[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G06F 13/10
G06F 3/16

[12] 发明专利说明书

[21] ZL 专利号 99123747.1

[43] 授权公告日 2003 年 2 月 5 日

[11] 授权公告号 CN 1101025C

[22] 申请日 1999.11.19 [21] 申请号 99123747.1

[71] 专利权人 清华大学
地址 100084 北京市海淀区清华园
[72] 发明人 郑 方 吴文虎 方棣棠
审查员 朱世菡

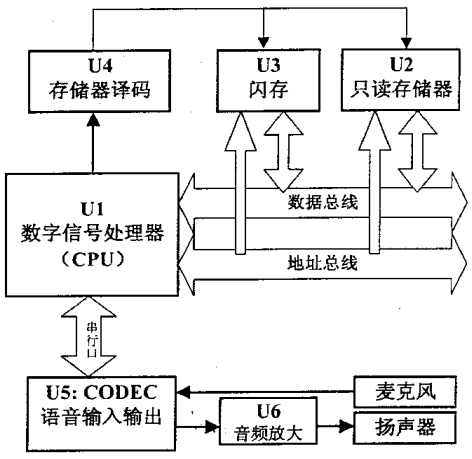
[74] 专利代理机构 北京清亦华专利事务所
代理人 廖元秋

权利要求书 1 页 说明书 7 页 附图 3 页

[54] 发明名称 语音命令控制器的训练与识别方法

[57] 摘要

本发明属于计算机语音识别技术领域，涉及语音命令控制器的训练与识别方法，包括定点数字信号处理器，只读存储器，闪存器，对存储器所存数据进行译码的译码器，编码译码器，音频放大器，扬声器与麦克风，以及存储在该闪存器中的语音命令的训练和识别软件；本发明可用于家电产品的控制以及声控电话簿、声控电话机、袖珍式声控拨号器等新产品上，可大大方便使用者，提高人们的工作效率和生活的质量。



ISSN1008-4274

1、一种用于语音命令控制器的语音命令的训练方法，其特征在于，包括以下步骤：

（1）启动 CODEC 采集过程：打开模数转换器件，开始对声音采样；

（2）采集一遍语音命令的有效发音：当自动检测到语音开始后，把采样到的语音数据逐一记录在内存中，检测到语音结束后，停止记录；

（3）对上一步记录的语音数据进行特征提取，即提取倒谱特征系数，并对语音按特征序列进行非线性分段；

（4）将倒谱系数及分段结果保存于存储器中，以便用于训练过程中的建模；

（5）如果训练未满三遍，转到 2，继续训练；否则，到下一步；

（6）建立该语音命令的模型并保存：利用提取的特征进行建模，将模型存到闪存，将来用于识别；

（7）结束。

2、一种用于语音命令控制器的语音命令的识别方法，其特征在于，包括以下步骤：

（1）启动 CODEC 采集过程：打开模数转换器件，开始对声音采样；

（2）采集一段有效发音：当自动检测到语音开始后，将采样到的语音数据记录在内存中，检测到语音结束后，停止记录；

（3）对上一步记录的语音数据进行特征提取，即提取倒谱系数，并对语音按特征序列进行非线性分段；

（4）暂存倒谱特征系数及分段结果，以便用于识别；

（5）将上一步得到的语音特征与所有已经存在的命令模型进行比较，记下最匹配的三个命令模型；

〈5.1〉 取一个已存的命令模型计算其匹配概率；

〈5.2〉 将该概率值以及相应的命令序号与保存三个最大概率值的结果数组比较，按情况更新结果数组；

<5.3> 命令比较未完，转到<5.1>；

（6）根据结果数组中三个最大概率值进行拒识判别：根据三个最匹配的模型的匹配概率判断是接受识别结果还是拒绝接受；

（7）将保存概率值以及命令序号的结果数组和识别接受/拒绝标志保存于参数交换区：保存识别结果；

（8）结束。

语音命令控制器的训练与识别方法

技术领域

本发明属于计算机语音识别技术领域，特别涉及一种用于家电控制以及声控电话簿、声控电话机、袖珍式声控拨号器等产品上的语音命令控制器的其训练与识别方法。

背景技术

目前家用电器的控制有两种方式：一是用手直接操作按钮，如电视机、洗衣机、微波炉、空调等；二是通过遥控器进行操作，如电视机、空调等。

随着家电的技术不断完善和发展，功能不断增多，家电的说明手册也越来越厚。由于家电的控制面板不能太大，按钮不能太多，因此很多按钮需要复用，对于某种功能就往往需要几个按钮操作结合起来才能完成。在这样的情况下，有时为了进行某种功能的操作，往往要翻半天的说明书，给人们带来很多不便。

