

基于HMM的语音识别

洪青阳 副教授

厦门大学信息科学与技术学院
qyhong@xmu.edu.cn

要点

- ▶ 语音识别分类
- ▶ 建模单元
- ▶ 发音过程与HMM状态
- ▶ 命令词识别

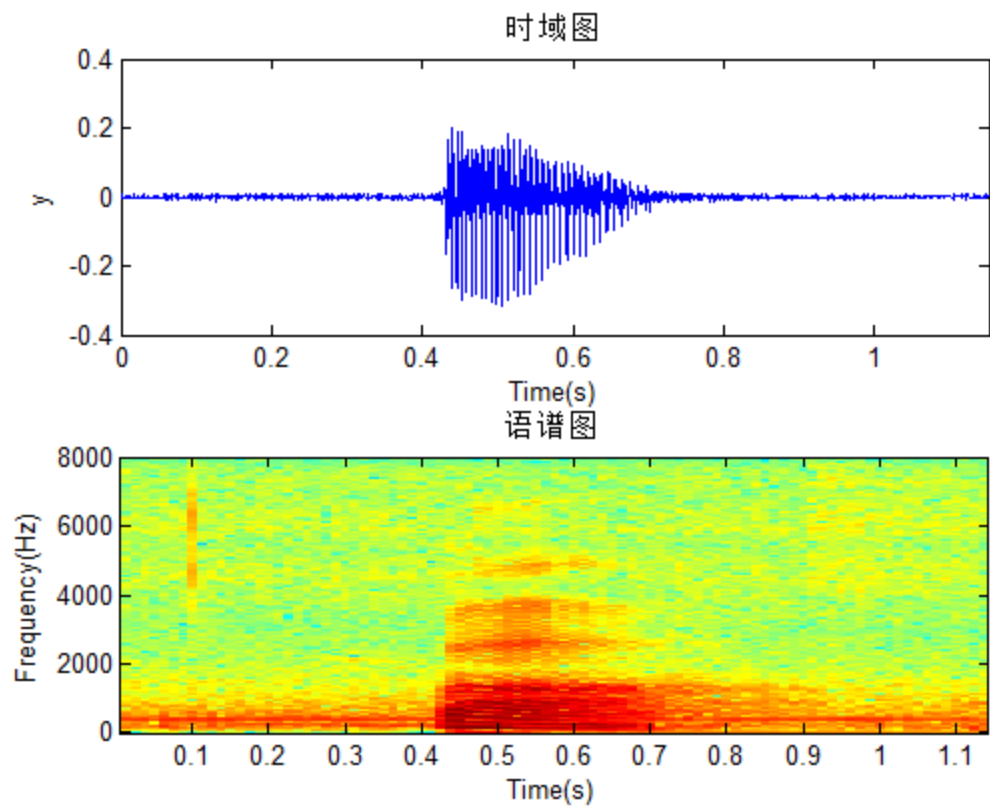
语音识别分类

- ▶ 按词汇量大小分类：
 - 小词汇量语音识别系统：几十个词
 - 中等词汇量语音识别系统：几百个到上千个词
 - 大词汇量语音识别系统：几千到几万个词
- ▶ 按说话的方式分类：
 - 孤立词语音识别系统
 - 连接词语音识别系统
 - 连续语音识别系统

