



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 97114917.8

[45] 授权公告日 2003 年 6 月 25 日

[11] 授权公告号 CN 1112669C

[22] 申请日 1997.4.30 [21] 申请号 97114917.8
[30] 优先权
[32] 1996. 5. 1 [33] US [31] 655273
[71] 专利权人 微软公司
地址 美国华盛顿
[72] 发明人 黄学东 米林德·V·马哈简
审查员 杨 叁

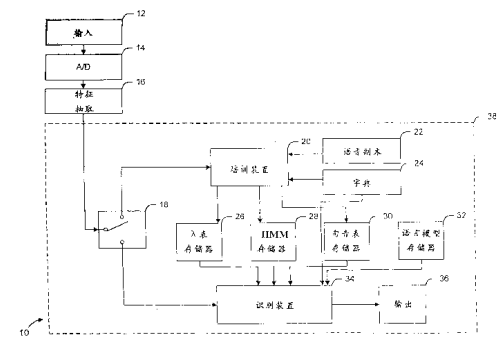
[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所
代理人 于 静

权利要求书 3 页 说明书 19 页 附图 8 页

[54] 发明名称 采用连续密度隐藏式马尔克夫模型的
语音识别方法和系统

[57] 摘要

本发明提供了用于改进语音识别系统的识别准确度的一种方法和系统，其利用连续密度隐藏式马尔克夫模型表示在所说的语言中存在的语音的音素单位。反映语音与一个模型化的语言表达式匹配的似然性的一个声评分依赖于与隐藏式马尔克夫模型的状态相关的输出概率。对于每个音素单位产生与前后音有关的和与前后音无关的连续密度隐藏式马尔克夫模型。通过根据一个加权因子将与前后音有关状态的输出概率和与前后音无关状态的输出概率加权确定与一个状态相关的输出概率。



I S S N 1 0 0 8 - 4 2 7 4

1、一种在计算机系统中使输入语音与语言表达式匹配的方法，该方法包括以下步骤：

对于语音的许多音素单位中的每一个，提供一组较为细致的声模型和一个不大细致的声模型表示该音素单位，较为细致的声模型所模拟的单个声概率比不太细致的声模型所模拟的声概率多，每个声模型具有一组状态，其后是一组变换，每种状态表示在某一时间点出现在该音素单位中的语音的一部分，并具有一个输出概率，表示输入语音的一部分在某一时间点出现在该音素单位中的似然性；

对于所选择的较为细致的声模型序列中的每一个，确定输入语音与这一序列匹配的接近程度，所说的匹配还包括以下步骤：

对于所选择的一序列较为细致的声模型的每种状态，确定一个累计的输出概率，作为该状态和代表相同音素单位的不大细致的声模型的相同状态的输出概率的组合；以及

确定与输入语音匹配最好的一个序列，这一序列代表语言表达式。

2、如权利要求1所述的一种方法，其特征在于：每个声模型都是一个连续密度的隐藏式马尔科夫模型。

3、如权利要求1所述的一种方法，其特征在于：确定输出概率的步骤还包括将不大细致模型和较为细致模型的输出概率加权的步骤，它们在组合时具有不同的加权因子。

4、如权利要求1所述的一种方法，其特征在于：提供一组较为细致的声模型的步骤还包括利用一定数量的语音培训数据培训每个声模型的步骤；和

确定输出概率的步骤还包括将不大细致模型和较为细致模型的输出概率相对于用于培训每个声模型的培训数据的数量加权的步骤。

5、根据权利要求1的方法，其特征在于提供较为细致的声模型包括：产生一组与前后音有关的句音，以及其中所说提供不大细致的声模型

包括:

产生与表示所说语言表达式的相同位置的一组与前后音有关的句音相关的一个与前后音无关的句音, 以及其中所说的匹配包括:

对于所说输入语音的每一个特征向量, 确定所说特征向量与在所说语言表达式中的与前后音有关的句音匹配的输出生率, 所说语言表达式与所说特征向量出现在同一时间间隔内, 所说输出概率的确定利用了与前后音有关的句音相关的与前后音无关的句音; 和

利用所说输出概率确定所说输入语音与所说语言表达式匹配的似然性。

6、如权利要求5所述的一种方法, 其特征在于: 所说输出概率包括一个连续密度概率函数。

7、如权利要求5所述的一种方法, 其特征在于: 产生一组与前后音有关的句音的步骤还包括利用表示语音的一定数量的培训数据培训所说与前后音有关的句音的步骤;

产生与前后音无关的句音的步骤还包括利用所说的一定数量的培训数据培训所说与前后音无关的句音的步骤; 和

确定输出概率的步骤还包括根据用于培训所说句音的培训数据的数量组合与前后音无关的和与前后音有关的句音的步骤。

8、如权利要求5所述的一种方法, 其特征在于: 产生一组与前后音有关的句音的步骤还包括以下步骤:

利用代表语音的一定数量的培训数据培训所说与前后音有关的句音的步骤;

对于每一个与前后音有关的句音产生一个表示用于估算所说句音的培训数据的数量的加权因子; 和

确定输出概率的步骤还包括根据所说加权因子组合所说的与前后音有关的句音和与前后音无关的句音的步骤。

9、如权利要求8所述的一种方法, 其特征在于: 产生一个加权因子的步骤还包括根据培训数据的数量利用一种删除插入技术产生所说加权因子的步骤。

10、如权利要求 8 所述的一种方法，其特征在于：产生一个加权因子的步骤还包括以下步骤：

产生所说培训数据的代表参数；和

对培训数据的数量的代表参数应用一种删除插入技术产生所说的加权因子。

11、如权利要求 8 所述的一种方法，其特征在于：产生一个加权因子的步骤还包括以下步骤：

产生所说培训数据的一个代表参数；

根据培训数据的代表参数产生一组数据点，所说的数据点表示所说的培训数据；和

通过向所说数据点应用删除内插技术产生所说的加权因子。

采用连续密度隐藏式马尔克夫 模型的语音识别方法

技术领域

本发明涉及计算机语音识别，更具体地说，本发明涉及采用连续隐藏式马尔克夫模型的计算机语音识别系统。

背景技术

语音识别领域正面临一种需要提供一种具有最小识别误差率的、与讲话者无关的连续语音识别系统的挑战。实现这个目标的关键在于语音识别系统所采用的识别算法。识别算法基本上就是语音信号，一种连续时间信号，向表示先前从培训数据中得到的语音的音素和音韵描述的一组参考模式的对应变换。为了实现这种对应变换，对于语音信号的数字形式采用信号处理技术，诸如快速富里叶变换（FFT）、线性预测编码（LPC）、或滤波器组（Fitter banks）进行处理，以获取这些语音信号的适合代表参数。一种共用的代表是在每个时间间隔包含表示包含在语音信号中的频带和/或能带的 FFT 或 LPC 系数的特征向量。一序列这种特征向量被对应变换到用于识别包含在语音信号中的语言单位、单词和/或句子的一组参考模式。

通常，语音信号不会与存储的参考模式精确匹配。实现精确匹配的困难是由于语音信号特征的极大的变化性使之不能利用存储的参考模式完全地和准确地获取。概率模型和统计技术比那些试图实现精确匹配的技术更加成功地应用于预测预期的信息。这类技术中有一种称为隐藏式马尔克夫模型（HMM）。这些技术更适合于语音识别，因为它们确定最有可能与语音信号匹配的参考模式，而不是寻找一种精确的匹配。

一个隐藏式马尔克夫模型由一序列与变换相连的状态构成。一个隐藏式马尔克夫模型可以表示语音的一个特定的音素单位，诸如一个音素或字词。与每个状态相关的是一个指示该状态与一个特征向量匹配的似然性的输出概率。对于每种变换，存在指示遵循变换的似然性的一个相关的变换概率。这种变换和输出概率是从先前的讲话语音模式，称之为“培训数据”，统计估算出来的。识别问题就是一个寻找具有与表示输入语音信号的特征向量匹配的最高概率的状态序列的问题。首先，这个搜索过程包括枚举每

一种已经模型化的可能的状态序列和确定该状态序列与输入语音信号匹配的概率。对应于具有最高概率的状态序列的语调被选择为识别的语调。

大部分基于隐藏式马尔科夫模型的语音识别系统是以向量量化的离散的隐藏式马尔科夫模型为基础的。离散的隐藏式马尔科夫模型具有有限组的输出符号，变换和输出概率是建立在离散的概率分布函数(pdf)基础上的。向量量化用于借助于被称为代码字的一种离散的代表参数表征连续语音信号。特征向量与使用失真量值的一个代码字匹配。该特征向量用具有最小失真量值的代码字的索引代替。识别问题转换降低为计算一个所关注的语音信号的离散的输出概率，就象一种查表运算，这只需要最少的计算量。

但是，语音信号是连续信号。尽管有可能利用代码字量化连续信号，但是这种量化可能会产生严重的音质降低，结果导致识别准确性很差。利用连续密度隐藏式马尔科夫模型的识别系统不会产生由于量化失真造成的不准确的问题。连续密度隐藏式马尔科夫模型能够直接利用估算连续密度概率分布函数对连续语音信号模型化，从而达到较高的识别准确率。但是，连续密度隐藏式马尔科夫模型需要大量的培训数据和需要较长时间的识别运算，这妨碍了它们在大部分商业语音识别系统中的应用。所以，存在于连续语音识别系统中的主要问题是采用连续密度隐藏式马尔科夫模型以实现较高的识别准确率。

