

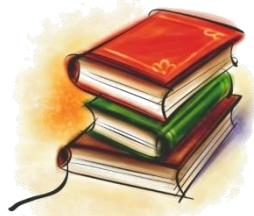
# آشنایی با زبان‌شناسی رایانشی

## کاربردها

هادی ویسی

[h.veisi@ut.ac.ir](mailto:h.veisi@ut.ac.ir)

دانشگاه تهران - دانشکده علوم و فنون نوین



## فهرست

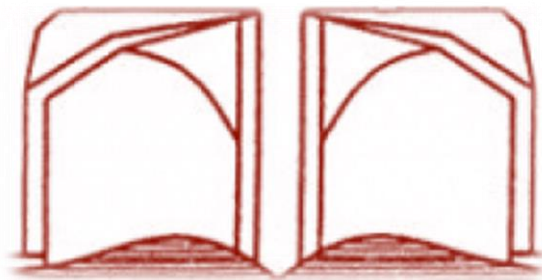
### ○ بازشناسی گفتار

- معرفی
- پیچیدگی‌ها
- تاریخچه
- ارزیابی
- روش‌ها

### ○ سنتز گفتار (تبدیل متن به گفتار)

- معرفی
- تاریخچه
- روش
- پردازش متن و مشکلات آن
- تحلیل نوایی
- تولید سیگنال گفتار

- ارزیابی



# آشنایی با زبان‌شناسی رایانشی

## کاربردها: بازشناسی گفتار

هادی ویسی

[h.veisi@ut.ac.ir](mailto:h.veisi@ut.ac.ir)

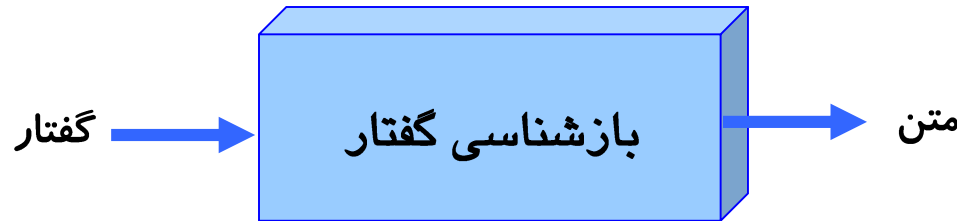
دانشگاه تهران - دانشکده علوم و فنون نوین



## بازشناسی گفتار: معرفی

### ○ بازشناسی خودکار گفتار (Automatic Speech Recognition (ASR)

- تبدیل گفتار انسان به متن یا دستورالعمل معادل
- گفتار = یک فایل صوتی، یا ورودی بر خط با میکروفون (یا تلفن)



- بازشناسی گفتار کار پیچیده ای است: ذات تصادفی دارد
  - یک کلمه خاص (مانند "آسمان") را ۱۰ بار مختلف بیان نمایید
  - در هیچ دو حالتی فایل های ضبط شده دیجیتالی آنها دقیقاً یکسان نیست!



## بازشناسی گفتار: پیچیدگی‌ها ...

انسان	سیستم بازشناسی گفتار	پارامترها
گفتار پیوسته کتابی و محاوره ای)	گفتار پیوسته کتابی	نوع گفتار
مستقل از گوینده	وابسته به گوینده (واژگان بزرگ با کارایی بالا) مستقل از گوینده (کاربردهای محدود)	گوینده
محیط شلوغ و نویزی	افت کارایی در شرایط نویزی	شرایط نویزی
بلادرنگ	بلادرنگ	سرعت
نامحدود	محدود به کلمات موجود در واژگان	دامنه واژگان
استفاده از اطلاعات مفهومی فراگیری آسان اطلاعات جدید	محدود به اطلاعات زبانی و معنایی یاد گرفته شده	اطلاعات زبانی





## بازشناسی گفتار: پیچیدگی‌ها ...

### ○ پیچیدگی‌های مطرح در سیستم‌های بازشناسی گفتار

- ① میزان وابستگی یا استقلال از گوینده
- ② پیوسته یا گسسته بودن گفتار
- ③ اندازه‌ی واژگان
- ④ محدودیت‌های زبانی
- ⑤ کارایی در حضور نویز و در محیط‌های کاربردی مختلف
- ⑥ ابهام آکوستیکی و میزان اشتباه بین کلمات



## بازشناسی گفتار: پیچیدگی‌ها ...

- میزان وابستگی یا استقلال از گوینده
  - تفاوت گوینده‌ها در مشخصات صدا، سن، لهجه، جنس، نحوه صحبت کردن و ...
  - سیستم‌های بازشناسی گفتار:
    - وابسته به گوینده (SD: Speaker Dependent) - دقت بالاتر
    - مستقل از گوینده (SI: Speaker Independent) - پیچیدگی بیشتر
- در عمل:
  - ۱- ایجاد سیستم مستقل از گوینده
  - ۲- تطبیق (Adaptation)، سیستم برای یک گوینده خاص در هنگام استفاده
- در بازشناسی گفتار تلفنی: وابسته به گوینده بودن معنی ندارد
  - گوینده از ابتدا مشخص نیست.



# بازشناسی گفتار: پیچیدگی‌ها ...

## ○ پیوسته یا گسسته بودن گفتار

فایل صوتی	توضیح	نوع
هر فایل صوتی یک کلمه	یک کلمه	بازشناسی کلمات مجزا IWR: Isolated Word Recognition
هر فایل چند کلمه و سکوت در بین کلمات	دنباله‌ای از کلمات که به وسیله سکوتی کوتاه از هم جدا شوند	بازشناسی گفتار متصل Connected Word Recognition
گوینده خبر تلویزیونی	گفتار دیکته کتابی و روان	بازشناسی گفتار پیوسته کتابی CSR: Continuous Speech Recognition
فایل سخنرانی	گفتار کاملاً طبیعی (شامل جملات ناقص، سرفه، تپق، مکث‌های طولانی و ...)	بازشناسی گفتار پیوسته محاوره‌ای Spontaneous Speech Recognition

