



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 99808603.7

[45] 授权公告日 2004 年 2 月 11 日

[11] 授权公告号 CN 1138252C

[22] 申请日 1999.5.3 [21] 申请号 99808603.7

[30] 优先权

[32] 1998. 5. 15 [33] DE [31] 19821900.8

[86] 国际申请 PCT/DE99/01322 1999.5.3

[87] 国际公布 WO99/60560 德 1999.11.25

[85] 进入国家阶段日期 2001.1.15

[71] 专利权人 西门子公司

地址 德国慕尼黑

[72] 发明人 B·凯梅雷尔

审查员 刘红梅

[74] 专利代理机构 中国专利代理(香港)有限公司

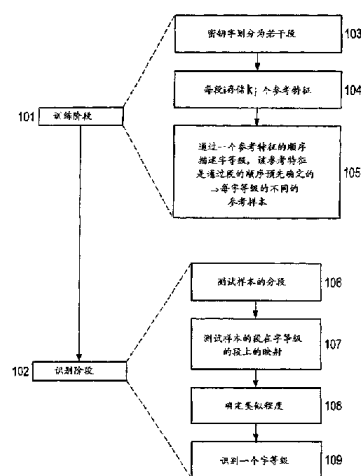
代理人 郑立柱 王忠忠

权利要求书 2 页 说明书 7 页 附图 4 页

[54] 发明名称 通过计算机在交谈语音中识别至少一个密钥字的方法和设备

[57] 摘要

为了在交谈语音识别一个密钥字,该密钥字,同样象一个必须识别的测试样本一样,划分为若干段。通过形成一个间隔度,关于这些段积累这个间隔度,密钥字和测试样本的各个段互相映射。在一个训练阶段多次存储密钥字,其中为密钥字的每个段分别确定并存储多个参考特征。在识别中实现同测试样本的附属段最好相配的参考特征的段方式分配。



1. 通过一个计算机在交谈语音识别至少一个密钥字的方法，
 - a) 在该方法中密钥字划分为若干段，其中分配给每一个段一定数量的参考特征，
 - 5 b) 存储所述密钥字的多个代表，以及存储不同代表的参考特征；
 - c) 在该方法中，在交谈语音中的测试样本划分为若干段，其中给测试样本的每个段分配由密钥字的相应段的一定数量的参考特征形成的类似于该段的参考特征，
 - d) 在该方法中测试样本作为密钥字识别，为如此密钥字确定一个
- 10 类似程度，对于密钥字的参考特征对测试样本的累积的段方式分配来说该类似程度低于一个预先确定的界限。
2. 按照权利要求 1 的方法，其中，测试样本是一个独立的声音单元。
3. 按照权利要求 1 的方法，其中，测试样本是一个字。
- 15 4. 按照权利要求 1 至 3 之一的方法，其中，对于密钥字和对于测试样本，段的数目总是相同的。
5. 按照权利要求 1 至 3 之一的方法，其中，测试样本与多个密钥字进行比较并且输出最类似的密钥字。
6. 按照权利要求 1 至 3 之一的方法，其中为了存储密钥字以及
- 20 测试样本的语音使用了特征矢量，其中在预先确定的取样时刻对语音数字化，并且以表征语音的数据存储每一个特征矢量。
7. 按照权利要求 5 的方法，其中为每个段存储一个对这个段的所有特征矢量取平均值的特征矢量，并且作为用于这个段的特征矢量继续使用。
- 25 8. 按照权利要求 1 至 3 之一的方法，其中如此确定预先确定的界限，即始终识别一个密钥字。
9. 在交谈语音识别至少一个密钥字的设备，在该设备中预先规定一个处理器单元，如此设计这个处理器单元，
 - a) 密钥字可以化为若干段，其中可以分配给每个段一定数量的参
 - 30 考特征，
 - b) 存储所述密钥字的多个代表，以及存储不同代表的参考特征；
 - c) 在交谈中的测试样本可以划分为若干段，其中可以给测试样本

