

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06Q 30/00 (2006.01)

G10L 15/26 (2006.01)



# [12] 发明专利说明书

专利号 ZL 02813480. X

[45] 授权公告日 2007 年 6 月 6 日

[11] 授权公告号 CN 1320499C

[22] 申请日 2002.6.20 [21] 申请号 02813480. X

[30] 优先权

[32] 2001. 7. 5 [33] EP [31] 01890203.1

[86] 国际申请 PCT/IB2002/002477 2002.6.20

[87] 国际公布 WO2003/005258 英 2003.1.16

[85] 进入国家阶段日期 2004.1.2

[73] 专利权人 皇家飞利浦电子有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 H·F·巴托斯克

[56] 参考文献

CN2167404Y 1994.6.1

CN-128198A 1999.2.17

US-5806078A 1998.9.8

EP-0932108A2 1999.7.28

CN1271131A 2000.10.25

US-6185514B1 2001.2.6

US-6173259B1 2001.1.9

审查员 吴芸

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 陈景峻 陈霁

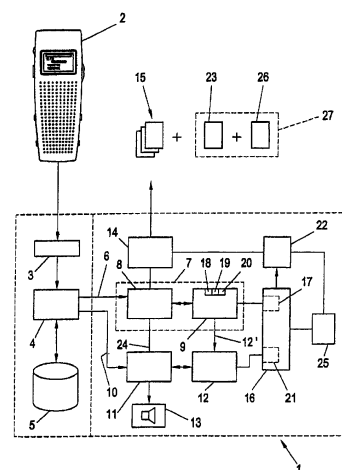
权利要求书 2 页 说明书 10 页 附图 3 页

[54] 发明名称

用来转录口述和检测人工动作的装置

[57] 摘要

为在先使用服务设施提供帐单信息的一种方法，由服务设施按输入信息自动获得输出信息，例如是口述的转录，利用自动语音识别系统将口述文件转换成文本文件，在提供帐单信息时考虑对自动获得的输出信息所检测到的人工处理工作量。



1. 用语音识别装置(4)转录口述的一种设备,将提供的口述文件自动转换成文本文件,用文本处理装置(7)人工处理文本文件,在其中为文本处理装置(7)配置有检测装置(16),检测装置(16)用来检测为了产生正确的文本文件所执行的人工处理工作量,并且检测装置(16)包括第一检测级(17),用来检测输入单元(9)的多个键中的至少一个特殊键的动作,和第二检测级(21),用来在与所述第一检测级(17)相同的时间,通过检测在所述文本文件中不再出现的那些口述文件中文字的时间,检测在控制单元(12)的控制下收听提供的口述文件所需的复杂性,其中对至少一个特殊键的动作的检测和对收听口述文件所需的复杂性的检测决定对所述文本文件所执行的人工处理工作量,所述转录口述的装置还包括汇报装置(25),所述汇报装置(25)与所述检测装置(16)相连,并通过输出单元(14)将关于所述复杂性的信息反馈给口述人。

2. 按照权利要求1的设备,其特征是第二检测级(21)检测一个脚踏开关的动作。

3. 按照权利要求1的设备,其特征是检测装置(16)包含比较装置,用来将用语音识别装置(4)获得的文本文件与人工处理后的文本文件自动比较。

4. 按照权利要求1的设备,其特征是检测装置(16)被分配给文本处理装置(7)的预定输入装置(18,19,20),并且用来检测这些预定输入装置(18,19,20)的操作。

5. 按照权利要求4的设备,其特征是所述至少一个特殊键包括用来在文本文件中删除数据的至少一个键(18)。

6. 按照权利要求4的设备,其特征是所述至少一个特殊键包括用来在文本文件中重写数据的至少一个键(19)。

7. 按照权利要求4的设备,其特征是所述至少一个特殊键包括用来在文本文件中插入数据的至少一个键(20)。

8. 按照权利要求1的设备,其特征是检测装置(16)被用来检测收听口述文件的各部分所需的时间。

9. 按照权利要求1的设备,其特征是为检测装置(16)配置有记录装

置(25),用来记录与检测到的人工处理工作量有关的数据。

## 用来转录口述和检测人工动作的装置

### 技术领域

本发明涉及到预先使用处理计算机软件的服务设备提供帐单信息的一种方法。

本发明还涉及到一种用来转录口述的装置,用语音识别装置将所提供的口述文件自动转换成文本文件,并且用文本处理装置人工处理文本文件,还用计算装置为执行的转录产生帐单信息。