发明内容

本发明提供了一种在计算机系统中使输入语音与语言表达式匹配的方法，该方法包括以下步骤：对于语音的许多音素单位中的每一个，提供一组较为细致的声模型和一个不大细致的声模型表示该音素单位，较为细致的声模型所模拟的单个声概率以不太细致的声模型所模拟的声概率多，每个声模型具有一组状态，其后是一组变换，每种状态表示在某一时间点出现在该音素单位中的语音的一部分，并具有一个输出概率，表示输入语音的一部分在某一时间点出现在该音素单位中的似然性；对于所选择的较为细致的声模型序列中的每一个，确定输入语音与这一序列匹配的接近程度，所说的匹配还包括以下步骤：对于所选择的一序列较为细致的声模型的每种状态，确定一个累计的输出概率，作为该状态和代表相同音素单位的不大细致的声模型的相同状态的输出概率的组合；以及确定与输入语音匹配最好的一个序列，这一序列代表语言表达式。

本发明涉及一种语音识别系统，这种系统改进了用对应于一种语言表

达式的连续密度隐藏式马尔克夫模型对语音信号的模型化。在优选实施例中，识别系统利用一个与前后音无关的和几个与前后音有关的隐藏式马尔克夫模型以不同的前后相关模式表示一种语音的音素单位。这些隐藏式马尔克夫模型中每一个的输出和变换概率利用培训数据估算。与对应于相同的模型化的音素的相同状态相关的输出概率被群聚形成各句音（senone）。对于每个表示预测未发生的数据的输出概率的稳定性的与前后音相关的各句音还产生一个加权因子。在该优选实施例中，通过在培训数据的所有数据点的删除内插估算加权因子。除了上述方式，还可以从数据点的代表参数或者从数据点的代表参数所产生的随机生成数据点估算加权因子。

识别装置接收一个输入的语音并产生最有可能与输入的语音的特征向量匹配的候选字序列。该字序列可以由对应于隐藏式马尔克夫模型的状态序列的各种句音排列构成。该识别装置利用一个声音和语言概率评分确定哪一个句音/状态排列与特征向量最匹配。声概率评分表示句音排列对应于特征向量的似然性，语言概率评分表示语音与出现在语言中的句音排列相对应的似然性。声概率评分基于输出和变换概率分析。输出概率分析通过将每个输出概率作为加权因子的函数加权而利用与前后音相关的和与前后音无关的句音的输出概率。具有最稳定估算的输出概率将支配分析过程，从而改进输出概率分析。输出概率分析的改进改善了声评分，进而改善了整个识别的准确率。

附图说明

本发明的上述和其它的特征以及优点通过以下对如附图所示的、本发明的优选实施例的更加详细的描述将变得十分清楚，在不同的附图中相同的参照标号表示相同的单元。附图不是按比例绘制的，重点在于解释本发明的原理。

图1为应用于优选实施例的一个语音识别系统的方框图。

图2为用于图1所示的系统的培训方法的流程图。

图3为计算用于图1所示系统中所用的加权因子的方法的流程图。

图4为计算用于图3所示系统中所用的 λ 的新值的优选实施例的流程图。

图5为计算用于图3所示系统中所用的 λ 的新值的第一变型实施例的流程图。

图6为计算用于图3所示系统中所用的 λ 的新值的第二变型实施例的流程图。

图 7A 和图 7B 表示隐藏式马尔克夫模型的一个实例和与一个音素相关的句音结构。

图 8 为用于图 1 所示系统的语音识别方法的流程图。

具体实施方式

通过本发明的优选实施例认识到通过对表示同一音素单位的不同输出概率针对每一输出概率能够预测未形成的数据的程度进行加权处理能够在采用连续密度隐藏式马尔克夫模型的语音识别系统中得到提高的识别准确率。权利要求中提出保护的发明中的语音识别系统接收以连续信号形式构成的输入语音，并产生相应于该语音的最可能的语言表达式。该优选实施例通过使构成语音信号的代表参数的一组特征向量与识别可能的语言表达式的一序列隐藏式马尔克夫模型相匹配而识别一个语言表达式。一个隐藏式马尔克夫模型可以表示一个音素，一序列隐藏式马尔克夫模型可以表示由许多音素构成的单词或句子。

由于连续密度概率分布函数，例如高斯概率分布函数的混合形式，在模型化一个语音信号时更加准确，所以可以利用它们表示一个状态的输出概率。这个输出概率函数由培训数据统计估算。通常为了准确地估算输出概率函数培训数据的数量不足。为了解决这个问题，对于预定组的音素构成了与前后音无关和与前后音有关的模型。然后用与前后音有关模型的输出概率对前后音无关模型的输出概率进行内插。这是通过一个加权或插值因子完成的，所说的加权或插值因子估算与前后音有关的隐藏式马尔克夫模型的输出概率函数能够预测先前在培训数据中不存在的数据的程度。因此，与前后音有关状态的新调整的输出概率函数是两种模型的输出概率函数根据估算的稳定性进行加权得到的组合函数。所以，在该优选实施例中，删除的插值用于平滑概率空间，而不是参数空间。

图 1 表示一个语音识别系统 10，该系统可以用于实施根据本发明的优选实施例的识别和培训方法。语音识别系统 10 包括一个输入装置 12，例如一个麦克风，但是并不限于麦克风，它接收一个输入语音，并产生相应的模拟电信号。或者，可以用存储在一个存储器中的语音作为输入语音。与该语音相应的模拟电信号被传输到模数（A/D）转换器 14，该转换器将模拟信号转换成一序列数字采样信号。然后这些数字采样信号传输到特征抽取器 16，该抽取器抽取数字化的输入语音信号的代表参数。这个代表参数获取输入语音的声特性。可取的是，特征抽取器 16 进行频谱分析以产生

一序列特征向量，每个特征向量包含表示输入语音信号的频谱的系数。进行频谱分析的方法是信号处理领域中所熟知的，可以包括快速傅里叶变换（FFT）、线性预测编码（LPC）、和倒谱系数，所有这些方法特征抽取器 16 都可以使用。特征抽取器 16 可以是能够进行频谱分析的任何常规处理器。频谱分析可以以 10 毫秒的间隔进行以将输入语音信号分割成表示 25 毫秒语音的特征向量。但是，本发明并不限于使用表示 25 毫秒语音的特征向量。还可以使用表示不同时间长度语音的特征向量。对整个输入语音信号重复这个过程，结果得到一序列特征向量，然后将这些特征向量传送到一个数据处理器 38 中。数据处理器 38 可以是任何常规的计算机，例如桌面个人电脑。数据处理器包含一个按规定路线发送这些特征向量的一个转换模块 18。转换模块 18 可以用硬件或软件实现。但是，语音识别系统并不局限于在数据处理器上运行。其它类型的可执行媒体也可以使用，诸如，但是并不限于，一个计算机可读的存储媒体，它可以是一个存储器、光盘，或软盘。

在最初的培训阶段，转换模块 18 转换到将特征向量导向培训装置 20 的位置。培训装置 20 利用这些特征向量估算表示存在于培训数据中的音素的隐藏式马尔克夫模型的参数，并计算识别装置 34 所用的一组加权因子。下面参照附图 2 - 6 更详细地描述培训装置 20 所使用的方法。简单地说，培训装置 20 通过从这些培训数据估算基于与前后音无关和与前后音有关音素的隐藏式马尔克夫模型的参数而产生这些模型。每一与前后音有关状态的输出分布群聚形成句音，存储在句音表存储器 30 中。一般来说，句音表存储器 30 保存与前后音有关和与前后音无关隐藏式马尔克夫模型的句音。对于每个隐藏式马尔克夫模型的句音标识符储存在隐藏式马尔克夫模型存储器 28 中。此外，计算对于每个与前后音有关句音的加权因子，并将其存储在用于识别装置 34 的 λ 表存储器 26 中。 λ 表存储器 26 存储用与前后音有关的隐藏式马尔克夫模型标引的 λ 值。培训装置 20 还利用了一个文本副本，其中包括培训数据的译文 22 和包含每个单词的语音描述的字典 24 以确保每个单词都被正确地模拟。在下面的讨论中将更加详细地描述培训装置 20 的操作。字典 24 包含每个单词的以音素形式体现的发音。例如，字

典中“add”的词条可能是“/AE DD”。

在最初的培训阶段之后，转换模块 18 进行转换将特征向量导向识别装置 34。识别装置 34 将这一组特征向量识别为由构成单词，这些单词又构成句子，的音素组成的语言表达式。现在参照附图 8 详细描述在该识别装置 34 中使用的方法。识别装置 34 使用存储在隐藏式马尔科夫模型存储器 28 中的与前后音无关的和与前后音有关的隐藏式马尔科夫模型、存储在句音表存储器 30 中的与前后音无关的和与前后音有关的句音、存储在 λ 表存储器 26 中的加权因子、和存储在语言模型存储器 32 中的语言模型以及字典 24。语言模型存储器 22 可以载明语法规则。在该优选实施例中，从识别装置 34 中产生的语言表达式显示在一个输出装置 36 上，例如一个常规的打印机、计算机监视器、或类似装置。但是，本发明并不局限于将语言表达式显示在一个输出装置上。例如，可以将语言表达式输入另一个程序或处理器以进行进一步的处理或者可以存储起来。

附图 2 - 6 是表示在系统的培训阶段中所执行的步骤的流程图，在这些步骤中估算隐藏式马尔科夫模型的参数和句音，并计算加权因子。简言之，培训方法开始时接收以单词、句子、短语、或类似形式构成的输入语音，并将它们转换为代表参数，例如已经知道的以特征向量的形式。形成隐藏式马尔科夫模型和句音的结构，并且利用这些培训数据计算这些数据结构的参数的估算值。然后利用删除内插技术确定加权因子。