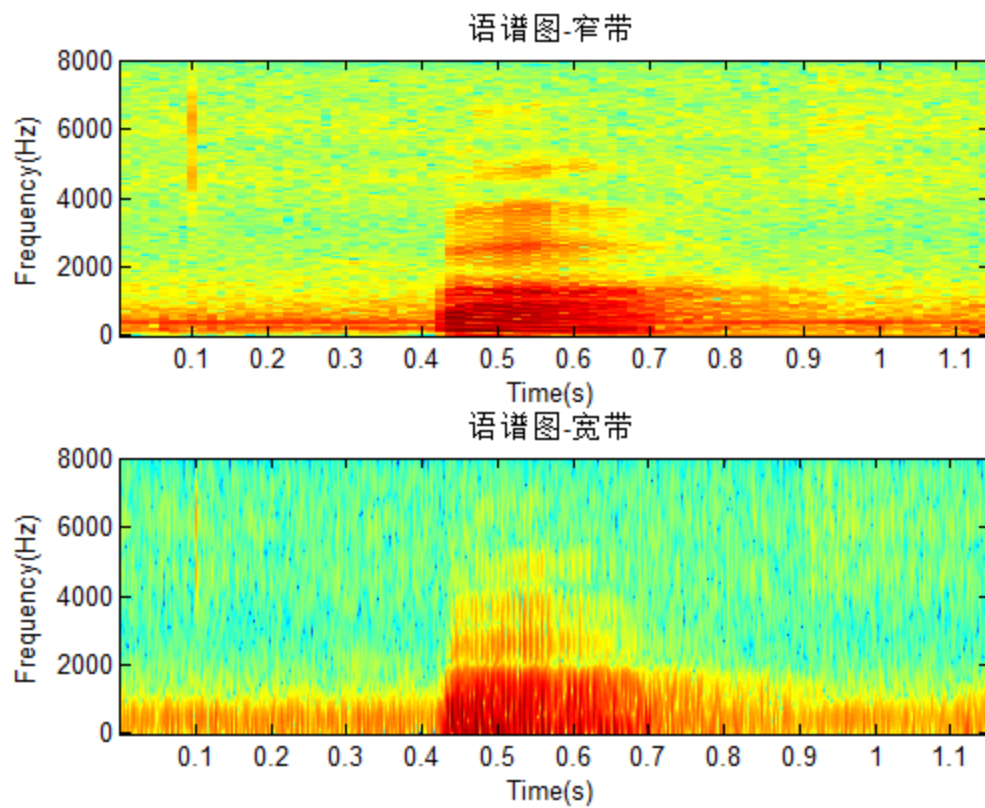
建模单元

- ▶ 整词(Word)
 - 音节(单字)
 - 词
- ▶ 音素(phone)
 - 辅音
 - 元音

音节-ba

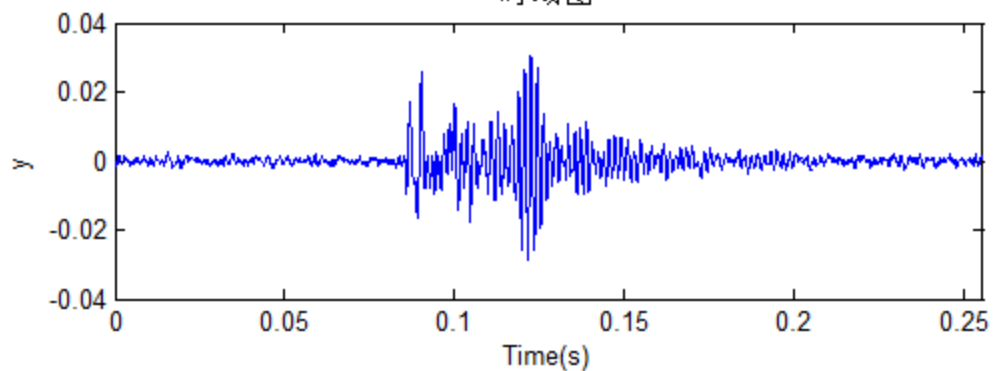


音节-ba

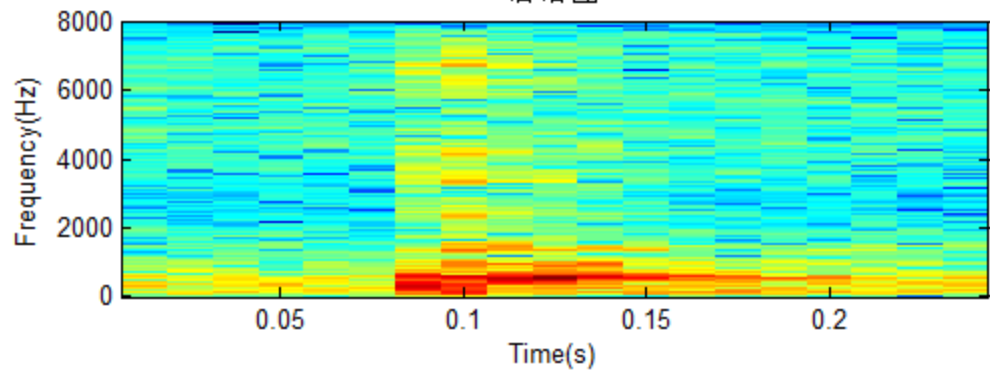


辅音-b