### 背景技术

自动语音识别系统越来越多地被用于口述的自动转录。转录的结果取决于有关语音识别系统在复杂性上的质量,这包括存储在程序库中专用来识别和有效地分析某一口述人的语音的“词典”。“词典”通常是按领域划分的,例如有医学领域,技术领域,法律领域和商业领域。考虑到高质量是来自高度复杂性的有效的语音识别软件和用于语音识别的数据存储,就出现了能使用完善的语音识别系统的转录服务公司,为各种顾客将口述转录成文本,并且为其相应的转录服务开发票。例如在专利文献 US6173259A 中就描述了这样一种口述转录系统。用语音识别软件对口述进行识别,并且转换成文本文件,文件能够由人工处理以纠正文本文件中因口述识别而产生以及口述中的缺陷造成的错误,例如是修改错字,插入丢字或删除多余的字。如果由口述文件转换而来的文本文件的文本质量差,还能够决定不用自动语音识别系统执行转录,也就是由口述人工输入文本。在使用中,如果顾客有要求,在这种情况下就能由顾客自己执行转录,并且也能按对应的方式提供转录的文本文件供顾客自己修改或校对。在难以为任何人产生准确文本的情况下可以提供由顾客自己来修改的转录,并且就不通顺的个别段落对作者提出询问。除去口述过程中的背景噪声之外,由于口述中的错误或缺乏经验的人口述的质量不满意,其他人就难以理解文本,这往往是因为使用了在存储的程序库中没有包含的比较生疏的专业术语。

由商业服务公司执行的口述转录通常附有发票,其数额会根据文本的数量即转录的行数而有所不同,还关系到按照副分类所应用的医学,

工程等等领域。然而,口述人(以下称为作者)在口述过程中的精确度的不同及其在“脱机转录”中的差别自然会受转录复杂性和转录文本质量的影响。如果一个书记员直接即“联机”口述一个文本,就能立即检测和校正任何错误。然而,如果是面对一个存储装置来口述,例如是模拟或数字文本记录器上的手持录音机,还可以通过通信链路连接到带自动语音识别系统的计算机,在这种“联机”使用中没有直接反馈,因而无法完成对作者的有关训练。然而,为了能够尽可能有效地使用自动语音识别系统,要随时识别不属于所要转录的文本的那些用清楚和便于理解的语言给出的指令,例如是制作某一特定文本部分的复本,用斜体字编排某些字句等等。可以为口述中的典型错误注释“要结束了”(在不清楚这是要转录的文本还是解释性注解的情况下)或是自动校正,例如是“不是左而是右侧”。许多问题会造成这种不清楚的措辞,例如是言语不清,背景噪声,口述过程的中断-比如说口述过程中有一个人进屋讲话,或是采用自动语音识别系统的自动转录遇到电话铃声,因而必须要人工修订或校正有关的文本文件,而这样做的成本是很高的。

在自动转录服务中有一种情况非常类似于口述转录时的情况,那就是用一种翻译系统将现有的一种文本自动翻译成另一种语言,在检查翻译文本时可能需要人工修订或校对。为这一修订所做的工作也应该记入要产生的帐单。此外还必须考虑到在为原始文本定格式时需要人工处理的工作量的多少,没有从句的简单句和采用日常用语代替稀有术语能便利自动翻译,这样能保证翻译文本的处理成本较低。

因而就迫切需要有一种能根据输入信息自动产生输出信息的人工处理所招致的成本的方法,例如是在产生帐单信息时必须考虑到在发票中记入服务操作。

#### 发明内容

本发明的目的就是要能够在使用适合有关的工作方法的服务设施时产生帐单信息。

本发明的具体目的是提供一种方法和装置,有可能用一个可变价目表自动替代产生文本的平均价格,如果为转录口述或翻译文本不需要或是几乎不需要人工文本处理,与需要更多努力人工处理转录或翻译文本的情况相比,能够支付一个合理的价目表。这样,作者就会主动学习和改进如何能以低发票额为其口述或是表达支付报酬的方法,并且刺激作

者加以改进,例如是在口述时采用清晰的发音并提供可清楚识别的指令。

为了实现本发明的这一目的,按照该方法第一实施例的特征如下:

为在先被用来处理计算机软件的服务设施提供帐单信息的一种方法,用计算机软件处理由服务设施接收到的输入信息,对由此获得的输出信息进行人工处理并检测所采用的人工复杂性,在为在先使用服务设施产生帐单信息时考虑检测到的人工复杂性。

在用计算机软件处理输入信息时获得的输出信息最好是一种文本文件。特别是在执行一种语音识别方法时使用属于一种自动语音识别系统的计算机软件,以及使用计算机软件来处理作为输入信息接收的口述文件,结果将接收的文本文件作为输出信息。

