入信息识别为文字(S1502)。将识别结果推入到图中未示出的文字识别结果栈中(S1503)。文字识别结果栈保存在 RAM 202 或 HDD 205 中。图 19 示出了推入到文字识别结果栈中的文字识别结果的示例。以这种方式，文字识别结果包括笔输入开始和结束时间戳(1901、1902)、以及用置信得分表示文字类型的 N 个最佳候选者

1903(在本例中, N=3)。识别结果的每个文字被提供了读音(或者在存在多种读法时, 全部读音)。当文字是片假名语音符号、平假名语音符号或字母时, 从标记(notation)获得该读音信息, 而当文字为汉字时, 从单汉字字典 1407 获得该读音信息(在诸如“?”和“#”的符号被视作汉字之一时, 单汉字字典可以处理符号)。

语音信息接收单元 1401 和语音识别单元 1402 的操作与图 4 中的相同。首先, 语音信息接收单元检查是否存在来自用户的语音输入(S401)。语言输入的检测是公知技术, 并且例如, 基于输入语音的幅度等而确定检测。当存在语音输入时, 接收并且在语音上识别语音信息(S402)。语音识别单元 1402 基于语音识别语法 1408 而识别语音。语音识别语法 1408 由无重叠地收集日语音节系统中的 50 个文字和包括在单汉字字典 1407 中的所有单汉字的读音的列表构成。因此, 语音识别结果是日语音节系统中的 50 个文字和包括在单汉字字典 1407 中的单汉字的读音其中的一个。然后, 将识别结果推入到在图中未示出的语音识别结果栈中(S403)。语音识别结果栈保存在 RAM 202 或 HDD 205 中。图 20 示出了推入到语音识别结果栈中的语音识别结果的示例。以这种方式, 语音识别结果包括发音开始和结束时间戳(2001、2002)、以及用置信得分表示单汉字读音的 N 个最佳候选者 2003(在本例中, N=3)。

图 17 是示出综合评价单元 1405 的操作的流程图。用于执行该流程图的程序保存在 ROM 203 或 HDD 205 中, 并且在 CPU 201 的控制下执行。

综合评价单元 1405 监视文字识别结果栈, 以检查是否有任何文字识别结果保存在该栈中(S1701)。如果保存了文字识别结果, 则检查与该文字识别结果相对应的语音识别结果是否保存在语音识别结果栈中(S1702)。这里, 可以一直监视或者可以定期检查文字识别结果栈。另外, 可以监视语音识别结果栈, 而非文字识别结果栈。首先基于时间戳而确定是否保存了与文字识别结果栈相对应的语音识别结果。该确定方法与第一实施例 1 的情况相同。

当发现与文字识别结果栈相对应的语音识别结果时, 还通过使用语

音识别结果来校正向文字识别结果中的一个或多个文字类型的每一个给予的置信得分。以下面方式执行置信得分校正。作为文字识别结果的第一候选者的“监”具有两种读法“kan”和“ken”。对于各个读音，计算语音识别结果中的三个候选读音“ei”、“sai”和“ai”的相似度。例如，作为“监”的读音的“kan”和语音识别结果中的“ei”根本没有相互匹配的组成音素，从而相似度是 0。因此，如下表达用“ei”对“kan”的校正识别置信得分  $MC(\text{“kan”}|\text{“ei”})$ 。

$$\begin{aligned} MC(\text{“kan”}|\text{“ei”}) &= C1(\text{“kan”}) + C2(\text{“ei”})/\text{相似度} \\ &= 70 + 60/0 \\ &= 70 \end{aligned}$$

其中 C1: 文字识别置信得分, C2: 语音识别置信得分。

在“kan”和“sai”的情况下，在“kan”的三个音素当中，仅仅一个音素与“sai”中的音素“a”相匹配，从而相似度是  $1/3=0.33$ 。因此，获得下面表达式。

$$MC(\text{“kan”}|\text{“ai”}) = 70 + 55/3 = 88.3$$

以相同的方式，获得下面表达式。

$$MC(\text{“kan”}|\text{“sai”}) = 70 + 30/3 = 80$$

在置信度当中，采用最大值，因此“kan”的校正识别置信得分  $MC(\text{“kan”})$  是 88.3。对于“监”的另一读音“ken”，类似地获得校正识别置信得分  $MC(\text{“ken”})$ ，其为  $70+60/3=90$ 。文字“监”的校正识别置信得分  $MC(\text{“监”})$  采用这些置信度当中的最大值，并且获得下面表达式。

$$MC(\text{“监”})=90$$

以相同的方式，获得下面表达式。

$$MC(\text{“兰”})=50+55/3=68.3$$

$$MC(\text{“蓝”})=40+55/1.0=95$$

如同以上述方式一样获得校正识别置信得分，选择具有最大校正识别置信得分的文字“蓝”作为最终识别结果(S1703)。应当注意，在 S1702 中，如果不存在对应的语音识别结果，则在文字识别结果中的一个或多个文字当中选择具有最大置信得分的文字(S1704)。此后，确定所选文字



作为所采用的文字。

如同以上述方式一样，完成用于在文字识别结果中的一个或多个文字当中选择一个文字的处理，并且全部清除文字识别结果栈和语音识别结果栈(S1705)。作为上述处理的结果，如图 21 所示基于综合评价结果而将文字确定为“蓝”，并且如图 22 所示以任意字体显示文字“蓝”。所选文字存储在 RAM 或 HDD 中。

在上述示例中，根据基于笔输入笔划的文字识别，识别结果的第一候选者是“监”，其不同于用户的意图。另一方面，语音识别结果的第一候选者是“ei”，其也不同于用户的意图。然而，这两个识别结果都经受综合评价，由此有可能显示文字“蓝”而符合用户的意图。以这种方式，文字识别结果和语音识别结果经受综合评价，从而获得提高按照用户意图的识别准确度的效果。

### 第三实施例

在上述实施例中，当笔输入时间戳在输入语音的输入时间段 D 中时，相互关联语音识别结果和图形识别结果或文字识别结果，但是本发明不限于上述方案。对于不擅长在书写时发音的用户，还可以考虑这样的情况更方便，其中在完成书写之后进行发音。因此，在本实施例中，在完成笔输入之后念出的语音与笔输入相关联。具体地说，记录笔输入的时间戳，并且将在该时间戳所表示的时间之后输入的语音与笔输入相关联。这里，可以提供这样的限制，其中语音输入仅仅与完成笔输入之后的预定时间内的笔输入相关联。当然，与此相对，可以将在完成语音输入之后进行的笔输入与语音相关联。

通过这样的配置，提供了念出在笔输入之后输入的、与文字或图形相对应的语音的实施例，以及在发音之后使用笔输入对应的图形或文字的实施例。

### 第四实施例

在上述实施例中，虽未明确表明，已经描述了这样的示例，其中笔输入和语音输入每次都经受综合评价，但是本发明不限于上述方案。可选地，可以重复执行用于使笔输入经受图形识别或文字识别以显示识别