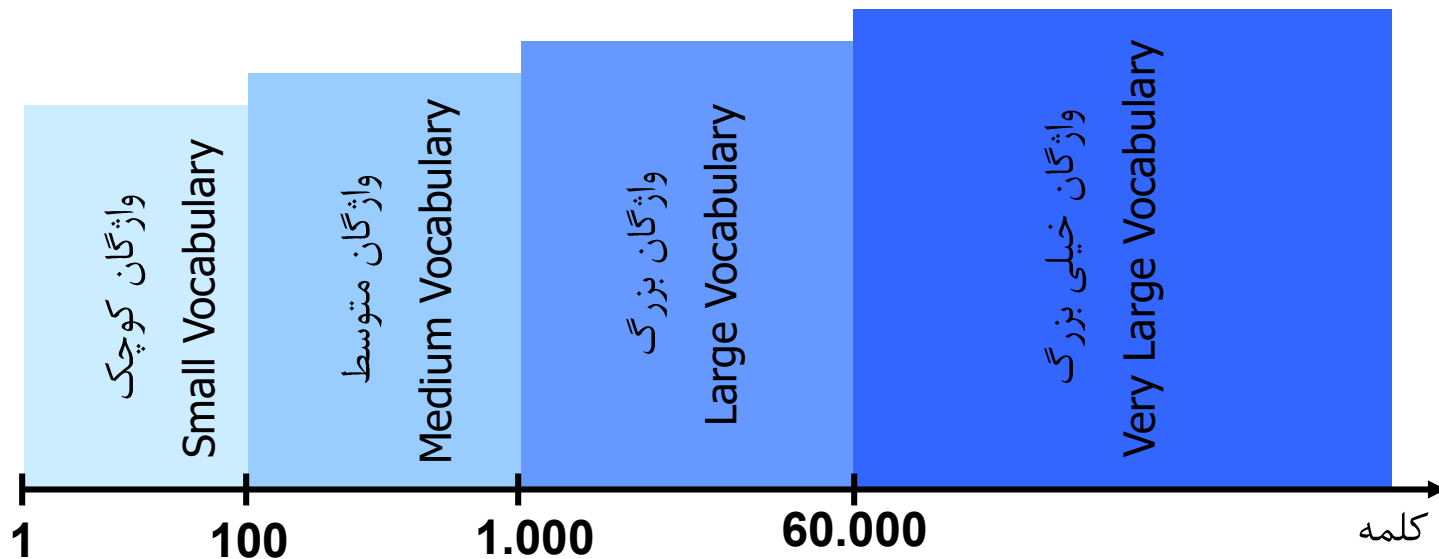




# بازشناسی گفتار: پیچیدگی‌ها ...

## ○ اندازه‌ی واژگان

- اندازه‌ی واژگان = تعداد کلمات = پارامتر موثر در دقت
- دقت بازشناسی با افزایش تعداد کلمات کم می‌شود
- به علت افزایش شباهت کلمات به هم



## بازشناسی گفتار: پیچیدگی‌ها ...

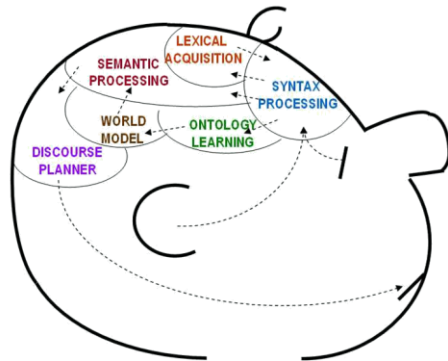
### ○ محدودیت‌های زبانی

- مدل زبانی (Language Model) بیانگر اطلاعات زبانی

- اطلاعات زبانی شامل دستور زبان و معنا

  - استفاده اندک از اطلاعات معنایی

  - استفاده فراوان از محدودیت‌های دستوری



- تاثیر اطلاعات زبانی در تشخیص گفتار در سیستم‌های تشخیص گفتار

تحقق یک سیستم تشخیص گفتار با واژگان بزرگ (یا خیلی بزرگ) با کاربرد واقعی بدون استفاده از اطلاعات زبانی غیر ممکن است.

## بازشناسی گفتار: پیچیدگی‌ها ...

- کارایی در حضور نویز و در محیط‌های کاربردی مختلف
- عدم حفظ کارایی سیستم در شرایط واقعی و عملی (محیط نویزی)
- از بزرگترین مشکلات سیستم‌های بازشناسی گفتار

- انواع:

- نویز و سرو صدای محیط، پژواک، تداخل صداها، اعوجاج و ...

- منابع:

- محیط، میکروفن، کانال انتقال، صدای گوینده و ...

- مقابله (روش‌های مختلف):

- ویژگی‌های مقاوم به نویز، تخمین گفتار تمیز و اصلاح مدل آکوستیکی





## بازشناسی گفتار: پیچیدگی‌ها

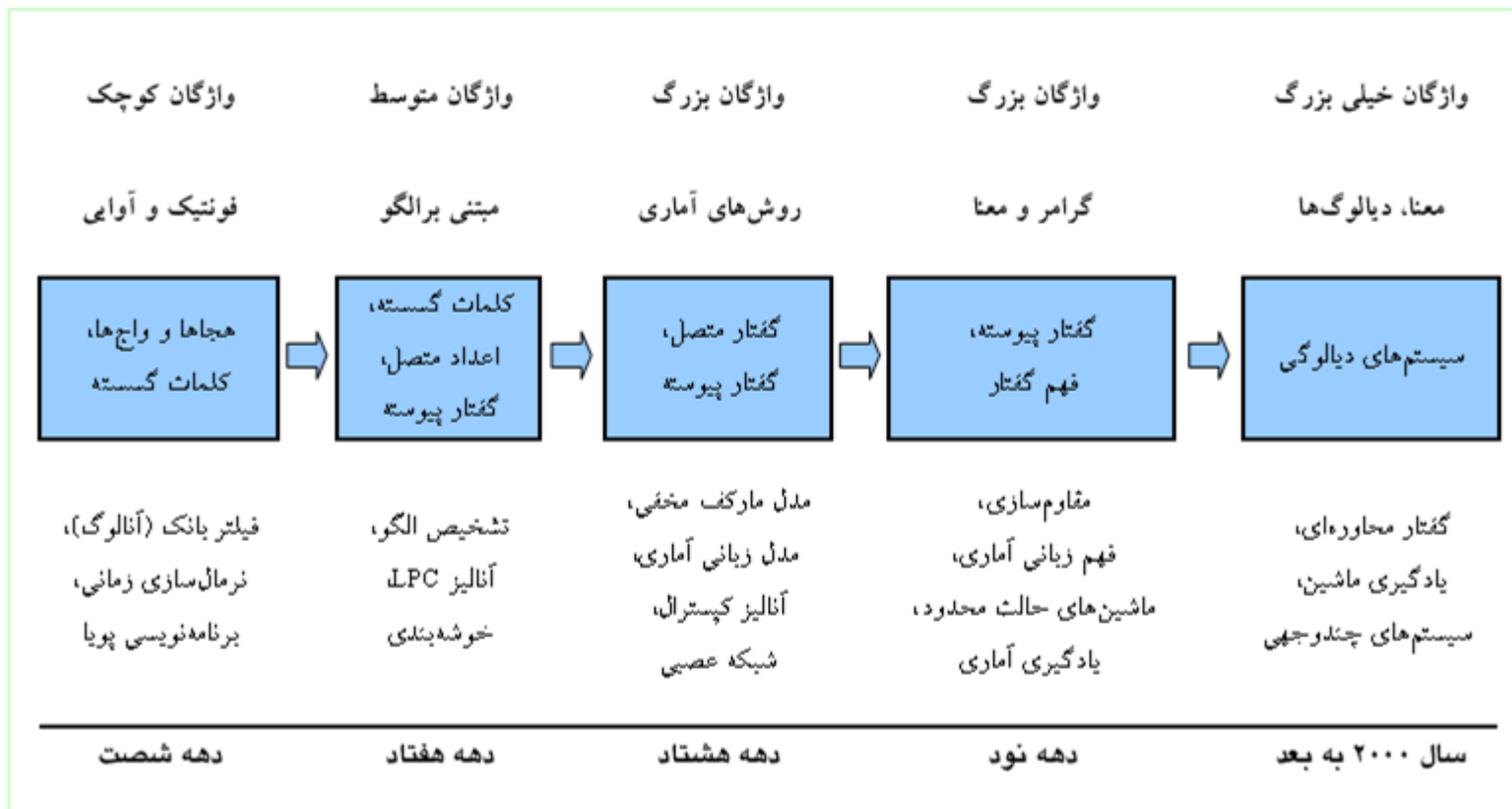
### ○ ابهام آکوستیکی و میزان اشتباه بین کلمات

کلمات هم آوا	
تعریف	شکل تلفظی یکسان و نوشتاری متمایز
نوع مشکل	تشخیص به جای همدیگر
مثال	«سمر ↔ ثمر» «خواستن ↔ خاستن» «خویش ↔ خیش» «ارز ↔ عرض ↔ ارض»
کلمات هم نگاره	
تعریف	شکل نوشتاری یکسان
نوع مشکل	پردازش متن و استخراج مدل زبانی
مثال	«نه - نه» «کرد - گرد» «مهر - مَهر» «گل - گِل»



