

# 基于改进的隐马尔科夫模型的语音识别方法

袁里驰<sup>1,2</sup>

- (1. 中南大学 信息科学与工程学院, 湖南 长沙, 410083;
2. 江西财经大学 信息管理学院, 江西 南昌, 330013)

**摘要:** 针对隐马尔可夫(HMM)语音识别模型状态输出独立同分布等与语音实际特性不够协调的假设以及在使用段长信息时存在的缺陷, 对隐马尔可夫模型进行改进, 提出马尔可夫族模型。马尔可夫族模型可看作一个数学上由多个马尔可夫链构成的多重随机过程, HMM模型则是双重随机过程, 因而, HMM模型可视为马尔可夫族模型的特例。马尔可夫族模型用条件独立性假设取代了HMM模型的独立性假设。相对条件独立性假设, 独立性假设是过强假设, 因而, 基于马尔可夫族模型的语音模型更符合语音实际物理过程。在马尔可夫族语音识别模型中引入状态段长信息, 能自动根据语速对语音单元段长进行调整。非特定人连续语音实验结果表明, 利用状态段长信息的改进语音识别模型比经典HMM模型的性能明显提高。

**关键词:** 隐马尔可夫模型; 马尔可夫族模型; 段长; 语音识别

**中图分类号:** TN912.34

**文献标识码:** A

**文章编号:** 1672-7207(2008)06-1303-06

## A speech recognition method based on improved hidden Markov model

YUAN Li-chi<sup>1,2</sup>

- (1. School of Information Science and Engineering, Central South University, Changsha 410083, China;
2. School of Information Technology, Jiangxi University of Finance & Economics, Nanchang 330013, China)

**Abstract:** In order to overcome the defects of the duration modeling of homogeneous hidden Markov model (HMM) in speech recognition and the unrealistic assumption that successive observations are independent and identically distribution within a state, Markov family model (MFM) was proposed. In the speech recognition model based on HMM, the time-sequence structure of speech signal was considered to be a double stochastic process, while Markov family model was a multiple stochastic process which consists of a few Markov chains, so HMM could be considered to be a special case of MFM. Moreover, independence assumption in HMM was placed by conditional independence assumption in MFM, and from the view of the statistics, the assumption of independence is stronger than that of conditional independence, so speech recognition model based on MFM is more realistic than HMM recognition mode. Markov Family model was applied to speech recognition, and duration distribution based MFM recognition mode which takes duration distribution into account and integrates the frame and segment based acoustic modeling techniques, was proposed. The speaker independent continuous speech recognition experiments show that this new recognition model has better performance than standard HMM recognition models.

**Key words:** hidden Markov model; Markov family model; duration; speech recognition

收稿日期: 2008-06-05; 修回日期: 2008-07-28

基金项目: 国家自然科学基金资助项目(60663007); 中南大学博士后科学基金资助项目(2007)

通信作者: 袁里驰(1973-), 男, 湖南邵阳人, 博士后, 副教授, 从事信息检索与语音识别研究; 电话: 0791-3076768; E-mail: yuan\_lich@hotmail.com

隐马尔可夫模型<sup>[1]</sup>(hidden Markov model, 即 HMM)被认为是语音识别领域中最成功的统计模型之一。HMM 对语音信号的时间序列结构建立统计模型, 将之看作一个数学上的双重随机过程: 一个是用具有有限状态数的 Markov 链来模拟语音信号统计特性变化的隐含随机过程, 另一个是与 Markov 链的每一个状态相关联的观测序列随机过程。HMM 模型可分为离散隐马尔可夫模型<sup>[1]</sup>(采用离散概率密度函数, 简称 DHMM)和连续隐马尔可夫模型<sup>[1]</sup>(采用连续概率密度函数, 简称 CDHMM)以及半连续隐马尔可夫模型<sup>[1]</sup>(SCHMM, 集中了 DHMM 和 CDHMM 的特点)。