的每个段分配由用于密钥字的相应段的一定数量的参考特征形成的类似于该段的参考特征，

- d) 如果对于密钥字的参考特征对测试样本的累积的段方式分配来说类似程度低于一个预先确定的界限，则测试样本可以作为密钥字识别。
- 5

通过计算机在交谈语音中识别
至少一个密钥字的方法和设备

5 技术领域

本发明涉及一个通过计算机在交谈语音识别至少一个密钥字的方法和设备。

背景技术

识别交谈语音的方法和设备公开于[1]。在那里也基本介绍了一部分用于语音识别以及重要的、在语音识别中常用技术的设备或者方法。

一个密钥字是一个确定的字，应当由一个语音识别设备以交谈的语音识别该密钥字。大多数的一个预先确定的事件结合如此的一个密钥字，也就是说在识别这个密钥字之后执行这个事件。

在[2]中也说明了识别交谈语音的方法和设备。由此表明，通过特别实施时间轴的压缩或者延伸，一个借助于隐藏的-马科夫（Markov）模型的仿真允许匹配于送话器速度的变化，并且因此实现在识别时预先给定的语音结构单元动态匹配于交换语音。这相当于一个动态匹配（也就是：动态的程序设计），例如通过维特比（Viterbi）算法保证这个匹配。

20 例如通过确定在特征矢量之间的一个（多维）间隔确定在声音或者声音串之间的间隔（时间间隔），这些特征矢量以数字的形式描述语音的声音。这个间隔是在声音或声音串之间类似程度一个实例。

发明内容

25 本发明的任务在于给出一个识别密钥字的方法和设备，其中这个识别是最强的并且对干扰不敏感。

根据本发明的特征解决了这个任务。

本发明的通过一个计算机在交谈语音识别至少一个密钥字的方法，

- a) 在该方法中密钥字划分为若干段，其中分配给每一个段一定数量的参考特征，
- 30 b) 存储所述密钥字的多个代表，以及存储不同代表的参考特征；
- c) 在该方法中，在交谈语音中的测试样本划分为若干段，其中给测试样本的每个段分配由密钥字的相应段的一定数量的参考特征形

成的类似于该段的参考特征，

d) 在该方法中测试样本作为密钥字识别，为如此密钥字确定一个类似程度，对于密钥字的参考特征对测试样本的累积的段方式分配来说该类似程度低于一个预先确定的界限。

- 5 如果对于密钥字的参考特征对测试样本的积累的段方式的分配来说，类似程度低于一个预先确定的界限，则测试样本识别为密钥字。如果该类似程度不低于一个预先确定的界限，则测试样本不识别为密钥字。对此，低的类程度表明密钥字的参考特征同测试样本的一个尽可能好的一致。

- 10 下面简短讨论不同的概念及其意义：

测试样本是在交谈语音中包含的样本，该样本与密钥字进行比较，并且也许可能识别为密钥字。类似程度表明测试样本与密钥字或者一部分测试样本与一部分密钥字的一致程度。段是测试样本或者密钥字的间隔，该间隔具有一个预先确定的持续时间。参考特征是密钥字的针对段的子特征。参考样本含有表明密钥字的表达形式的参考特征，
15 字等级含有全部的参考样本，可以通过参考特征的不同组合产生全部参考样本，其中对于密钥字来说每段存储了特别的多个参考特征。在训练阶段，为各自密钥字的参考特征确定并存储代表，而在识别阶段进行测试样本与密钥字的可能的参考样本的比较。

- 20 在训练阶段，主要为参考特征存储一个预先确定数量 M 个代表。如果提供比参考特征多的位置作为自由位置 M 使用，则例如以一个浮动的平均值的形式可以实现参考特征的取平均，以便因此在代表中考虑附加的参考特征的信息。

本发明的一个继续发展在于，测试样本（和/或密钥字）是一个已
25 存储的声音单位，特别是一个字。测试样本和/或密钥字也可以是每一个音位、一个复音、一个通常由多个音位组合的声音或者一定数量的字。

一个另外的继续发展在于，对于密钥字和对于测试样本来说段的数目总是相同的。

- 30 在一个附加的继续发展的范围内，测试样本与多个密钥字比较，并且得出类似于测试样本的密钥字。这相当于一个单字识别的系统，其中多个密钥字说明在交谈语音中必须识别的单字。分别得出密钥字，

