

## 5- سنتز آماری

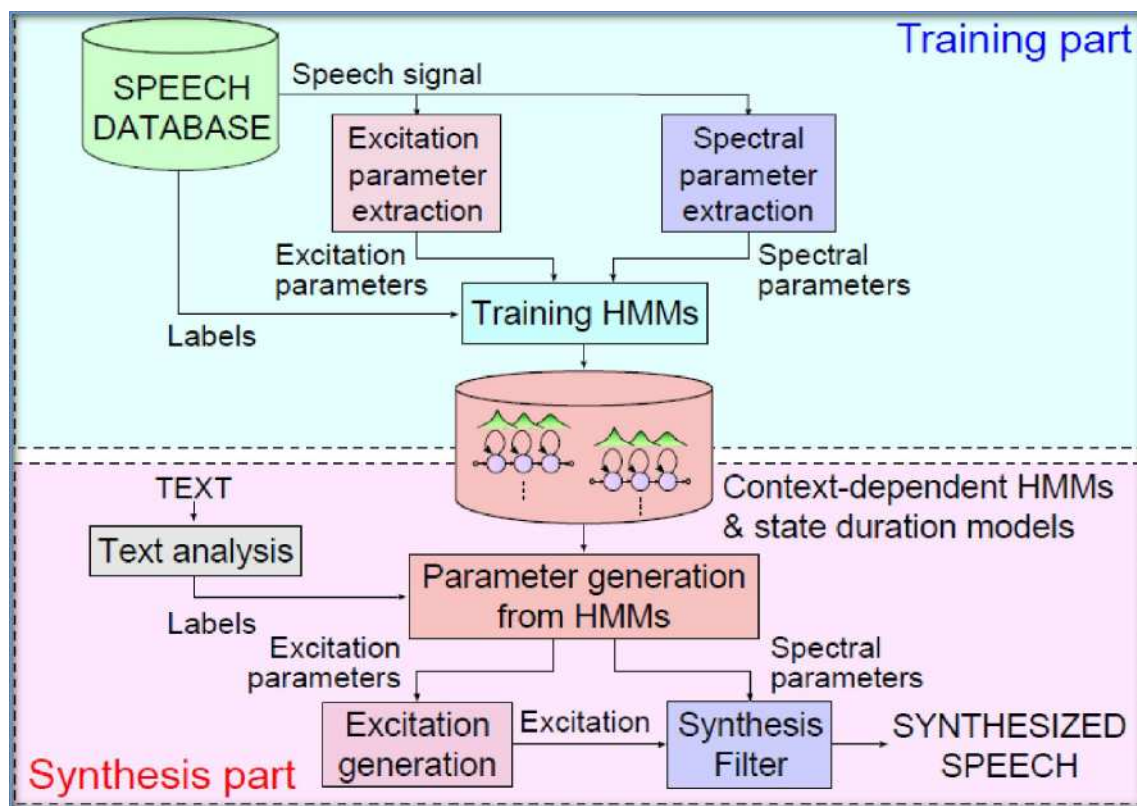
این سنتز مبتنی بر دادگان زیاد می باشد.

پارامترهای زیادی را از روی این دادگان آموزشی، آموزش می دهند.

مدل منبع-فیلتر + مدل صوتی آماری برای این روش مورد نیاز است.

معمولاً از مدل‌های مخفی مارکوف به عنوان مدل صوتی آن استفاده می کنیم.

کل فرآیند سنتز مبتنی بر مدل مخفی مارکوف را در تصویر 7 مشاهده می کنید.



تصوی 7 – سنتز مبتنی بر مدل مخفی مارکوف

ابتدا بازنمایی های پارامتری گفتار (شامل پارامترهای طیف و پارامترهای تحریک) را از دیتابیس گفتار به دست می آوریم.

بوسیله مجموعه ای از مدل های تولیدکننده (مانند HMM) آن ها را مدل می کنیم.