近几十年来, 有关语音识别的研究取得了许多进展, 尤其是在出现了隐马尔可夫模型以后, 大词汇量连续语音识别<sup>[2]</sup>取得了重大突破, 但其识别性能仍有待提高, 体现在: 首先, 经典的 HMM 是齐次的马尔可夫模型, 其状态转移概率为常数, 状态驻留长度服从指数分布<sup>[3]</sup>, 这与语音的实际物理过程不符; 其次, 由于人在自然发音过程中发音器官的运动惯性, 相邻帧的语音特征之间有很强的时间相关性。而在现有 HMM 应用中, 为了处理方便, 引入了一个重要的状态输出特征矢量独立分布的假设。针对 HMM 语音识别模型现有的一些缺陷, 本文作者提出马尔可夫族模型(Markov family model, 即 MFM)。马尔可夫族模型(MFM)是由多个马尔可夫连构成的多重随机过程, 随机过程相互之间有一定的概率关系, 该模型用条件独立性假设取代隐马尔可夫模型的独立性假设。

## 1 隐马尔可夫模型与马尔可夫族模型

### 1.1 隐马尔可夫模型

**定义 1** (隐马尔可夫模型) 隐马尔可夫模型<sup>[1]</sup>是一个五元组 $(S, A, V, B, \pi)$ 。其中:  $S = \{s_1, \dots, s_N\}$ , 是状态集;  $V = \{v_1, \dots, v_M\}$ , 是输出符号集合;

$$\pi = \{\pi_1, \dots, \pi_N\}, \quad 1 \leq i \leq N, \quad (1)$$

是初始状态概率分布;  $A = (a_{i,j})_{N \times N}$ , 是状态转移概率分布矩阵;

$$a_{i,j} = P(X_{t+1} = s_j | X_t = s_i), \quad (2)$$

是从状态  $s_i$  转移到状态  $s_j$  的概率;  $B = (b_{j,k})_{N \times M}$ ,

是状态符号发射的概率分布矩阵;

$$b_{i,k} = P(o_t = v_k | X_t = s_i), \quad 1 \leq k \leq M, \quad 1 \leq i \leq N, \quad (3)$$

表示在状态  $s_i$  时输出符号  $v_k$  的概率。

从隐马尔可夫模型的定义可以看出, HMM 模型是一个数学上的双重随机过程; 建立在以下 3 个基本假设的基础上: 马尔可夫性假设; 不动性假设; 输出独立性假设。

### 1.2 马尔可夫族模型

**定义 2** 马尔可夫族模型(Markov family model) 令  $X = \{X_1, X_2, \dots, X_m\}$  表示  $m$  维随机向量, 其中分量  $X_i (1 \leq i \leq m)$  取值于有限状态集  $S_i (1 \leq i \leq m)$ 。分量  $X_i (1 \leq i \leq m)$  构成马尔可夫族模型, 若它们满足下列条件:

a. 每一个分量  $X_i (1 \leq i \leq m)$  都是一个  $n_i$  阶马尔可夫链:

$$P(x_{i,t} | x_{i,1}, \dots, x_{i,t-1}) = P(x_{i,t} | x_{i,t-n_i+1}, \dots, x_{i,t-1}). \quad (4)$$

b. 分量在时刻  $t$  出现某一个状态的概率只与该分量在时刻  $t$  以前状态和时刻  $t$  其他分量的状态有关:

$$\begin{aligned} P(x_{i,t} | x_{1,1}, \dots, x_{1,t}, \dots, x_{i,1}, \dots, x_{i,t-1}, \dots, x_{m,1}, \dots, x_{m,t}) = \\ P(x_{i,t} | x_{i,t-n_i+1}, \dots, x_{i,t-1}, x_{1,t}, \dots, x_{i-1,t}, x_{i+1,t}, \dots, x_{m,t}). \end{aligned} \quad (5)$$

c. 条件独立性假设:

$$\begin{aligned} P(x_{i,t-n_i+1}, \dots, x_{i,t-1}, x_{1,t}, \dots, x_{i-1,t}, x_{i+1,t}, \dots, x_{m,t} | x_{i,t}) = \\ P(x_{i,t-n_i+1}, \dots, x_{i,t-1} | x_{i,t}) \cdot P(x_{1,t} | x_{i,t}) \cdots P(x_{m,t} | x_{i,t}). \end{aligned} \quad (6)$$