参见附图 2，培训方法开始时接收一序列的语音（步骤 42），并如前面参照图 1 所述将其转换成一序列的特征向量（步骤 44）。完整的一组特征向量被称为“培训数据”。在优选实施例中，利用 LPC 倒谱分析以使语音信号模型化，并得到一个特征向量，该特征向量包含下列 39 个表示信号中包含的频率和能谱的倒谱和能量系数：（1）12 个 LPC 美 - 频倒谱系数， $x_k(t)$ ， $1 \leq K \leq 12$ ；（2）12 个 LPC Δ 美 - 频倒谱系数 $\Delta x_k(t)$ ， $1 \leq K \leq 12$ ；（3）12 个 LPC $\Delta\Delta$ 美 - 频倒谱系数 $\Delta\Delta x_k(t)$ ， $1 \leq K \leq 12$ ；以及（4）能量、 Δ 能量、和 $\Delta\Delta$ 能量系数。使用 LPC 倒谱分析模拟语音信号在语音识别领域是众所周知的。

在步骤 46，生成句音和隐藏式马尔科夫模型数据结构。句音是语

音识别领域中熟知的数据结构，有关句音以及用于构成句音的方法的详细介绍可以在 M.Huang 等人所写“用句音预测未产生的三音素”（triphone）（Proc.ICASSP '93 Vol.II,pp.311-314,1993）一文中找到。在优选实施例中，一个隐藏式马尔克夫模型可以用于对一个音素的语音单位模型化。隐藏式马尔克夫模型也可以被称为声模型。选择这个语音单位是为了适应大词汇识别。对单个的单词模型化需要较长的培训时间和附加的存储空间以存储相关的参数。这对于小词汇系统是可行的，但是对于使用大词汇的系统是不实用的。然而，本发明不限于以音素为基础的隐藏式马尔克夫模型。其它语音单位，例如单词、双音素、和音节都可以用作隐藏式马尔克夫模型的基础。

可以使用两种类型的隐藏式马尔克夫模型。与前后音有关的隐藏式马尔克夫模型可以用于模拟一个音素及其左右音素。这种类型的模型化获取到通常在单词模型化中存在的前后音的相关性。与前后音无关的隐藏式马尔克夫模型可以用于任何出现在培训数据中处于前后音的音素模型化，所以使得它与任何特定的前后音无关。选择包含一组音素及其相关的左右音素的预定模式，用与前后音有关的隐藏式马尔克夫模型进行模型化。所选择的这些模式表示最常出现的音素和最常出现的这些音素的前后音素。培训数据可以提供有关这些模型的参数的估算值。与前后音无关的模型可以基于所选择的音素，并由任何出现在培训数据中的音素上下文进行模拟。同样，培训数据将提供对于这些与前后音无关模型参数的估算结果。

使用与前后音无关的和与前后音有关的模型都有益于提高识别准确度。每种模型的稳定性与用于估算其参数的培训数据的数量有关，所说参数还能够使其预测培训数据中没有的数据。两种模型相互结合，由于得益于两种模型的培训数据，可以提供更加稳定的估算结果。例如，与前后音有关的模型在共同发音效应模型化时是有益的，但是由于有限的培训数据可能培训得不够（尽管一个讲话者可能努力将各个单词发音成一组连接的音素，但是讲话者的发音器官不能同时动作以产生相互不发生影响的音素。结果，一个音素被在一个单词中位于它前面的音素和位于它后面的音素强烈地影响着。这种效应就是“共同发音效应”）。相反，与前后音无

关的模型是经过很好培训的，从而产生更加稳定的估算结果，对此不再详述。识别装置可以结合使用两种模型，并以适当的方式加权，以产生更加准确的声概率评分。

进一步考虑讲话者之间的差别，例如男性和女性声域中共振频率（即声域谐振频率）的差别，隐藏式马尔科夫模型可以利用这些输出概率分布函数（在本申请中称之为“输出概率分布函数”）的单峰分布的混合。可取的是，可以使用高斯概率密度函数的混合。但是，本发明并不局限于这种特定限制。其它众所周知的连续密度函数的混合，例如拉普拉斯和 K_0 型密度函数也可以使用。

此外，为了获取在不同的与前后音有关的音素状态之间的相似性和增加可用于每个句音的培训数据的数量，对于相同的与前后音无关的音素，不同的与前后音有关的音素隐藏式马尔科夫模型模型的相同状态的输出分布群聚在一起构成句音。

附图 7A 表示了用于音素/aa/114 的一个与前后音无关的隐藏式马尔科夫模型结构的实例。该与前后音无关的隐藏式马尔科夫模型包括三种状态，标记为状态 1（111）、状态 2（112）和状态 3（113）。附图 7A 中表示的隐藏式马尔科夫模型对伴随有任何出现在培训数据中的左右音素的音素/aa/模型化，所说的左右音素在附图 7A 中用符号（*,*）标记。括号中的第一个位置表示在指定音素之前的音素，第二个位置表示在指定音素之后的音素。句音在对于与相同音素相应的每种类型的模型（例如与前后音有关相对与前后音无关的）相同的状态（例如状态 1）范围中分类。在这个实例中，与前后音无关的隐藏式马尔科夫模型分别包括与状态 1、2 和 3 对应的句音 10、55 和 125。

附图 7B 表示一个与音素/aa/相应的与前后音有关的隐藏式马尔科夫模型的实例。在附图 7B 中有 5 个与前后音有关的模型，这些模型以 5 种不同的音素前后关系（115 - 119）模拟音素/aa/。例如，与前后音有关的模型/aa/（/dh/，/b/）115 在左边或前面的音素为/dh/，右边的音素为/b/的前后范围内对音素/aa/模型化。句音在不同隐藏式马尔科夫模型的相同状态范围内分类。在状态 1 中，有两个与前后音有关的句音，标记为句音 14 和 25。

总之,对于音素/aa/,在状态1有两个与前后音有关的句音14和35和一个与前后音无关的句音10,在状态2有两个与前后音有关的句音25和85和一个与前后音无关的句音55;在状态3有一个与前后音有关的句音99和一个与前后音无关的句音125。

因此,用于优选实施例中的基于音素的连续密度隐藏式马尔克夫模型可以表征为下列数学定义:

(1) N , 模型中状态的数目;可取的是,使用三个状态。但是,本发明并不局限于三个状态,而是可以使用多达5个状态。

(2) M , 在输出概率分布函数中混合的数目。

(3) $A=\{a_{ij}\}$, 状态转变概率分布,从状态 i 转变到状态 j 。

(4) $B=\{b_i(x)\}$, 输出概率分布;当处于状态 i 时输出特征向量 x 的概率,其中

$$b_i(x) = \sum_{k=1}^M c_k N(x, \mu_k, V_k) \quad (1)$$

其中 $N(x, \mu_k, V_k)$ 表示利用平均向量 μ_k 和协方差矩阵 V_k 定义的多维高斯密度函数;

混合部分的数目通常为1到50;

c_k 是第 k 个混合部分在状态 i 中的加权因子。

与每状态 i 相关的输出概率分布用句音, sd_i 表示,可以表示为 $P(x_i, sd_i)$ 。

(5) $\pi = \{\pi_1\}$, 初始状态分布。

为了方便,用压缩符号 $\alpha = (A, B, \pi)$ 表示模型的一组完整的参数,其也被称为一个隐藏式马尔克夫模型的参数空间。

在附图2的步骤48中,估算有关句音、与前后音有关的隐藏式马尔克夫模型和与前后音无关的隐藏式马尔克夫模型参数。一个隐藏式马尔克夫模型的培训阶段包括利用培训数据、语音22的音素、和单词24的音素拼读字典估算这些参数。输出和转变概率可以利用众所周知的 Baum-Welch 或向前-向后算法估算。由于 Baum-Welch 算法使得培训数据可以更好地得到利用,所以较为可取。有关内容在 Huang 等人撰写的“用于语音识别的隐藏式马尔克夫模型”(爱丁堡大学出版社, 1990)一书中予以了介绍。

但是，本发明并不局限于这种特殊的培训算法，其它算法也可以使用。通常利用培训数据的大约5次迭代就可以得到较好的参数估算结果。

在附图2的步骤50中，产生每一与前后音有关的句音的加权或内插因子，并用数学符号， λ 表示。加权因子将用于将与前后音有关的隐藏式马尔克夫模型的输出概率内插入到与前后音无关的隐藏式马尔克夫模型的输出概率中。这些加权因子表示与前后音有关的输出概率分布函数在预测未产生数据方面的适合程度。输出概率分布函数利用培训数据估算，并且接近地预测与培训数据类似的数据。但是，利用表示每一个可能的输入语音的培训数据不可能估算输出概率分布函数，或者利用足够的培训数据正确地预测所有未产生的数据。加权因子的作用是指示对于预测未产生数据输出pdf的适合程度，这是用于估算与前后音有关的与前后音无关的模型的培训数据的函数。当用于与前后音有关的模型的培训数据的数量变大时， λ 将趋近于1.0，输出概率分布函数将急剧地加权。利用少量的用于与前后音有关的模型的培训数据， λ 将趋近于0.0，输出概率分布函数将加权较少。用于每一与前后音有关的句音的 λ 最佳值通过删除内插技术确定。