# بازشناسی گفتار: تاریخچه (مختصر) ...

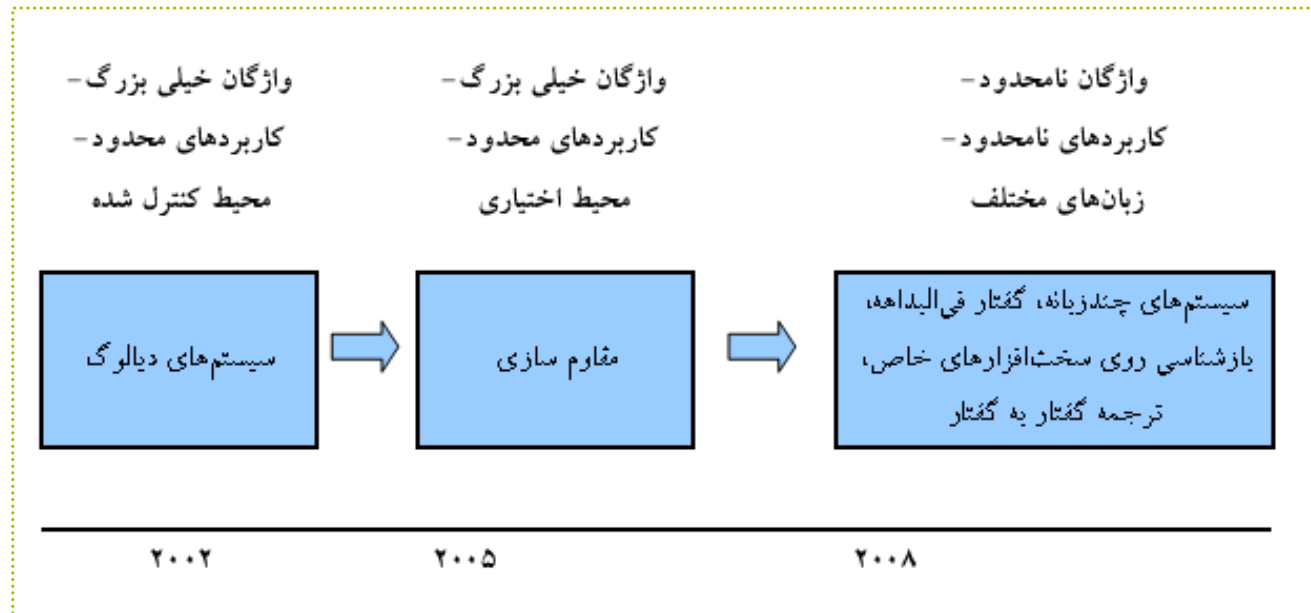
## ○ بازشناسی گفتار از گذشته تاکنون در ۵ دهه اخیر





# بازشناسی گفتار: تاریخچه (مختصر)

## وضعیت کنونی سیستم‌های بازشناسی گفتار





## بازشناسی گفتار: نمونه

وضعیت کارایی سیستم‌های بازشناسی گفتار امروزی در کاربردهای مختلف

WER	اندازه واژگان	نوع	دادگان
0.3%	11 (صفر تا نه و oh)	گفتار فی البداهه	رشته اعداد پیوسته - دادگان TI
2%	1,000	گفتار کتابی	RM
2.5%	2,500	گفتار فی البداهه	ATIS
6.6%	64,000	گفتار کتابی	NAB/WSJ
13-17%	210,000	اخبار تلویزیونی	انتشار اخبار (Broadcast News)
25-29%	45,000	مکالمه تلفنی محاوره‌ای	Switchboard

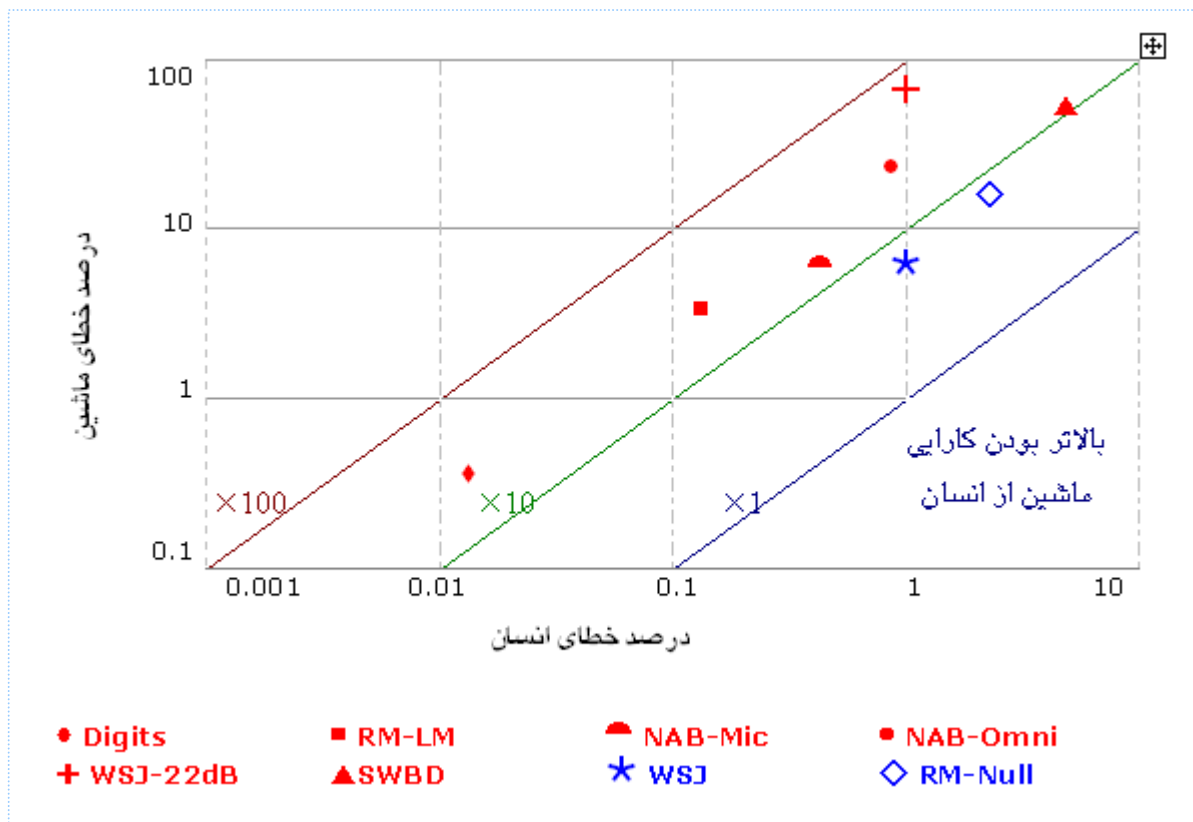




# بازشناسی گفتار: مقایسه با انسان

○ مقایسه کارایی سیستم‌های بازشناسی خودکار گفتار با انسان

• مربوط به سیستم‌های قبل از سال ۲۰۰۵



خطای کامپیوتر 10 تا 100 برابر بیشتر از انسان است!







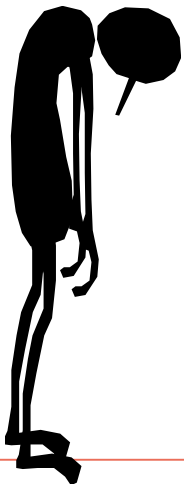
## بازشناسی گفتار: ارزیابی ...