其最好地相配于在交谈语音中包含的测试样本。

这也是一个继续发展，即特征矢量应用于密钥字以及测试样本的存储，其中在预先确定的取样时刻语音被数字化，并且以表明语音的数据存储每一个特征矢量。在一个预处理的范围内进行语音信号的数字化。主要是全部的 10ms 从语音信号中确定一个特征矢量。

一个另外的继续发展在于，为每个段存储一个对该段的所有特征矢量取平均值的特征矢量，并且作为对于该段表征的特征矢量继续使用。例如全部 10ms 产生的数字化的语音数据主要预处理成为具有 25ms 的时间扩展的重叠的时间窗口。为此可以使用一个线性预测编码(LPC)分析、一个频谱分析或一个倒频谱分析。对于每个 10ms 间隔提供具有 n 个系数的特征矢量作为各自分析的结果使用。主要对一个段的特征矢量取平均，因此每段提供一个特征矢量使用。在识别密钥字的训练的范围内，来自交谈语音的若干源的每段存储多个不同的参考特征，因此提供多个已取平均的参考特征（密钥字的特征矢量）使用。

此外给出一个用于在交谈语音识别至少一个密钥字的设备，在该设备中预先规定一个处理器单元，如此设计这个处理器单元，

a) 密钥字可以化为若干段，其中可以分配给每个段一定数量的参考特征，

b) 存储所述密钥字的多个代表，以及存储不同代表的参考特征；

c) 在交谈中的测试样本可以划分为若干段，其中可以给测试样本的每个段分配由用于密钥字的相应段的一定数量的参考特征形成的类似于该段的参考特征，

d) 如果对于密钥字的参考特征对测试样本的累积的段方式分配来说类似程度低于一个预先确定的界限，则测试样本可以作为密钥字识别。—如果类似程度不低于预先确定的界限，则不识别密钥字。

后文中还给出了本发明的继续发展。

该设备特别适合于实施根据本发明的方法或实施一个其前面阐述的继续发展。

发明说明

根据下面的图详细描述本发明的实施例。

图示：

图 1 一个简图，描述了在交谈语音识别至少一个密钥字的方法的步骤

骤,

图 2 一个具有二种可能的识别密钥字的实施例的简图,

图 3 一个简图, 其阐明测试样本在密钥字上的映射和类似程度的确定,

5 图 4 在交谈语音识别一个密钥字的设备。

具体实施方式

图 1 指出了—个简图, 其描述了在交谈语音识别至少一个密钥字的方法的步骤。

10 在 25ms 持续时间的重叠的时间窗口内首先所有 10ms 数字化交谈语音, 并且也许可以预处理(滤波)。为此或者应用一个 LPC 分析、一个频谱分析或者应用一个倒频谱分析。在每个 10ms 间隔内提供一个具有 n 个系数的特征矢量作为预处理的结果使用。

根据在字之间依据间隔能量或间隔频谱断定的间隔确定交谈语音的—主要是字—各个成分。按这种方式辨认在交谈语音内的一个单独的字。

15 在图 1 中粗略区别两个组成部分, 一个训练阶段 101 和一个识别阶段 102。不仅在训练阶段 101 而且也在识别阶段 102 断定的字是一个密钥字或是一个测试样本, 划分为一个预先确定的数目段。首先对段的特征矢量取平均值。具有已取平均的特征矢量的段的顺序提供一个字样本。

20 一个在交谈语音中识别的字(密钥字或测试样本)分别被划分为预先确定数目的段。对在一个段的内部的多个特征矢量取平均值, 其中这个已取平均值的特征矢量全部说明这个字(字样本)。存储一个密钥字用于此后的识别, 对此存储这个密钥字的多个代表。特别有用地表明, 多次录取多个送话器的密钥字, 并且分别存储最好描述密钥字的录取。对此每段以各一个已取平均值的特征矢量的形式存储最好的录取。25 因而得出涉及密钥字的各自段的参考特征的预先确定的数目。根据以这种方式存储的参考特征, 由通过段的顺序确定的次序形成的字可以组合成不同的参考样本。对此密钥字的不同代表的参考特征组合成为一个参考样本。因此得出参考样本的多种可能性作为密钥字的原始代表存储。30 在这个接着训练阶段 101 的识别阶段 102 中最近的参考特征(涉及段)分别被分配给测试样本的相应段。