另外以目前在家庭中最常用的电话为例，由于人们的活动范围不断扩大、工作、学习和生活上的需要，每天都要给不同的人或单位打电话，这就需要记忆和查找电话号码，而记忆大量电话号码是一个令人心烦的过程，如果不去记忆，就要每次去翻阅电话号码簿，既费时又费事。

众所周知，目前大家记录电话号码的方法无外乎以下两种：(1) 用笔记录到一个电话簿(下称“纸张电话簿”)上；(2) 记录到类似个人数字助理等的电子产品(下称“一般电子电话簿”)上。不管哪种方法，一个最大的问题是号码的输入、修改和查询。

对于纸张电话簿，虽然一些生产厂家生产了带有A－Z标签的纸张电话簿，但由于无法预计到不同用户的实际情况，印刷时一般总是让每个标签的页数相同。但在实际使用时，大部分的情形是，在有的标签可能一个姓名没有时有的标签却已经用完了(如Z标签有“张”，“郑”，“周”，...等姓氏非常多)，用完的标签只好用其他标签来补充。这样既比较混乱，又导致查找时困难。对没有设计标签的一般纸张电话簿来说，查询就更加麻烦了，有时为了找到一个人的电话号码要翻找好多页。纸张电话簿的另外一个缺点是号码修改不方便，有时由于号码改变了，不得不把相应的地方用笔涂掉，很不雅观。

对一般电子电话簿来说，它可以很好地解决纸张电话簿中的人名排序、电话号码修改等问题，而且可以通过键盘较好地解决人名的查询问题。但其不方便之处是人名的查询仍然不是特别方便。类似于手机，查询的方法一般是先输入姓名的首字母，然后再用前翻或后翻键去找需要的人名。

因此，随着家用电器高技术含量和功能的不断提高，操作简便灵活的要求就提到了日程。

本发明突破了传统控制家电的方法，提出了用语音控制家电的构想。

语音是最自然的一种人机交互方式，随着计算机的飞速发展和语音处理技术的

日益成熟，人们希望把语音识别技术应用到实际产品中的愿望正一步一步地成为现实。尤其是特定人中词表孤立词语音识别技术的成熟以及低成本、高性能的单片机数字信号处理器的出现，使得人们在日常工作和生活中应用高新技术——语音识别技术——成为可能。

发明内容

本发明的目的是为克服已有技术的不足之处，提出一种语音命令控制器的其训练与识别方法，将语音识别技术用于家电产品的控制以及声控电话簿、声控电话机、袖珍式声控拨号器等新产品上，可大大方便使用者，提高人们的工作效率和生活的质量。