按照本发明第二方面的特征在于以下方法:

转录口述的一种装置,用语音识别装置将提供的口述文件转换成文本文件,用文本处理装置人工处理文本文件,并且用计算装置为所执行的转录提供帐单信息,如果文本处理装置配置有检测装置,检测装置就用来检测为了产生正确的文本文件所执行的人工处理工作量,并且将检测装置连接到计算装置来提供帐单信息。

按照本发明的方案,为输出信息的人工处理工作量特别配备了自动检测,在采用自动语音识别系统或自动翻译系统获得文本文件时,正如下文还要具体解释的,可以用各种方式执行这种工作量检测,无需处理人员具体操作。可以根据所需的处理工作量执行不同的计算,对那些实际上不需要人工处理的输出文本或一般输出信息可以支付低价帐单,反之对大量处理支付较高价格。对于口述的情况,在评估自动获得的处理后文本文件时,使用连接到文本处理装置的检测装置能够自动检测有多少识别有误的文字,以及口述中缺少多少诸如“另起一段”,“开头”,“下划线”,“句号”,“逗号”,“问号”等等指令。还能通过反复收听口述的段落时所需的工作来检测口述的主题意思。意思含糊就会带来相应的帐单信息,还会促使口述人在产生输入信息时努力做到更精确,也就是在口述中说话更加精确并且给出指令,而得到的回报就是帐单信息即低价格发票,并且最终的输出信息质量会更好。

专利文件 US 6,185,514B1 揭示了一种计算机系统,用于检测进行通信的雇员的行动,以便确认由雇员操作的时间。已知的计算机系统

对键盘的击键进行检测，并存储 log 文件。已知的计算机系统不能用于转录口述和人工处理识别的文本文件，也不能通过聆听口述检测用户校正文本文件的动作。此外，已知的装置看到来自键盘的每一次击键。仅仅需要知道的信息是键盘产生的动作，而不必知道是什么特定的键在操作。实际的键不被记录。记录的是这样一个事实，即在一秒的间隔内用户输入键盘的内容。

为了检测人工处理工作量，根据要求的检测精度有各种可能性可以用来实现反映真实费用的检测。最简单的选择是执行口述转录的情况，可以将用自动语音识别系统转录的原始文本文件与最终校对过的文本文件相比较，用两个文本文件之间的差别构成处理复杂性的尺度。添加和删除的文字部分以及纠正即重写的文字或文字部分被用做计算的指示器。特别是有可能通过联系着文本文件的文字总数并通过产生的百分数设置纠正的，删除的，插入的文字或文字部分，由此获得帐单信息的产值。

按逻辑方式检测人工处理工作量的另一种方法是自动检测在处理文本文件过程中执行的某些人工输入。经常要通过文本处理装置的键盘人工输入，对经常操作的键盘组合也就是功能键和相应的字符键进行记录。例如是对要删除的字符串做标记，然后删除或是按下功能键“插入”来插入一个字符，而后用有关的键输入要插入的字符。如果要用其它字符替代某些字符也就是“重写”，首先，系统用一个预定的功能键变换到“重新打字”模式，然后输入所需的字符。为了检测文本处理所需的人工工作量，可以检测和记录上述键的操作。特别是不断监视上述功能键的操作，在操作时连同后续的字符键操作对键盘操作计数。记录这些受监视键的操作的另外一种可能性是同时检测收听口述的时间，获得有关删除，插入或重写文字的信息作为人工处理工作量的尺度，并且在产生帐单信息时加以考虑。

可以将收听各部分口述文件所需的时间与单独但同样是自动检测的键盘操作加以组合。在转录口述时精确地确定复杂性，还能知道口述的一部分必须收听几次-例如是要几次才能听懂一个难以理解的词-并且仅仅是比较文本或仅仅检测键盘操作而不考虑实际所需的时间。为了额外检查插入频率或仅仅指出在哪里检查转录的文字，需要收听口述以便正确理解后续的文本部分，为了精确计算可以在许多点上采用自动

检测,也就是在人工修订文本文件时需要收听口述文件的各部分。

所有上述的检测都可能获得类似于打字的信息,检测出口述中“无用”部分的长度,并且适当地提供帐单信息。例如,发票中可能的帐单信息是打字中有“15%的口述长度与文本无关”。为了能够传递这种帐单信息并且得到对处理复杂性的论证而适当提供了检测到的人工处理工作量的自动汇报,在其中存储个别的自动检测,并且能组合成上述的总帐单信息。然而,如果发票额有疑问,必要时能够读出并打印处理汇报的细节。这样就能产生通用的汇报信息来描述人工处理的工作量,并且可以将这种汇报信息发送给服务设施的用户以便更加有效地使用服务设施。