训练阶段 101 包含密钥字划分为预先确定数目的段(参见方框

103)。在步骤 104 中为每段 i 存储 k_i 个参考特征, 其中 k 表明为密钥字确定的代表的数目。在步骤 105 中通过参考特征的顺序描述字等级, 通过段的顺序预先确定该参考特征。通过参考特征同参考样本的不用组合说明字的等级, 该字的等级以所有存储的变化说明密钥字, 该组合描述了字等级的参考样本层次。

在识别阶段 102 中确定, 是否可以把一个称作测试样本的字分配给密钥字。为此根据上面的论述在步骤 106 中把测试样本分段。在步骤 107 中测试样本的段映射到密钥字的段上, 其中给测试样本的每个段分别分配密钥字的最类似的参考特征。对所有的段实施这个分配, 为每个段计算的类似程度积累成为一个总的类似程度 (参见步骤 108)。如果积累的类似程度的值低于一个预先确定的界限, 则该类似程度满足测试样本和密钥字的高度类似, 测试样本识别为密钥字 (参见步骤 109)。

特别根据一个间隔确定这个类似。如果两个样本类似, 则这两个样本彼此有一个较小的间隔, 相应地特征矢量的差值是低的。为各自的段确定的类似程度因此按照特征矢量的间隔, 再者对于一个在段到段的映射中执行的映射误差来说也满足这样的间隔。类似程度的累积相当于段方式引起的映射误差的相加, 总的类似程度因而是一个在测试样本分配给密钥字的情况下总共引起的误差的值。由于特别多个测试样本应当识别为一个密钥字, 所以测试样本被映射在多个密钥字上, 其中分别段方式地确定类似程度, 并且为对每个密钥字的每个分配计算一个累积的类似程度。识别这一个密钥字, 在该密钥字中累积的总类似程度具有对多个密钥字的所有分配的最小值。

图 2 指出了具有两种可能的识别密钥字的实施例的简图。在图 1 的情况下描述的、对于每个密钥字来说类似程度的确定 (参阅步骤 201) 导致最类似的密钥字的识别或者得出 (参阅步骤 202 或者对图 1 的说明)。如果最好的累积类似程度、也就是在测试样本映射到各自密钥字上的情况下最低的误差, 高于一个预先确定的界限, 则第二实施例不识别字等级或不得出字的等级。在如此情况下一个分配、也就是测试样本在一个密钥字上的映射是如此差, 以此为出发点, 即测试样本不适密钥字。中止对最好相配的密钥字的被迫分配, 该字始终是差的, 并且几乎不相配。

在图 3 中描述了一个简图, 该简图阐明了测试样本在密钥字上的映

射和类似程度的确定。

一个测试样本 TEM 的五个段 Sg_iT ($i=1, 2 \dots 5$) 示范地映射在一个参考样本 RMU 的五个段 SG_iS 上。可以通过参考特征 RM_i 的不同组合描述必须识别的密钥字 (字等级), 通过段的次序决定参考特征。正如上面说明的, 参考特征确定为密钥字的段的特别好的代表 (训练阶段)。

开始时把测试样本的第一段 $SG1T$ 分配给密钥字的第一段 $SG1S$ 。对此描述测试样本 TEM 的第一段 $SG1T$ 的、已取平均值特征矢量映射到参考特征 $RM1$ 和 $RM2$ 的最好一个上。接着测试样本 TEM 的第二段 $SG2T$ 映射到密钥字的一个最近的参考特征上。对此三个不同的途径 $W1$ 、 $W2$ 和 $W3$ 是可能的。途径 $W1$ 向右增加 0 是一个段, 途径 $W2$ 向右增加 1 是一个段, 并且途径 $W3$ 向右增加 2 是一个段。也就是确定最好的类似程度, 其中测试样本 TEM 的第二段 $SG2T$ 与参考特征 $RM1$ 、 $RM2$ (对于途径 $W1$)、 $RM3$ 、 $RM4$ 、 $RM5$ (对于途径 $W2$) 和 $RM1$ 、 $RM2$ (对于途径 $W3$) 进行比较, 并且确定最类似的。相应地, 依赖于在从第一到第二段的转变过程中, 也在从第二到第三段的转变过程中采取的途径继续进行。