$$\hat{\lambda} = \arg \max_{\lambda} p(O | \mathcal{W}, \lambda)$$

آموزش: تخمین پارمترها

$$\hat{o} = \arg \max_o p(o | w, \hat{\lambda})$$

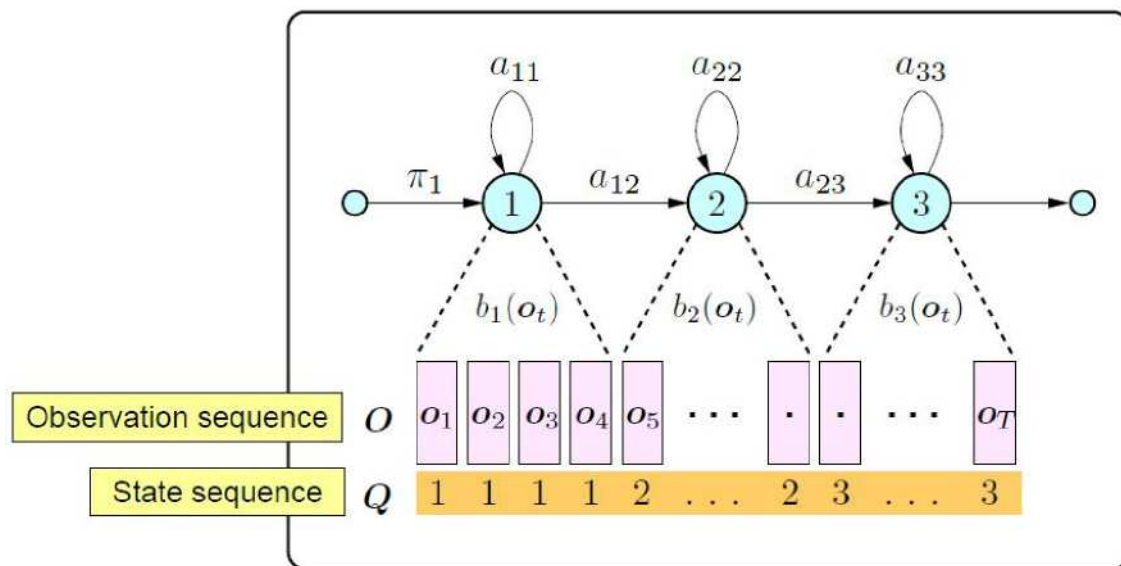
ستز: به دست آوردن پارمترهای طیف و تحریک از روی دنباله واجی

سه مدل کردن:

- مدلینگ پارامترهای طیف
- مدلینگ پارامترهای تحریک
- مدلینگ مدت زمان حالت ها

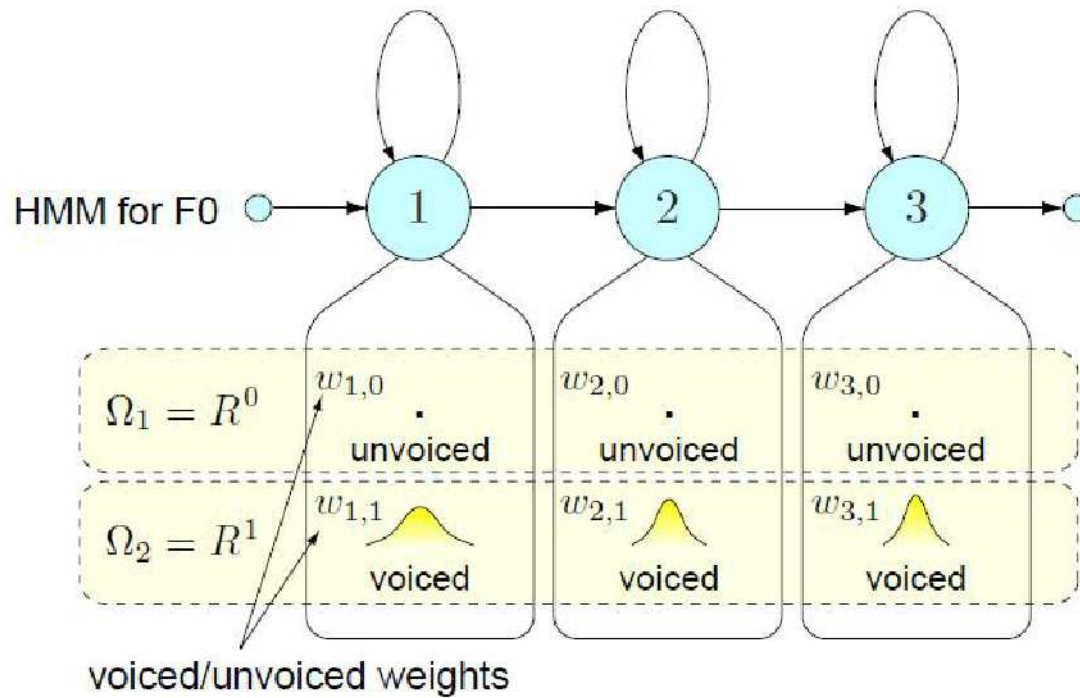
از آنالیز مل-کپسترال برای تخمین طیف استفاده می شود.

