

# 第六章 语音识别-隐马尔可夫模型

---

- 概述
- 模型引入
- 模型定义
- 基本问题
- 实际问题



# 第六章 语音识别-隐马尔可夫模型

---

- 概述
- 模型引入
- 模型定义
- 基本问题
- 实际问题



# 概述

---

- HMM (Hidden Markov Model), 统计模型
  - Markov Chain, 100年前
  - HMM, 60~70年代, Baum, 数学基础
  - For Speech Recognition, 70年代, Baker (CMU), Jelinek (IBM)
  - Widespread, 80年代中, Rabiner (Bell Lab.)
- 主流技术、技术基础
  - Time Alignment: 语音信号的时变性
  - Data-Driven Method: 语音特征的多样性

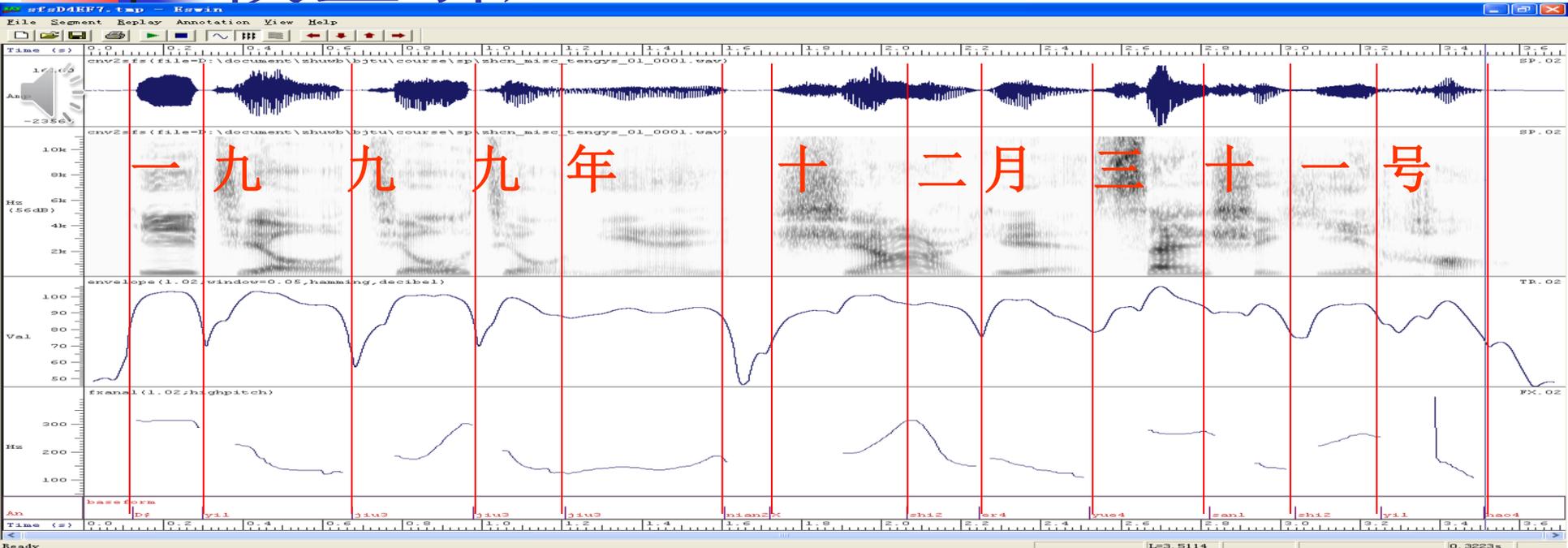


# 第六章 语音识别-隐马尔可夫模型

---

- 概述
- 模型引入
- 模型定义
- 基本问题
- 实际问题

# 模型引入



## ■ 问题

- 信号的时变性
- 特征的多样性

## ■ 对策

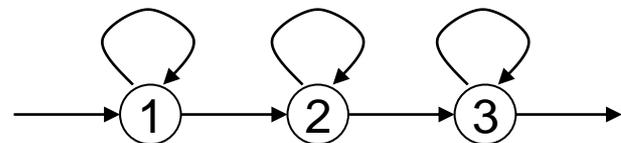
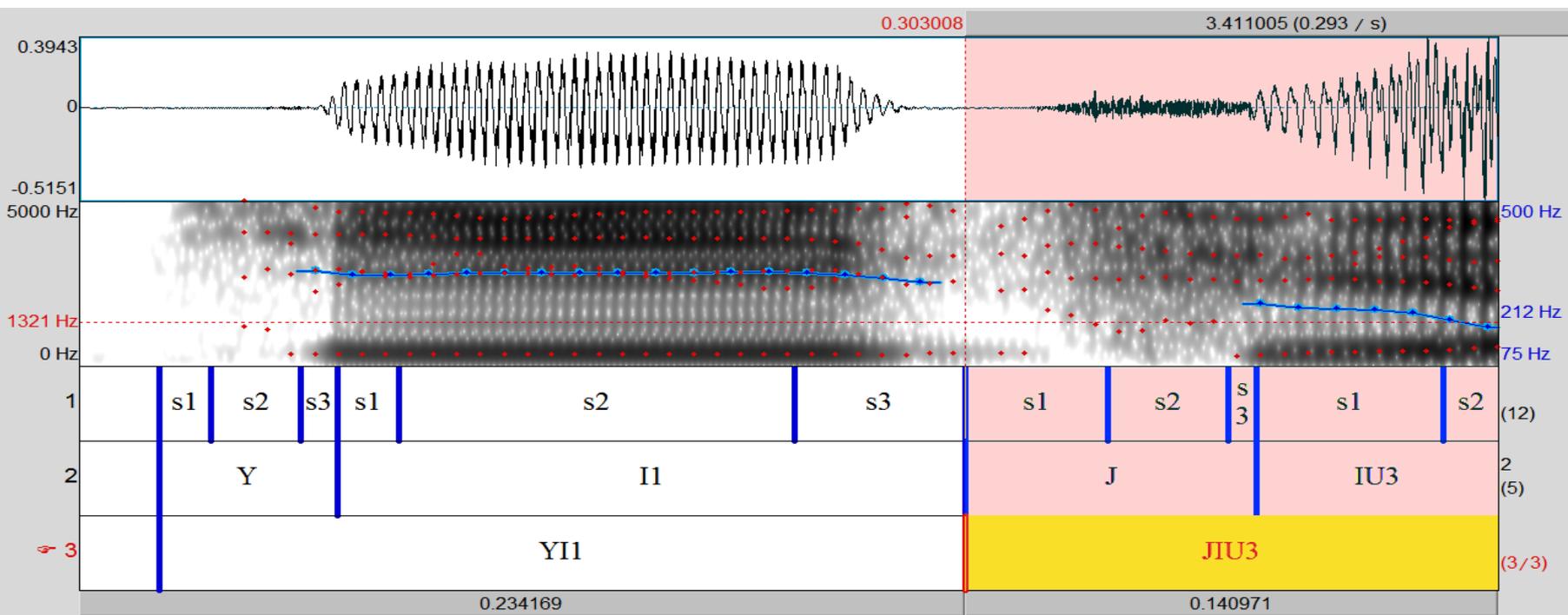
- 动态模型
- 涵盖多样性变化