本发明用于语音命令控制器的语音命令的训练方法，其特征在于，包括以下步骤：

（1）启动 CODEC 采集过程：打开模数转换器件，开始对声音采样；

（2）采集一遍语音命令的有效发音：当自动检测到语音开始后，把采样到的语音数据逐一记录在内存中，检测到语音结束后，停止记录；

（3）对上一步记录的语音数据进行特征提取，即提取倒谱特征系数，并对语音按特征序列进行非线性分段；

（4）将倒谱系数及分段结果保存于存储器中，以便用于训练过程中的建模；

（5）如果训练未满三遍，转到 2，继续训练；否则，到下一步；

（6）建立该语音命令的模型并保存：利用提取的特征进行建模，将模型存到闪存，将来用于识别；

（7）结束。

本发明用于语音命令控制器的语音命令的识别方法，其特征在于，包括以下步骤：

（1）启动 CODEC 采集过程：打开模数转换器件，开始对声音采样；

（2）采集一段有效发音：当自动检测到语音开始后，将采样到的语音数据记录在内存中，检测到语音结束后，停止记录；

（3）对上一步记录的语音数据进行特征提取，即提取倒谱系数，并对语音按特征序列进行非线性分段；

（4）暂存倒谱特征系数及分段结果，以便用于识别；

（5）将上一步得到的语音特征与所有已经存在的命令模型进行比较，记下最匹配的三个命令模型；

<5.1> 取一个已存的命令模型计算其匹配概率；

<5.2> 将该概率值以及相应的命令序号与保存三个最大概率值的结果数组比较，按情况更新结果数组；

<5.3> 命令比较未完，转到<5.1>；

（6）根据结果数组中三个最大概率值进行拒识判别：根据三个最匹配的模型的匹配概率判断是接受识别结果还是拒绝接受；

（7）将保存概率值以及命令序号的结果数组和识别接受/拒绝标志保存于参数交换区：保存识别结果；

（8）结束。

本发明的训练和识别方法两个部分的原理说明如下：

语音命令的训练和识别的基本特征参数是经典的基于线性预测编码(LPC)的 10 阶倒谱系数，采样率为 8KHz，是电话信道带宽的两倍多，因此适应于类似于电话信道的窄带应用中。

语音识别模型是中心距离连续概率模型(CDCPM)，其输出观察概率密度为：

$$p(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\|x-\mu\|^2/2\sigma^2) \tag{1}$$

其中 x 是随机特征向量 ξ 的一个取值，μ 为 ξ 的均值。如果记 ξ 与 μ 之间的距离为另外一个随机变量 η，那么 η 的 p.d.f. 可以表示为：

$$p(y;\mu,\sigma)=\frac{2}{\sqrt{2\pi}\sigma}\exp(-y^2/2\sigma^2),y\geq 0 \tag{2}$$

其中η的均值为：

$$\mu_\eta=\int_0^\infty yp(y;\mu,\sigma)dy=\frac{2\sigma}{\sqrt{2\pi}}, \tag{3}$$

该分布的参数μ和σ可以由下面的公式从训练数据库中估计出来：

$$\mu=\frac{1}{S}\sum_{n=1}^{S}x_n,\sigma=\frac{\sqrt{2\pi}}{2}\mu_\eta=\frac{\sqrt{2\pi}}{2}\left[\frac{1}{S}\sum_{n=1}^{S}d(x_n,\mu)\right] \tag{4}$$

其中 S 是样本的数目。

为了加速运算，在识别时把(2)重写为：

$$\begin{aligned}\log p(y;\mu,\sigma)&=-y^2/2\sigma^2-\log\sigma+const\\&=-y^2*\left[\frac{1}{2\sigma^2}\right]-\log\sigma+const\\&=-y^2*S_r-S_l+const\end{aligned} \tag{5}$$

其中的常数项可以省略，而把μ、S_r和S_l作为模型参数。这样做的结果是使模型的时间复杂度降低。

上面是一个状态中参数的估计方法。对一段有效语音，我们要根据其中特征参数的具体分布情况，将其划分为若干状态(段)，这个过程是为了解决不同长度的发音之间进行匹配时的时间对准问题。我们这里使用的是算法复杂度很低，但顽健性很高的非线性分段(NLS)算法。

根据语音识别的统计模型对状态的假设，NLS 算法要保证在同一状态(段)内的特征向量应该变化很小，或说同一状态内的特征向量的变化应该比较平稳。NLS 算法的步骤是这样的。

首先计算出特征向量序列中相邻两个向量之间的距离(特征变化量)

$$y_i \overset{def}{=} y(o_{t+1}, o_t), \quad 1 \le t \le T-1 \tag{6}$$

以及段内平均总变化量:

$$\Delta y \overset{def}{=} \frac{1}{N}\sum_{t=1}^{T-1} y_t \tag{7}$$

其中 N 是状态数。以 L_n 表示前 n 段中向量总个数，对 $1 \le n \le N-1$ 如果

$$i\sum_{t=1}^{k-1} y_t < n\Delta y \le \sum_{t=1}^{k} y_t \tag{8}$$

则有 $L_n = k$, (9)

同时令 $L_N = T$。 (10)

显然 L_n 就是第 n 和第 $n+1$ 段的分界点。由 NLS 算法可以看出，每段内的特征变化量总和大致相等。

NLS 算法比较简单地给出了在“等特征变化量”的意义下“最好的”状态序列，而且不管状态和识别基元的驻留时间如何变化，它总能比较一致地把变化较小的那些序列分到一起。从某种意义上讲，它对语速的变化有相当好的鲁棒性。

在我们的算法实现中，我们所选的状态数目是 6。模型存储开销是每个语音命令 78×16bit，训练和识别的时间开销均是数十毫秒级的(<0.01 秒)，完全达到实时。