结果的处理，并且当用户确定了识别结果不正确时，进行语音输入，从而整形图形或文字。

图 16 是这种情况下的处理的流程图。用于执行该流程图的程序保存在 ROM 203 或 HDD 205 中，并且在 CPU 201 的控制下执行。在本例中描述识别图形的示例。

首先，检查是否存在来自用户的笔输入(S1601)。当存在笔输入时，从笔划识别图形(S1602)。然后，将识别结果推入到图形识别结果栈中(S1603)。此后，显示作为识别结果的、具有最大置信得分的图形(S1604)。此时，用户确认所显示的图形。当图形与所意欲的图形相同时，不输入语音。因而，在 S1605 中确定否，清除图形识别结果(S1606)，并且该流程返回到 S1601。应当注意，用于清除图形识别结果栈的定时可以为开始下一个笔输入的定时，或者可以在给定时间流逝之后清除该栈。

作为用户的图形确认的结果，当用户确定了图形不同于所意欲的图形时，用户进行发音，以将图形校正成所意欲的图形。本系统接受发音(S1605)，在语音上识别所接受的发音(S1607)，并且将语音识别结果推入到语音识别结果栈中(S1608)。后续处理与图 5 的相同。

应当注意，当连续进行笔输入而非逐文字地进行笔输入时，用户可能在注意到前面输入不正确之前输入下一个图形。因此，可以考虑，在 S1604 中显示的图形中，指定要被校正的图形，并且相对于该图形念出所意欲的图形名称。

通过这样的配置，从用户的视点，基本上整形使用笔的图形输入，并且当确定了图形不正确时，念出所意欲的图形的名称，从而使得有可能将图形校正成期望图形。

应当注意，已经描述了这样的示例，其中通过使用笔来进行输入，并且当输入不正确时，如上进行语音输入。作为可选方案，可以通过使用语音来进行输入，并且当输入不正确时，可以通过笔输入来进行校正。

另外，已经描述了如上输入图形的示例。然而，当然，本发明可以

应用于与第二实施例类似地输入文字的情况。

### 第五实施例

在上述实施例中，已经描述了在进行笔输入时进行发音的情况，但是本发明不限于上述方案。本发明也适用于这样的情况，其中当通过扫描来阅读手写纸件文档等时，整形所书写内容。

在本实施例中，将作为示例描述多功能设备，其中多功能设备具有扫描在其上形成了多个手写图形的纸张并且整形所扫描的图形的功能。图 24 是根据本实施例的多功能设备的硬件配置图。标号 2401、2402、2403 和 2404 分别表示 CPU、RAM、ROM 和硬盘驱动器。标号 2405 表示 A/D 转换器，其用于将输入语音转换成数字信号。标号 2406 表示麦克风。标号 2407 表示图像读取单元，其用于从纸张读出信息。标号 2408 表示触摸屏。标号 2409 表示打印机单元。

图 25 是示出根据本实施例的处理流程的流程图。用于执行该流程图的程序保存在 ROM 203 或 HDD 205 中，并且在 CPU 201 的控制下执行。首先，图像读取单元 2407 读出在其上绘画了多个手写图形的纸张，以将该信息转换成图像数据(S2501)。此后，从图像数据提取图形(S2502)。所提取的图形经受图形识别(S2503)，并且将识别结果推入到图形识别结果栈中(S2504)。从识别结果中，选择具有最大置信得分的图形，并且基于所选图形而从在 S2503 中提取的图形提取特征量(S2505)。基于所提取的特征量而生成图形(S2506)，并且将该图形显示在触摸屏 2408 上(S2507)。由于在所阅读的纸张上绘画了多个图形，因此在本例中显示多个图形。用户确认所显示的图形，并且如果存在非意欲的图形，则指定该图形。这里，在触摸屏上直接指定所显示的图形。多功能设备确定指定了由用户指示的图形(S2508)。用户念出所意欲的图形，然后多功能设备接收该发音作为输入语音(S2509)。在识别接收语音(S2510)之后，语音识别结果和在 S2504 中推入到图形识别结果栈中的图形识别结果经受综合评价(S2511)。综合评价处理与第一实施例的相同。基于作为综合评价的结果而选择的图形，从在 S2503 中提取的图形提取特征量，使用该特征量来生成图形，并且显示所生成的图形

(S2512)。

通过上述配置，可以通过使用语音来不仅将实时笔输入而且将先前绘画的图形校正成所意欲的图形。应当注意，文字当然如同在第二实施例中一样可以是目标。

#### 第六实施例

根据第二实施例，处理一个文字的认识，但是本发明可以应用于一次识别两个或更多个文字的短语的情况。在这种情况下，使用短语词典来代替单汉字字典 1407，并且语音识别语法 1408 由无重叠地收集短语的读音的列表组成。短语词典保存各个短语标记和读音的配对数据。然后，文字识别单元 1404 识别多个文字，并且在短语词典中搜索所识别的文字串，以获得读音。此后，将包括读音信息的结果推入到文字识别栈中。随后，与第二实施例中的算法相同的算法可以处理该处理。

通过上述配置，不仅可以提高一个文字的认识准确度，而且可以提高短语的认识准确度。

#### 第七实施例

根据第六实施例，已经描述了一次识别多个文字的情况。然而，作为可选方案，多个文字可以通过使用根据第二实施例的方法来一次识别一个文字，以确认结果，然后可以识别下一个文字。在这种情况下，通过按例如安装在信息处理设备上的预定按钮，执行每个文字的认识结果的确认。此外，当要识别一个文字时，使用紧靠在这个文字之前的已确认文字的认识结果。因此，图 14 的配置还包括短语词典。将描述在触摸屏上绘画短语“忧郁”(由两文字组合“忧”和“郁”构成)的情况。“忧郁”的读音为“yuuutsu”，并且“忧”的读音为“yuu”并且“郁”的读音为“utsu”。首先，用户在使用笔书写文字“忧”时念出“yuu”，并且通过根据第二实施例的方法输入文字“忧”。在通过确认按钮确认之后，在触摸屏上以任意字体显示文字“忧”。在紧靠在该文字的右边，在以书写文字“郁”的意图使用笔书写由图 23 中的标号 2303 表示的这样的笔划时，用户念出“utsu”。在基于文字认识和语音认识而通过综合评价识别该手写文字之前，从短语词典中搜索以紧靠之前的文字“忧”开始的短语。然

后，获得“忧郁(yuu/utsu)”、“忧国(yuu/koku)”、“忧伤(yuu/shuu)”、“忧愁(yuu/shoku)”以及“忧虑(yuu/ryo)”。这样，第二文字有可能是“郁(utsu)”、“国(koku)”、“伤(shuu)”、“愁(shoku)”和“虑(ryo)”。在语音识别语法 1408 中的识别词汇表当中，增大“utsu”、“koku”、“shuu”、“shoku”和“ryo”的权重，并且帮助挑选这些文字作为识别结果。另一方面，文字识别单元 1404 还对笔划 2303 设置适当的权重，以挑选诸如“郁”、“国”、“伤”、“愁”以及“虑”的文字作为该识别的识别结果。通过这样的配置，获得按照“短语”的可能识别结果的效果。