# 动态模型

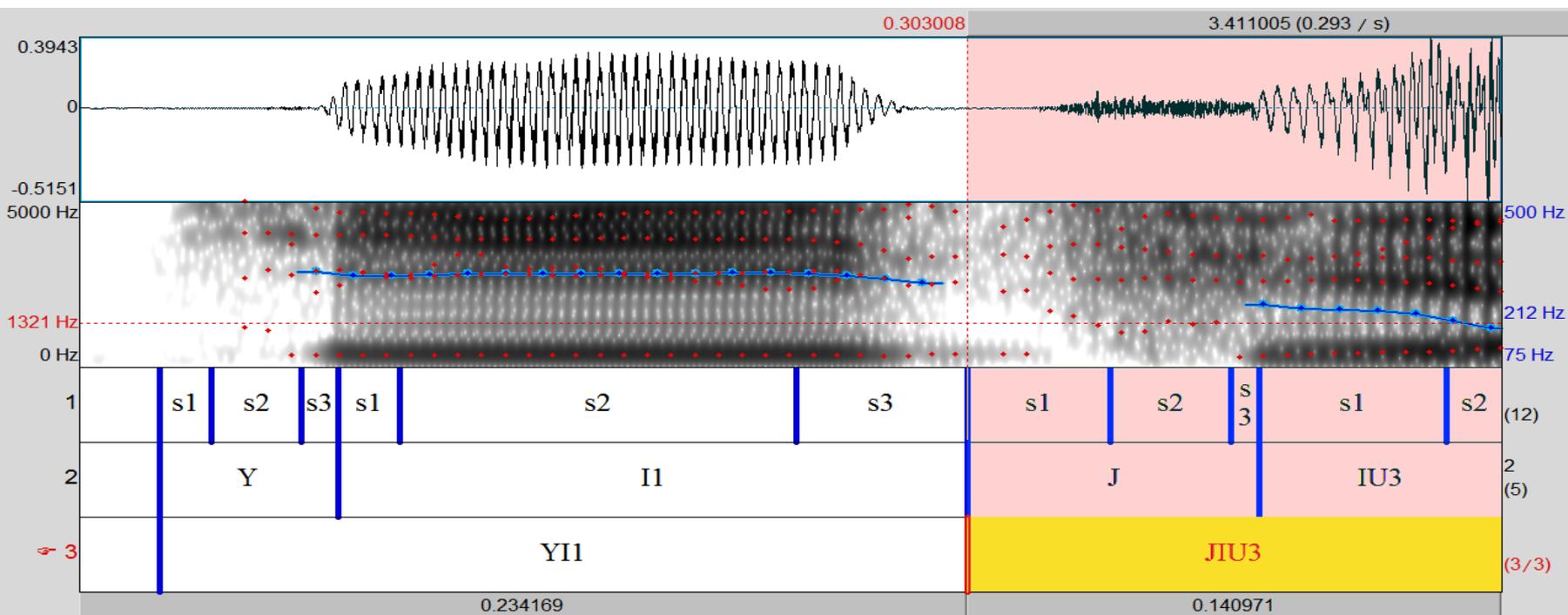
---

- 形成过程
  - 连续语音划分为为识别单元—音段序列
  - 每个音段采用一组数量有限的状态加以刻画
  - 状态的变化反应语音的变化
  - 各时刻所处状态依概率而存在/跳转

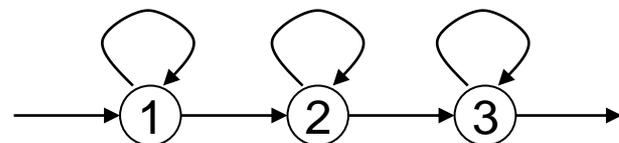


三状态自左到右无跳转HMM模型

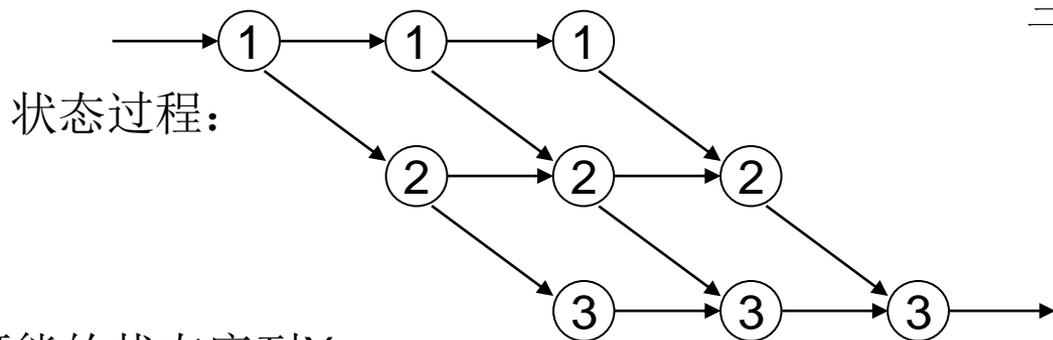




特征向量O:  $\boxed{o_1}$     $\boxed{o_2}$     $\boxed{o_3}$     $\boxed{o_4}$     $\boxed{o_5}$