### ○ پارامترهای موثر

- کلمات موجود در واژگان سیستم
- کلمات خارج از دادگان مجموعه آزمون = OOV
- سازگاری/عدم سازگاری داده‌های آموزش و آزمون
- محیط آکوستیکی آزمون و میزان سیگنال به نویز سیگنال گفتار
- وابسته/مستقل از گوینده

### ○ خطاها

- حذف (Deletion)، درج (Insertion) و جایگزینی (Substitution)





## بازشناسی گفتار: ارزیابی ...

### ○ معیارها

• دقت (Accuracy)

تعداد کل کلمات

$$Accuracy \% = \frac{N^{All} - (N^{Del.} + N^{Ins.} + N^{Sub.})}{N^{All}} \times 100$$

تعداد کلمات حذف، درج و جایگزین شده

• صحت (Correctness)

$$Correctness \% = \frac{N^{All} - (N^{Del.} + N^{Sub.})}{N^{All}} \times 100$$

• نرخ خطای کلمات (WER)

$$WER \% = 1 - Accuracy \% = \frac{N^{Del.} + N^{Ins.} + N^{Sub.}}{N^{All}} \times 100$$



## بازشناسی گفتار: ارزیابی

○ مثال هایی از خطاهای بازشناسی گفتار

مرتبه جواب	جملات بازشناسی شده	درج	حذف	جایگزینی	درست	درصد دقت	درصد صحت
جواب درست	یک قوطی میخ پرروز خریدم				5		
1	یک قوطی میخ پرروز خریدم	0	0	0	5	100	100
2	یک قوطی میخ پرروز خریدم <b>هم</b>	1	0	1	4	60	80
3	یک قوطی میخ پرروز خرید <b>در</b>	1	0	1	4	60	80
4	یک قوطی میخ پرروز <b>خرید</b>	0	0	1	4	80	80
5	یک قوطی میخ <b>تایید</b> روز خریدم	1	0	1	4	60	80
6	یک قوطی میخ <b>تایید</b> روز خریدم <b>هم</b>	2	0	2	3	20	60
7	یک قوطی میخ <b>تایید</b> روز خرید <b>در</b>	2	0	2	3	20	60



## روش‌های بازشناسی گفتار ...

### ○ رویکرد مبتنی بر تشخیص الگو ...

- دو فاز آموزش (Train) و آزمون (Test)

- آموزش

- الگوهای مربوط به هر کلاس (واحدهای آوایی مانند کلمه، واج و ...) با استفاده از روش‌هایی مدل می‌شوند.

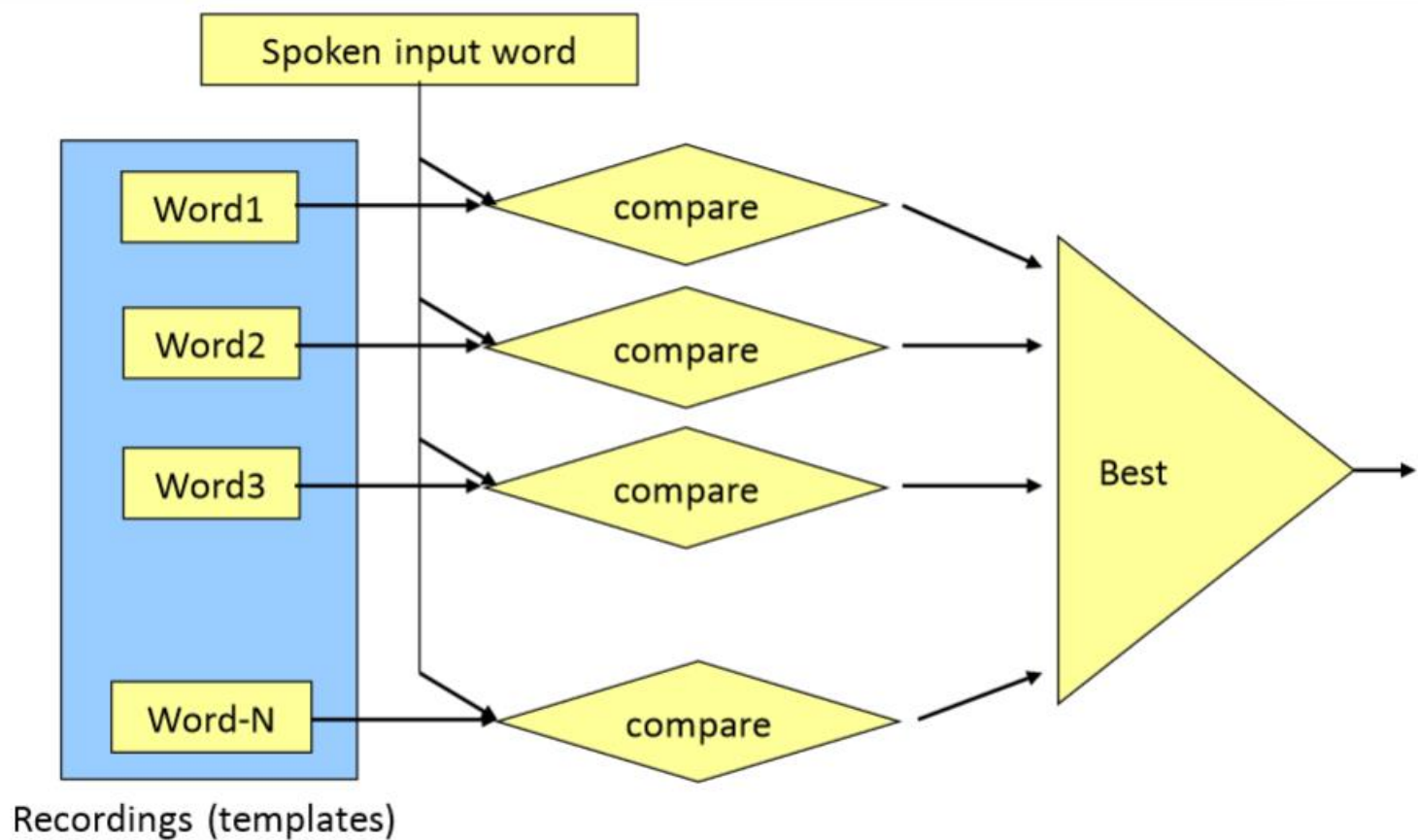
- آزمون (استفاده)

- مقایسه گفتار ورودی با الگوهای آموزش داده شده جهت تشخیص واحدهای آوایی موجود در گفتار ورودی



# روش‌های بازشناسی گفتار ...

## ○ بازشناسی کلمات مجزا (IWR)





## روش‌های بازشناسی گفتار ...

### ○ مدل‌سازی آوایی

- مدل کردن واحدهای آوایی گفتار (کلمه، هجا، سه واجی، واج)

- روش‌ها

- انطباق زمانی پویا (DTW)

- ساده و قدیمی - مورد استفاده در Voice dialing گوشی‌های تلفن همراه

- شبکه عصبی مصنوعی (ANN)

- ساده و کارا، سرعت تشخیص بالاتر، نسبتاً مقاوم به نویز، فرایند آموزش زمان‌بر

- مدل مخفی مارکوف (HMM)

- موفق‌ترین رویکرد، مدل کردن آماری گفتار

- مدل‌های ترکیبی (ترکیب شبکه‌های عصبی و HMM)