#### 第八实施例

根据第二实施例，语音识别单元 1402 基于语音识别语法 1408 而执行语音识别。语音识别语法 1408 由无重叠地收集 50 个文字日语音节和单汉字字典 1407 中的所有单汉字的读音的列表组成。然而，本发明不局限于上述方案。准备了用于接受任何音素的组合的串联语音识别语法作为语音识别语法 1408，并且允许语音识别单元 1402 执行串联语音识别处理。这样，该单元可以不加限制地将任意的音素串识别为单汉字字典 1407 中的单汉字的读音。

#### 第九实施例

根据第二实施例，语音识别单元 1402 与文字识别单元 1404 的识别处理相独立地执行语音识别。与此相对，也可以采用下面配置。等待文字识别单元 1404 的识别，并且在 N 个最佳候选者列表中包括文字读音的列表。为了有可能在语音识别结果中挑选读音，在语音识别语法 1408 上设置适当的权重，以执行语音识别处理。例如，当获得如图 19 所示的文字识别结果时，可以执行加权，使得有可能在语音识别结果中挑选“kan”、“ken”、“ran”和“ai”。

#### 第十实施例

与第九实施例相对，还可以采用下面配置。等待语音识别单元 1402 的识别结果，并且在 N 个最佳者中包括文字读音的列表。为了有可能在文字识别结果中挑选具有该读音的文字，在文字识别单元 1404 上设置适当的权重，以执行该处理。例如，当获得如图 20 所示的语音识别结

果时，可以执行加权，使得有可能在文字识别结果中挑选具有读音“ei”、“ai”和“sai”的文字。

### 第十一实施例

在上述实施例中，已经作为示例描述了这样的示例，其中当语音识别结果和图形识别结果或文字识别结果经受综合评价时，对置信度进行求和，但是也有可能对结果进一步执行加权。

例如，当在喧闹的周围噪声环境中执行该处理时，可以考虑，语音识别的准确度由于噪声的影响而降低。鉴于上述情形，检测周围噪声，并且可以根据周围噪声的幅度而调整语音识别结果上的权重。当噪声电平高时，如果减小语音识别结果上的权重，则有可能减小噪声影响。作为实施例模式，例如，图 1 的配置还包括用于检测噪声的噪声检测单元(未示出)，并且由综合评价单元 105 接收由噪声检测单元检测的噪声。当进行综合评价时，可以根据所接收的噪声幅度来调整语音识别结果上的权重。

另外，在通过诸如笔的定点设备输入的情况下，当输入速度高时，可以建立所书写图形或文字粗略的假设。因此，可以考虑检测笔输入速度，并且根据输入速度来调整图形识别结果或文字识别结果上的权重。当速度高时，减小图形识别结果或文字识别结果上的权重，由此可以预期整个识别准确度的提高。作为实施例模式，例如，图 1 的配置还包括用于检测笔输入速度的笔输入速度检测单元(未示出)，并且由综合评价单元 105 接收由笔输入速度检测单元检测的笔输入速度。当进行综合评价时，可以根据所接收的笔输入速度来调整图形识别结果或文字识别结果上的权重。

### 第十二实施例

在上述实施例中，已经描述了这样的示例，其中图形识别或文字识别和语音识别经受综合评价，但是目标可以是面部识别或姿势识别。具体地说，为了识别人，在语音上识别名称的发音，并且随同面部识别一起进行综合评价，有可能提高该人的识别率。另外，在表现姿势时，念出对应于该姿势的语音发音，从而使得有可能提高姿势的识别率。

### 第十三实施例

应当注意，可以如下实现本发明。也就是，将存储用于实现上述实施例功能的软件的程序代码的存储介质提供给系统或设备。然后，该系统或设备的计算机(可选地，CPU 或 MPU)读出并且执行存储在存储介质中的程序代码。同样以这种方式，当然可以实现本发明。

在这种情况下，从存储介质读出的程序代码本身实现上述实施例功能，因此存储该程序代码的存储介质构成本发明。

例如，软盘、硬盘驱动器、光盘、磁光盘、CD-ROM、CD-R、磁带、非易失性存储卡、ROM 等可以用作提供该程序代码的存储介质。

另外，根据本发明的实施例不局限于执行由计算机读出的程序代码以实现上述实施例功能的情况。例如，基于程序代码的指令，在计算机等上运行的操作系统(OS)执行实际处理的部分或全部，并且通过该处理实现上述实施例功能。当然，这种情况也在本发明的范围内。

此外，可以以下面方式实现根据本发明的实施例的功能。也就是，将从存储介质读出的程序代码写入到插入到计算机的功能扩展板或者连接到计算机的功能扩展单元中提供的存储器。然后，基于程序代码的指令，向功能扩展板或功能扩展单元提供的 CPU 执行实际处理的全部或部分。当然，通过该处理实现了上述实施例功能。

在上面描述中，程序和硬件的组合实现本发明，但是可以减小程序的比重，并且可以增加硬件的使用。此外，在这些实施例中描述了从输入步骤到输出步骤的流程，但是本发明适用于这些实施例的一部分。

此外，在上述实施例中，作为信息处理方法的步骤，为了描述起见而给出了包括语音识别步骤的示例和包括图形识别步骤或文字识别步骤的示例，但是本发明不限于上述方案。代替执行诸如语音识别、图形识别或文字识别的识别，可以实现包括接收识别结果的步骤和对所接收的识别结果执行处理的步骤的流程。如果 OS 接收语音识别、图形识别、文字识别等的结果，并且包括对这些结果执行综合评价的步骤，则可以执行本发明。

在本发明中，给定目标是诸如图形识别或文字识别的模式识别的目

标, 例如, 第一实施例的用笔绘画的图形、第二实施例的用笔绘画的文字、或者第五实施例的扫描图形。图形识别装置例如是根据第一实施例的图形识别单元 104。语音识别装置例如是根据第一实施例的语音识别单元 102。上述实施例中的选择装置从综合评价结果选择图形或文字。提取装置例如是根据第一实施例的特征量提取单元 106。生成装置例如是根据第一实施例的生成单元 107。显示控制装置例如是根据第一实施例的显示控制单元 108。置信得分是表示作为识别结果而获得的候选者是各个目标时的确信度的值, 例如, 图 7 所示的值。文字识别装置例如是根据第二实施例的文字识别单元 1404。用于识别模式的信息在图形的情况下为图形名称/类型, 并且在文字的情况下为文字读音。图形类型是表示图形种类的信息, 换句话说, 图形名称等。

虽然参考示例性实施例描述了本发明, 但是应当理解, 本发明不局限于所公开的示例性实施例。所附权利要求的范围应当被给予最宽的解释, 以便囊括所有变型、等效结构和功能。



