

- مقدمه

اهداف درس:

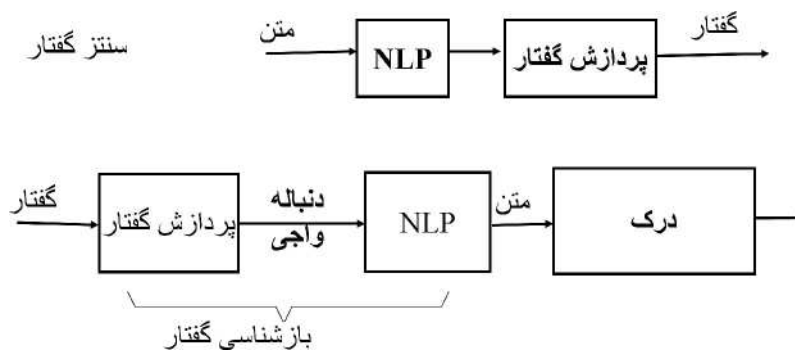
آشنایی با مفهوم اولیه بازشناسی گفتار.

آشنایی کلی با روش های حل مسئله بازشناسی گفتار

2- مفاهیم اولیه

مسئله بازشناسی گفتار مسئله تبدیل ورودی صوتی به متن است.

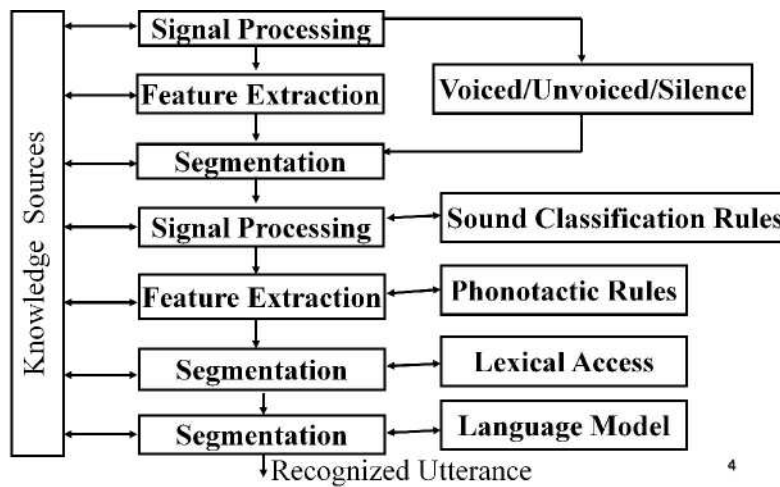
مسئله سنتز گفتار درست عکس این مسئله است. یعنی تبدیل متن به خروجی صوتی.



تصویر 1 - سنتز گفتار در مقابل بازشناسی گفتار

سه روش اصلی برای حل مسئله بازشناسی گفتار ارائه شده است.

- پایین به بالا (bottom-up): پردازش از سیگنال شروع می شود و تا به دست آوردن نتایج نهایی له ترتیب ادامه می یابد. در تصویر 2 مراحل این کار شرح داده شده است.

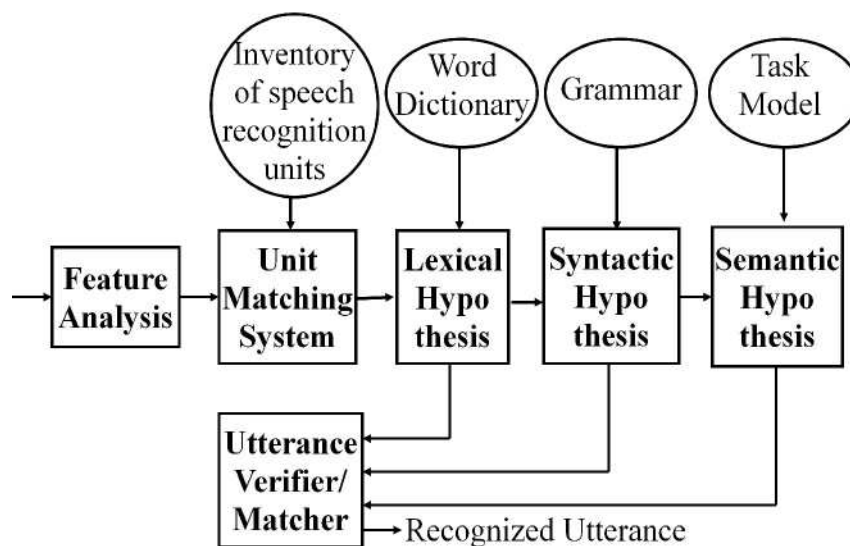


4

خلاصه ای از این

• بالا به پایین (n)