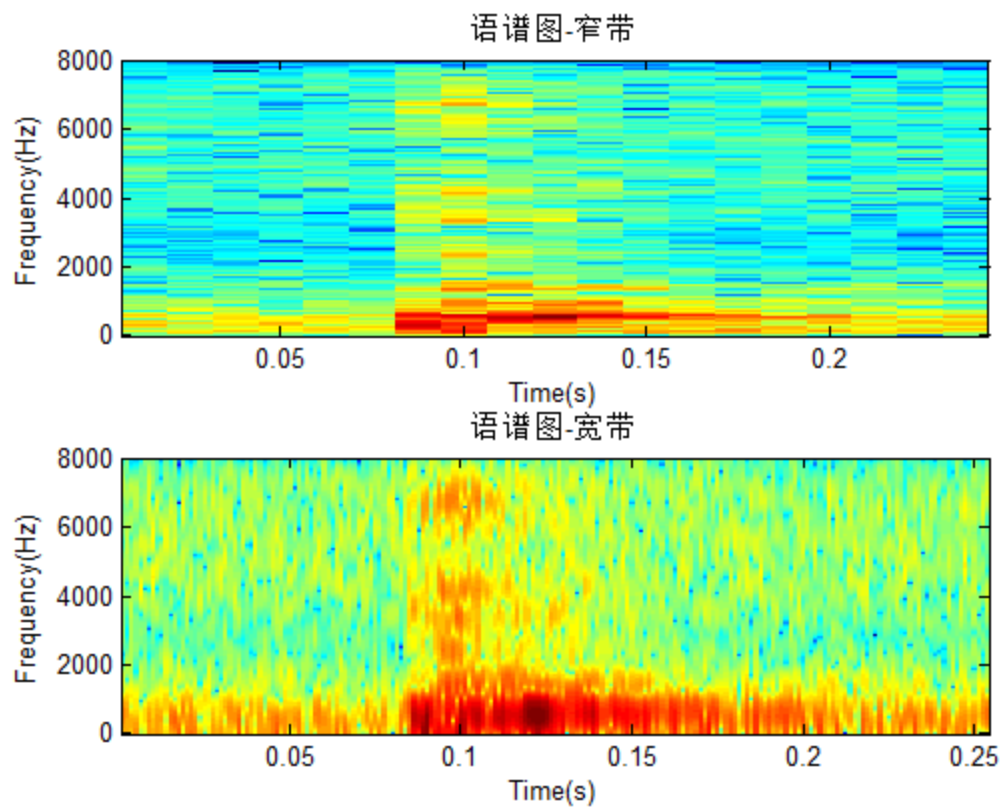
时域图



语谱图

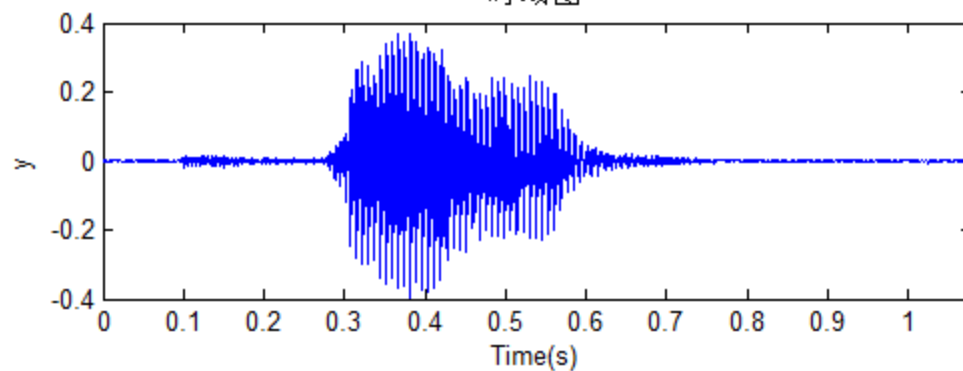


辅音-b

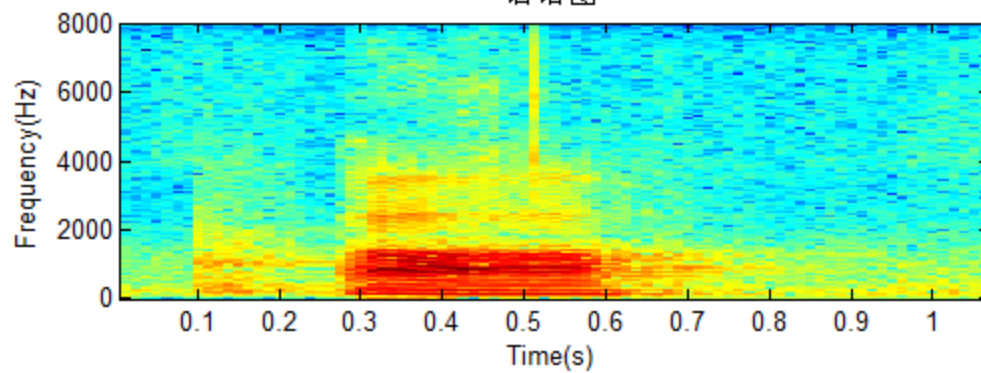


元音-a

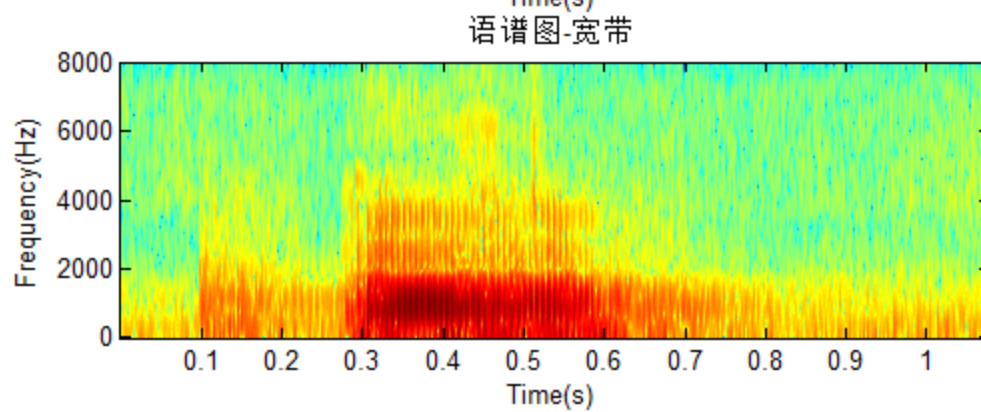
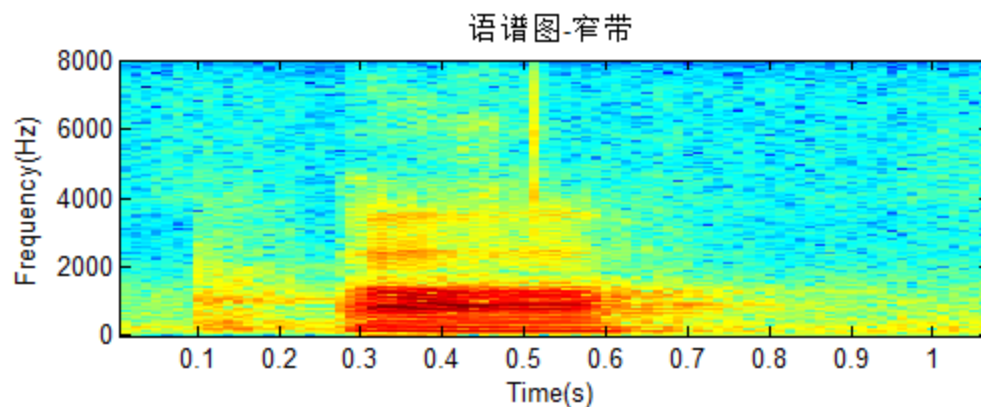
时域图