采用本发明所述方法的一种语音命令控制器，包括用于 进行主要控制和计算的定点数字信号处理器CPU U1； 用于存放程序和初始化数据的只读存储器U2； 用于永久存放语音识别模型/模板以及其他需要保存数据的闪存器U3；对存储器所存数据进行译码的译码器U4； 用于进行语音输入和输出的CODEC编码译码器U5； 用于进行音频放大的音频放大器U6，扬声器和麦克风； 以及存储在该闪存器中的语音命令的训练和识别软件。图1中各元件的连接关系为：只读存储器和闪存器通过数据总线和地址总线与数字信号处理器相连；译码器一端与数字信号处理器相连，另一端与只读存储器和闪存器相连；编码译码器一端连于数字信号处理器的串行接口，另一端连于音频放大器输入端和麦克风，音频放大器输出端与扬声器相连。

本发明的语音处理工作过程结合图1说明如下：

CPU U1在上电后自动将只读存储器U2中的程序(包括不同用户定制的控制程序和本发明编制的核心程序)调入CPU 并进行控制。在CPU的控制下，用户的语音经过“麦克风”转变为电信号后，进入“CODEC语音输入输出”U5并被转换为数字信号。这些数字信号经过CPU的串行口传送到CPU中，CPU会根据不同的功能进行不同的处理。此时的处理有以下两种。

1. 训练，即用户训练预先设定的语音命令并存储其语音模型。核心处理程序把语音经过特征抽取和建模后转换为高度压缩的语音模型，并把这些模型存储在“闪存器”U3中，以便以后控制时使用。

2. 识别，把用户所说的语音与闪存器中所存的语音命令模型进行比较，找到正确的语音命令。在用户说出命令后，核心处理程序对经过模数转换U5后的数字语音进行预处理并从闪存U3中调出表征命令的语音模型，并让它们逐个与刚刚输入的语音进行模式匹配，通过实时的识别，得到最终的匹配结果。CPU可以先把识别出来的命令复述一遍(这个步骤是可选的，不是必需的)，然后将识别结果反馈给上层控制电路。识别的结果有几种可能：一是特别肯定的正确识别结果；二是特别肯定的拒识结果(可能录入的语音是噪音或非命令)；三是不很肯定的识别结果。上层控制电路可以根据

识别结果进行相应的处理。

在上述的过程中，对只读存储器 U2 和闪存器 U3 的地址译码是通过存储器译码 U4 进行的。

本发明具有如下技术特点：

核心程序对于 6 状态的 CDCPM 模型，每个模型的空间开销是 78W；对 200 个模型(人名)的规模，识别响应时间是毫秒级的。性能可与传统的 HMM 媲美。

另外，在一般使用环境中，由于不能保证信噪比很高，识别率也不会达到 100%；而打电话最令人尴尬的是把电话拨错。因此，如何保证错误率很小成了关键问题之一。在本实施例的核心代码中，采用了两项措施来保证这一点。措施之一是让用户进行确认；措施之二是增加拒识功能。

用户一次口呼的识别结果共有三个候选。如果识别器对第一候选特别有把握，那么它将直接给出结果；如果判断这是噪音，那么就拒识。如果用户说得不清楚，或有噪音干扰，或对识别结果没有把握，那么可以从第一候选开始依次用语音报出候选结果，由用户用语音或键盘进行确认，直至找到正确结果。

上面的措施非常有效地降低了错误率，使得电子电话簿的满容测试，无论环境噪音如何，无论训练完之后间隔了多长时间以后再使用，都把错误率有效地控制在一定的范围之内。

本发明具有如下应用范围：

本发明在控制诸如洗衣机、电视机、空调、微波炉等家电时，人们可以不必去阅读繁琐而复杂的说明书，只需要口述语音命令就可以对家电进行控制，既直接又高效。无论白天黑夜，无论盲人还是正常人，可谓方便之至。

本发明用于“声控电话簿”使人们不必再记电话号码，而把记忆电话号码的工作交给电子电话簿来完成。在需要打电话的时候，只要口呼姓名或单位名，它就会把相应的电话号码调出来。此外，把本发明集成到普通电话机电路中可实现“声控电话机”方案；也可以把本发明嵌入到个人数字助理(PDA, Personal Digital Assistant)产品中当作电子电话簿使用，再附加DTMF拨号功能，实现可移动的袖珍式声控拨号器。因此，本发明有极其广泛的应用前景。