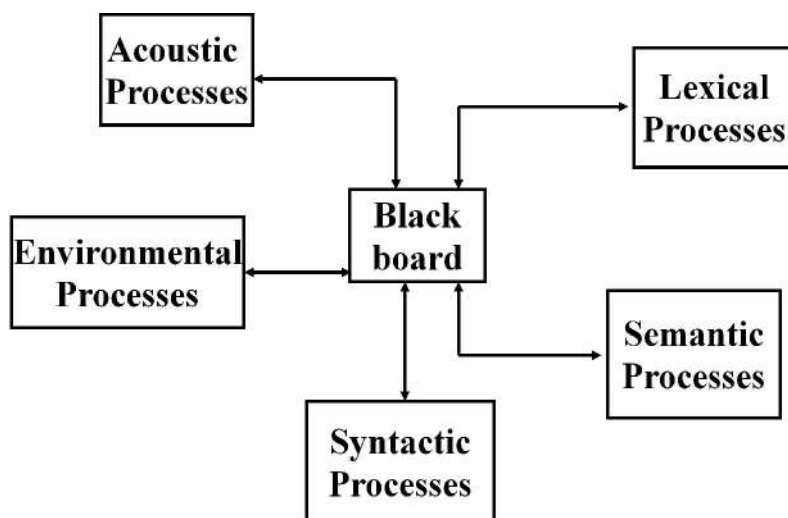
روش را در تص



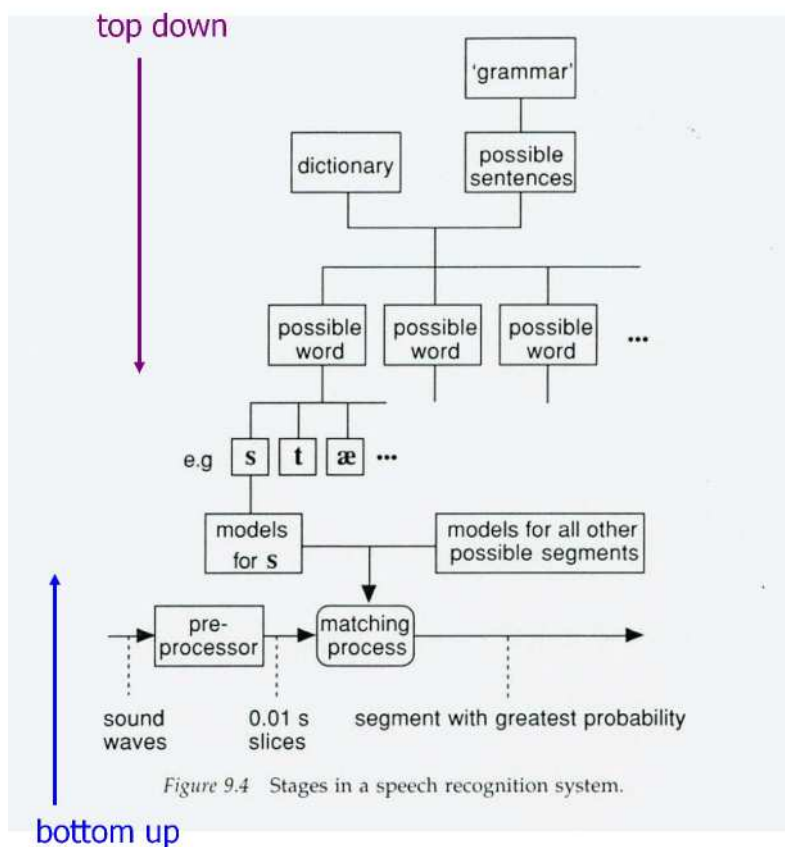
تصویر 3- روش بالا به پایین برای حل مسئله بازشناسی

• تخته سیاه (blackboard): در این روش بلوک های پردازشی به صورت ترتیبی با هم متصل نیستند بلکه به یک کنترل

کننده مرکزی هستند و اطلاعات را به آن می دهند. خلاصه این روش را در تصویر 4 مشاهده می کنید.