语谱图



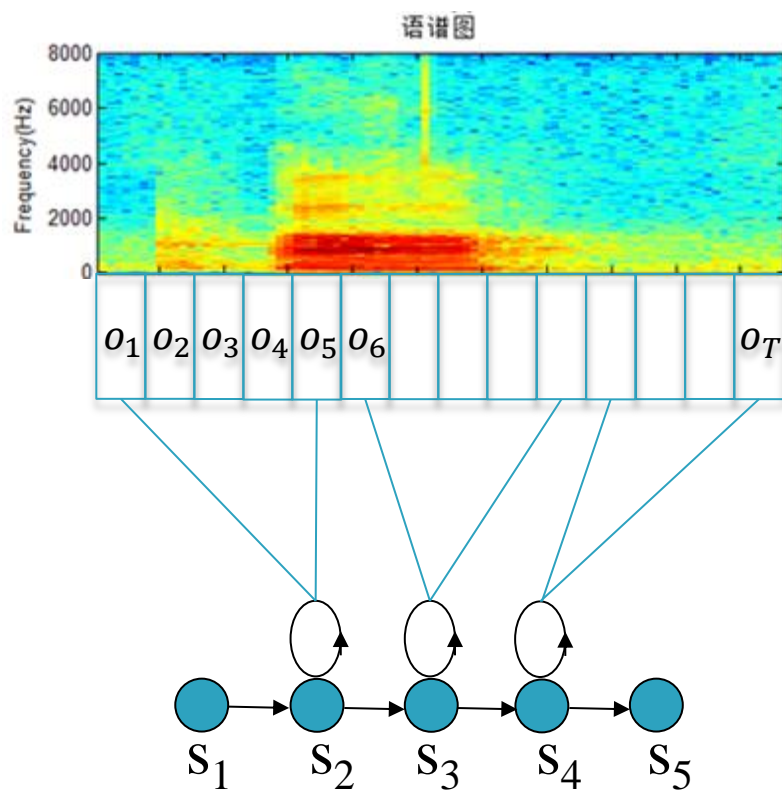
元音-a



语谱图分析

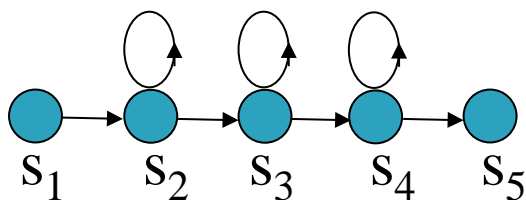
- ▶ **窄带语谱图：**具有良好的频率分辨率，但是时间分辨率较差。有利于显示基音频率及其各次谐波，不利于观察共振峰的变化。
- ▶ **宽带语谱图：**具有良好的时间分辨率，但是频率分辨率较差，不能很好反映声音的纹理特性；宽带语谱图能给出语音的共振峰频率及清辅音的能量汇集区，在语谱图里共振峰呈现为黑色的条纹。

发音过程与HMM状态



发音过程与HMM状态

发音过程的抽象描述：



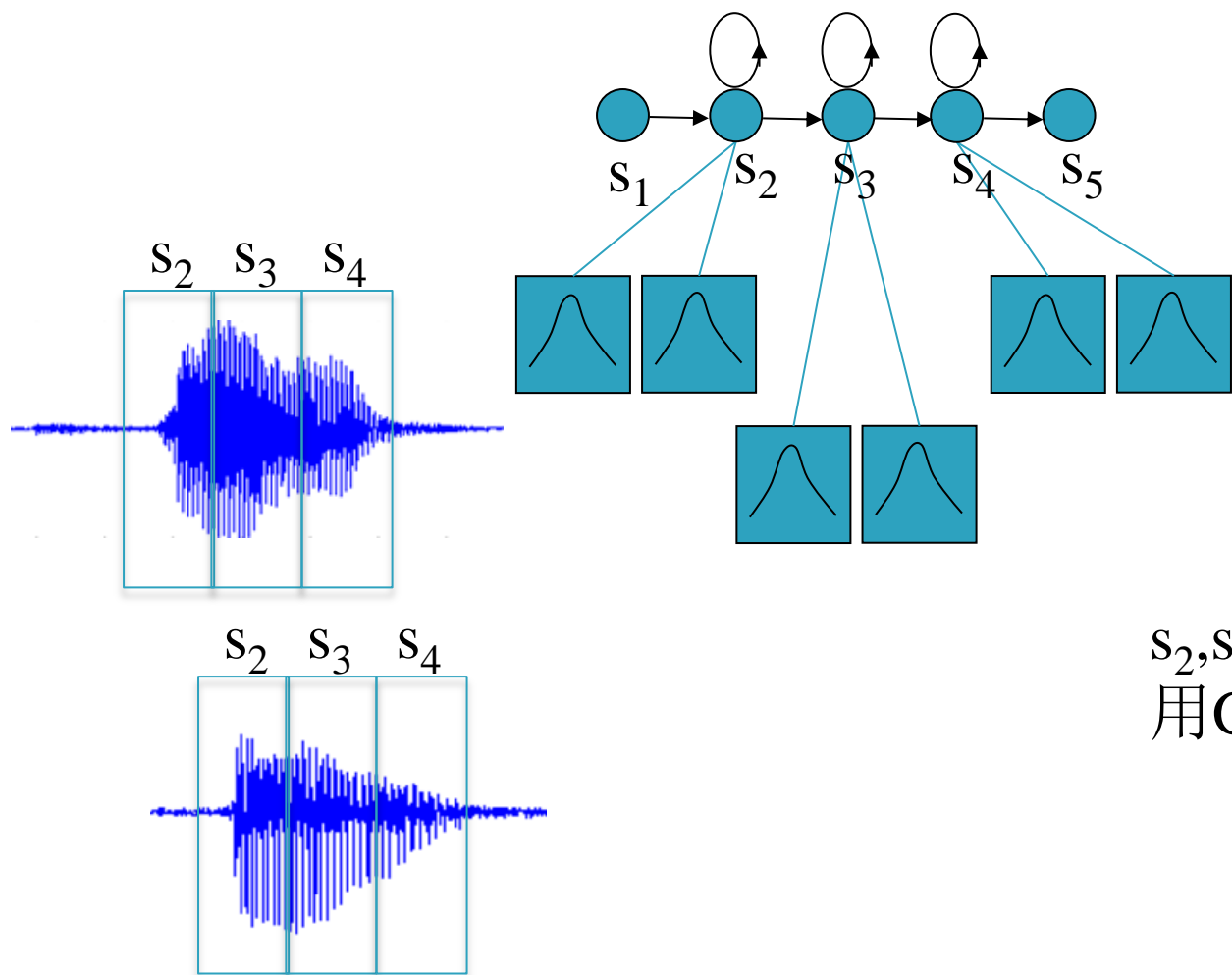
S_1 模拟发音基元的起始； S_2 、 S_3 、 S_4 模拟发音基元的发声过程；