简而言之，删除内插技术将培训数据划分为不同的两组。一组用于估算模型的参数，另一组用于确定加权因子，加权因子表示输出概率分布函数能够预测未产生数据的适合程度。这种方法是反复进行的，在每次迭代过程中循环不同的各组，并产生一个新的模型和加权因子。在所有迭代结束时，计算加权因子的平均值，并用于识别阶段。

附图3-6表示用于计算加权因子的步骤。参见附图3，培训数据在步骤60被分成K块。可取的是，有两块数据。但是，本发明不局限于这些数目的数据块，根据培训数据存储的限制和培训时间可以使用其它数目的数据块。

对于每个与前后音有关的句音计算加权因子（步骤62），这种计算是通过首先利用句音表求出 sens_I （步骤63）实现的，该 sens_I 是该与 sens_D （即与前后音有关的句音）对应的与前后音无关的句音。计算公式是通过一种迭代方法在步骤64推导出的，当新的 λ 值，标记为 $\lambda_{\text{新}}$ 之间的差值满足一定的阈值时，该计算公式收敛。可取的是，当 $|\lambda - \lambda_{\text{新}}| < .0001$ 时，该方法收

敛或结束。该方法在步骤 66 从选择一个初始的 λ 值开始。开始时，为了进行一个句音的第一次迭代，使用者预先选择了一个初始值。可取的是，这个初始值可以是一个估算的假设值，例如 0.8。对于所有其它迭代，初始值可以是以前计算出的新值， $\lambda = \lambda_{\text{新}}$ 。在步骤 68 中，该方法迭代 K 次。在步骤 70 的每次迭代中，选择一块数据作为删除的数据块，所选择的删除数据块是以前没有选择的。

然后该方法利用除去删除的数据块以外的所有数据块的培训数据估算每个与前后音有关的（标记为 b_1 ）句音和与前后音无关的（标记为 b_2 ）句音的输出概率（步骤 72）。这些参数是利用如上所述在培训阶段估算隐藏式马尔克夫模型参数的相同技术（即 Baum-Welch 算法）进行估算的。

接着在步骤 74，计算新值 $\lambda_{\text{新}}$ 。计算假定要求“强制排列”。在培训过程中，如果使用了 Viterbi 算法，可以用一个特定的句音识别培训数据中的每个特征向量。这种利用句音的对应变换或向量就是所知的“强制排列”。 $\lambda_{\text{新}}$ 根据以下的数学公式计算：

$$\lambda_{\text{new}} = \frac{1}{N} \sum_{i=1}^N \frac{(\lambda * b_1(x_i))}{(\lambda * b_1(x_i) + (1 - \lambda) * b_2(x_i))} \quad (2)$$

其中

N = 在相应于句音 sen_{SD} 的删除数据块中利用强制排列的数据点或特征向量的数目

x_i = 特征向量 i ， $1 \leq i \leq N$

$b_1(x_i)$ = 由上述方程（1）定义的与前后音有关的输出概率分布函数

$b_2(x_i)$ = 由上述方程（1）定义的与前后音无关的输出概率分布函数

$\lambda * b_1(x_i) + (1 - \lambda) * b_2(x_i)$: 简称总概率。

为 K 次迭代的每一次迭代确定一个 $\lambda_{\text{新}}$ 值。在步骤 76 中，当完成所有 K 次迭代时，计算一个平均值，该平均值可以用下列数学表达式表示：

$$\lambda_{new} = \frac{\sum_{j=1}^K N_j \lambda'_{new}}{\sum_{j=1}^K N_j} \quad (3)$$

其中

j = 删除数据块的标号

K = 数据块的数量

λ'_{new} = 利用删除数据块 j 估算的 λ 值

N_j = 在删除数据块 j 中利用强制排列对应于 sen_{SD} 的数据点的数量。

如果 λ_{new} 值不满足预定的阈值，则重新执行步骤 66 到步骤 76。当该方法对于一个特定的与前后音有关的句音收敛时，将 λ_{new} 的当前值存储到与该特定的与前后音有关的句音相关的 λ 表 26 中。

附图 4 表示根据上述的方程 (2) 和 (3) 计算加权因子的新值 λ_{new} 的步骤的流程图。将对应于删除数据块中的每个数据点的总概率的与前后音有关的输出概率分布函数的有影响部分求和可以计算出所说新值。因此，在步骤 79，利用在步骤 48 中生成的模型和强制排列找出在对应于 sen_{SD} 的删除数据块中的所有点。在步骤 80，对于用 sen_{SD} 排列的删除数据块中的每一个数据点 x_i 重复该方法。在步骤 82 根据下列数学公式确定数据点 x_i 的与前后音有关的输出概率分布函数相对于总概率的有影响部分：

$$\frac{(\lambda * b_1(x_i))}{(\lambda * b_1(x_i) + (1 - \lambda) * b_2(x_i))} \quad (4)$$

在步骤 84 将这样计算得出的所有数据点的有影响部分之和相加。当迭代计算完成，即在删除数据块中用 sen_{SD} 排列的所有数据点都经过处理时，在步骤 86 根据上述的方程 (2) 计算有影响部分的平均值， λ_{new} 。

上述加权因子的计算利用了删除数据块中的数据点。这种方法以增加培训时间以及培训装置执行计算程序所需的存储量为代价得到了更加准确的计算结果。在某些情况下，生成相应的删除数据块中数据点的代表参数，并用适合的参数代替可能更加有利。另一种方法是利用从数据点的代表参

数到 sen_{SD} 重构的数据点。这些方法只能提供粗略近似的数据点，但是却具有计算效率高的优点。

附图 5 和附图 6 表示这些计算加权因子的选择实施例。附图 5 表示第一选择实施例。参见附图 5，如图所示在步骤 90 中生成删除数据块中的数据点的代表参数。在这种情况下，该代表参数是高斯函数的混合表示。这种表示可以利用如上所述的 Baum-Welch 算法求得。所产生的参数包括相应于每一个混合部分 j 的平均值 μ_j ，和加权因子 c_j 。 λ 新值 $\lambda_{\text{新}}$ 的计算可以根据下列数学公式相对于删除数据块 μ_i 进行：

$$\lambda_{\text{new}} = \sum_{j=1}^M \frac{c_j * (\lambda * b_1(\mu_j))}{(\lambda * b_1(\mu_j) + (1 - \lambda) * b_2(\mu_j))} \quad (5)$$

其中

M = 正常混合分量的数量；

C_j = 第 j 个正常混合分量的加权；

记住 $\sum_{j=1}^M C_j$

μ_j = 第 j 个正常混合分量的平均值；

步骤 92 - 98 按照下列方式进行计算。步骤 92 对于每个混合分量进行迭代计算，并确定与前后音有关的输出概率相对于具有相应的平均值和加权参数的混合的总概率的有影响部分。对于混合部分，它在数学上的表示为：

$$\frac{c_j * (\lambda * b_1(\mu_j))}{(\lambda * b_1(\mu_j) + (1 - \lambda) * b_2(\mu_j))} \quad (6)$$

在步骤 96，对于所有的混合分量求出这些有影响部分的和。在步骤 98，将在步骤 96 中求得的最后的和作为当前的 sen_{SD} 和删除数据块的 $\lambda_{\text{新}}$ 值存储起来。

参见附图 3，当 K 次迭代完成时，过程进入步骤 76，根据上述的方程 (3) 计算 $\lambda_{\text{新}}$ 的平均值。该过程如上所述继续进行，参见图 3，直到该方

法收敛和当前的平均值 $\lambda_{\text{新}}$ 存储在对应于特定的与前后音有关的句音的 λ 表 26 中。

在计算加权因子的第二个选择实施例中，使用了选择数目的数据点，它们是从句音的代表参数中随机产生的。附图 6 表示了所说的第二个选择实施例，它可以根据上述提出的方程 (2) 对一个删除数据块进行数学描述，除了 $\{x_i\}$ = 所产生的数据点和 N = 所产生的数据点的数量。

如图 3 所示，这个选择实施例与优选实施例不同之处在于 $\lambda_{\text{新}}$ 新值的确定方式 (步骤 74)。流程顺序保持如图 3 所示。参见图 6，在步骤 100，为删除数据块中的数据点产生一个代表参数。这个代表参数可以由高斯混合函数构成。这个代表参数可以使用 Baum-Welch 算法根据删除数据块中的培训数据推导出来。根据这个代表参数，如图所示在步骤 102，利用一个具有平均值和加权参数的随机数发生器重构预定数量的数据点。重构数据点的数量折衷考虑了计算 $\lambda_{\text{新}}$ 所需的精确度和计算需求两个方面。较大数量的数据点提高了 $\lambda_{\text{新}}$ 的精确度，但是增大了计算需求。每个混合部分的适合的重构数据数量是 100。

在步骤 104，对于在步骤 104 中成组数据点中的每个数据点执行步骤 106 和 108。在步骤 106，确定数据点相对于数据点的总概率与前后音有关的输出概率的有影响部分。这可以用数学公式表示为：

$$\frac{(\lambda * h_1(x_i))}{(\lambda * h_1(x_i) + (1 - \lambda) * h_2(x_i))} \quad (9)$$

在步骤 108，对于集合中所有数据点求出这些有影响部分的和。当对所有数据点完成迭代运算时，将所有有影响部分的平均值返回作为 $\lambda_{\text{新}}$ 值，(步骤 110)。参见附图 3，当完成 K 次迭代时，过程进入步骤 76 根据上述方程 (3) 计算 $\lambda_{\text{新}}$ 的平均值。该计算程序如上参照图 3 所述继续进行，直到该计算过程收敛和当前的平均值 $\lambda_{\text{新}}$ 存储在特定的与前后音有关的句音的 λ 表 26 中。