图1

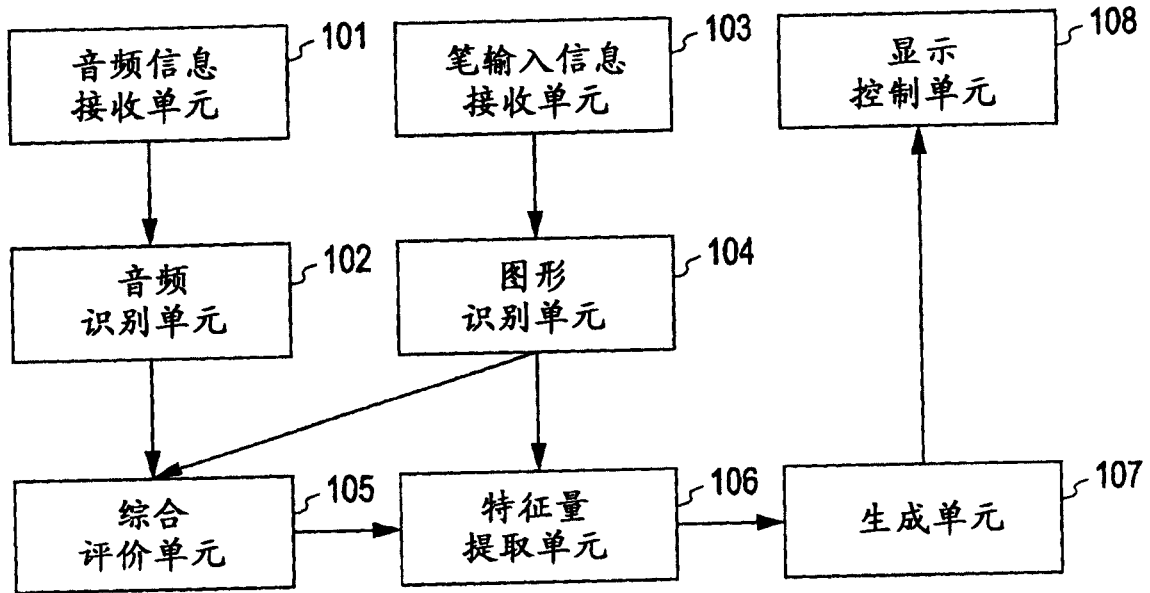


图2

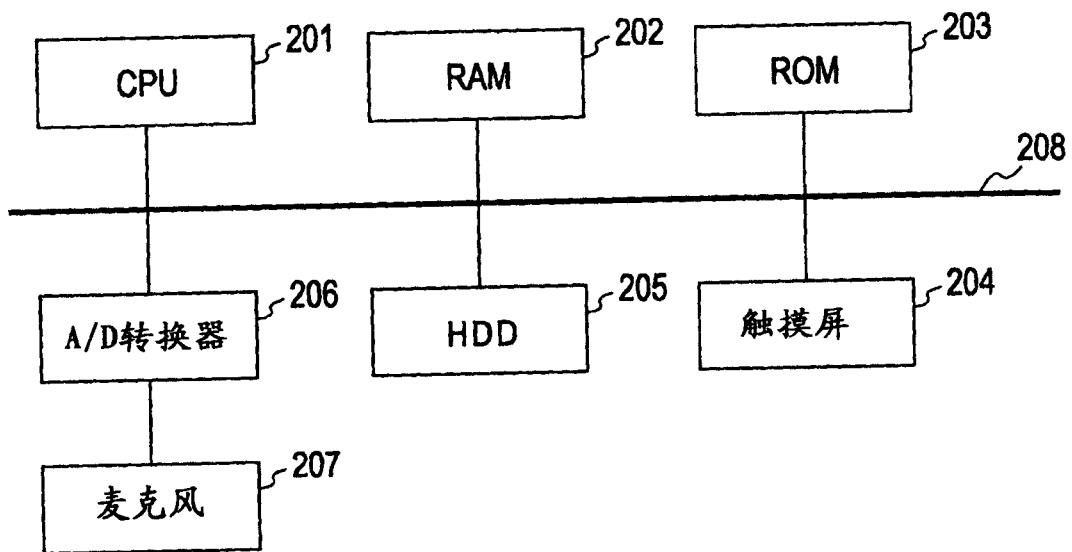


图 3

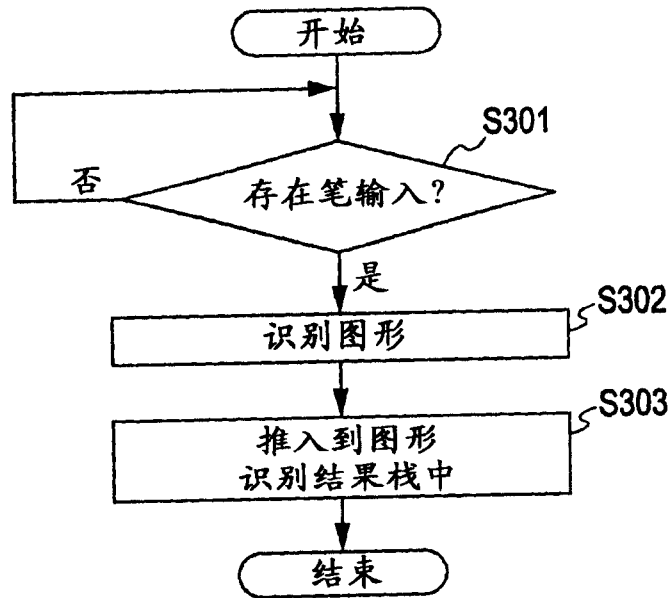


图 4

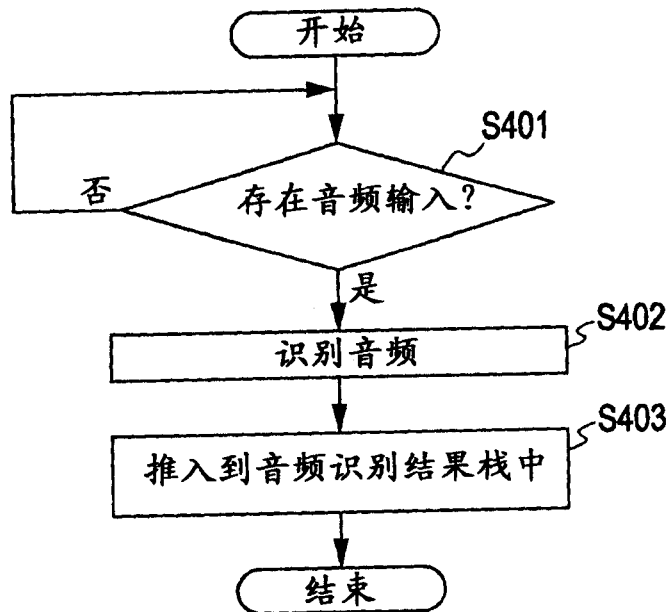


图5

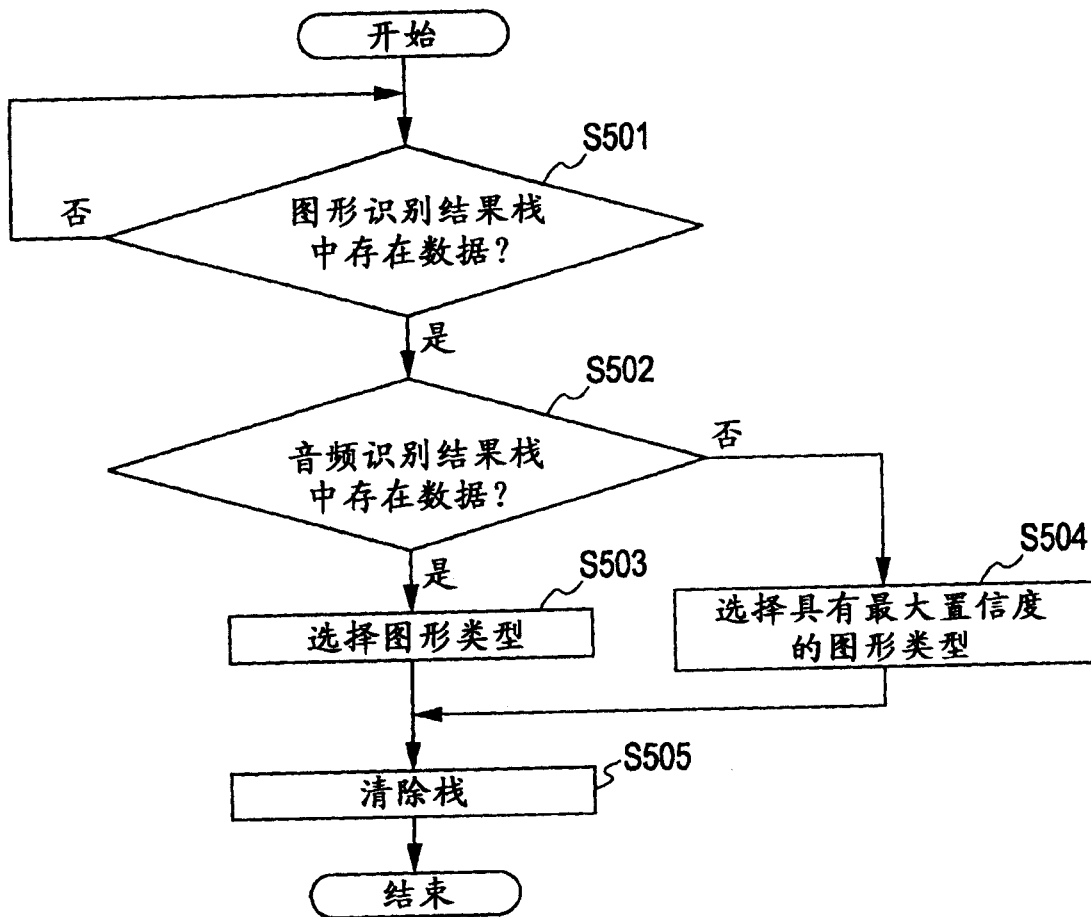


图6

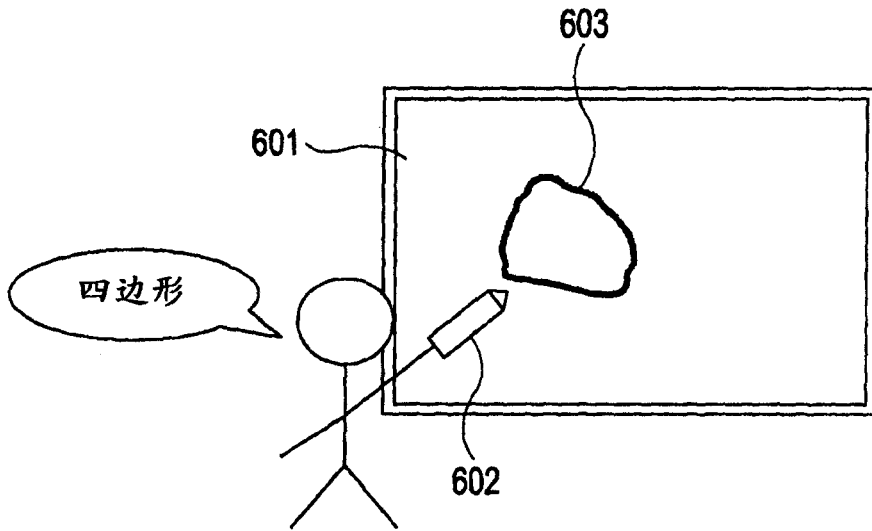


图7

开始	00:00:04:08			702
结束	00:00:05:20			
候选者	第一	图形类型	椭圆	703
		置信度	60	
	第二	图形类型	四边形	
		置信度	50	
	第三	图形类型	三角形	
		置信度	20	

701

图 8

开始	00:00:05:10			801
结束	00:00:06:20			
候选者	第一	图形类型	三角形	803
		置信度	70	
	第二	图形类型	四边形	
		置信度	60	
	第三	图形类型	椭圆	
		置信度	10	