S_5 模拟发音基元的结束。

状态起始于 S_1 ，结束于 S_5 ，且只能向自身或向右转移。

HMM的输出是与转移弧相联系的，允许不产生输出的转移，即从一个状态转移到另一个状态时，无观察值输出。这样的转移称为空转移。

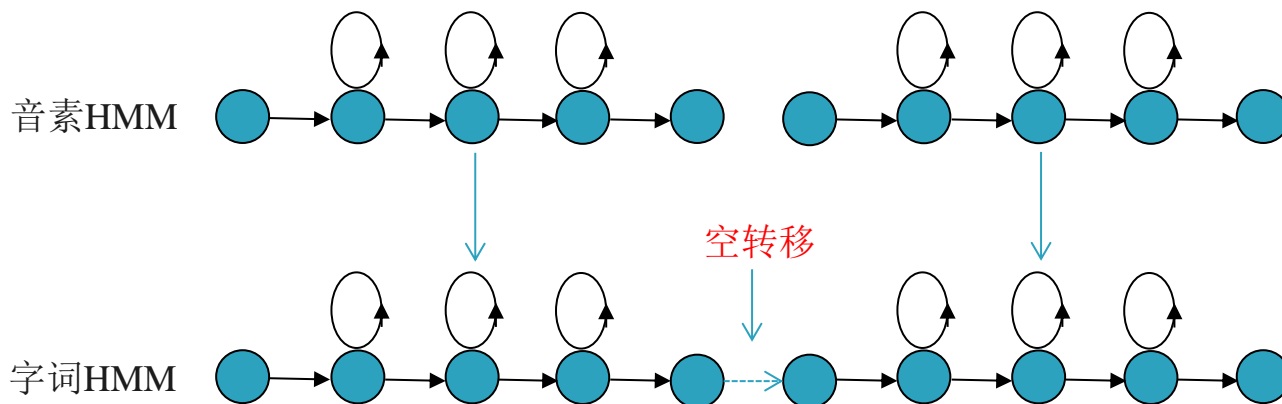
观察值分布用GMM表征



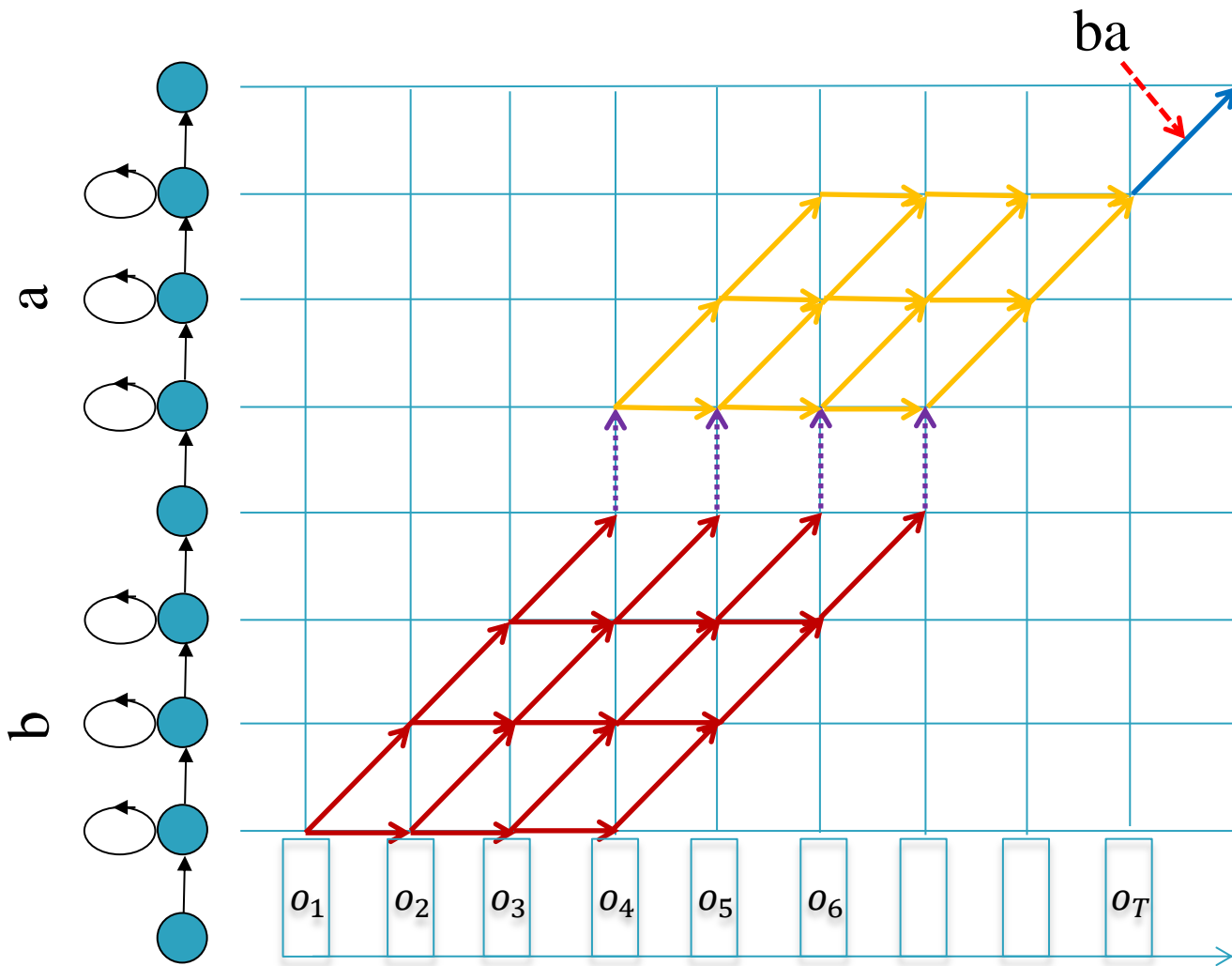