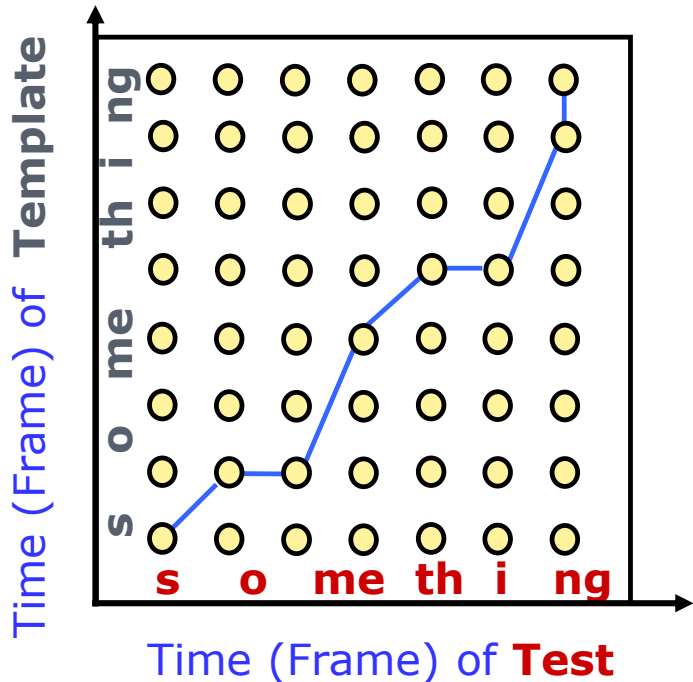


# روش‌های بازشناسی گفتار: انطباق زمانی پویا (DTW) ...

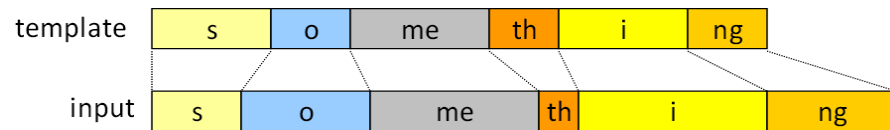
## ○ انطباق زمانی پویا (DTW: Dynamic Time Warping)

### • ایده

- ضبط کردن یک یا چند الگو از کلمه‌های (یا واحدهای) مورد نظر
- مقایسه گفتار ناشناس با آنها
- یافتن الگوی با کم‌ترین فاصله



### • روشی برای مقایسه دو الگوی غیرهم طول



### • برنامه‌ریزی پویا (dynamic programming)



## روش‌های بازشناسی گفتار: انطباق زمانی پویا (DTW)

### ○ کاربردها

- Voice dialing در گوشی‌های تلفن همراه
- فرمان‌های صوتی ساده
- تشخیص گوینده

### ○ مزایا/معایب

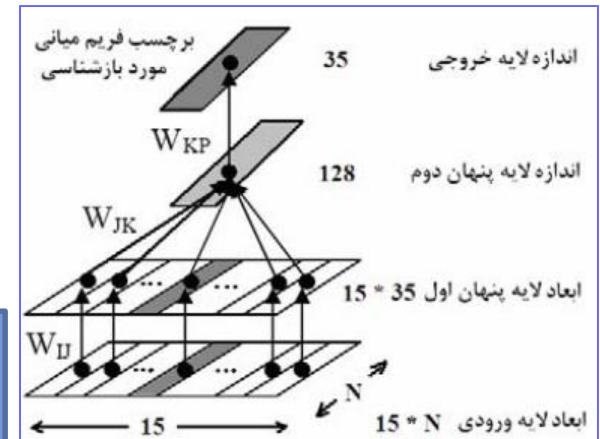
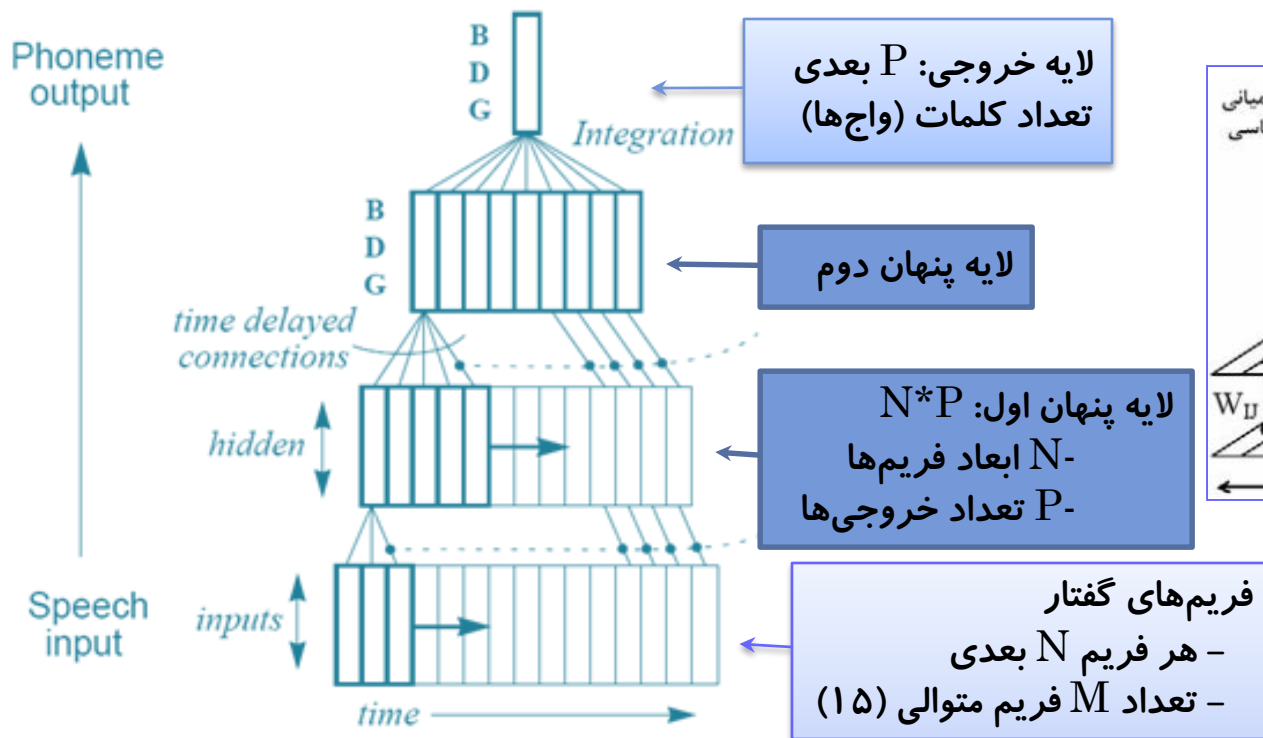
- وابسته به گوینده
- ساده و دارای کارایی مناسب برای تعداد الگوهای پایین (حدود ۲۰)
- مستقل از زبان
- آموزش و اختصاصی کردن آسان



# روش‌های بازشناسی گفتار: شبکه عصبی (ANN) ...

## ○ تشخیص گفتار با TDNN: Time-Delay Neural Network

- ورودی: دنباله متوالی از فریم‌های سیگنال گفتار
- خروجی: واحد موردنظر در تشخیص (کلمه، واج)