条件 a 表明马尔可夫族模型是多重随机过程, 而隐马尔可夫模型可看作一个数学上的双重随机过程。从这个意义说, 隐马尔可夫模型可视为马尔可夫族模型的特例; 条件 b 明确了马尔可夫族模型的多重随机过程相互之间的关系, 利用该特性简化马尔可夫族模型的计算。根据条件 c, 某分量在时刻  $t$  的值已知的条件下, 该变量在时刻  $t$  以前的  $n_i-1$  个取值, 与其他分量在时刻  $t$  取值是相互独立的, 即马尔可夫族模型用条件独立性假设取代了隐马尔可夫模型中的独立性假设。从统计学的角度来说, 相对条件独立性假设, 独立性假设是过强假设, 与语音、语言现象也不尽符合。因而, 可以说, 基于马尔可夫族模型的语音、语言模型比基于隐马尔可夫模型的语音、语言模型更符合语音、语言的实际物理过程。

## 2 基于段长分布的 MFM 语音识别模型

在连续语音中, 不同说话者在不同语境下说话的速度差异是很大的。偏离正常语速过大往往会造成识别错误, 过快的语速会使删除错误增加, 过慢的语速会造成插入错误增加, 从而使识别性能下降。目前, 人们对这个问题的研究主要是先按照某种方法得到识别语料的语速度量, 然后, 根据语速的快慢更新转移概率, 将慢速情况下的状态自转移概率增大, 离开转移概率变小, 快速时则反之, 从而调整每个语音单元的持续时间以适应语速<sup>[4]</sup>。

基于段长的语音识别模型<sup>[3, 5-8]</sup>, 直接从段长出发, 说话速度的变化直接反映为段长的变化, 同时, 语速变化对段长的影响是同步增长或同步下降的, 即在慢速情况下, 前一个语音单元长于平均段长, 后一个语音单元也会以相同的趋势长于其平均段长, 而在快速下情况正好相反。并且在一个较短的时间段内, 1个说话者的说话速度会比较稳定, 即一个短时间段内这种语速对段长的影响可以认为是基本一致的。这样, 就可以用前一个语音单元段长对其均值的偏差来预测后一个语音单元段长的变化趋势<sup>[9-13]</sup>。

语音识别统计模型通常以音子<sup>[1]</sup>为最基本的识别单位<sup>[14]</sup>。假设音子模型允许出现的状态有  $L$  种, 记为  $s_l (l=1, \dots, L)$ , 在时刻  $n (n \geq 1)$  所处的状态用  $x_n$  表示,  $y_n$  表示状态  $x_n$  的观察特征, 系统在状态  $x_n$  连续驻留的时间长度(简称为段长)用  $\tau_n$  表示。记有关概率为:

$$a_l = a_{s_l} = P(x_n = s_l), \quad l=1, \dots, L;$$

$$a_{i,j} = a_{s_i, s_j} = P(x_{n+1} = s_j | x_n = s_i), \quad i, j=1, \dots, L;$$

$$b_l(y_n) = b_{s_l}(y_n) = P(y_n | x_n = s_l), \quad l=1, \dots, L.$$

假设识别系统的词汇表容量为  $V$ , 其中每一个词条表示为  $w_v, v=1 \sim V$ , 设每一词条  $w_v$  中包含  $L_v$  个状态, 记为  $s_{l_v}^v, l_v=1 \sim L_v$ 。现假定一个完整的句子所相应的特征序列, 即观察序列为  $O=\{o_1, o_2, \dots, o_T\}$ , 待识别句子的词序列为  $W=\{w_1, w_2, \dots, w_N\}$ , 词序列中的第  $i$  ( $1 \leq i \leq N$ ) 个词  $w_i$  对应的第  $j$  ( $1 \leq j \leq L_i$ ) 个状态记为  $s_{i,j}^i$ , 系统在状态  $s_{i,j}^i$  连续驻留的时间长度(段长)记为  $\tau_{i,j}$ , 则系统所经历的状态序列为:

$$S = \underbrace{\{s_1^1, \dots, s_1^1\}}_{\tau_{1,1} \uparrow}, \dots, \underbrace{\{s_{L_1}^1, \dots, s_{L_1}^1\}}_{\tau_{1,L_1} \uparrow}, \dots, \underbrace{\{s_j^i, \dots, s_j^i\}}_{\tau_{i,j} \uparrow}, \dots, \underbrace{\{s_{L_N}^N, \dots, s_{L_N}^N\}}_{\tau_{N,L_N} \uparrow}.$$