示范地对于第一段 $SG1S$ 和第三段 $SG3S$ 来说体现密钥字的参考特征是相同的, 因为各自段描述相同的语音。为了不浪费不必要的存储位置, 分别多次存储参考特征 RM_i 并且逐个存储在每个密钥字中, 则编制一个表, 该表包含参考特征 (参阅表)。就密钥字的段来说因此在表 TABELLE 的存储范围内仅仅保存指针, 其中指针介绍参考特征的各自数据。一个指针 (也就是在一个表内的偏移) 的存储位置需求比属于各自参考特征的数据明显低。

图 4 指出了用于语音识别至少一个密钥字的设备。根据预处理 402 从交谈语音 401 中确定特征矢量 (参见 [2])。接着实施一个字开始 / 字结束确定, 并且已鉴别的字划分为 N 个段 (参见方框 404)。在训练阶段 (通过连接 405 确定) 期间在一个步骤 406 中这些段存储在一个数据库 407 中, 其中特别是按照一个浮动的平均值, 一个段的多于 M 个的代表导致对在数据库 407 中的代表取平均。为此一个段的代表经过一个连接 408 被供给取平均的过程 (也就是聚合)。在一个识别阶段 (通过连接 409 说明), 随着最佳相配于一个测试样本的段选择进行非线性的映射, 其中为每个段从数据库 407 中确定一个代表 (参见方框 410)。接着在一个方框 411 中实现分级, 并且得出已识别的字 (参见

方框 411)。

在该文献的范围内引用了如下的出版物：

5 [1]A. Hauenstein: “用于自动语音识别的一个处理器的算法和方案的最佳化”，慕尼黑工程学院，集成电路讲座，论文，19.07.1993，第二章，13 至 26 页。

[2]N. Haberland, 及其他人的: “语音教学 - 如何在以计算机为基础的语音识别中起作用? ”, C' t 5/98 , Heinz Heise 出版社, 汉诺威 1998, 120 至 125 页。