当培训数据已经产生，并存储在适合的存储位置时，识别系统就准备工作。语音识别系统的基本任务是检测在输入的语音信号中包含的语音信息。这个任务是一个多级解码问题，因为它需要使一序列特征向量与一序

列音素匹配，使一序列音素与一序列单词匹配，和使一序列单词与一个句子匹配。这是通过构成所有已经模型化的可能的语言表达式和计算该表达式与一序列特征向量匹配的概率来实现的。由于一个语言表达式包含一序列音素，所以所说的确定可能包括计算一种似然性即构成表达式的音素与特征向量匹配和该表达式可能出现（即语法正确）的似然性。构成表达式的音素与特征向量匹配的概率可以简称为声评分，该表达式出现的概率可以简称为语言评分。语言评分考虑了语言的句法和语义，例如语法，并且指出与这一组音素相应的这一组单词是否构成一个语法正确的语言表达式。

在优选实施例中，音素用隐藏式马尔克夫模型表示，其中相同状态的输出概率分布函数群聚在一起构成句音。使一个特征向量与一个音素匹配的方法必然使得一个特征向量与表示该音素的一个隐藏式马尔克夫模型的状态相关的句音匹配。因此，语言表达式可以由对应于一组隐藏式马尔克夫模型的状态的句音构成。

在本发明的优选实施例中，识别系统的任务可以是找出使概率 $P(W/X)$ 达到最大的单词序列 W 。概率 $P(W/X)$ 表示语言表达式 W 出现给定的输入语音信号 X 的概率。 W 可以是一个单词串，表示为 $W = w_1, w_2, \dots, w_i, \dots, w_n$ ，其中 w_i 表示各个单词，每个单词用一组音素表示， $w_i = p_1, p_2, \dots, p_q$ ， X 是用一组特征向量表示的输入语音信号，表示为 $X = x_1, x_2, \dots, x_n$ 。这个最大化问题可以利用众所周知的 Bayes 公式的改进公式求解，其数学表示为：

$$P(W/X) = (P(X/W)P(W))/P(X) \quad (10)$$

$P(X/W)$ 是输入语音信号 X 与单词串 W 匹配的概率，简称为声评分。 $P(W)$ 为单词串 W 出现的概率，简称为语言评分。由于 $P(X)$ 与 W 无关，使 $P(W/X)$ 最大化等价于使分子，即对于所有单词串 W 的， $P(X/W)P(W)$ 最大。

识别工作考虑各种单词组以努力确定最佳匹配。对于每个单词组，识

别工作计算声评分和语音评分。语音评分指示在语音中该单词组的似然性如何，在上述方程（10）中由 $P(W)$ 项表示。声评分指示一序列声特征向量与单词组 W 的声模型匹配的程度如何。在上述公式中声评分由 $P(X/W)$ 项表示。

在计算一个给定单词组的声评分过程中，识别工作考虑各种句音排列。一种句音排列就是从这一组声特征向量到句音的一个对应变换，其为每个声特征向量指定唯一的一个句音。只有能够得出所考虑的单词组的句音排列才会被识别工作所考虑。计算在每一种句音排列的约束下单词串的声评分。单词串的声评分是在所有可能的句音排列中最好的声评分。

数学上，它可以表示为

$$P(X/W) = \text{Max}[P(X/(W, A_i))]$$

从 $i = 1$ 到 q

其中

$A_1 \cdots A_q$ ：对于单词串 W 的所有可能的句音排列

在给定的句音排列 A 的约束下对于单词串 W 的声评分的计算还可以表示为：

$$P(X/(W, A)) = \left(\prod_{j=1}^n P(x_j / sd_j) \right) * P(A) \quad (11)$$

其中句音排列 A 指定或标记第 i 个声特征向量 x_i 对应于与前后音有关的句音 sd_i 。 $P(A)$ 表示句音组 $sd_1 \cdots sd_n$ 的状态转变概率。 $P(x_i/sd_i)$ 表示特征向量 x_i 与与前后音有关的句音 sd_i 匹配的概率。

声评分的实质是计算输出概率 $p(x|sd)$ 。这表示特征向量， x ， 与与前后音有关的隐藏式马尔克夫模型状态相应的句音， sd ， 匹配的似然性。但是，估算不准确的输出概率分布函数会造成声评分计算的不准确性。这通常是由于培训数据不足引起的。随着使用更多的培训数据估算输出概率分布函数增加了分布的稳定性。

解决这个问题的一种方法是利用多个在若干清晰度水平上模拟相同音素的隐藏式马尔克夫模型。然后可以利用在各种清晰度水平上的输出概率分布函数和将它们结合在一起构成特定状态的输出概率分布函数。这种结合是在培训过程中预测未产生数据的能力基础上进行的。更加适合于预测未产生数据的稳定的输出概率分布函数在组合的输出概率分布函数中将得到更多的加权，而估算不准确的输出概率分布函数只得到较少的加权。在优选实施例中，利用多个与前后音有关的隐藏式马尔克夫模型和一个与前后音无关的隐藏式马尔克夫模型模拟一个音素。对于相应于以前在培训阶段计算的与前后音有关状态的每个句音，用一个加权因子 λ 指示每个句音被指定的加权。 λ 越大（趋近于 1.0），与前后音有关的句音越占优势，而与前后音无关的句音的加权越少。当 λ 较小时（趋近于 0.0），与前后音无关的句音占主导地位。因此，输出概率 $p(x|sd)$ 的计算可以用下列的数学公式表示：

$$p(x|sd) = \lambda * p(x|sd_d) + (1 - \lambda) * p(x|sd_i) \quad (12)$$

其中

λ 是句音 sd 的在 0 到 1 之间的加权因子；

x 是特征向量；

sd_d 是与与前后音有关的隐藏式马尔克夫模型的一个状态相关的句音；

sd_i 是与与前后音无关的隐藏式马尔克夫模型的相应状态相关的句音；

$p(x|sd_d)$ 是特征向量 x 与句音 sd_d 匹配的概率；

$p(x|sd_i)$ 是特征向量 x 与句音 sd_i 匹配的概率。

因此，输出概率， $p(x|sd)$ ，是作为与前后音有关的和与前后音无关的句音的输出概率的函数线性插入的。加权因子或内插因子 λ 指示每个句音内插的程度。

附图 8 表示语音识别方法的工作。参见附图 8，该方法从接收一个输入语音开始（步骤 122），并将输入语音转变成特征向量（步骤 124），

有关特征向量前面已经参照附图 1 作了详细描述。在步骤 126，该方法对于每个能表示输入语音的单词序列执行步骤 128 - 136。该单词序列可以由各种不同的句音排列构成，其中每个音素排列相应于一序列隐藏式马尔克夫模型状态。在步骤 128 - 134，确定可以表示单词序列的每一种可能的句音排列的组合识别评分。组合识别评分可以根据改进的 Bayes 公式，即如上所述的方程 (10) 确定。组合识别评分包括一个声评分和一个语言评分。声评分在步骤 130 确定，语言评分在步骤 132 确定，在步骤 134 计算组合评分。然后在步骤 136 选择具有最高组合识别评分的句音排列表示单词序列。在步骤 138，该方法将输入语音识别为具有最高组合识别评分的单词序列。