语音识别的任务就是由观察序列  $O=\{o_1, o_2, \dots, o_T\}$  来求最佳的状态序列  $S$ , 进而求得最佳的词序列  $W=\{w_1, w_2, \dots, w_N\}$ , 即搜索最可能的词序列:

$$\arg \max_W P(S, W | O) = \arg \max_W \frac{P(O | W) \cdot P(W)}{P(O)} = \arg \max_W P(O | W) \cdot P(W). \quad (7)$$

其中:  $P(W) = \prod_{i=1}^N p(w_i | w_{i-K+1}, w_{i-K+2}, \dots, w_{i-1}), K \geq 2$ 。

当  $K$  为 2, 3 时, 分别称为双词文法、三词文法。

设  $\Gamma$  表示对应于词序列  $W$  的可能状态序列  $S$  的集合, 则有

$$P(O | W) = \sum_{\Gamma} P(O, S | W) = \sum_{\Gamma} P(O | S, W) \cdot P(S | W). \quad (8)$$

设

$$T_1 = 1, T_i = \sum_{u=1}^{i-1} \sum_{v=1}^{L_u} \tau_{u,v} + 1, \quad 2 \leq i \leq N,$$

$$T_{i,1} = T_i, T_{i,j} = T_i + \sum_{v=1}^{j-1} \tau_{i,v}, \quad 2 \leq j \leq L_i,$$

表示段的分割点, 并假定  $O=\{o_1, o_2, \dots, o_T\}$  是一个  $M$  阶马尔可夫链, 则有

$$P(O | S, W) \approx \prod_{i=1}^N \prod_{j=1}^{L_i} \prod_{k=1}^{\tau_{i,j}^{-1}} P(o_{T_{i,j}+k} | o_{T_{i,j}+k-M+1}^{T_{i,j}+k-1}, s_{i,j}^i, w_i) \approx \prod_{i=1}^N \prod_{j=1}^{L_i} \prod_{k=1}^{\tau_{i,j}^{-1}} P(o_{T_{i,j}+k} | o_{T_{i,j}+k-M+1}^{T_{i,j}+k-1}, s_{i,j}^i). \quad (9)$$

而式(8)中概率  $P(S | W)$  的计算如下:

$$P(S | W) = P(\underbrace{s_1^1, \dots, s_1^1}_{\tau_{1,1} \uparrow} | w_1) \cdot \prod_{j=2}^{L_1} P(\underbrace{s_j^1, \dots, s_j^1}_{\tau_{1,j} \uparrow} | \underbrace{s_{j-1}^1, \dots, s_{j-1}^1}_{\tau_{1,j-1} \uparrow}, w_1) \cdot \prod_{i=2}^N [P(\underbrace{s_1^i, \dots, s_1^i}_{\tau_{i,1} \uparrow} | \underbrace{s_{L_{i-1}}^{i-1}, \dots, s_{L_{i-1}}^{i-1}}_{\tau_{i-1,L_{i-1}} \uparrow}, w_{i-1}, w_i) \cdot \prod_{j=2}^{L_i} P(\underbrace{s_j^i, \dots, s_j^i}_{\tau_{i,j} \uparrow} | \underbrace{s_{j-1}^i, \dots, s_{j-1}^i}_{\tau_{i,j-1} \uparrow}, w_i)]. \quad (10)$$

其中:

$$\begin{aligned}
 & P(\underbrace{s_j^i, \dots, s_j^i}_{\tau_{i,j} \uparrow} | \underbrace{s_{j-1}^i, \dots, s_{j-1}^i}_{\tau_{i,j-1} \uparrow}, w_i) = \\
 & P(x_{n+1} = s_j^i, \tau_{n+1} = \tau_{i,j} | x_n = s_{j-1}^i, \tau_n = \tau_{i,j-1}, w_i) \approx \\
 & P(\tau_{n+1} = \tau_{i,j} | x_{n+1} = s_j^i, x_n = s_{j-1}^i, \tau_n = \tau_{i,j-1}, w_i) \cdot \\
 & P(x_{n+1} = s_j^i | x_n = s_{j-1}^i, w_i). \quad (11)
 \end{aligned}$$