附图说明

图 1 为本发明的总体逻辑组成框图。

图 2 为本发明软件程序之一——命令的训练流程框图。

图 3 为本发明软件程序之一——命令的识别流程框图。

图 4 为本发明的实施例电路原理图。

具体实施方式

本发明用于语音命令控制器的训练和识别方法实施例程序流程结合图 2、3 分别说明如下：

(一) 命令训练过程，如图 2 所示：

（1）启动 CODEC 采集过程：打开 ADC(模数转换)器件，开始对声音采样；

（2）采集一遍语音命令的有效发音：当自动检测到语音开始后，把采样到的语音数据逐一记录在内存中，检测到语音结束后，停止记录；

（3）对上一步记录的语音数据进行特征提取，即提取倒谱特征系数，并对语音按特征序列进行非线性分段；

（4）将倒谱系数及分段结果保存于存储器中，以便用于训练过程中的建模；

（5）如果训练未满三遍，转到2，继续训练；否则，到下一步；

（6）建立该语音命令的模型并保存：利用提取的特征进行建模，将模型存到闪存，将来用于识别；

（7）结束。

(二) 命令识别过程如图3所示：

（1）启动CODEC采集过程：打开ADC(模数转换)器件，开始对声音采样；

（2）采集一段有效发音：当自动检测到语音开始后，将采样到的语音数据记录在内存中，检测到语音结束后，停止记录；

（3）对上一步记录的语音数据进行特征提取，即提取倒谱系数，并对语音按特征序列进行非线性分段；

（4）暂存倒谱特征系数及分段结果，以便用于识别；

（5）将上一步得到的语音特征与所有已经存在的命令模型进行比较，记下最匹配的三个命令模型；

<5.1> 取一个已存的命令模型计算其匹配概率；

<5.2> 将该概率值(含命令序号)与保存三个最大概率值的结果数组比较，按情况更新结果数组；

<5.3> 命令比较未完，转到<5.1>；

（6）根据结果数组中三个最大概率值进行拒识判别：根据三个最匹配的模型的匹配概率判断是接受识别结果还是拒绝接受；

（7）将结果数组(概率值以及命令序号)和识别接受/拒绝标志保存于参数交换区：保存识别结果；

（8）结束。

本发明的实施例是一个声控电话簿。结合各附图详细说明如下。

本发明的实施例电路原理如图4所示，其中主要部件的实施例为：

U1：ADSP-2186。

U2：AT27C010。

U3：AT29C020。

U4：74HC139。

U5：AD73311。

U6：4083B。

其中核心处理芯片的实施例为：

核心程序所采用的主CPU是美国Analog Devices, Inc.(ADI)公司的定点处理芯片ADSP-218x，它具有速度快(30n)，内部RAM较大(16K word)，I/O口丰富等优势，既满足了系统程序的运行，价格又合适。在这样的处理器上，对于6状态的CDCPM模型，每个模型的空间开销是78W；对200个模型(人名)的规模，识别响应时间是毫秒级的。性能可与传统的HMM媲美。

本发明的实施例工作过程结合图4说明发如下：

在上电后，CPU(U1，ADSP-2186)自动将只读存储器(U2，AT27C010)中的程序(包括不同用户定制的控制程序和我们编制的核心程序)调入 CPU 并进行控制。在 CPU 的控制下，用户的语音经过“麦克风”转变为电信号后，进入“CODEC 语音输入输出”(U5，AD73311)并被转换为数字信号。这些数字信号经过 CPU 的串行口传送到 CPU 中，CPU 会根据不同的功能进行不同的处理。主要是：

(一) 训练，即用户录入一个新的人名。核心处理程序把语音经过特征抽取和建模后转换为高度压缩的语音模型，并把这些模型存储在“闪存”(U3，AT29C020)中，供以后查询时使用。

(二) 识别，即用户查询人名。在用户说出一个待查人名后，核心处理程序对经过模数转换(U5)后的数字语音进行预处理并从闪存(U3，AT29C020)中调出表征人名的语音模型，并让它们逐个与刚刚输入的语音进行模式匹配，通过实时的识别，得到最终的匹配结果。识别结果有三种可能。一是核心程序对正确识别结果非常肯定。一是核心程序非常肯定是环境噪音或所说的人名不在电话簿中。一是核心程序对识别结果不很肯定，但它可以提供三个候选供用户选择。CPU可以有不同的方式将识别结果反馈给用户。其一是把识别结果所对应的信息(它们在用户录入人名时被存放在闪存(U3)中)的语音数据由CPU(U1)送到“CODEC语音输入输出”(U5)转换成模拟信号，再送到“音频放大器”(U6，4083B)经“扬声器”回报出来，由用户进行确认。其二是在“LCD液晶”上显示出来。