三状态自左到右无跳转HMM模型



可能的状态序列X:

s1,s1,s1,s2,s3; s1,s1,s2,s2,s3; s1,s1,s2,s3,s3;  
s1,s2,s2,s2,s3; s1,s2,s2,s3,s3; s1,s2,s3,s3,s3



# 动态模型

---

- 形成过程
  - 连续语音切分为识别单元—音段
  - 每个音段采用一组数量有限的状态加以刻画
  - 状态的变化反应语音的变化
  - 各时刻所处状态依概率而存在/跳转
- 自左到右无跳转HMM模型
  - 转移概率矩阵 $A=[a_{ij}]$ ,  $i+1 \geq j \geq i$
  - 观察概率矩阵 $B=[b_i(o_t)]$
- HMM (Hidden Markov Model)
  - Hidden: 状态state为一抽象对象, 不可直接观察



# 问题的解决

---

- 短时平稳性
  - 每个状态对应于特征的平稳段
- 信号的时变性
  - 状态间的变化，反映特征的变化
  - 通过调整各个模型中在每个状态的驻留次数（时间），实现模型与信号之间的时间对齐
- 特征的多样性
  - 搜集训练数据，覆盖音段特征的变化，进而用于模型训练
  - 模型中，状态所对应的特征统计分布，反映了特征的多样性
- 对应于两个随机过程
  - 观察概率矩阵**B**：状态与特征间的统计关系
  - 状态转移矩阵**A**：状态转移统计描述平稳段间的如何转移



# 第六章 语音识别-隐马尔可夫模型

---

- 概述
- 模型引入
- 模型定义
- 基本问题
- 实际问题



# 模型定义

---

- 模型包含L个有限状态 $S_j$ 构成状态集；在时刻 $t$ ，模型输出观察(声学特征)向量为 $\mathbf{o}_t$ ，所处状态为 $\mathbf{x}_t$ 。由三个参数构成：
  - 初始状态概率向量  $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_L]$   
$$\pi_j = P_r[x_1 = S_j]$$
  - 状态转移概率矩阵 $\mathbf{A}$   $A = [a_{ij}]$   
$$a_{ij} = P_r[x_t = S_j \mid x_{t-1} = S_i]$$
  - 观察概率矩阵 $\mathbf{B}$   $B = [b_i(o_t)]$   
$$b_i(o_t) = P_r[o_t \mid x_t = S_i]$$
- HMM模型  $\lambda$  构成  $\lambda = f(A, B, \boldsymbol{\pi})$



# HMM对信号的分析过程

---

给定模型 $M$ , 计算特征序列 $O$ 与状态序列 $X$ 的联合概率

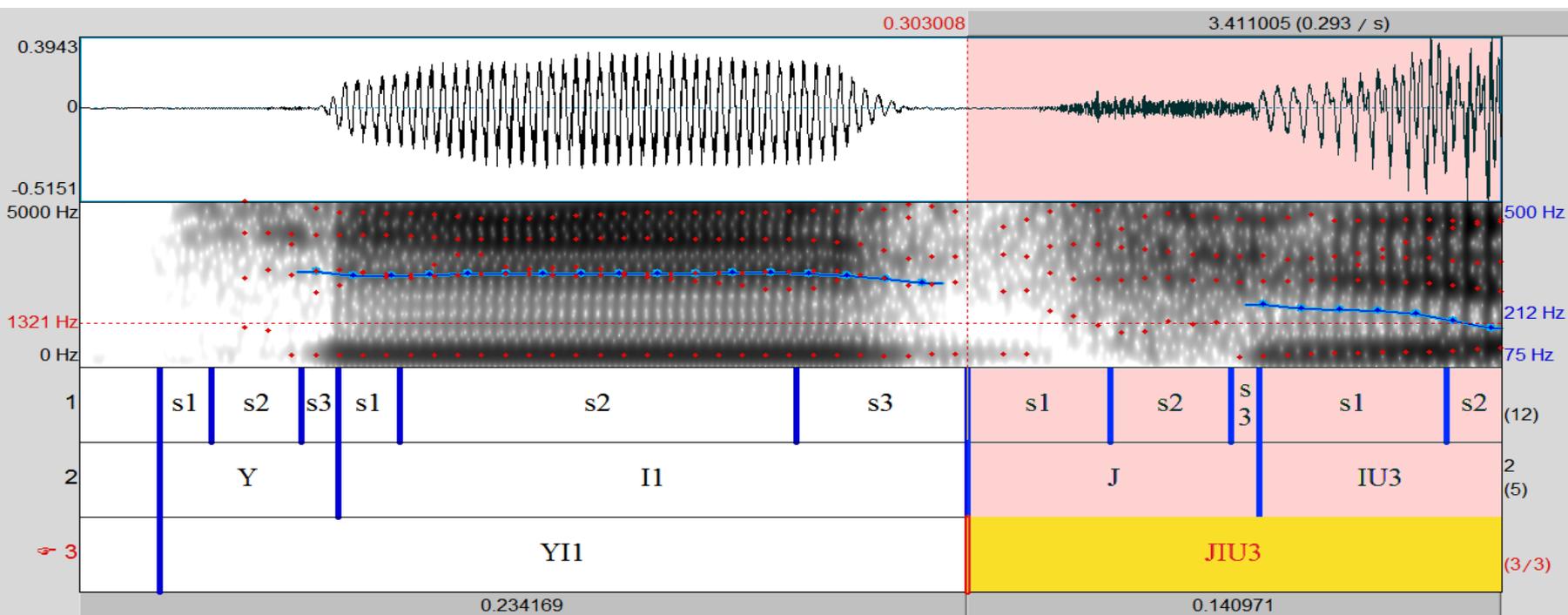
(1)  $t = 0$ 时, 根据 $\pi$ , 选一初始状态 $S_{x_0}$ 出现概率 $\pi_{x_0}$