条件概率  $P(x_{n+1} = s_j^i | x_n = s_{j-1}^i, w_i)$  可取近似值  $P(x_{n+1} = s_j^i | x_n = s_{j-1}^i) = a_{s_{j-1}^i, s_j^i}$ , 由马尔可夫族模型的条件独立性假设, 有

$$\begin{aligned}
 & P(\tau_{n+1} = \tau_{i,j} | x_{n+1} = s_j^i, x_n = s_{j-1}^i, \tau_n = \tau_{i,j-1}, w_i) = \\
 & P(\tau_{n+1} = \tau_{i,j} | x_{n+1} = s_j^i, \tau_n = \tau_{i,j-1}, w_i) = \\
 & \frac{P(x_{n+1} = s_j^i, \tau_n = \tau_{i,j-1} | \tau_{n+1} = \tau_{i,j}, w_i) P(\tau_{n+1} = \tau_{i,j} | w_i)}{P(x_{n+1} = s_j^i, \tau_n = \tau_{i,j-1} | w_i)} \cdot \\
 & \frac{P(x_{n+1} = s_j^i | \tau_{n+1} = \tau_{i,j}, w_i) P(\tau_n = \tau_{i,j-1} | \tau_{n+1} = \tau_{i,j}, w_i) P(\tau_{n+1} = \tau_{i,j} | w_i)}{P(x_{n+1} = s_j^i, \tau_n = \tau_{i,j-1} | w_i)}. \quad (12)
 \end{aligned}$$

由贝叶斯定理, 有

$$\begin{aligned}
 & P(x_{n+1} = s_j^i | \tau_{n+1} = \tau_{i,j}, w_i) = \\
 & \frac{P(\tau_{n+1} = \tau_{i,j} | x_{n+1} = s_j^i, w_i) \cdot P(x_{n+1} = s_j^i | w_i)}{P(\tau_{n+1} = \tau_{i,j} | w_i)}; \quad (13) \\
 & P(\tau_n = \tau_{i,j-1} | \tau_{n+1} = \tau_{i,j}, w_i) = \\
 & \frac{P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}, w_i) \cdot P(\tau_n = \tau_{i,j-1} | w_i)}{P(\tau_{n+1} = \tau_{i,j} | w_i)}. \quad (14)
 \end{aligned}$$

将式(13)和(14)代入式(12)可得:

$$\begin{aligned}
 & P(\tau_{n+1} = \tau_{i,j} | x_{n+1} = s_j^i, \tau_n = \tau_{i,j-1}, w_i) \approx \\
 & \frac{P(\tau_{n+1} = \tau_{i,j} | x_{n+1} = s_j^i, w_i) \cdot P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}, w_i)}{P(\tau_{n+1} = \tau_{i,j} | w_i)}. \quad (15)
 \end{aligned}$$

式(10)中的其他参数也可通过类似的计算得到。  
条件概率

$$P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}, w_i)$$

的计算利用了相邻 2 个语音单元间的段长相关信息, 称这样实现的系统为段长二元概率系统。当然, 也可利用相邻  $r$  个语音单元间的段长相关信息, 则相应的系统称为段长  $r$  元概率系统。

由于数据稀疏,  $P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}, w_i)$  可取近似值  $P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1})$  或用平滑方法得到:

$$P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}, w_i) \approx (1 - \lambda_{w_i}).$$

$$P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}) +$$

$$\lambda_{w_i} P(\tau_{n+1} = \tau_{i,j} | \tau_n = \tau_{i,j-1}, w_i). \quad (16)$$

其中:  $\lambda_w$  为平滑参数,  $0 < \lambda_w < 1$ ;  $\tau_{i,j}$  为状态段长, 但也可以是描写半音节(声母和韵母)或音节的段长。因此, 这里提供的模型和算法具有很大的灵活性。在经典的 HMM 语音识别模型中, 状态  $i$  的驻留概率  $a_{i,j}$  为常数, 系统进入状态  $i$  后在该状态连续驻留的时间  $\tau$  即段长服从几何分布<sup>[1]</sup>:

$$P(\tau) = a_{i,i}^{\tau-1} (1 - a_{i,i}), \quad \tau \geq 1. \quad (17)$$

实验统计结果表明, 经典 HMM 模型这种段长的几何分布形式不能很好地描述语音的段长特征。为此, 许多研究者对模型进行改进, 将状态段长概率直接引入统计模型中。常用的段长分布形式有 Gamma 分布、高斯分布、泊松分布和均匀分布等。