本发明提供了用语音识别装置转录口述的一种装置,将提供的口述文件自动转换成文本文件,用文本处理装置人工处理文本文件,在其中为文本处理装置配置有检测装置,检测装置用来检测为了产生正确的文本文件所执行的人工处理工作量,并且检测装置包括第一检测级,专用来检测输入单元的多个键的动作,和第二检测级,用来检测在控制单元的控制下收听口述文件所需的复杂性。

#### 附图说明

以下要参照附图中所示的最佳实施例进一步描述本发明,然而本发明并不受实施例的限制。

图1用电路框图的形式表示一种用来转录口述的装置,在其中检测人工处理的工作量。

图2表示在修改文本文件时用来确定人工复杂性的流程图。

图3表示在收听各部分口述时用来确定复杂性检测程序的进一步流程图。

#### 具体实施方式

图1表示用来转录口述的装置1的示意图,在这种情况下为存储口述材料提供了一个数字口述装置2。可以通过电信连接,互联网连接或直接通过存储卡等等简单数据载体将口述文件发送到装置1,而装置1具有一个接口3,按相应的形式将口述文件提供给具有有关数据库5的自动语音识别装置4。由语音识别装置4利用语音识别软件首先访问存储在数据库5中的语音图形,然后访问字库,通过翻译自动产生一个文本文件。文本文件通过通信线路6被传送到文本处理装置7,具体说



就是文本处理装置 7 的文本处理单元 8。文本处理单元 8 配置有一个输入单元 9,也就是属于文本处理装置 7 的一个普通键盘。

同时通过通信线路 10 将口述文件提供给一个音频单元 11,音频单元 11 配置有一个控制单元 12 用来控制通过一个扬声器或头戴送受话器单元 13 收听口述指示。除了诸如开始,重放,回跳,停止,快进等控制功能之外,还能执行快倒等功能。在图 1 中仅仅示意性表示了输入装置 12'可以是脚踏开关,但这仅仅是为了简化图 1,还可以连接到键盘 9 用键盘 9 补充所需的控制功能。

文本处理单元 8 连接有一个输出单元 14,具体说在此处是一个打印机,能获得文本打印输出 15。还可能有其它可能的输出,例如是在口述过程中通过对应的接口和通信线路将文本文件输出到现有的计算机系统,这种传输经常使用 e-mail。

然而,在使用自动语音识别系统发送转录的文本之前,要提供人工检查并执行输入单元 9 和控制单元 12 的应用程序。输入单元 9 和控制单元 12 被连接到用来自动检测工作量的检测装置 16,特别是检测使用文本处理装置 7 人工处理原始文本文件期间的工作量。

检测装置 16 包含第一检测级 17,专用来检测输入单元 9 中特殊键的动作,例如是用来删除字符的键 18,用来重新打字的键 19,以及用来插入字符的键 20 的动作。为了自动检测处理复杂性的目的还可以监视其它键的动作,对键盘动作的这种监视和记录还取决于所使用的文本处理系统。对键 18,19 或 20 动作的记录从技术上可以直接由硬件来实现,但是最好是检测这种键盘操作的代码输出。

检测装置 16 还包含第二检测级 21,用来检测在控制单元 12 的控制下收听口述文件所需的复杂性。

在区域 17 和 21 内检测的复杂性所包含的信息可以但是不必要组合。这一信息被提供给计算装置 22,在提供有关口述文件的转录及文本文件处理的帐单信息时加以考虑。用这种方式获得的帐单信息被提供给上述输出单元 14 打印输出一个发票 23。代替打印输出可以用电子文件形式发送至少一件帐单信息,为此要相应地提供不同结构的输出单元 14。

为了在收听口述文件的同时处理文本文件并且能将所建立的工作范围(与检测装置 16 的区域 17 和 21 相比)的有关信息加以组合,在图

中用连接 24 示意性地表示按常规方式将文本处理单元 8 连接到音频单元 11 (例如是通过图中没有具体表示的这些单元的处理器或是各单元共享的一个处理器)。

为了验证检测到的处理费用或是能够将作者发出的口述通知作者, 汇报装置 25 被连接到检测装置 16, 记录人工文本处理中的个别处理, 并且在处理帐单信息也就是产生发票时将对应这些记录的信息传送到输出单元 14, 以便输出单独的汇报 26, 或是如图 1 中用虚线表示的框 27 所示的一个随汇报的组合帐单。口述文件的作者就这样接收对转录文本的处理复杂性的反馈, 或是他的口述究竟有多差。还可以发送细节信息, 例如是口述中难以理解的  $x\%$ , 口述中多余或是自我纠正因素等等的  $y\%$ 。对错误的细节信息可以令作者知道可能需付出较高的价格, 也就是可能没有价格折扣。