(2) 根据 $B$ 和状态 $S_{x_t}$ 确定 $o_t$ 的概率分布 $b_{x_t}(o_t)$

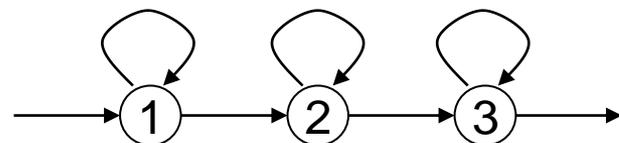
(3) 根据 $A$ 和 $S_{x_{t-1}}$ 确定转移到下一状态 $S_{x_t}$ 的转移概率 $a_{x_{t-1}x_t}$

(4)  $t = t + 1$ , 如果  $t < T$  goto(2), 否则结束

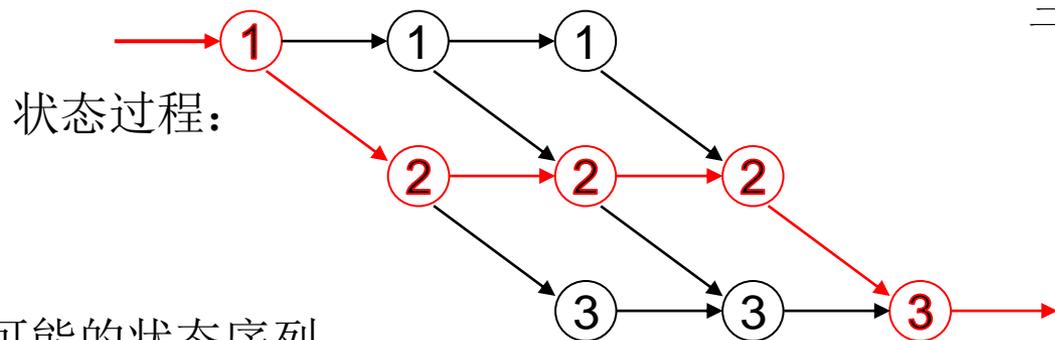
$$P(O, X | M) = \pi_{x_0} \prod_{t=1}^T a_{x_{t-1}x_t} b_{x_t}(o_t)$$



特征向量:  $\mathbf{o}_1$     $\mathbf{o}_2$     $\mathbf{o}_3$     $\mathbf{o}_4$     $\mathbf{o}_5$