### 3 实验结果

为了验证本文提出的语音识别方法, 对连续语音进行试验。大词汇量连续语音试验所采用的语音数据是“863”计划提供的男女声各 83 人的连续语音录音数据。每个说话人对应一个 520 句话到 650 句话不等的文件, 其中 9 个文件用于识别(慢速, 适中语速, 快速 3 类各 3 个文件), 另外 74 个文件用于训练。采用的特征<sup>[15-17]</sup>是 14 维 MFCC 特征及其一阶差分和二阶差分, 归一化能量及其一阶差分和二阶差分, 共 45 维特征, 结果见表 1。

表 1 非特定人连续语音识别的实验结果

Table 1 Experimental results of speaker-independent continuous speech recognition

模型	替代错误率/%	插入错误率/%	删除错误率/%	总错误率/%
HMM 语音识别模型	22.83	3.35	0.78	26.96
基于段长的 MFM 模型	20.22	2.14	0.57	22.93

表 1 中的 HMM 语音识别实验基于连续的隐 Markov 模型，利用英国剑桥大学提供的 HTK(HMM Tool Kit) 工具包中编码、训练、识别等相关工具。从表 1 可以看出：替代错误率从 22.83%下降到 20.22%，插入错误率从 3.35%下降到 2.14%，删除错误率从 0.78%下降到 0.57%，总错误率从 26.96%下降到 22.93%，相对下降近 15%。对大词汇量连续语音的试验结果表明：基于段长分布的马尔可夫族语音识别模型(DDBMFM)克服了传统 HMM 不合理的独立性假设，对识别性能有明显的改善。该模型同时对偏离正常语速的语料进行识别时，能自动根据语速对语音单元段长进行调整，从而降低由语速带来的插入错误和删除错误，获得了更精确的分割点，并因此减少了替代错误，从而提高了系统的性能。

4 结 论

- a. 在隐马尔可夫模型的基础上，提出了一种新的统计模型即马尔可夫族模型。马尔可夫族模型是多重随机过程，而隐马尔可夫模型可看作一个数学上的双重随机过程，因而隐马尔可夫模型可视为马尔可夫族模型的特例。马尔可夫族模型用条件独立性假设取代隐马尔可夫模型中的独立性假设，从统计学的角度来说，相对条件独立性假设，独立性假设是过强假设，与语音、语言现象也不尽符合。因而，基于马尔可夫族模型的语音、语言模型比基于隐马尔可夫模型的语音、语言模型更符合语音、语言的实际物理过程。
- b. 将马尔可夫族模型应用于语音识别，同时在语音识别模型中直接引入状态段长信息。基于段长的语音识别模型直接从段长出发，说话速度的变化直接反

映为段长的变化，能自动根据语速进行语音单元段长的调整。结果表明，利用状态段长信息的 MFM 语音识别模型比经典 HMM 模型的性能明显提高。

c. 马尔可夫族模型是一种新的统计模型，有关马尔可夫族模型的理论及其在语音识别和自然语言处理等领域中的应用有待进一步研究。

参考文献：

[1] Rabiner L, Juang B H. Fundamentals of speech recognition[M]. New Jersey: Prentice Hall, 1993.

[2] Chang E, ZHOU Jian-lai, SHOU Di, et al. Large vocabulary mandarin speech recognition with different approaches in modeling tones[C]//Proceedings of the 6th International Conference on Spoken Language Processing(ICSLP 2000). San Jose: IEEE Press, 2000: 983-986.

[3] Mitchell C D, Jamieson L H. Modeling duration in a hidden Markov model with the exponential family[C]//Proceedings of the IEEE International conference on Acoustic, Speech, Signal Process (ICASSP 1993). San Jose: IEEE Press, 1993: 331-334.