图 9

综合评价	第一	图形类型	四边形
		置信度	110
	第二	图形类型	三角形
		置信度	90
	第三	图形类型	椭圆
		置信度	70

图10

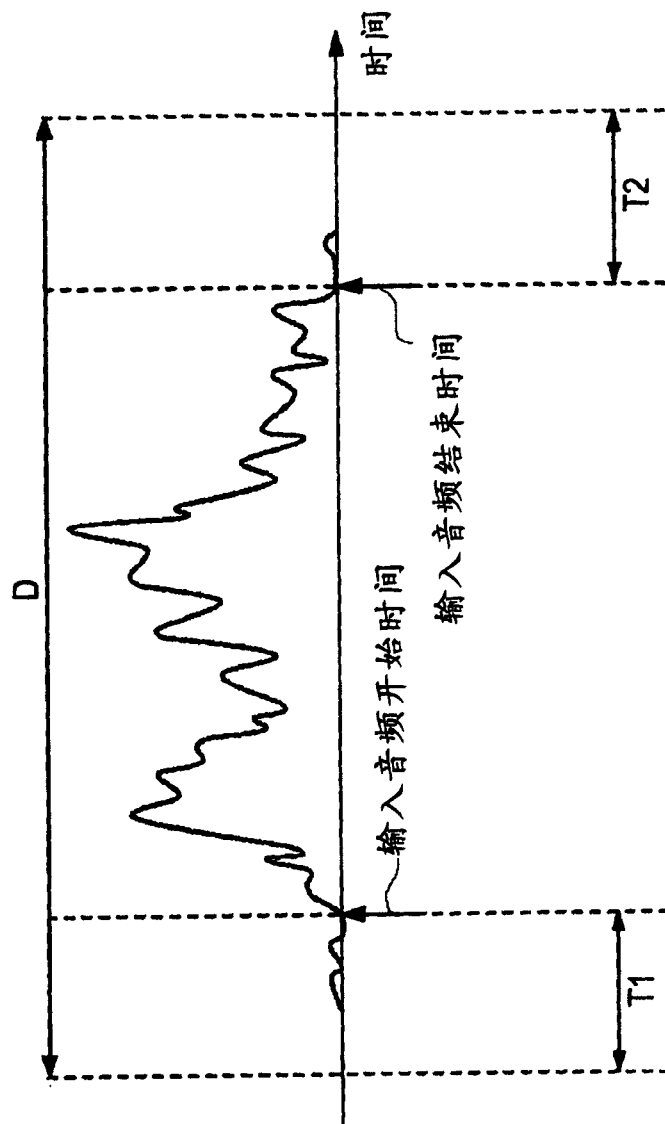


图 11

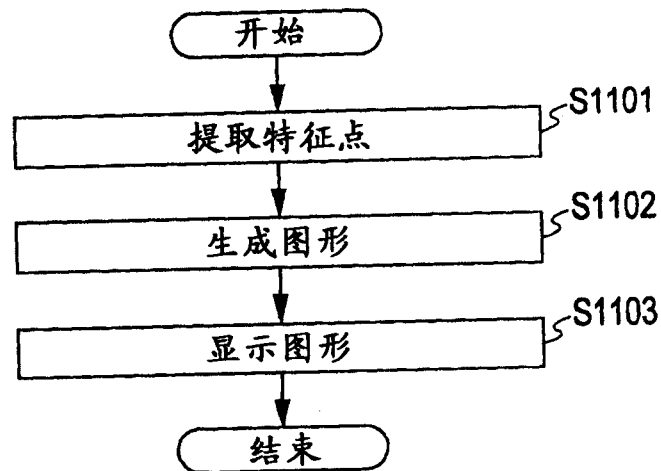


图 12

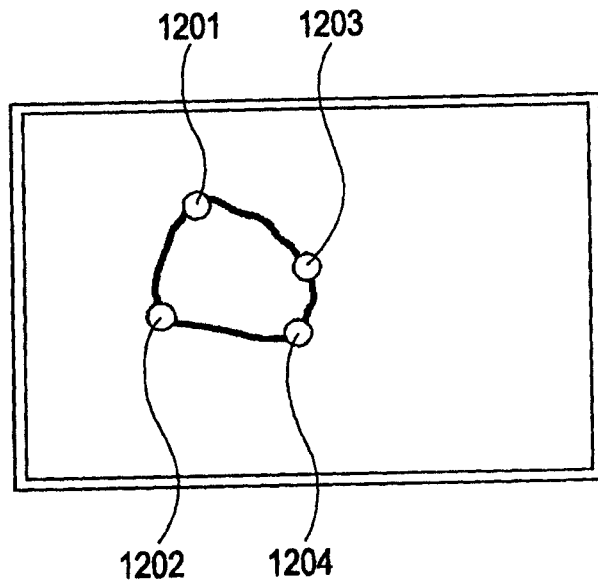


图 13

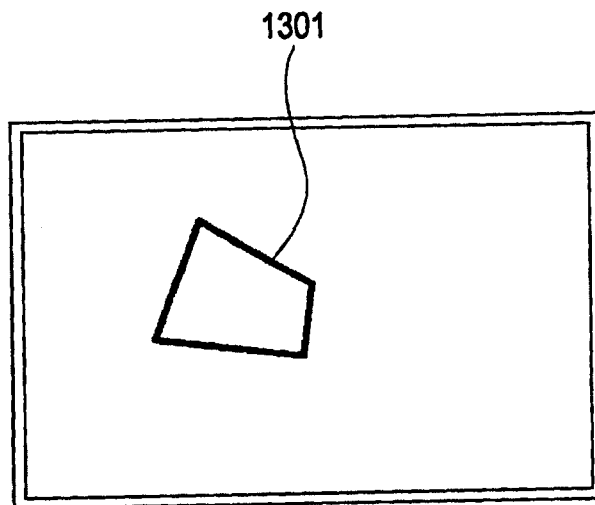




图 14

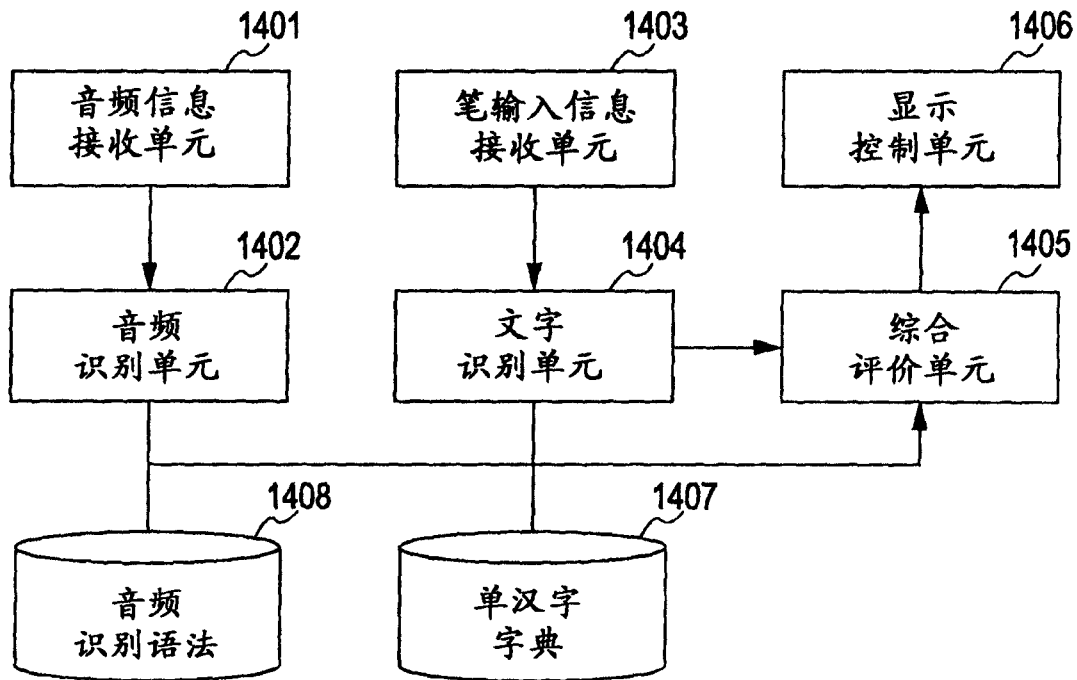


图 15

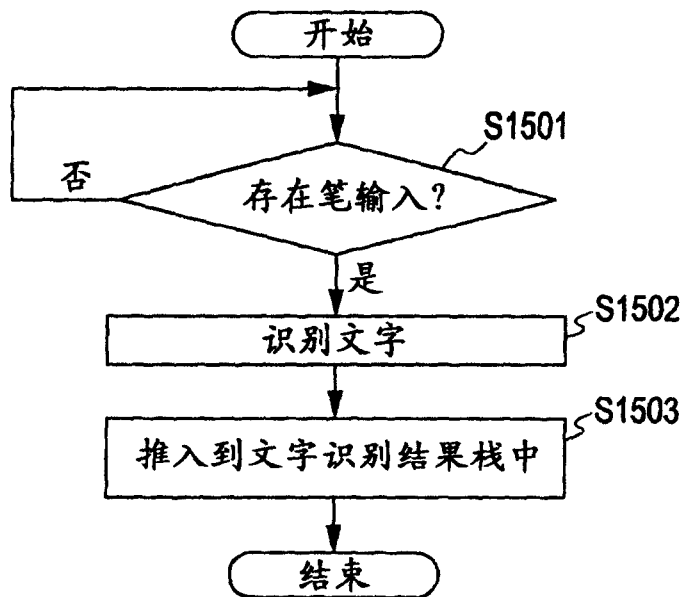