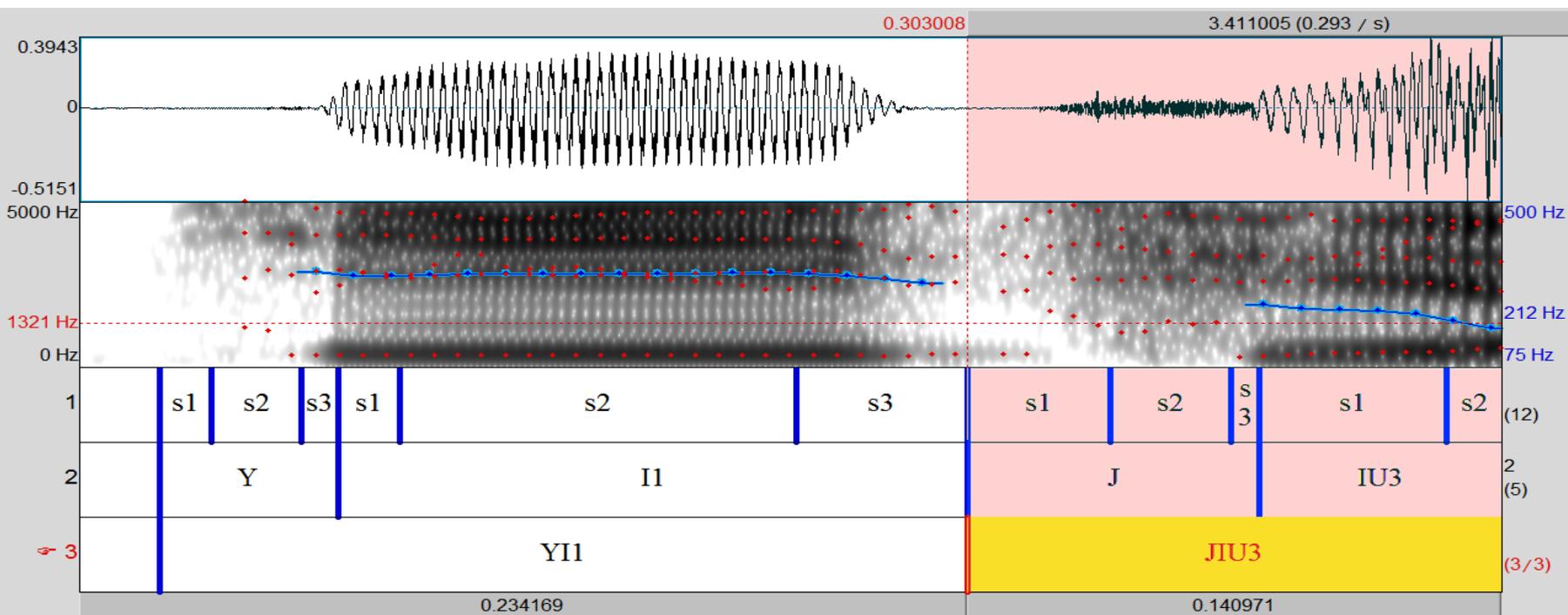


三状态自左到右无跳转HMM模型

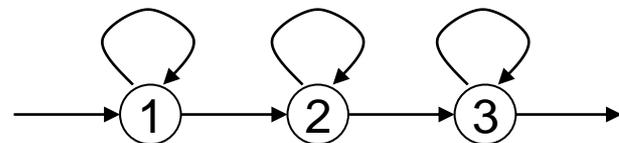


可能的状态序列

s1,s1,s1,s2,s3; s1,s1,s2,s2,s3; s1,s1,s2,s3,s3;  
**s1,s2,s2,s2,s3;** s1,s2,s2,s3,s3; s1,s2,s3,s3,s3