[4] Shinoda K, Lee C. A structural Bayes approach to speaker adaptation[J]. IEEE Transaction on Speech and Audio Processing, 2001, 9(3): 276-287.

[5] Vaseghi S V. State duration modeling in hidden Markov models[J]. Journal of Signal Processing, 1995, 41(1): 31-41.

[6] Lai W H, Chen S H. Analysis of syllable duration models for mandarin speech[C]//Proceedings of the IEEE International conference on Acoustic, Speech, Signal Process (ICASSP 2002). San Jose: IEEE Press, 2002: 497-500.

[7] WANG Zuo-ying, XIAO Xi. Duration distribution based HMM speech recognition models[J]. Chinese Journal of Electronics, 2004, 32(1): 46-49.

[8] Hon H W, Wang K S. Unified frame and segment based models for automatic speech recognition[C]//Proceedings of the IEEE International conference on Acoustic, Speech, Signal Process (ICASSP 2000). San Jose: IEEE Press, 2000: 1017-1020.

[9] GONG Yi-fan. Stochastic trajectory modeling and sentence searching for continuous speech recognition[J]. IEEE Transactions on Speech Audio Processing, 1997, 5(1): 33-44.

[10] WANG W J, CHEN S H. The study of prosodic modeling for mandarin speech[C]//Proceedings of the International Computer Symposium (ICS). Hualien: IEEE Computer Society Press, 2002: 1777-1784.

[11] 严斌峰, 朱小燕, 张智江, 等. 基于邻接空间的鲁棒语音识别

- 方法[J]. 软件学报, 2007, 18(11): 878–883.
- YAN Bin-feng, ZHU Xiao-yan, ZHANG Zhi-jiang, et al. Robust speech recognition based on neighborhood space[J]. Journal of Software, 2007, 18(11): 878–883.
- [12] 刘宇红, 刘 桥, 任 强. 基于模糊聚类神经网络的语音识别方法[J]. 计算机学报, 2006, 29(10): 1894–1900.
- LIU Yu-hong, LIU Qiao, REN Qian. Speech recognition based on fuzzy clustering neural network[J]. Chinese Journal of Computers, 2006, 29(10): 1894–1900.
- [13] 唐 赟, 刘文举, 徐 波. 基于后验概率解码段模型的汉语语音数字串识别[J]. 计算机学报, 2006, 29(4): 635–641.
- TANG Yun, LIU Wen-ju, XU Bo. Mandarin digit string recognition based on segment model using posterior probability decoding[J]. Chinese Journal of Computers, 2006, 29(4): 635–641.
- [14] 张 辉, 杜利民. 汉语连续语音识别中不同基元声学模型的复合[J]. 电子与信息学报, 2006, 28(11): 2045–2049.
- ZHANG Hui, DU Li-min. Combination of acoustic models trained from different unit sets for Chinese continuous speech recognition[J]. Journal of Electronics & Information Technology, 2006, 28(11): 2045–2049.
- [15] 颜 龙, 刘 刚, 郭 军. 基于小波分析的大词汇汉语连续语音识别系统鲁棒性的研究[J]. 中文信息学报, 2006, 20(2): 60–65.
- YAN Long, LIU Gang, GUO Jun. A study on robustness of large vocabulary Chinese continuous speech recognition system based on wavelet analysis[J]. Journal of Chinese Information Processing, 2006, 20(2): 60–65.
- [16] 刘耦耕, 贺素良, 龙永红. 语音信号变速算法及其 TMS320C5402 实时实现[J]. 中南大学学报: 自然科学版, 2004, 35(1): 117–121.
- LIU Ou-geng, HE Su-liang, LONG Yong-hong. An algorithm for altering voice speed and its real-time realization based on TMS320C5402[J]. Journal of Central South University: Science and Technology, 2004, 35(1): 117–121.
- [17] 覃爱娜, 杨 勇, 陈明义. 基于 CELP 的语音编解码系统及其 DSP 实现[J]. 中南工业大学学报: 自然科学版, 2003, 34(4): 416–419.
- QIN Ai-na, YANG Yong, CHEN Ming-yi. CELP based speech coding/decoding system and its DSP realization[J]. Journal of Central South University of Technology: Natural Science, 2003, 34(4): 416–419.