S_2, S_3, S_4 是发射状态，
用GMM表征分布

串接HMM

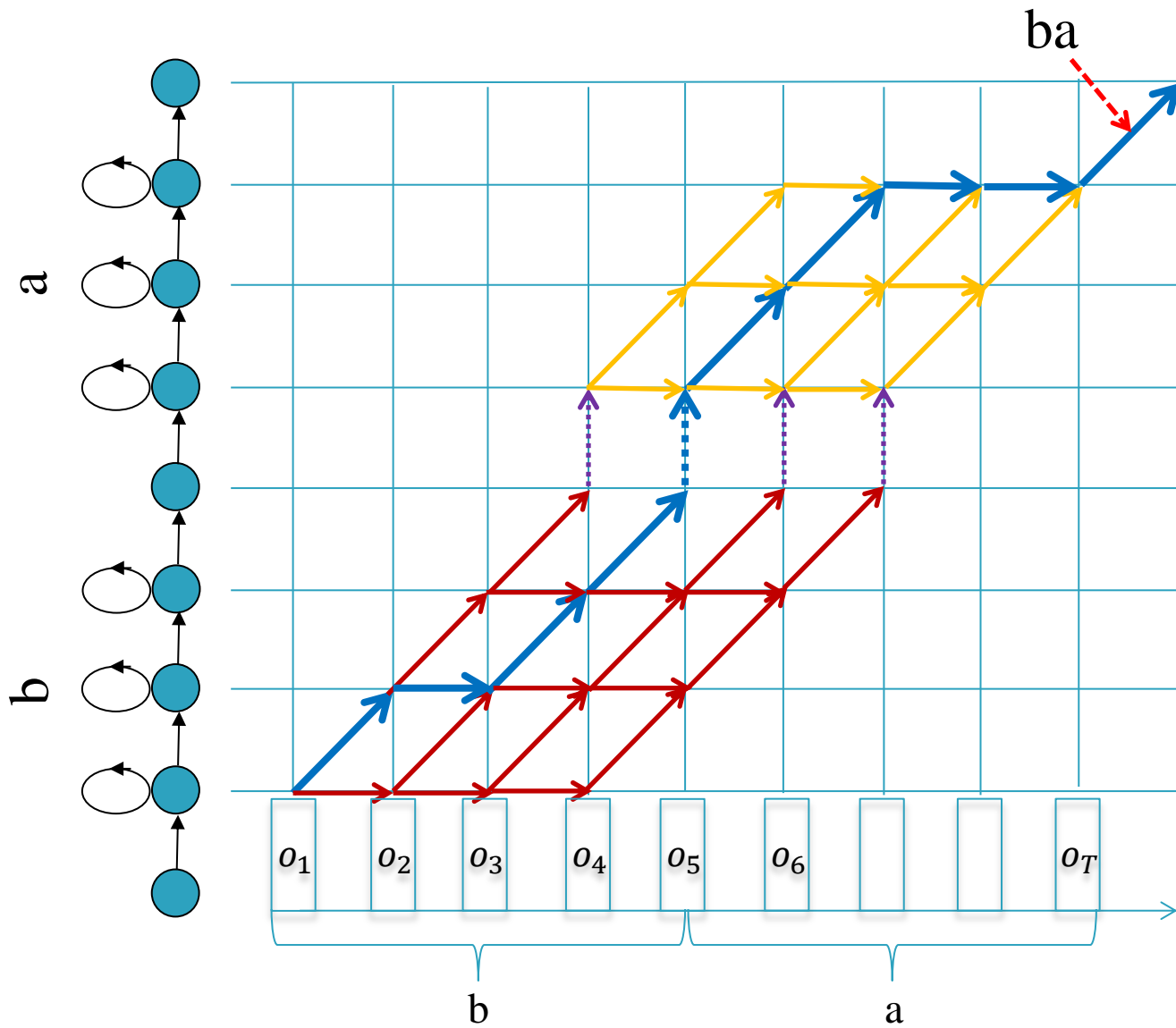
- 在连续语音识别中，字词的HMM都是由子词单元(音素)的HMM连接形成的。一般在连接时，一个子词单元HMM的终止状态和相邻基元HMM的初始状态相连接，这种连接产生的转移弧就是空转移，如下图所示：