图16

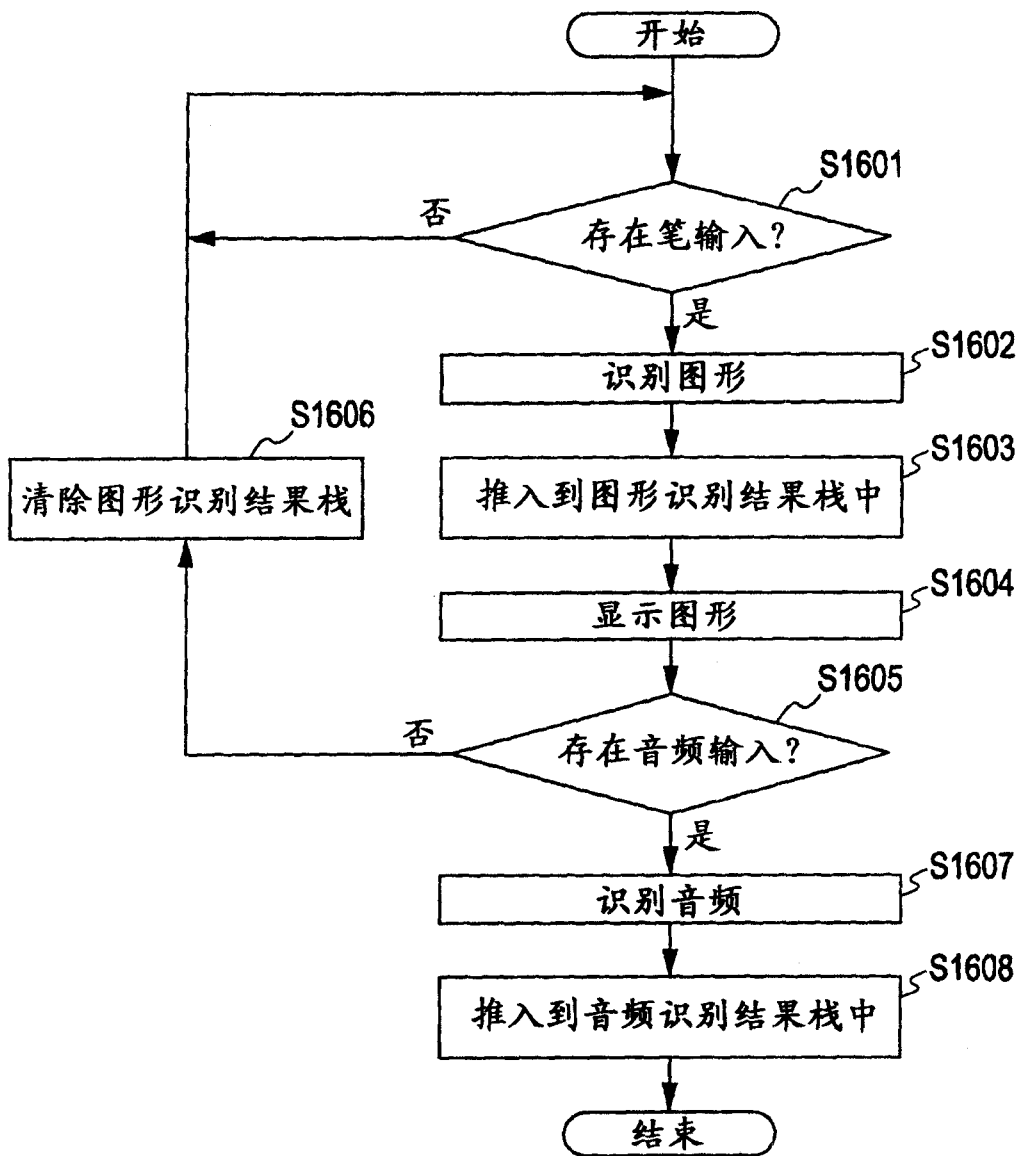


图17

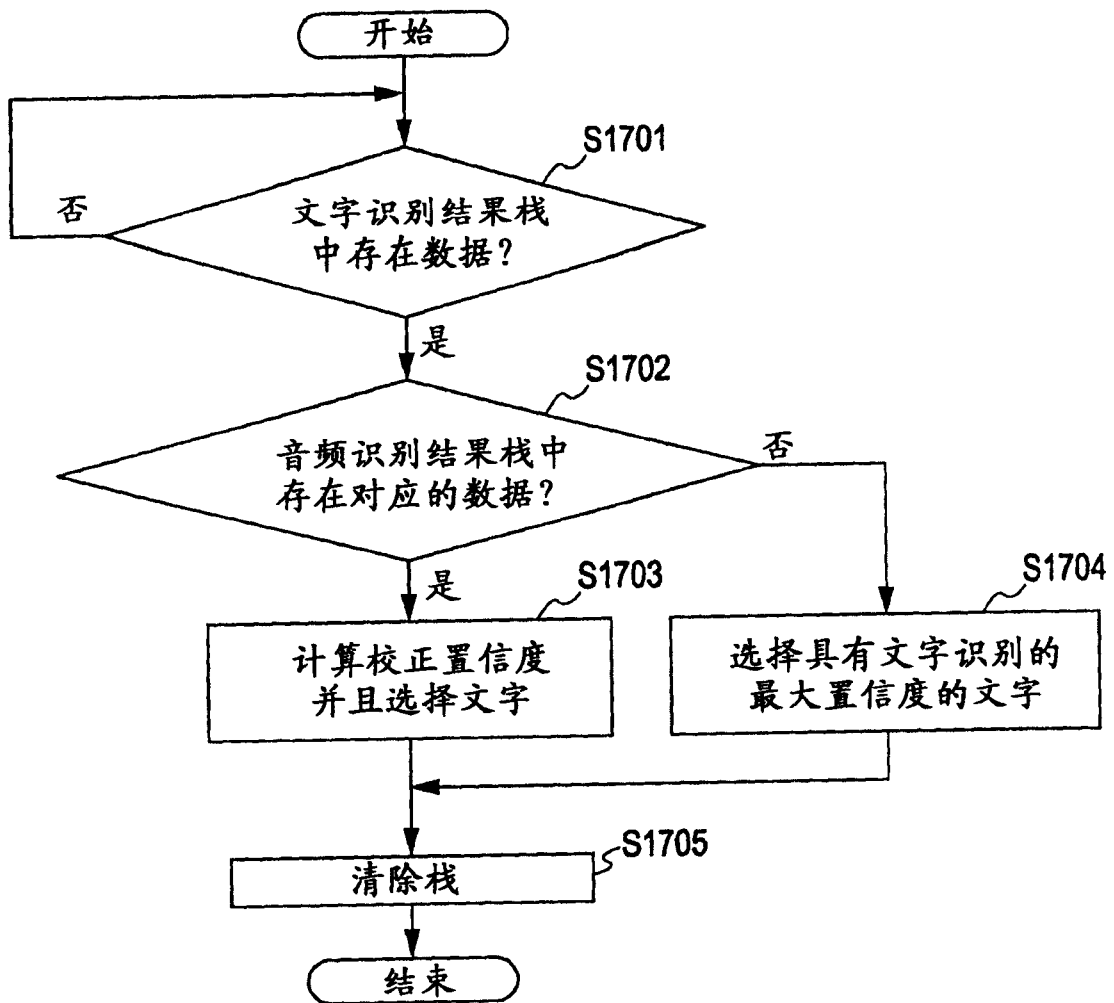


图18

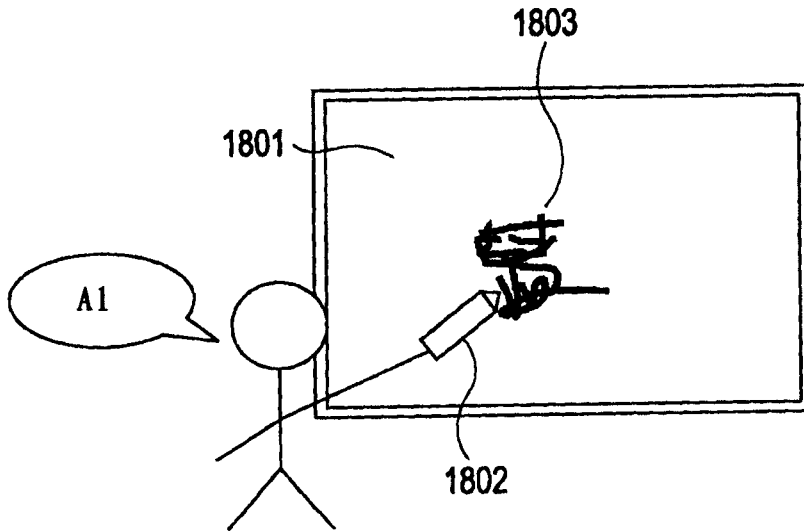


图19

开始	00:00:04:80			1901	1902
结束	00:00:05:20				
候选者	第一	文字	監 MONITOR (KAN, KEN)	1903	
		置信度	70		
	第二	文字	蘭 ORCHID (RAN)		
		置信度	50		
	第三	文字	藍 INDIGO (AI, RAN)		
		置信度	40		

图 20