最好是利用常规转录系统中为检测到的各个字提供的现成的事项, 现有的时间信息, 也就是各个字在口述文件(语音文件)中的起始时间和结束时间。

图 2 表示的流程图是在收听口述文件的同时处理文本文件的一个例子。从框 30 开始之后, 在框 31 收听部分口述文件。在框 32 检查文本文件与口述文件相比有没有多余文字, 如果有就在框 33 中选择这些多余文字然后在框 34 中用图 1 中的删除键 18 删除(或是用为这种多余文字设置的标记键做标记)。在框 35 中存储这些多余文字所持续的时间。

接着在框图 36 中检查是否有重复的文字部分, 如果有, 就在框 37 中选择这些文字随后在框 38 中删除或是按下一个重复键做标记。然后在框 39 中存储这些重复文字所持续的时间。

在框 40 中检查在口述文件中是否包含作者的纠正也就是自我纠正。如果有, 就在框 41 中选择有关的字, 然后在框 42 中用自我纠正键删除或是做标记。然后在框 43 中存储这些文字所持续的时间。

在框 44 中检查有没有所谓的背景文字, 也就是在背景中有没有另一个人说的话中的文字也被转换成了文本。如果有, 就在框 45 中选择这些背景文字并在框 46 中删除或是按下一个背景噪声键做标记, 然后在框 47 中存储这些背景文字所持续的时间。

在框 48 中检查有没有错误文字, 如果有, 就在框 49 为错误文字做

标记,并且在框 50 中纠正,用图 1 中的键 19 通过重新打字也就是输入正确的字来纠正。然后在框 51 中存储打字字符的数量。

在框 52 中检查文本文件与口述相比是否有丢字。如果有,就在框 53 中用图 1 的键 20 按插入模式输入丢失的字(或是部分字)。然后在框 54 中存储打字字符的数量。

在框 55 中检查是否已经达到口述文件的结尾。如果没有,系统就返回到框 31 收听口述文件的下一部分。然而,如果发现口述文件的结尾,系统就在程序的结尾转向框 56。

图 3 的流程图表示在复制口述文件即收听口述的同时检查对文本文件进行处理的位置来检测复杂性。在框 60 中检查是否按下了重放键。如果是,就在框 61 开始重放并且对重放文字做标记。框图 62 中存储重放时间,然后在框 63 中检查是否按下了停止键。如果不是,就在框 63 中重复这一检查直至发现停止键被按下,而后在框 64 中停止由图 1 中音频单元 11 的重放。然后在框 65 中存储重放的结束时间。

在图 3 所示的流程图中还要在框 66 内检查是否按下了倒退键。如果是,就在框 67 中开始倒退也就是快速倒退,并在框 68 中检查是否已经按下了停止键。如果不是,就继续重复框 68 中的检查直至最终发现停止键已被按下,这时就在框 69 中停止倒退。(在此处应该注意到,除了因对口述文件中的各个字检测到某种时间标记,还可以按公知方式逐字执行重放和快进,并且在特殊字处停止)。

然后在框 70 中检查是否按下了快进键。如果是,就在框 71 中开始快进,而后在框 72 中检查是否按下了停止键。重复这一检查直至发现按下了停止件,然后就在框 73 停止快进。

然后在框 74 中检查是否按下了位置指示器。如果是,就将位置指示器设置在位置指示器指定的位置,然后系统返回到框 60 中搜索,就象在框 74 中检查发现已经按下了位置指示器时一样。

为了分析按这种方式检测到的信息,从存储的重放时间目录中检测哪一时间段曾被收听多次。由此可以按百分数根据重放重复时间和总口述时间计算出一个比例值。

例如是检测在确定的文本中不再出现的那些口述文字的时间,按照被删除文字的类型加以区别:

-多余文字的时间值

-重复文字的时间值

-自我纠正的文字的时间值。

还要根据总口述时间的百分数值设置这些时间值。

另外,同样作为与总口述时间有关的百分数,还要检测对应着噪声或是背景中的话音产生的那些文字的时间值。

最后检测与文本文件中的字符总数有关的输入字符数,为纠正错误识别的文本而输入的字符数,以及为插入丢字而输入的那些字符。

检查与总文本有关的输入字符总和,以及各种时间值的总和是否超过了预定的最大值,例如对每一种情况都是 25%,这些时间值有 a) 重放重复时间, b) 多余文字造成的时间值, c) 重复文字造成的时间值, d) 因删除自我纠正文字造成的时间值,以及 e) 因删除声音背景信号而删除文字造成的时间值。如果属于这种情况,就可以支付起价的最高价格,它等于最低价格的二倍。

或是将上述的参数(字符总数和总时间的百分数)与最小值例如是 5%相比较,如果这些值(字符数,时间)低于这一最小值,就支付最低价格,也就是从最高价起折扣 50%。

对处于上下限之间的费用值,在从最高价格起按百分数折扣提供帐单信息时,可以考虑采用逐步或是逐渐跟随一条曲线的中间值。

当然还可以按不同方式来检测处理复杂性,可能不够精确,并且要考虑按其它方式检测的成本,例如是需要按各种校对为起价增加一个固定增值。特别是在使用检测装置 6 检测人工处理费用时有可能直接将原始文本也就是由语音识别装置 4 转换的文本文件与最终提供的文本文件相比较,以便根据两个文件的差别做出对处理所需工作量的决定。在这种简单的情况下,可以省略对口述文件收听范围的检测,而图 1 的检测装置 16 可以用做比较文本文件的比较装置,这样就能检测出文本文件之间的差别,例如是对不同字符计数,将最终文本中增加的字符和省去的字符与初始文本相比较。同样可以按照最佳文本的百分数来检测偏差的范围。

本发明是参照口述的转录来描述的。然而,对于由起始数据或信息开始的其它自动数据或信息产生系统显然也能采取类似的方式;例如是按第一语言产生的输入文本文件到用第二语言表达的一个输出文本文件的自动翻译,然后按类似的方式检查错误,并通过插入,删除或重写字