10

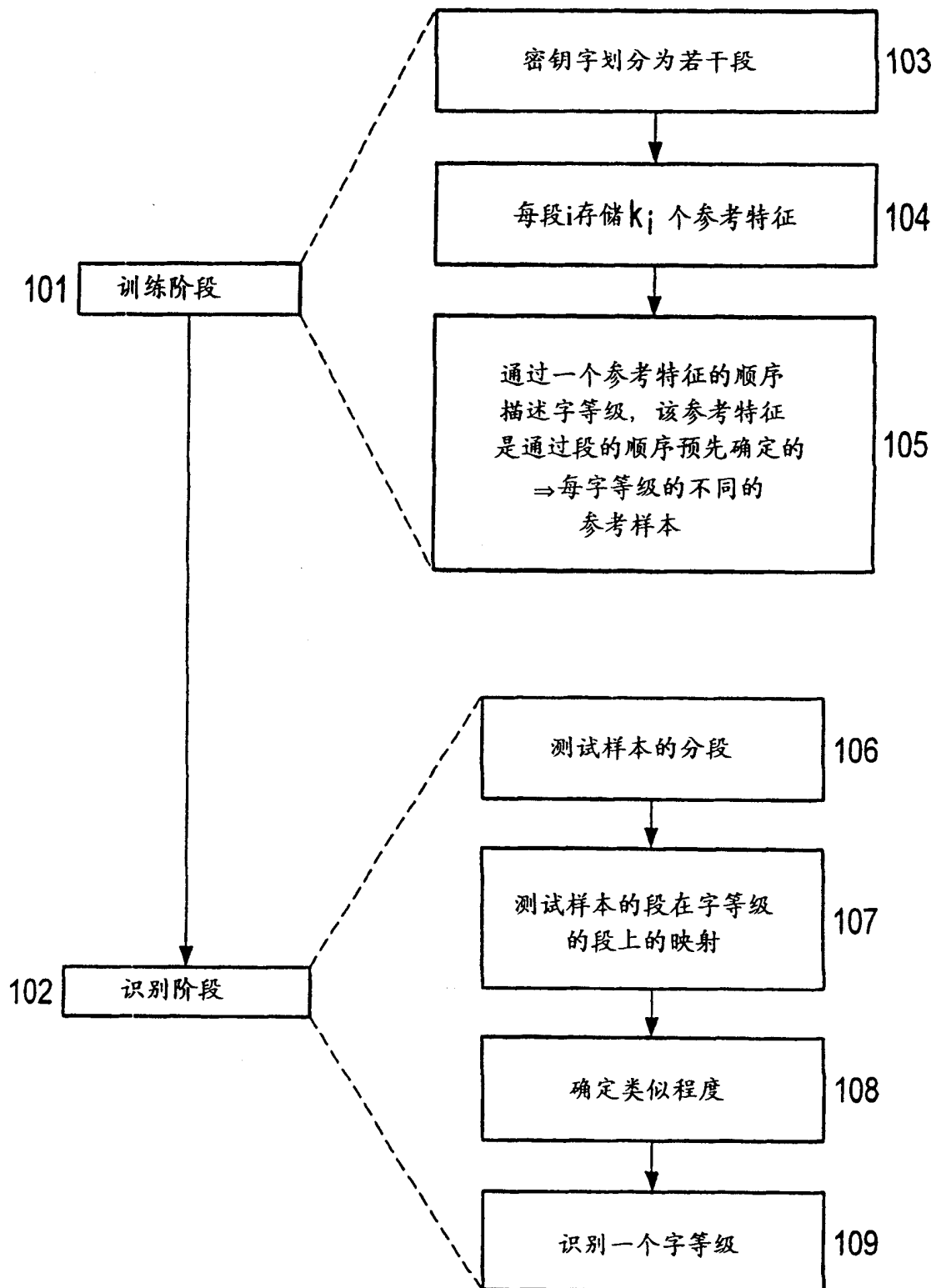


图 1