开始	00:00:05:10			2001	2002
结束	00:00:06:20				
候选者	第一	读音	EI	2003	
		置信度	60		
	第二	读音	AI		
		置信度	55		
	第三	读音	SAI		
		置信度	30		

图 21

综合评价	第一	文字	藍 INDIGO
		置信度	95
	第二	文字	監 MONITOR
		置信度	90
	第三	文字	蘭 ORCHID
		置信度	68.3

图 22

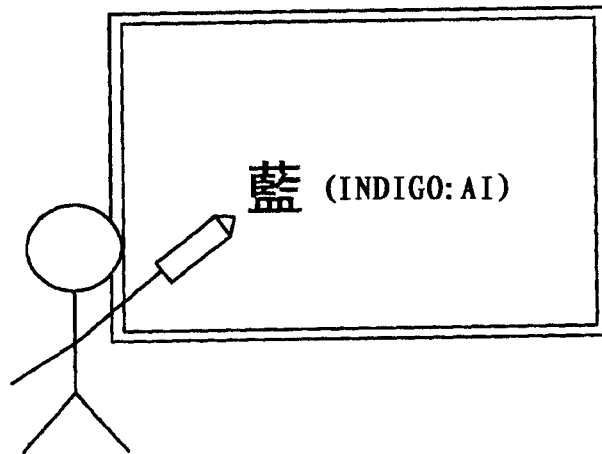


图 23

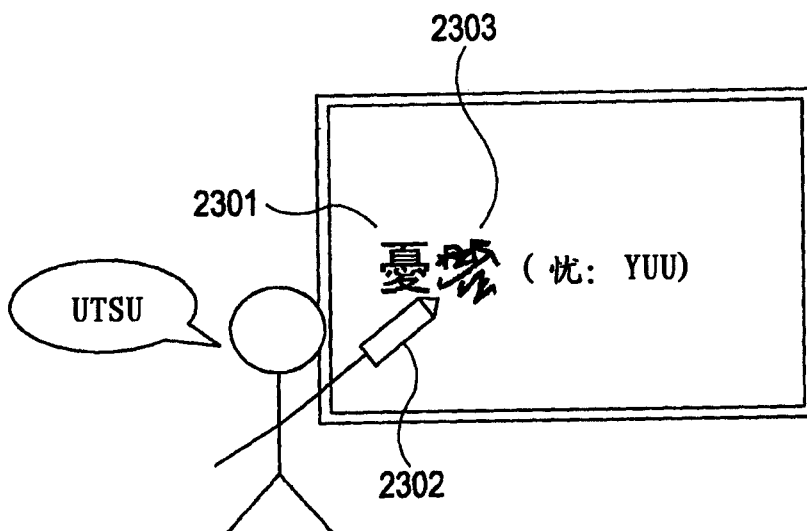


图 24

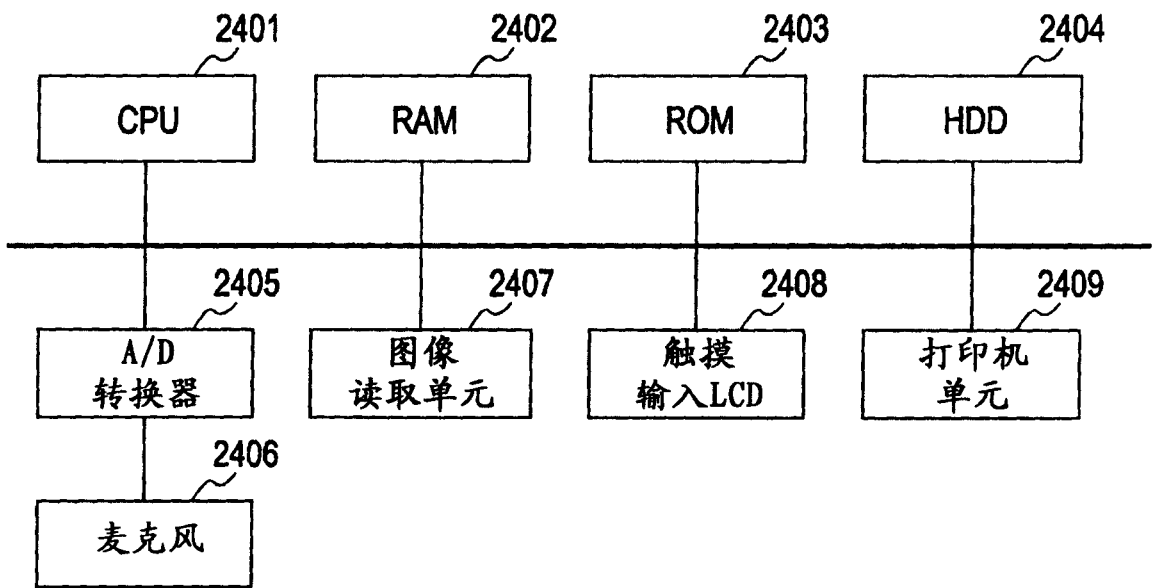


图 25