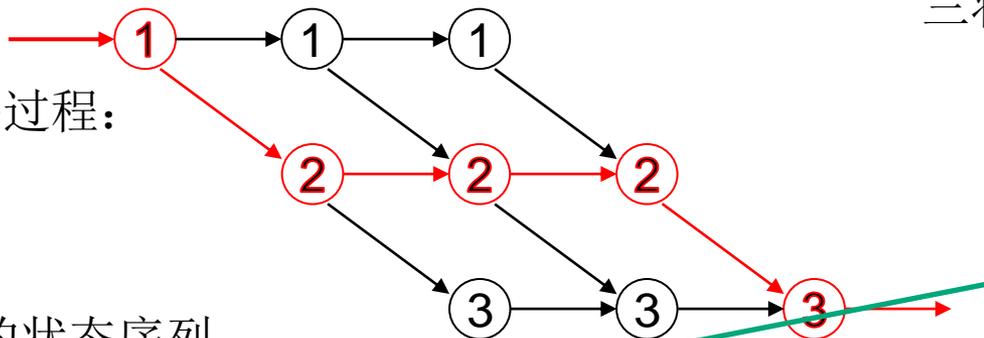


特征向量:  $\mathbf{o}_1$   $\mathbf{o}_2$   $\mathbf{o}_3$   $\mathbf{o}_4$   $\mathbf{o}_5$



三状态自左到右无跳转HMM模型

状态过程:



可能的状态序列

s1,s1,s1,s2,s3; s1,s1,s2,s2,s3; s1,s1,s2,s3,s3;  
**s1,s2,s2,s2,s3**; s1,s2,s2,s3,s3; s1,s2,s3,s3,s3

只是一种可能的序列



# 第六章 语音识别-隐马尔可夫模型

---

- 概述
- 模型引入
- 模型定义
- 基本问题
- 实际问题



# 基本问题

---

(1) 已知模型  $\lambda = f(A, B, \pi)$ , 计算指定特征矢量序列  $O$  的出现概率  $P_r(O | \lambda)$ 。 (评价问题)

“前向 - 后向”算法

(2) 已知特征矢量序列  $O$  及模型  $\lambda = f(A, B, \pi)$ , 计算最大似然状态序列  $X$ 。 (识别问题)

Viterbi算法

(3) 获得若干  $O$  后, 如何修正  $A$ 、 $B$  和  $\pi$ , 以使  $O$  出现概率最大。 (学习问题)

Baum - Welch算法



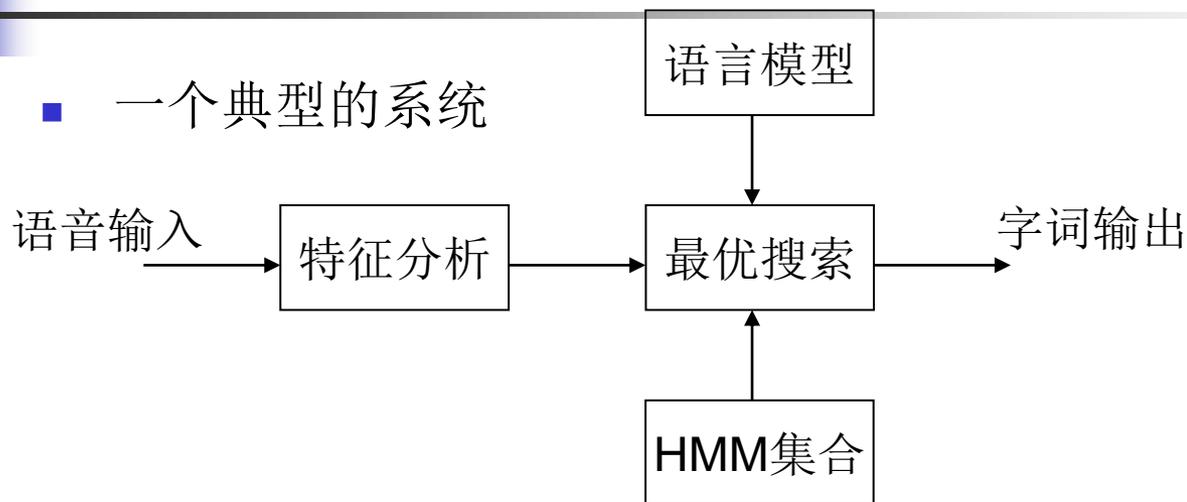
# 第六章 语音识别-隐马尔可夫模型

---

- 概述
- 模型引入
- 模型定义
- 基本问题
- 实际问题

# 实际问题

- 一个典型的系统



- 一个HMM模型对应于一个音子（phone）
- 自左到右无跳转HMM模型
- B矩阵：GMM（Gaussian-Mixture-Model）



# 思考

---

- 从语音生成的角度，理解HMM中状态与声学特征变化间的关系