在步骤 130，可以如上所述根据方程 (11) 确定该声评分，其中输出概率按照如上所述的方程 (12) 计算。

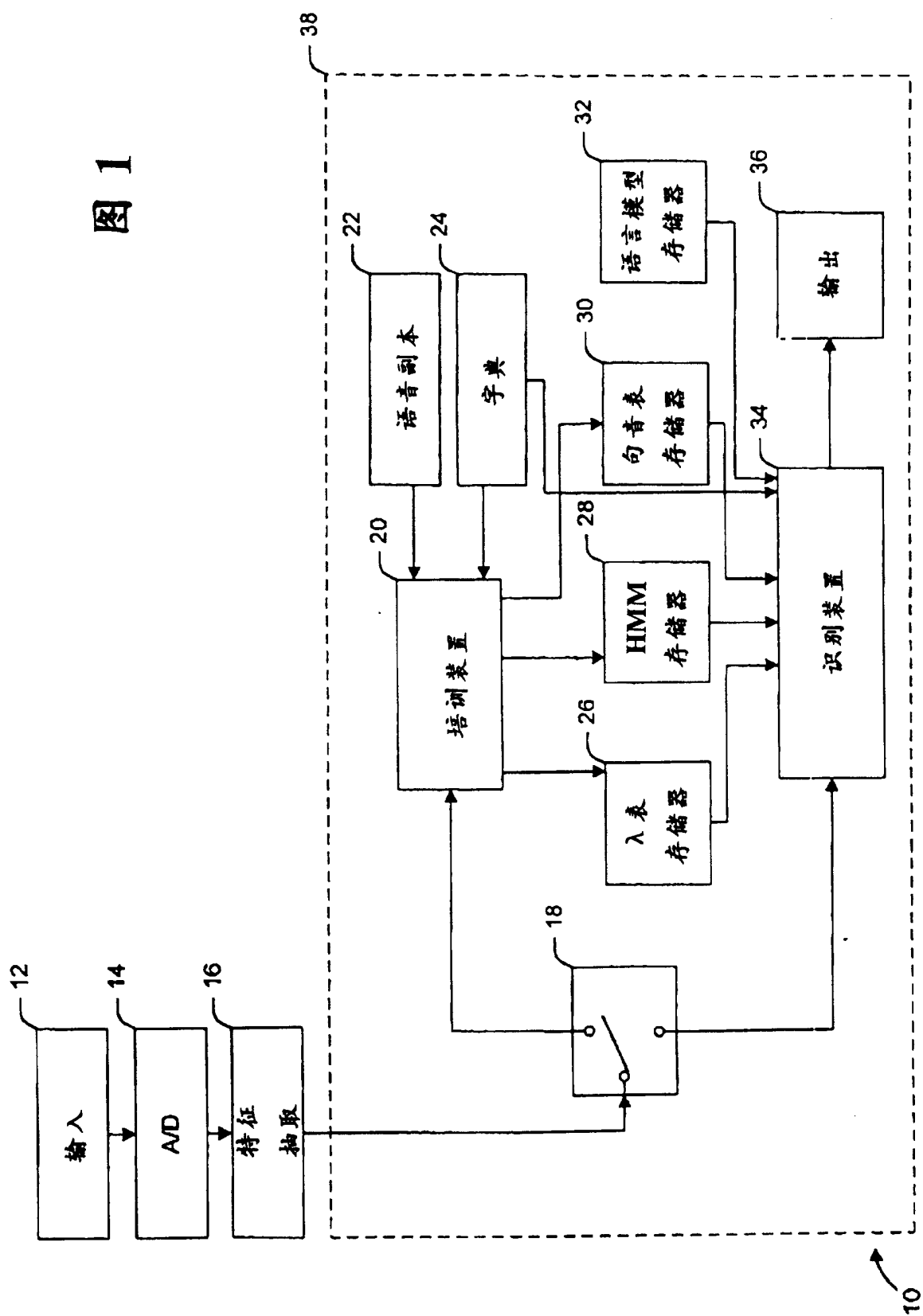
在步骤 132，该方法在表示存储在语音模型存储器 32 中的代表语言表达式的语言模型基础上计算语言评分。语言模型利用语言的结构和语义知识预测在以前发音的单词中一个单词出现的似然性。语言模型可以是一种双词式模型，其中语言评分基于一个单词被一个特定的第二个单词跟随的概率。另一方面，语言模型也可以基于 N-词式而不是双固定模型，或者各自基于子字语言概率。此外，可以使用其它词法知识，例如句法和语法规则创造语言模型。创造和使用语言模型的方法在本领域中是众所周知的，在上面参照的 Huang 等人撰写的书中有详细的描述。

上面详述的本发明通过利用对应于在不同的前后语音关系中相同的发音过程的多个连续密度输出概率提高了语音识别系统的识别能力。由于它改进了模型在模型没有被训练过的预测语音过程的性能，因而提高了特征向量与隐藏式马尔克夫模型的对应变换。在这种水平上的改进是极其有利的，因为在这种水平上的对应变换是识别方法进一步建立的基础。

但是，应当指出本发明不限于语音识别系统。任何需要使语音与语言表达式匹配的应用都可以利用该提出权利要求的发明。语音可以是任何形式的声数据，例如，但并不限于，声音、语声波形、和类似者。这样一种应用的实例是语音合成系统，这种系统利用概率化模型从代表一个语言表

达式的字符串产生一个语音波形。

尽管上面已经详细地描述了本发明的优选实施例，但是需要强调指出，这样做是为了解释本发明，从而能够使本领域技术人员将本发明用于各种需要改进上述装置的不同应用中，因此，本说明书中的具体描述除了说明与本发明相关的现有技术所需内容以外，并不是对本发明范围的必要限制。



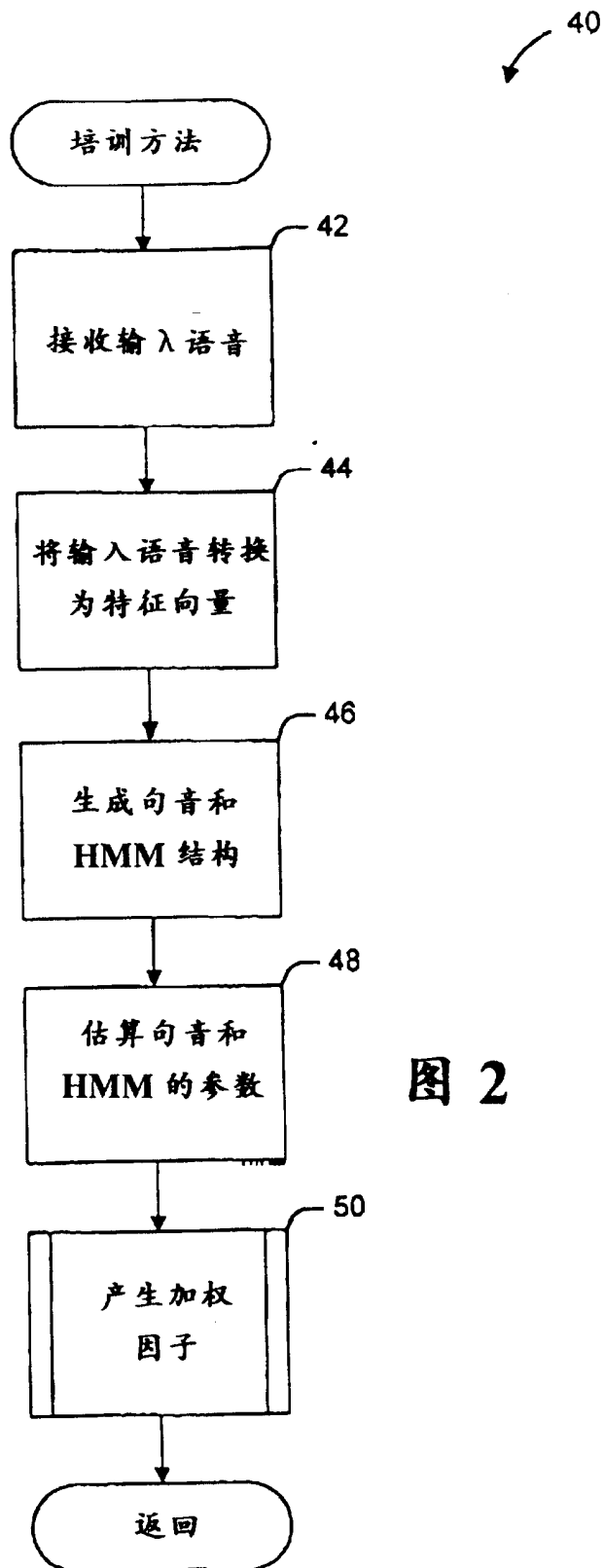
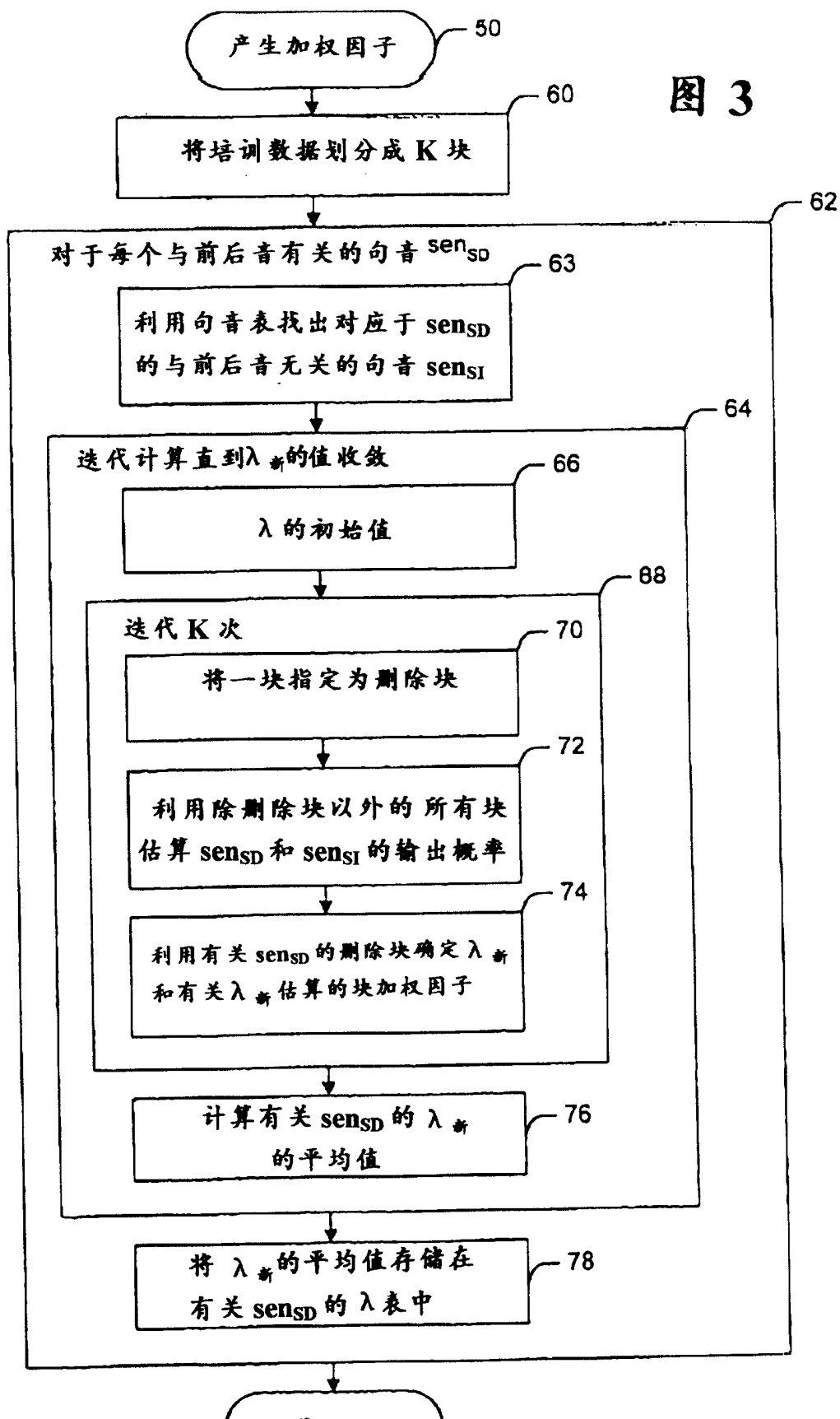