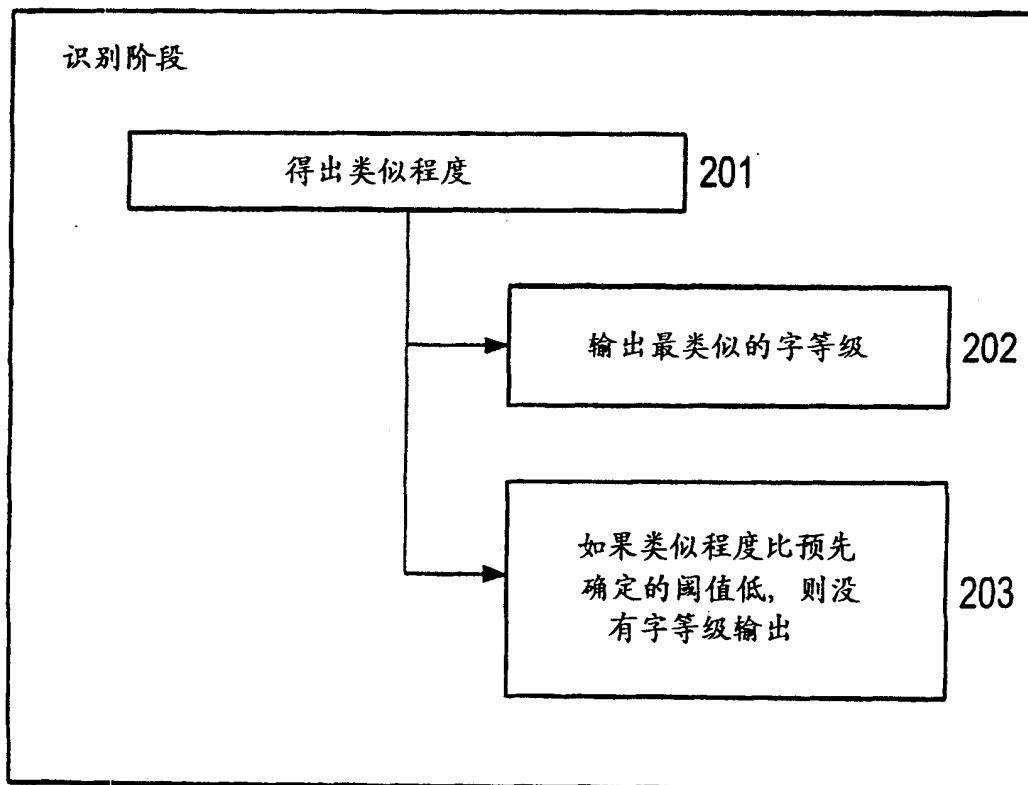
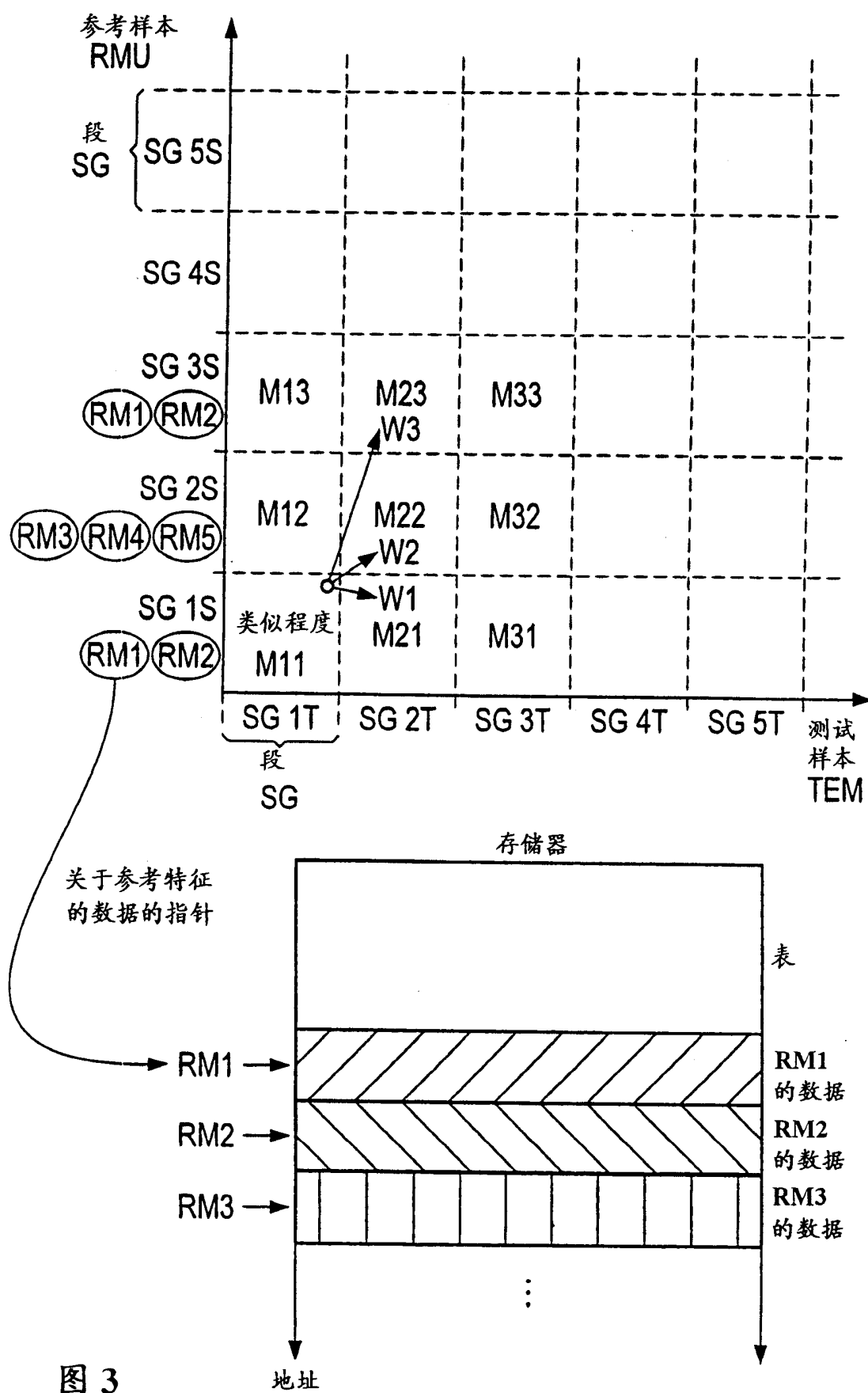