符或文字来进行纠正,同样能检测为此所需的费用。

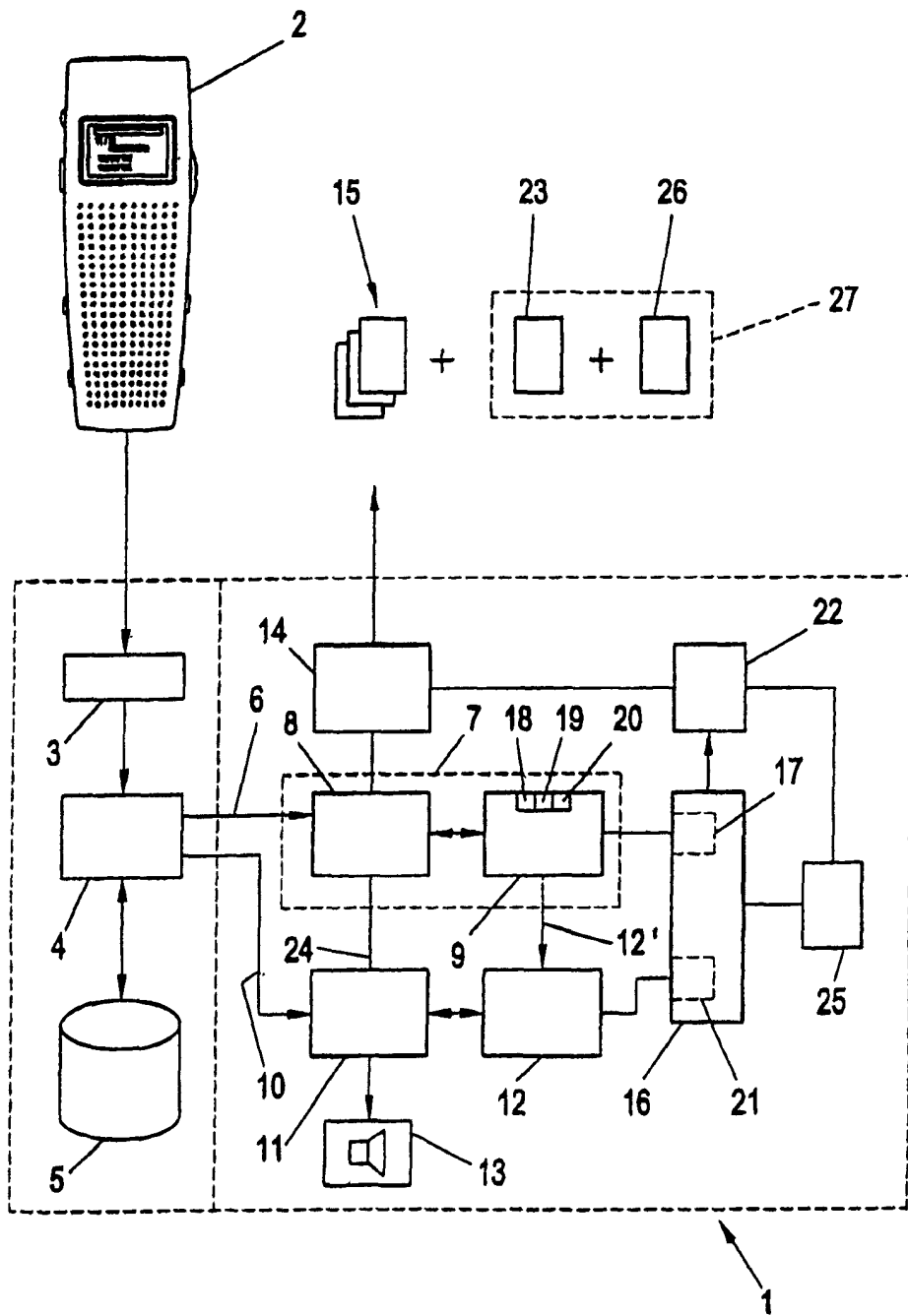


图 1