图 3



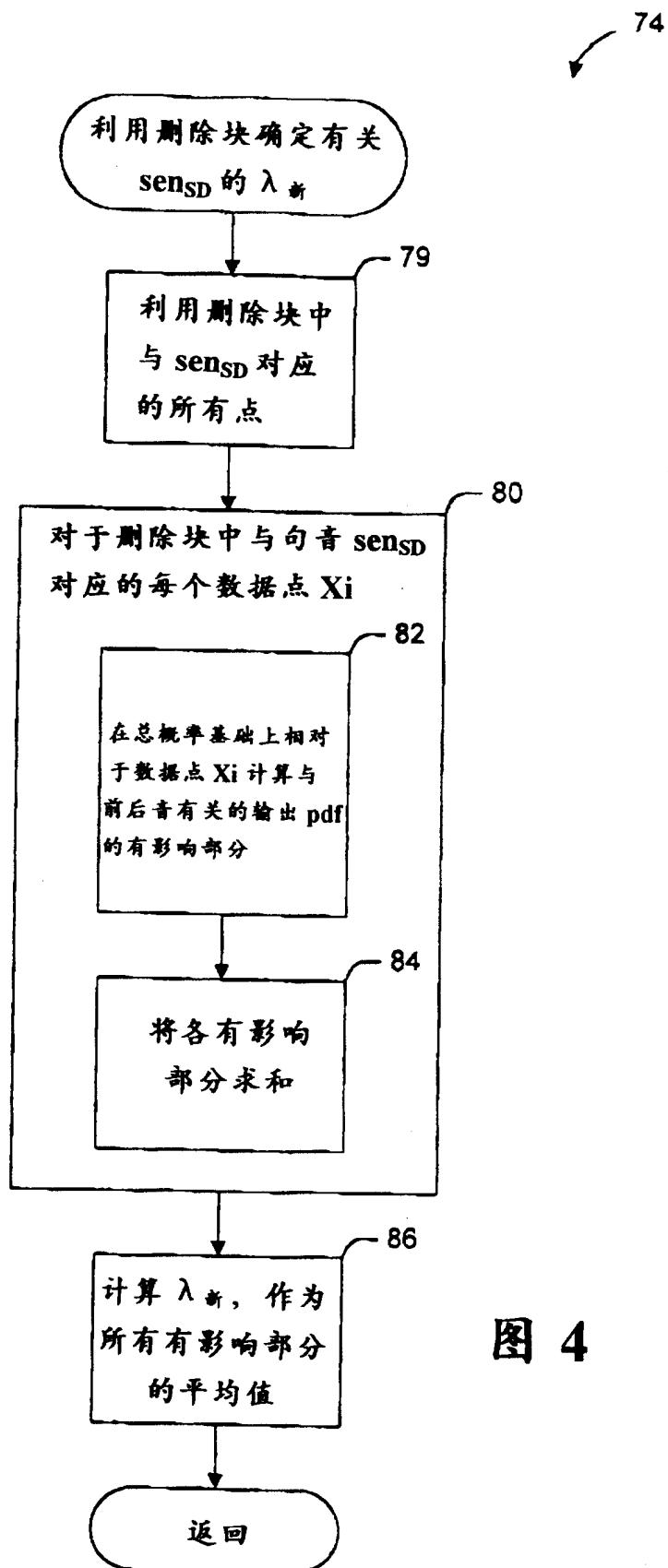


图 4

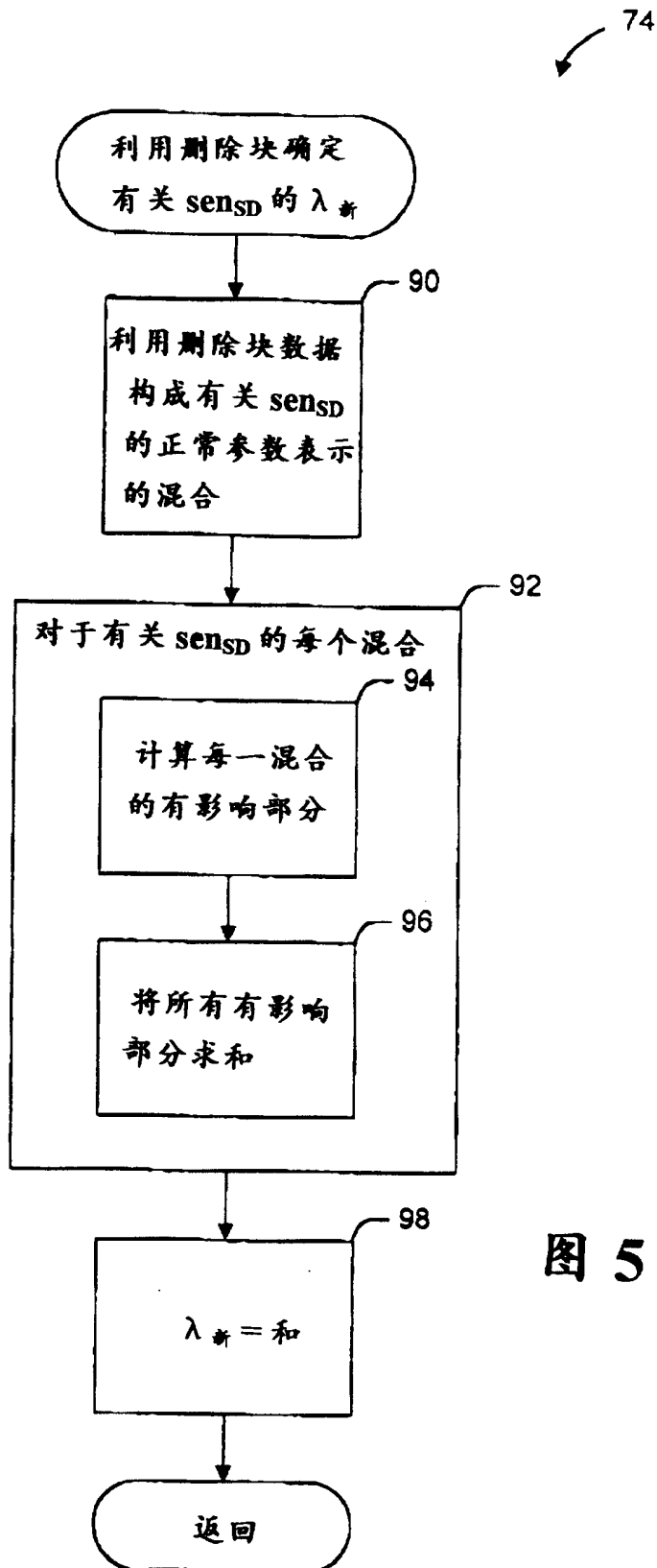


图 5

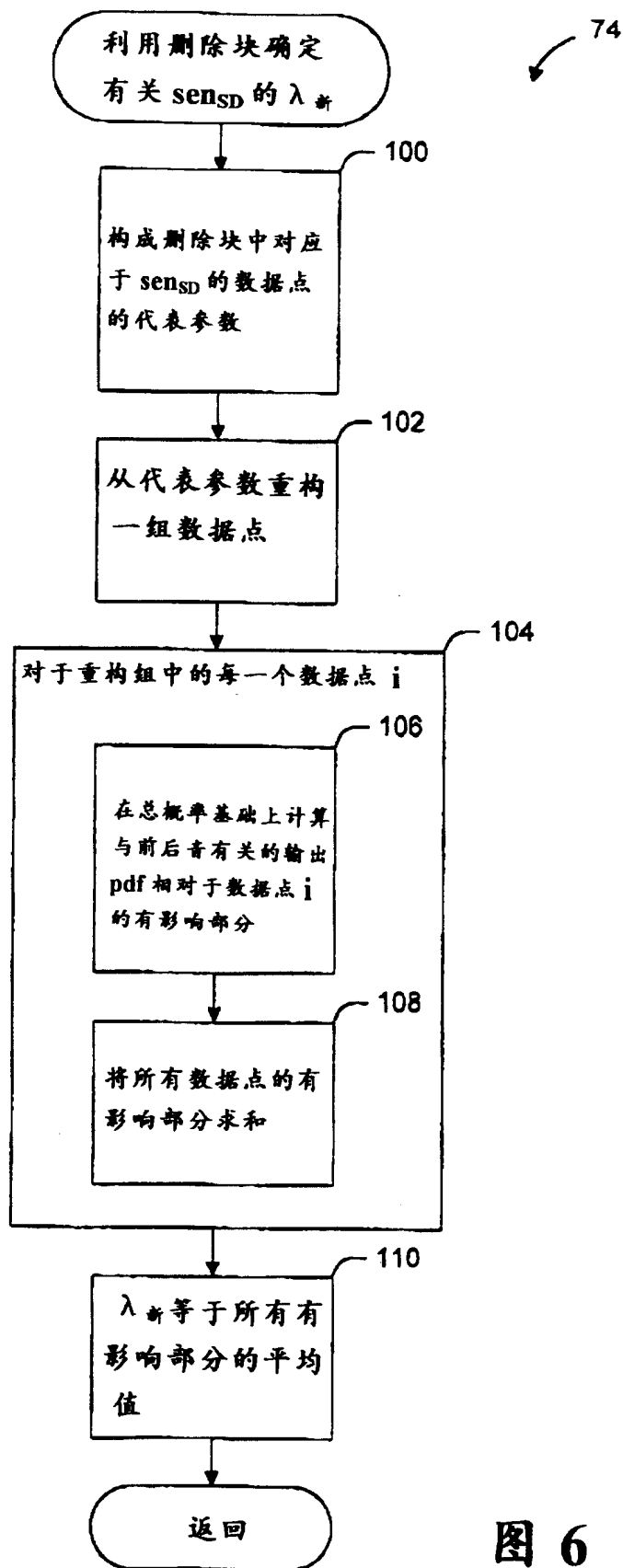