• سیستم کلی: یک س



تصویر 5 - سیستم کلی بازشناسی گفتار

چهار تئوری کلی بازشناسی گفتار وجود دارد.

- بازشناسی مبتنی بر مفصل (articulatory-based): استفاده از سیستم مفصلی برای بازشناسی. این تئوری موفق ترین روش تا به حال بوده است.
- بازشناسی مبتنی بر شنوایی: استفاده از سیستم شنوایی برای بازشناسی.
- بازشناسی ترکیبی: ترکیبی از دو روش بالا.
- تئوری موتور: سعی در مدل کردن هدف گوینده دارد.

2- مسئله باز شناسی

مسئله بازشناسی به این صورت است که دنباله ای از نمادهای صوتی داریم و می خواهیم کلماتی که گوینده تلفظ کرده است را از آن ها استخراج کنیم.

یافتن مهمت مل ترین دنباله کلمات با داشتن نمادهای صوتی.

- A: دنباله های صوتی

$$P(\underline{w}) \approx \prod_{i=1}^n P(w_i | w_{i-1} w_{i-2})$$

- W: دنباله کلمات

- یافتن فرمول زیر را پیشنهاد کند.

$$P(\hat{w} | A) = \max_{\underline{w}} P(\underline{w} | A)$$

$$P(x | y)P(y) = P(x, y)$$

$$P(x | y) = \frac{P(y | x)P(x)}{P(y)}$$

$$\Rightarrow P(\underline{w} | A) = \frac{P(A | \underline{w})P(\underline{w})}{P(A)}$$

$$P(\hat{w} | A) = \max_{\underline{w}} P(\underline{w} | A) = \max_{\underline{w}} \frac{P(A | \underline{w})P(\underline{w})}{P(A)}$$

- قانون بیزین:

$$\hat{w} = \underset{\underline{w}}{\operatorname{Arg\,max}} P(\underline{w} | A)$$

- فرمول نهایی:

$$= \underset{\underline{w}}{\operatorname{Arg\,max}} P(A | \underline{w})P(\underline{w})$$

با توجه به فرمول بالا، برای محاسبه محتمل ترین دنباله کلمات، باید مقدار

$$P(A | w) \quad \circ$$



○ $P(w)$

را محاسبه نماییم.

• محاسبه $P(w)$

○ این مقدار بوسیله «مدل زبانی» محاسبه می شود.

○ مفهوم این مقدار این است که احتمال رخداد دنباله کلمات w در زبان مورد نظر چقدر است. یعنی در زبان

مورد نظر به چه احتمالی دنباله کلمات w ظاهر می شود.

○ یک مدل زبانی ساده این است که محاسبه کنیم احتمال رخداد دنباله $w=w_1....w_n$ چقدر است.

○ به عبارتی

$$\underline{w} = w_1 w_2 w_3 \cdots w_n$$

○ محاسبه احتمال بالا بسیار پیچیده و نیازمند اندازه خیلی زیادی داده است. به همین دلیل از فرم های ساده تر

مدل زبانی به نام bigram و trigram استفاده می شود.

$$P(\underline{w}) \approx \prod_{i=1}^n P(w_i) \quad \text{Monogram} \quad \circ$$

$$P(\underline{w}) \approx \prod_{i=2}^n P(w_i | w_{i-1}) \quad \text{Bigram} \quad \circ$$

$$P(\underline{w}) = \prod_{i=1}^n P(w_i | w_{i-1} w_{i-2} \cdots w_1) \quad \text{Trigram} \quad \circ$$

○ روش محاسبه $P(w_3 | w_2 w_1)$:

$$P(w_3 | w_2 w_1) = \frac{\text{Number of happening } W_3 \text{ after } W_1 W_2}{\text{Total number of happening } W_1 W_2}$$

$$P(w_3 | w_2 w_1) \approx \lambda_1 f(w_3 | w_2 w_1) + \lambda_2 f(w_3 | w_2) + \lambda_3 f(w_3) \quad \text{adHoc} \quad \text{روش محاسبه} \quad \circ$$