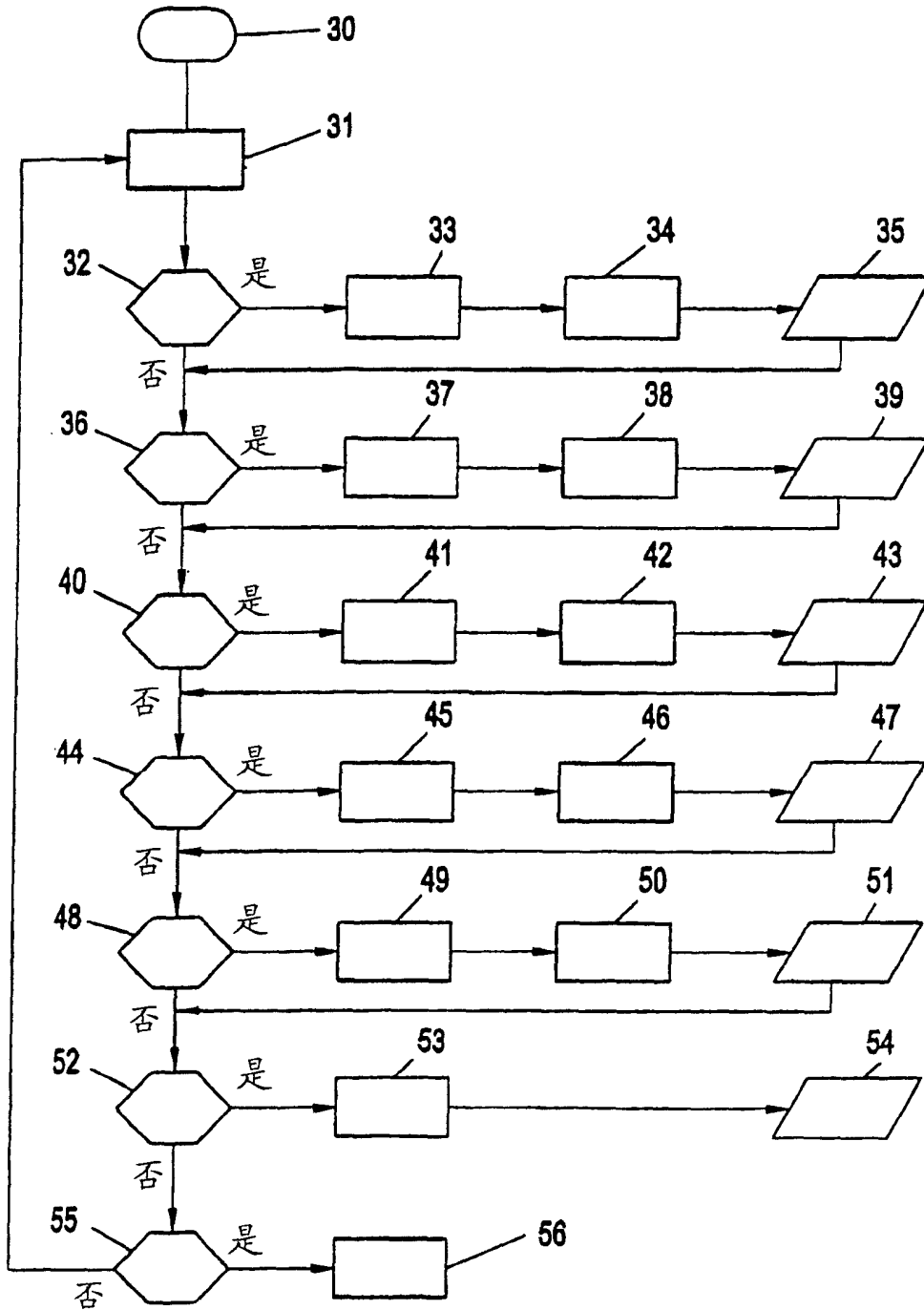


图 2

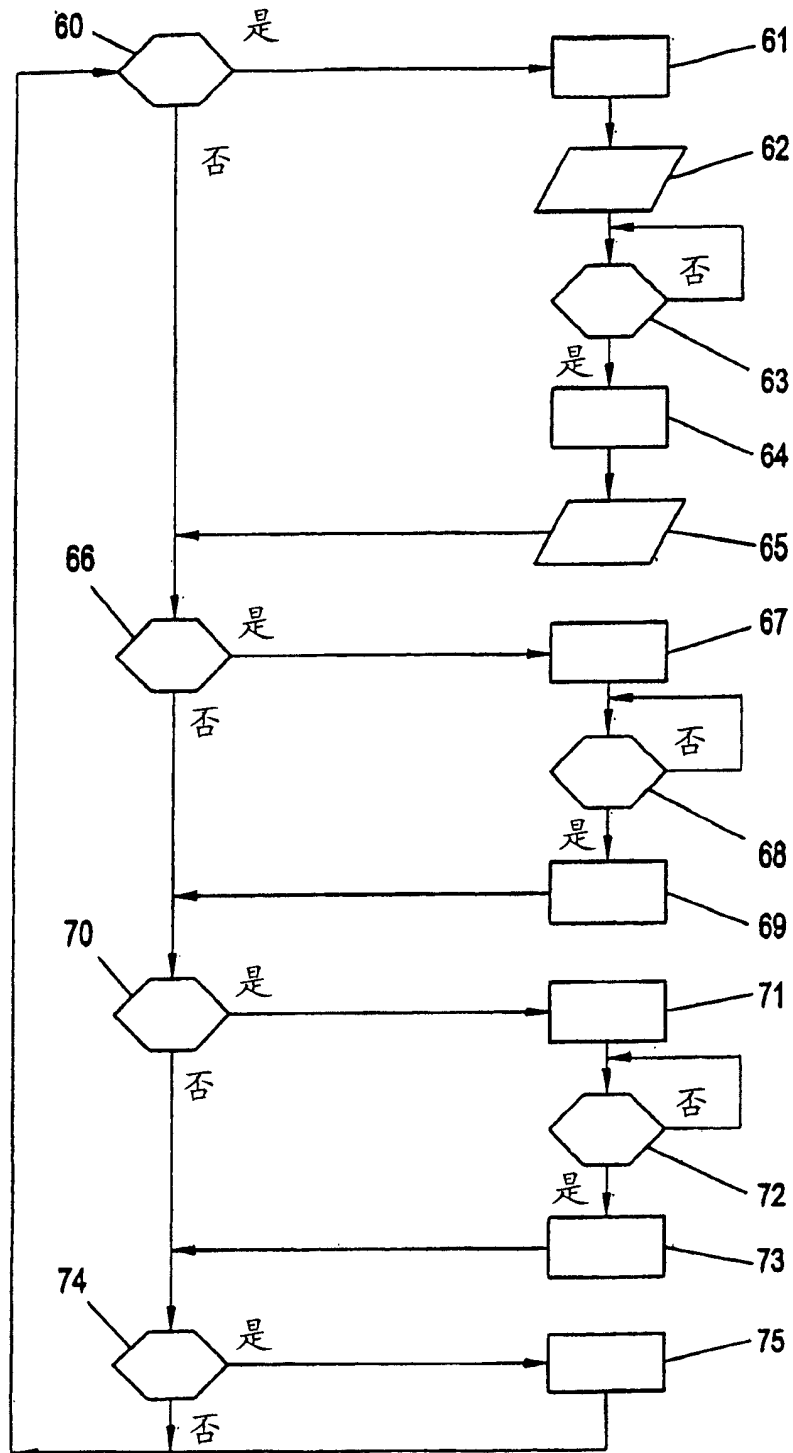


图 3