图 6

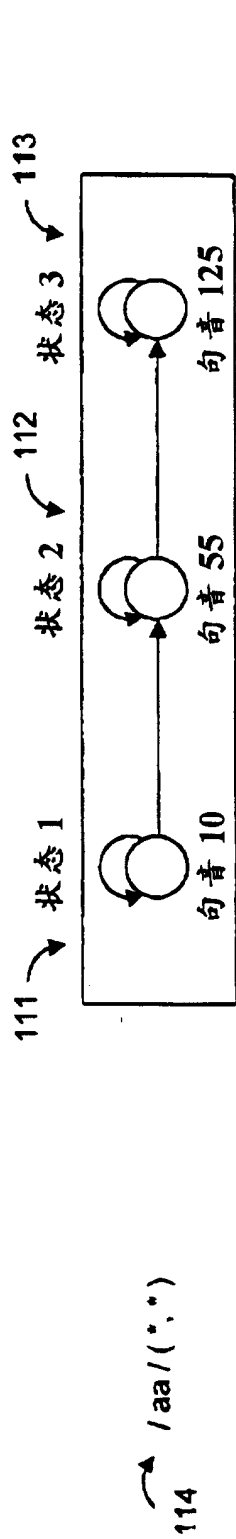


图 7A

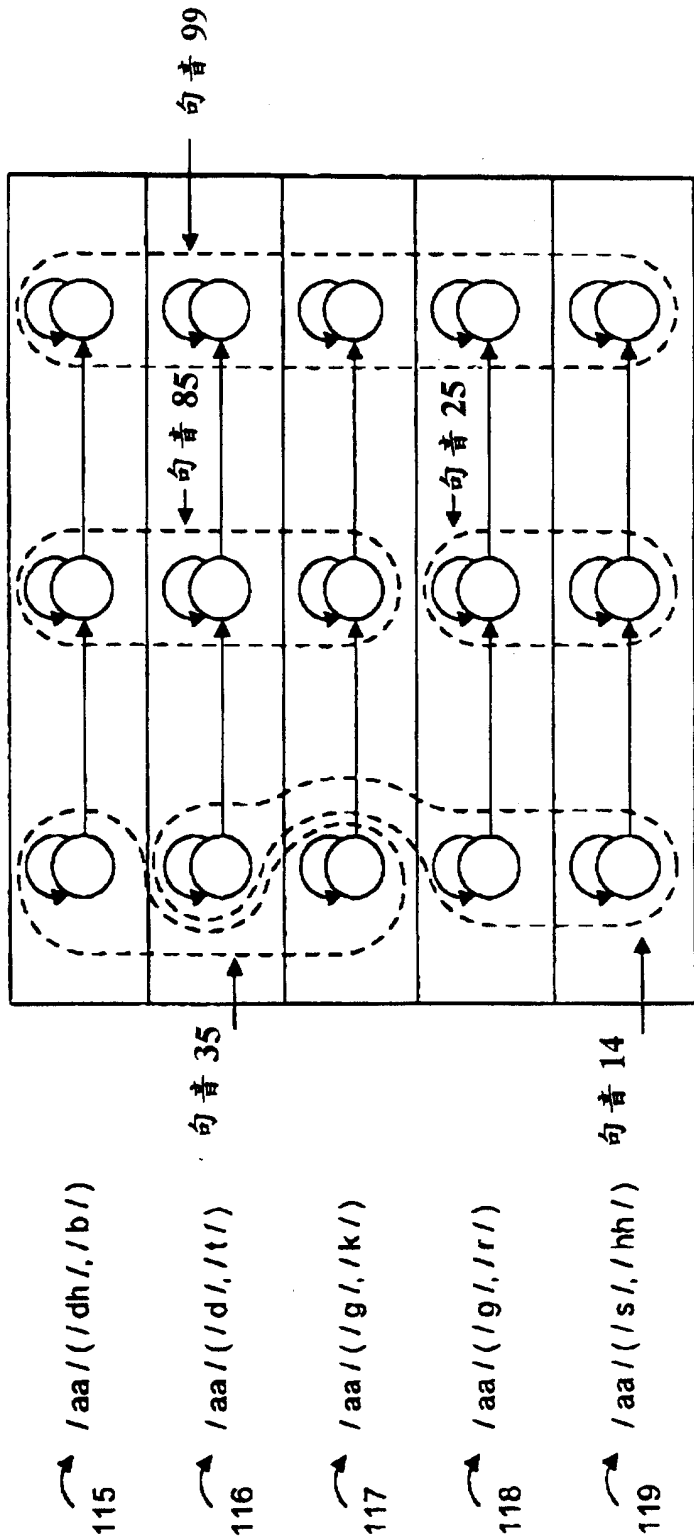


图 7B

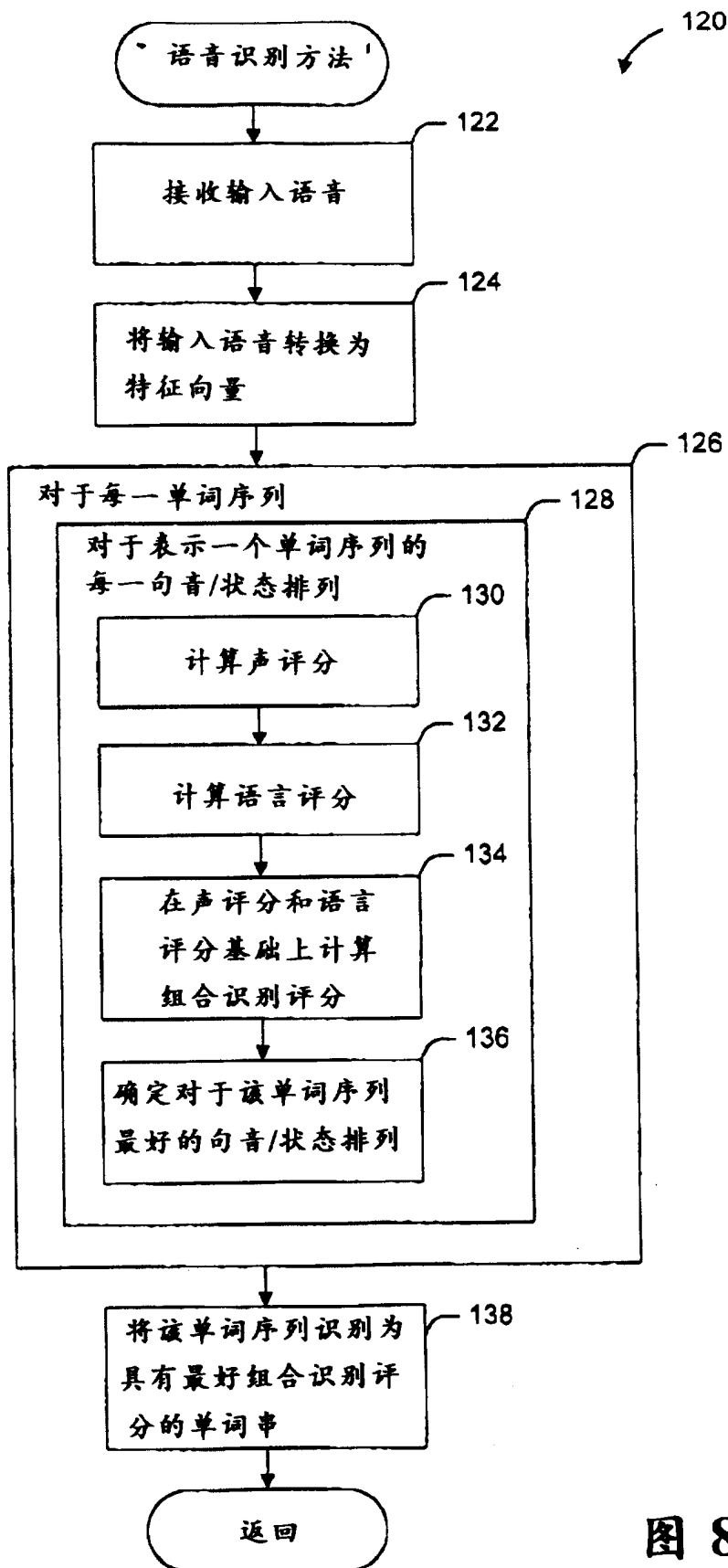


图 8