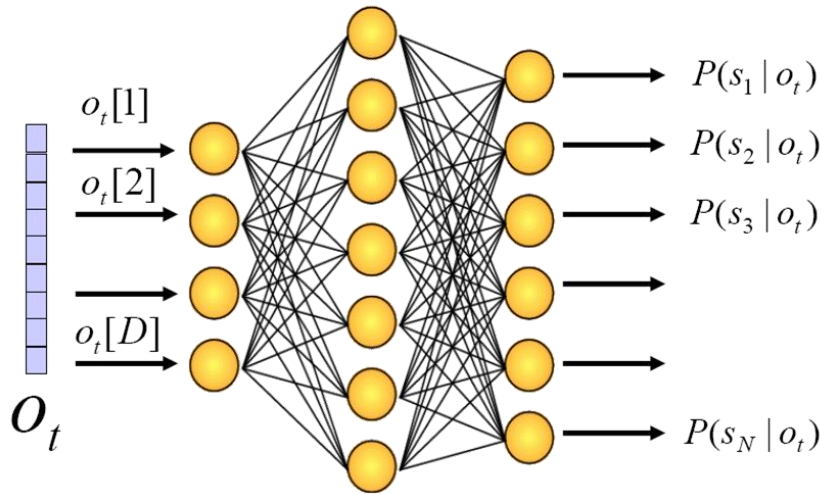


تشخیص واج‌های فارسی

# روش‌های بازشناسی گفتار: شبکه عصبی (ANN)

## ○ شبکه یادگیری عمیق (Deep NN)

- استفاده در تشخیص گفتار
- ترکیب با روش‌های آماری



## ○ حافظه کوتاه-مدت ماندگار (LSTM)

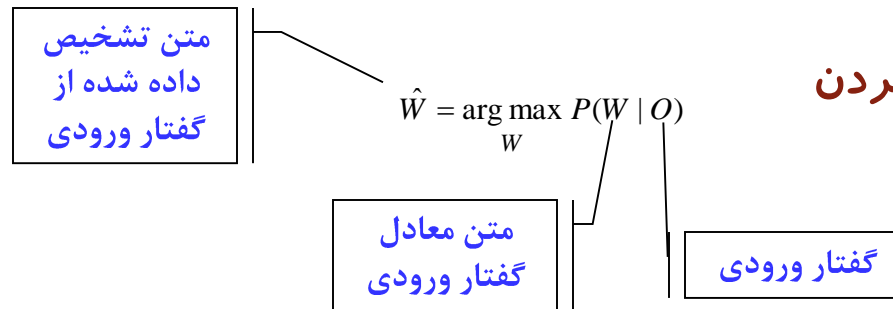
- Long Short-Term Memory
- شبکه بازگشتی



## روش‌های بازشناسی گفتار: مدل مخفی مارکوف (HMM) ...

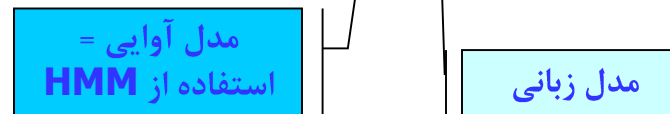
### ○ مدل مخفی مارکوف (HMM: Hidden Markov Model)

- گفتار یک منبع تولید اطلاعات آماری است که باید به دنباله‌ای از کلمات تبدیل شود.



- هدف: بیشینه کردن

$$\hat{W} = \arg \max_w \frac{P(O | W).P(W)}{P(O)} = \arg \max_w P(O | W).P(W)$$



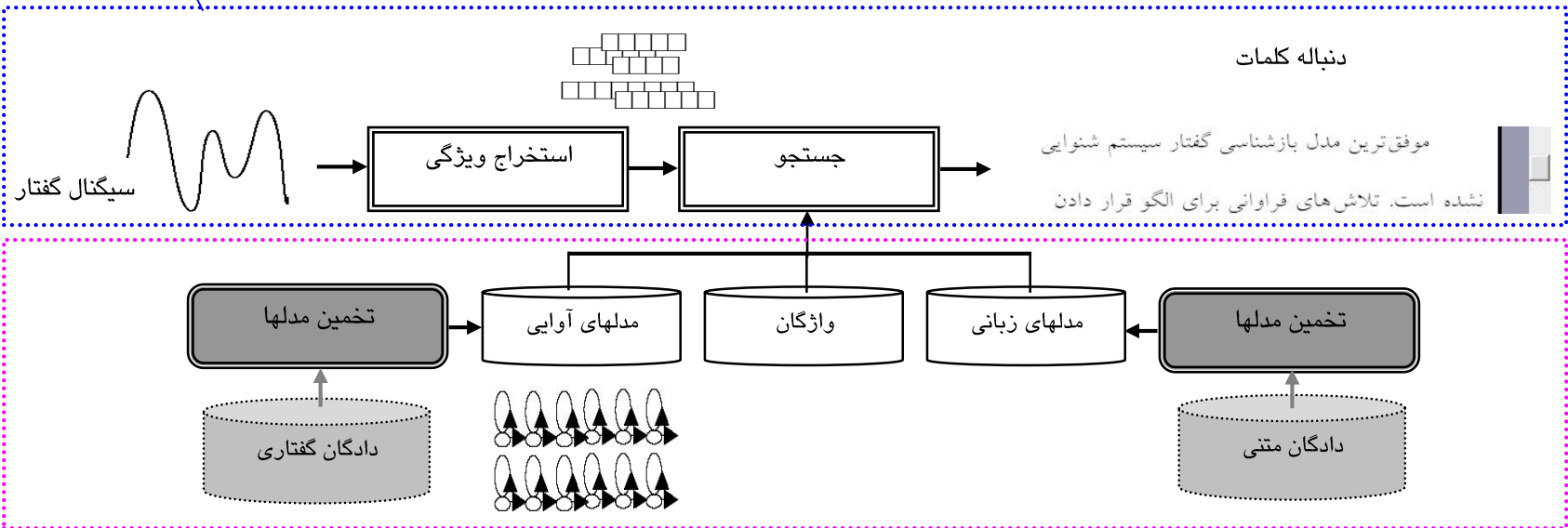
- با استفاده از قانون بیز



# روش‌های بازشناسی گفتار: مدل مخفی مارکوف (HMM) ...

- رویکرد مبتنی بر تشخیص الگو
- تشخیص گفتار پیوسته با واژگان بزرگ

آزمون



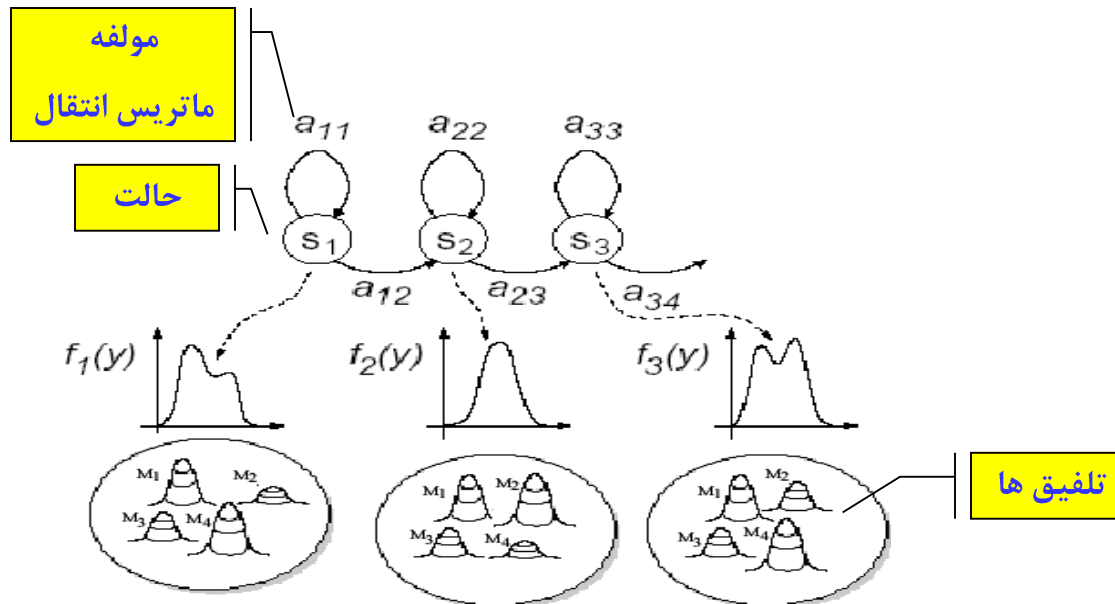
آموزش



# روش‌های بازشناسی گفتار: مدل مخفی مارکوف (HMM) ...

## ○ مدل مخفی مارکوف (HMM)

- فرض: توزیع بردارهای ویژگی در گفتار گاوسی است.
- برای هر واحد آوایی از تعدادی حالت (State) و در هر حالت از تعدادی تابع گاوسی که آنرا تلفیق (Mixture) استفاده می‌شود.
- توپولوژی HMMها به صورت چپ به راست است