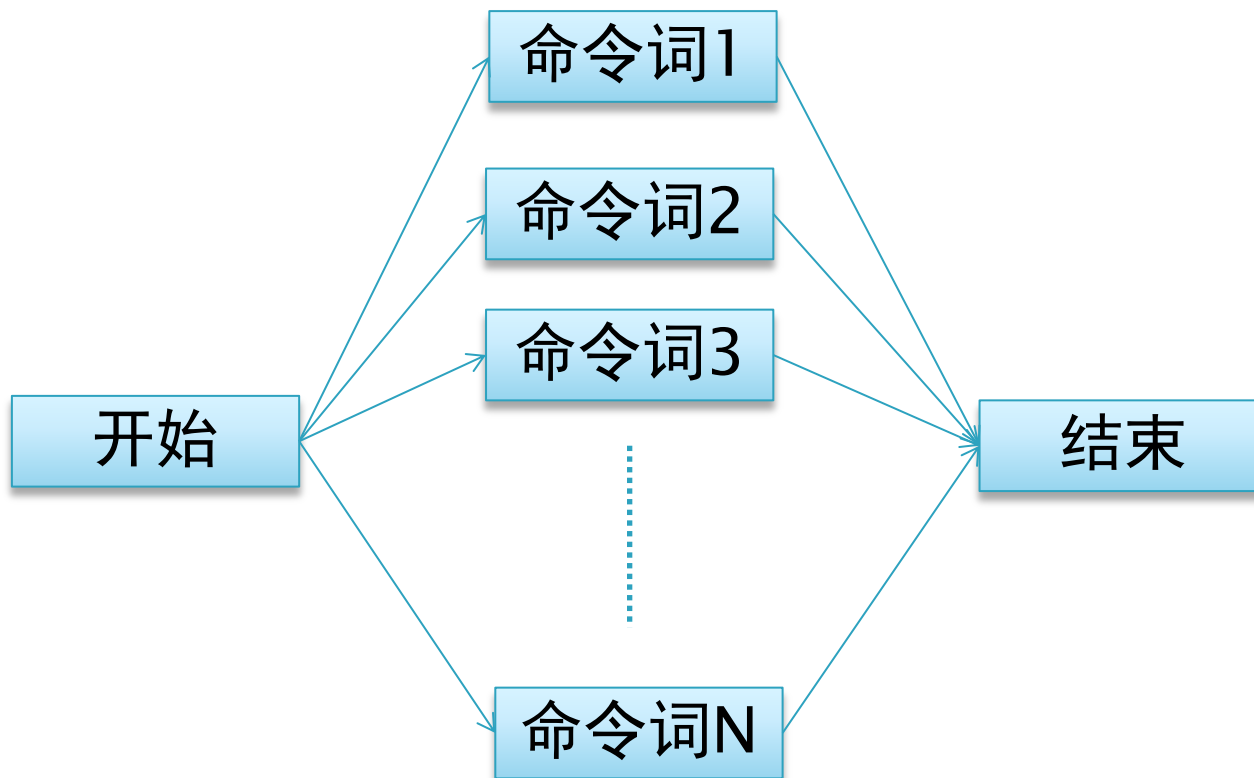
串接HMM的Viterbi识别



串接HMM的Viterbi识别



命令词识别

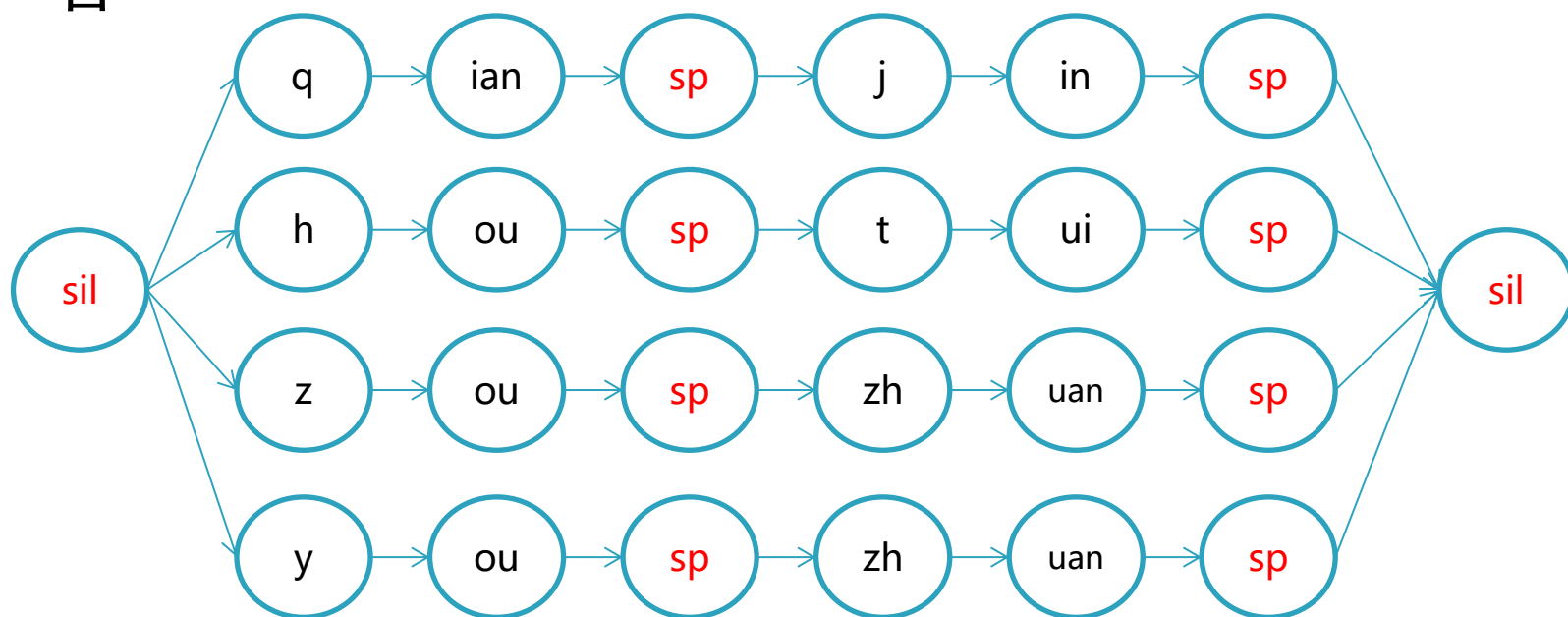


命令词

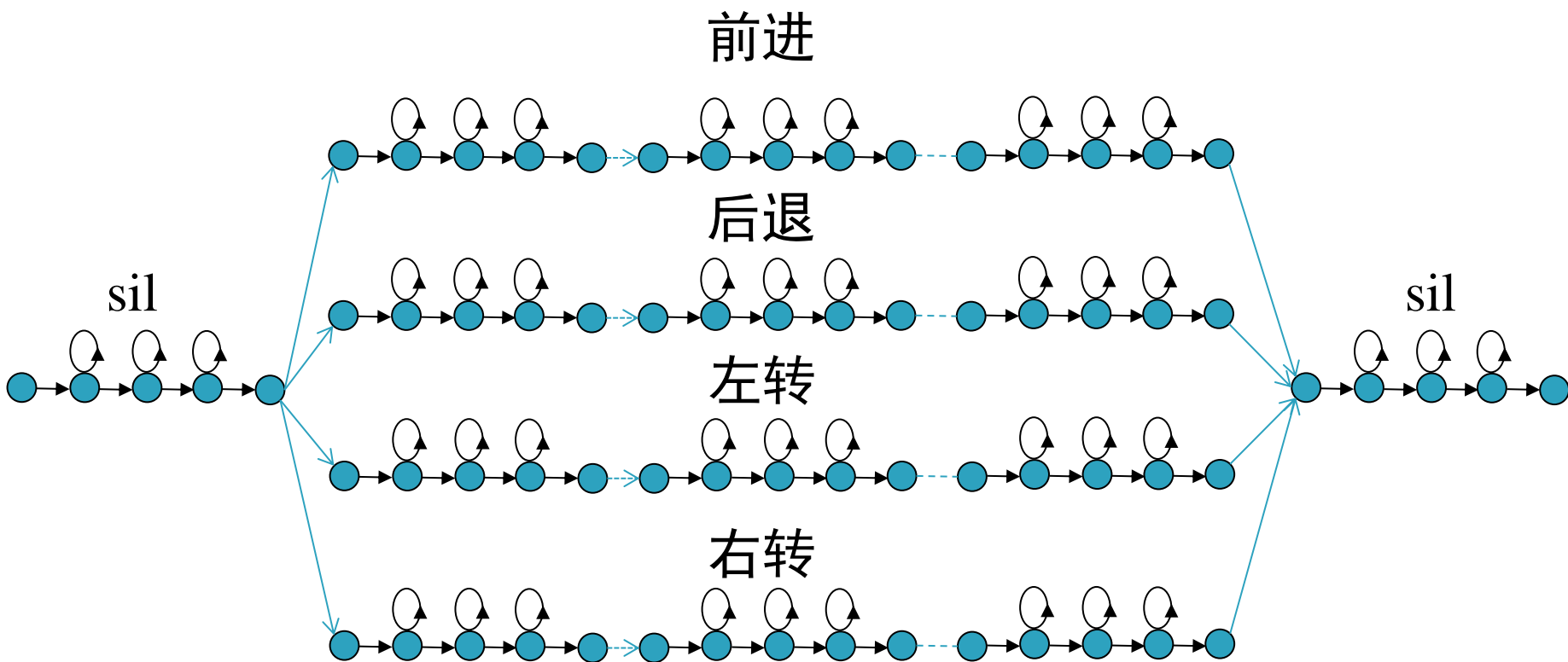
- ▶ 前进 (qian jin)
- ▶ 后退 (hou tui)
- ▶ 左转 (zou zhuan)
- ▶ 右转 (you zhuan)

识别网络的构建

- ▶ 以语法的形式限定待识别的内容
\$action = 前进 | 后退 | 左转 | 右转;
- ▶ 将字、词转化为对应的发音序列
- ▶ 增加**静音 (sil)**和**停顿 (sp)**节点，以适应语音前后的静音



对应的HMM网络



Thank you!

Any questions?