在上述的过程中，对只读存储器(U2)和闪存(U3)的地址译码是通过存储器译码(U4，74HC139)进行的。

本发明的实施例有如下效果：

(1) 声控命令容量大：最多可以存储200个语音命令(人名)，这对大多数应用是足够了。

(2) 声控存取识别率高：在进行200个人名(二、三字的人名各占一半)的满容量测试时，对数十次的测试结果统计得到，人名识别正确率高，平均达到97%以上。声控电话簿具有一定的智能：如果它对识别结果非常有信心，则直接给出识别结果；如果它认为使用人说的人名在号码簿中没有或者它认为噪音进入了麦克风，则直接拒绝；对不肯定的部分，则给出其他候选由用户选择。

(3) 声控有拒识功能：对于不在集合中的命令(人名)，语音命令控制器有拒识功能，这可以保证错误的结果不会被错误地接收。我们进行过大量的测试，当电话簿满容(200个人名)时，随机说一些不在电话簿中的人名或有噪音进入，则声控电话簿可以以很高的准确性拒绝，正确拒识率高达90%以上。

(4) 适合各种语言：对语音命令的训练和识别不受语言限制，可以用普通话、方言，甚至是外语，只要和训练时所用语音一致就行。

本发明的实施例主要部件功能说明：

U1：用于进行主要控制和计算的定点数字信号处理器(DSP)芯片U3，是系统的中央处理部件(CPU)部件。

U2：用于存放程序和初始化数据的只读存储器EPROM。

U3：用于永久存放语音识别模型/模板以及其他需要保存数据的闪存(Flash Memory)。

U4：进行存储器译码的芯片。

U5：用于进行语音输入(模数转换ADC)和输出(数模转换DAC)的CODEC编码译码器。

U6：用于进行音频放大的芯片。

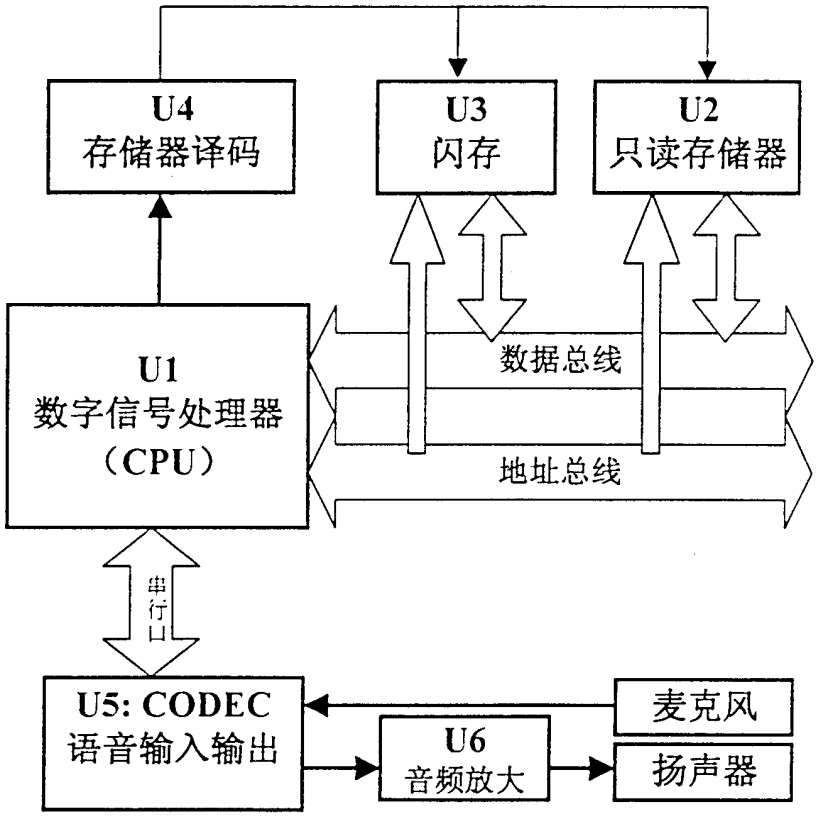

U4
存储器译码
U3
闪存
U2
只读存储器
U1
数字信号处理器
（CPU）
数据总线
地址总线
串行口
U5: CODEC
语音输入输出
U6
音频放大
麦克风
扬声器