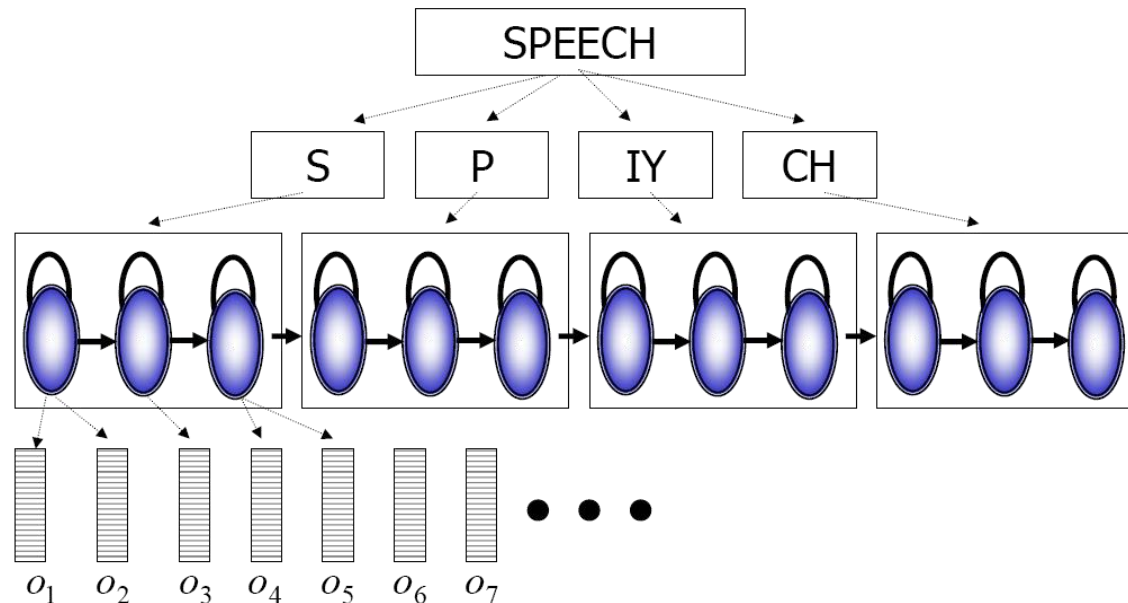


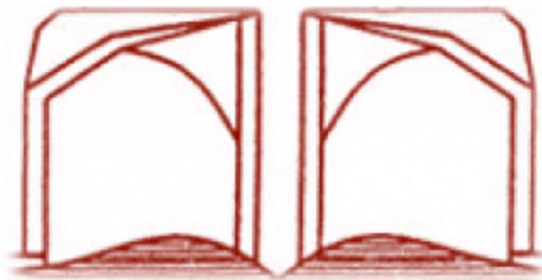


# روش‌های بازشناسی گفتار: مدل مخفی مارکوف (HMM)

## ○ مدل مخفی مارکوف (HMM)

- هر واج (phoneme) یک HMM دارد
- تشکیل کلمه با پشت‌سرهم قرار دادن مدل واج‌ها





# آشنایی با زبان‌شناسی رایانشی

## کاربردها: سنتز گفتار

هادی ویسی

[h.veisi@ut.ac.ir](mailto:h.veisi@ut.ac.ir)

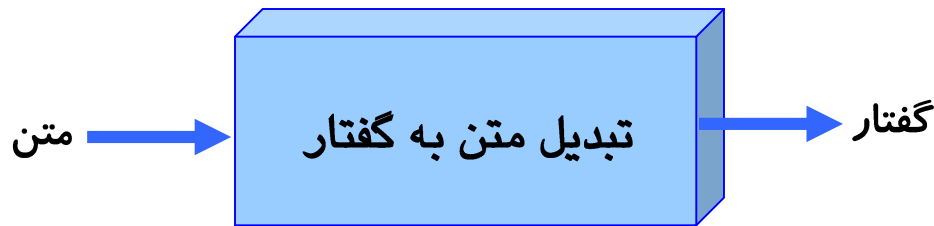
دانشگاه تهران - دانشکده علوم و فنون نوین



## سنتز گفتار: معرفی ...

### ○ تبدیل متن به گفتار (TTS: Text-to-Speech)

• Speech Synthesis



### ○ هدف سنتز گفتار

- تبدیل متن به گفتار انسان
- خواندن متن به صورت خودکار

### ○ تولید صدا از متن در ظاهر ساده به نظر می‌رسد

- مثال: شماره خوان سیستم نوبت دهی بانکها
- "شماره" + "یک" + "هزار و" + "دویست و" + "دو" + "به" + "باجه" + "چهار"

### ○ اما تولید صدای طبیعی بسیار مشکل است

- قابل فهم بودن و خوشایند بودن





## سنتز گفتار: معرفی ...

### ○ سیستم متن به گفتار به عنوان رمزکننده

• متن = حالت رمز شده گفتار

○ نسبت فشردگی بالا

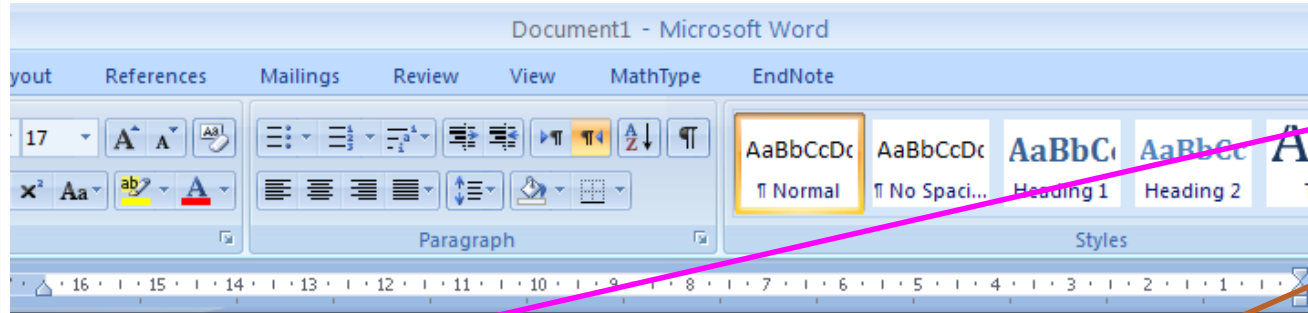
○ انعطاف‌پذیری بالا

○ تغییر آسان جنسیت، متوسط زیرویمی، دامنه زیرویمی، و ...