از یک HMM با چگالی احتمال پیوسته برای مدل کردن مسیر صوتی استفاده می شود (درست مانند سیستم بازشناسی گفتار) (تصویر 8).



تصویر 8 – استفاده از HMM برای مدل کردن مسیر صوتی

برای مدل کردن فرکانس گام، این مشکل وجود دارد که در قسمت های صدادار فرکانس گام وجود دارد ولی در قسمت های بدون صدا وجود ندارد. می توان فرض کرد که مقدار پیوسته قسمت صدادار از یک فضای یک بعدی و قسمت بدون صدا از یک فضای صفر بعدی آمده است (تصویر 9).



تصویر 9- در نظر گرفتن فرکانس گام درون HMM

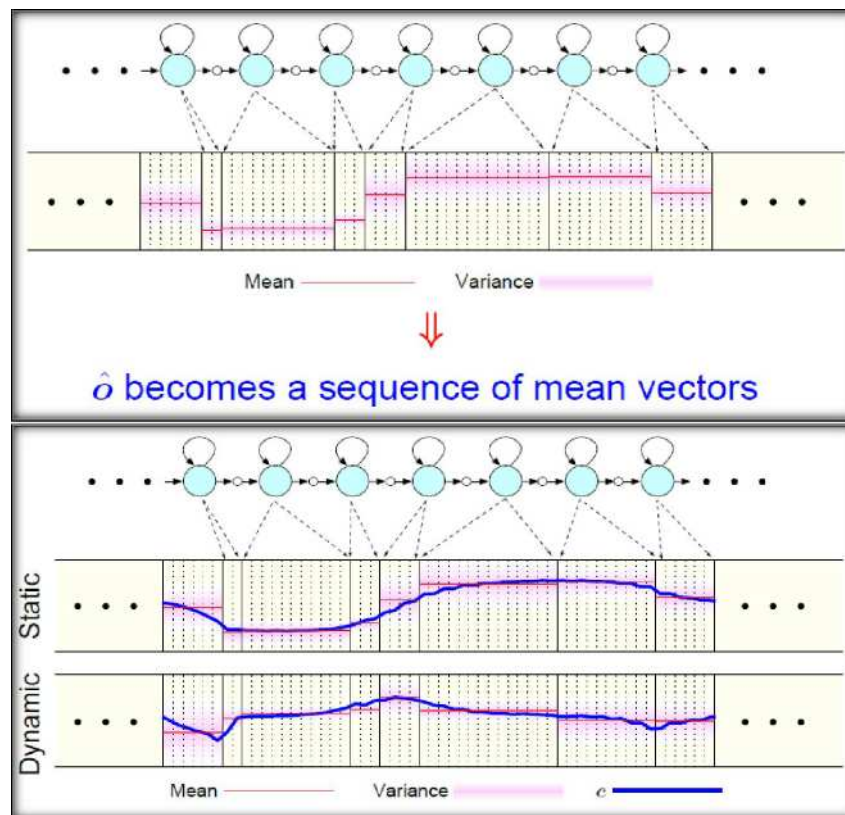
مانند کاربردهای بازشناسی، در اینجا نیز از ویژگی های پویا استفاده می شود.

$$\Delta c_t = \frac{\partial c_t}{\partial t} \approx 0.5(c_{t+1} - c_{t-1})$$

$$\Delta^2 c_t = \frac{\partial^2 c_t}{\partial t^2} \approx c_{t+1} - 2c_t + c_{t-1}$$

فرمول 2

این ها باعث پیوسته شدن طیف نهایی می شوند (تصویر 10).



تصویر 10 – به کار گیری ضرایب پویا در HMM

فیلتر مل-کپستروم استفاده شده فیلتر MLSA نام دارد.

## 7 – خلاصه و نتیجه گیری

در این فصل بحث سنتز را شروع کردیم.

تبدیل متن به گفتار

- مبتنی بر قانون
- مفصلی
- چسبانندنی (concatenative)
- انتخاب واحد
- آماری (مبتنی بر مدل مخفی مارکوف)



## 8 – منابع درس

- 1- Rabiner, “Fundamentals of Speech Recognition”
- 2- Huang, Acero, “Spoken Language Processing”
- 3- Deller, “Discrete-time processing of speech signals”