图 1

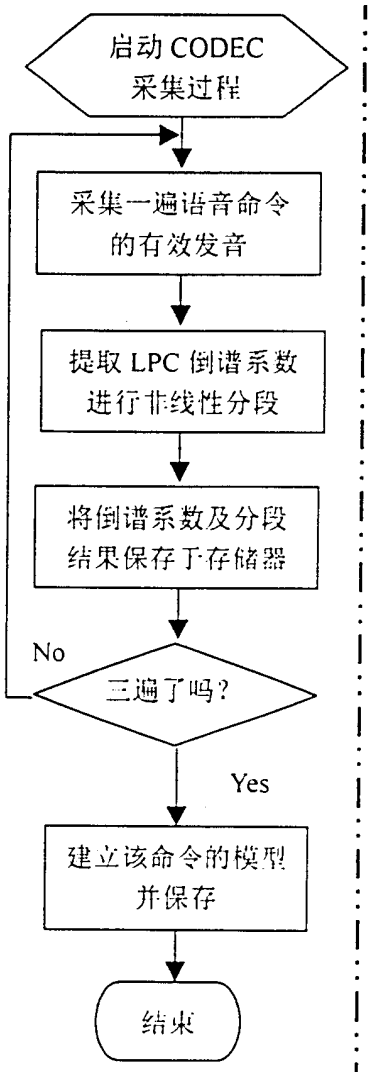

启动 CODEC
采集过程
采集一遍语音命令
的有效发音
提取 LPC 倒谱系数
进行非线性分段
将倒谱系数及分段
结果保存于存储器
No
三遍了吗?
Yes
建立该命令的模型
并保存
结束

图 2

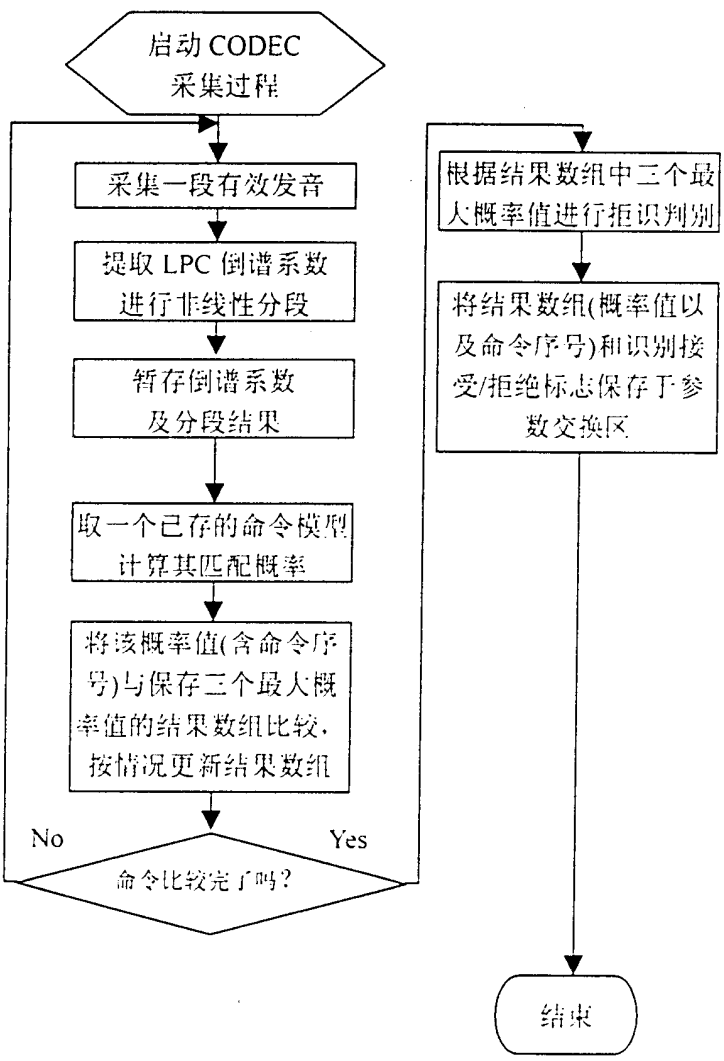

启动 CODEC
采集过程
采集一段有效发音
提取 LPC 倒谱系数
进行非线性分段
暂存倒谱系数
及分段结果
取一个已存的命令模型
计算其匹配概率
将该概率值(含命令序
号)与保存三个最大概
率值的结果数组比较,
按情况更新结果数组
No
Yes
命令比较完了吗?
根据结果数组中三个最
大概率值进行拒识判别
将结果数组(概率值以
及命令序号)和识别接
受/拒绝标志保存于参
数交换区
结束