图 2



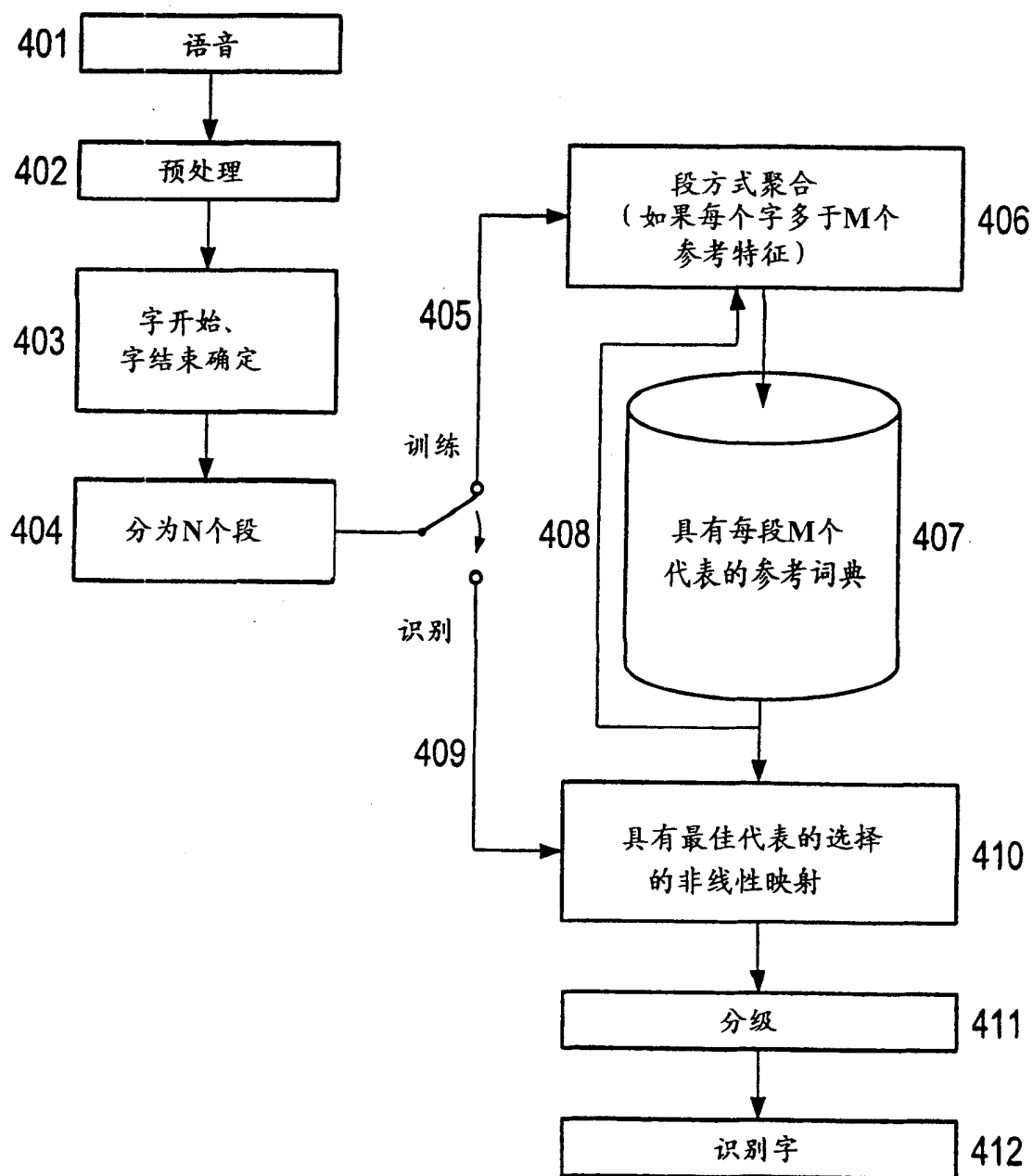


图 4