图 3

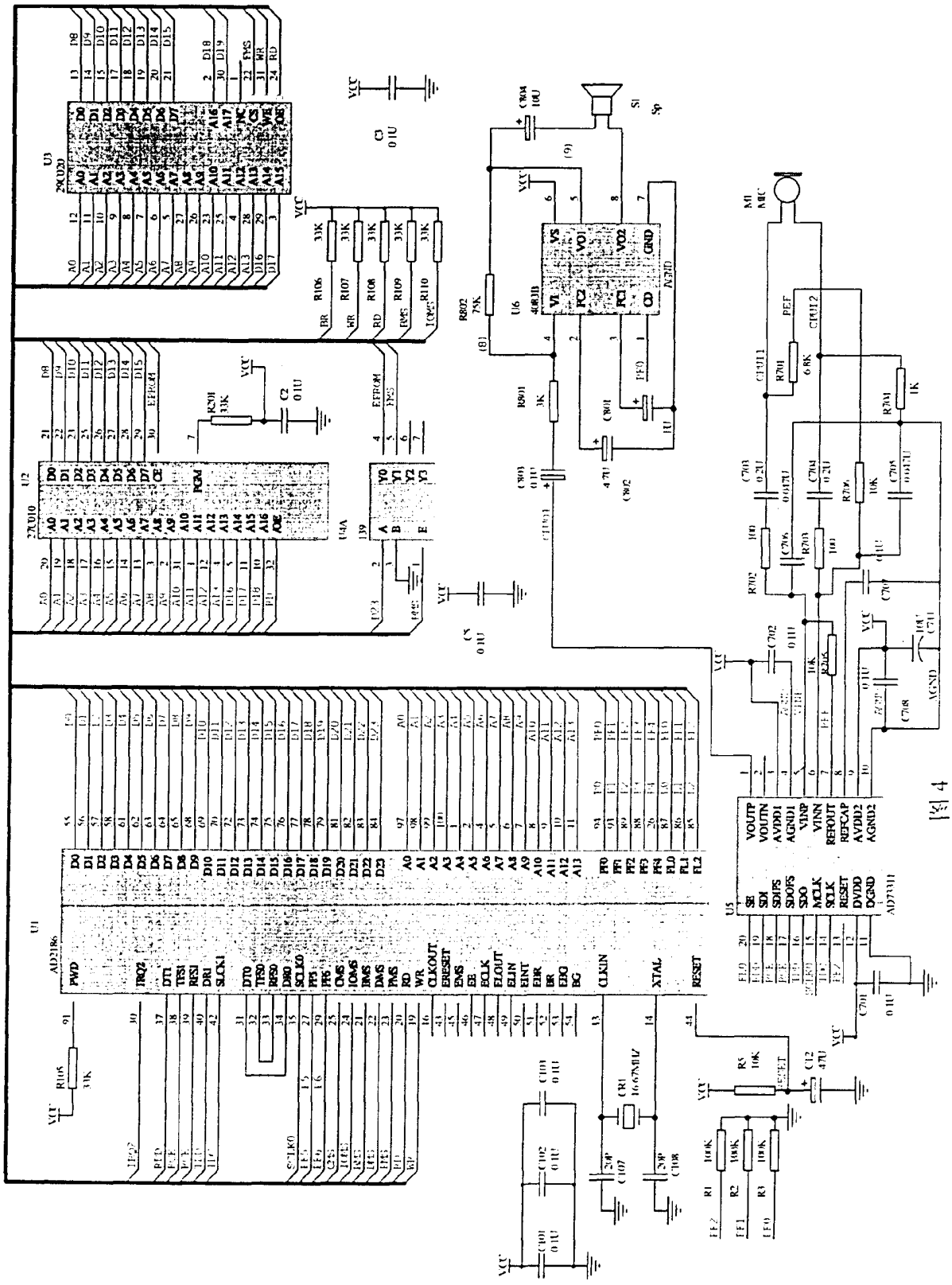


图 4