○ تغییر آسان محتوای پیام



# سنتز گفتار: معرفی ...



پرانتر باز یا بسته

عدد- خواندن ساده

نقطه، کاما و ...

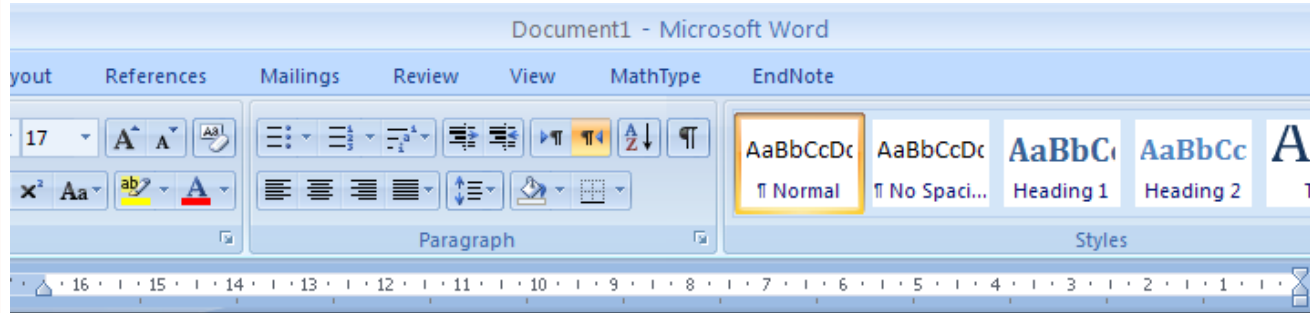
عدد- خواندن پیچیده

کلمات غیرفارسی

سامانه من خوان فارسی آر بانا به دست متخصصین ایرانی (از دانشگاه صنعتی شریف) در سال ۱۳۸۹ طراحی و ساخته شد. این سامانه، قادر است متون مختلف الکترونیکی مانند فایل‌های Word و PDF را بخواند. طراح آریانا گفت: شکر خدا، کار مطابق برنامه پیش می رود. علاقه‌مندان می توانند برای کسب اطلاع بیشتر به وب سایت [www.farsireader.com](http://www.farsireader.com) یا شماره تلفن ۰۲۱-۶۶۵۵۱۵۲۵ تماس بگیرند.



## سنتز گفتار: معرفی ...



سامانه یا سامانه ی؟

متن خوان یا متنخوان  
یا متنخوان

سامانه متن خوان فارسی آر بانا به دست متخصصین ایرانی (از دانشگاه صنعتی شریف) در سال ۱۳۸۹ طراحی و ساخته شد. این سامانه، قادر است متون مختلف الکترونیکی مانند فایل‌های Word و PDF را بخواند. طراح آریانا گفت: شکر خدا، کار مطابق برنامه پیش می رود. علاقه‌مندان می‌توانند برای کسب اطلاع بیشتر به وب سایت [www.farsireader.com](http://www.farsireader.com) یا شماره تلفن ۰۲۱-۶۶۵۵۱۵۲۵ تماس بگیرند.

شکر یا شِکر!؟

می رود یا میروَد



## سنتز گفتار: معرفی ...

### ○ تبدیل متن به صورت آوایی

سامانه متن خوان فارسی آریانا به دست متخصصین ایرانی (از دانشگاه صنعتی شریف) در سال ۱۳۸۹ طراحی و ساخته شد.

sAmAneye matnxAne fArsie AriAna be daste motaxasesine irani  
**parantez baz** az dane\$gahe sanatie \$arif **parantez baste** dar sAle  
**hezaro sisado ha\$tdo noh** tarAhi va sAxte \$od **noGte**



## سنتز گفتار: تاریخچه ...

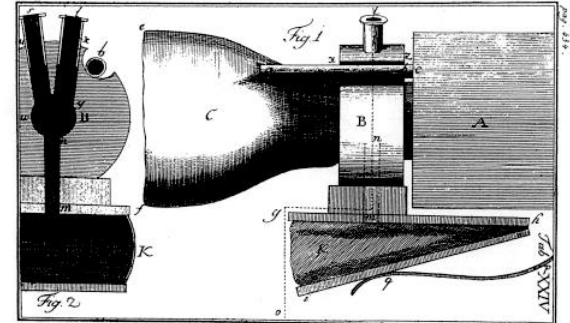
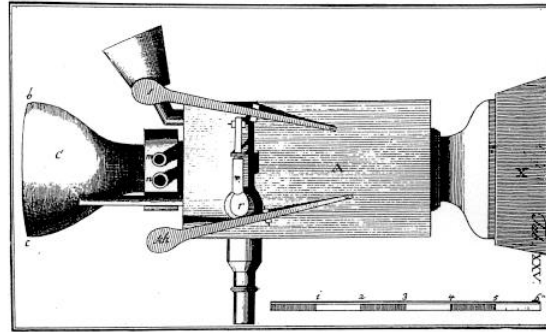
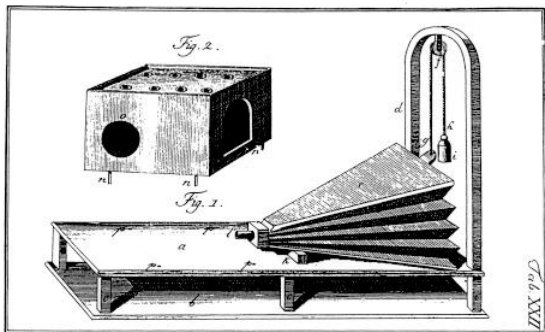
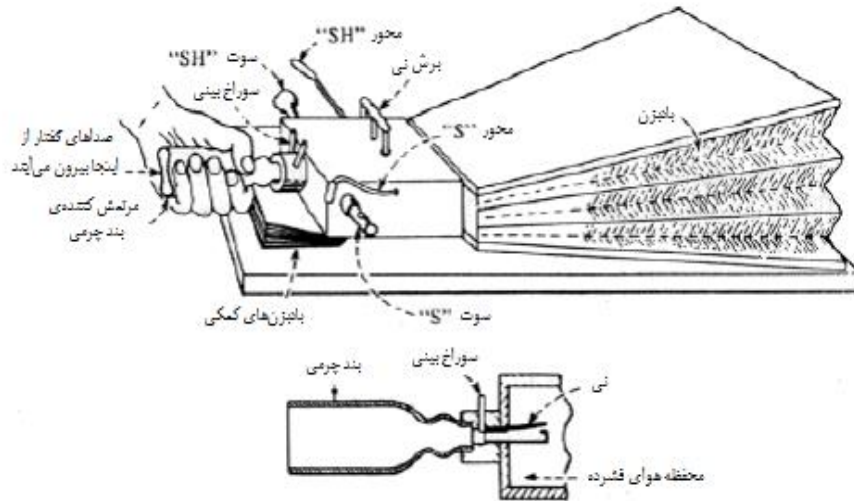
### ○ سنتزکننده‌های مکانیکی ...

روش و صدای تولیدی	دستگاه	توسط	سال
پنج واکه	اولین سنتز کننده مکانیکی	کریستین کراتزنستین در سن پترزبورگ روسیه	1779
تغییر شکل یک لوله چرمی صداهاى تکی و تعدادی صدای ترکیبی	ماشین گفتار مکانیکی	ولفگانگ وان کمپلن	1791
تولید واکه‌ها و تعداد زیادی از همخوان‌ها	ماشین سخنگوی مکانیکی	چارلز ویتستون	1800
لوله‌های تشدیدساز - تولید صدای واکه‌ها		ویلیس	1838

# سنتز گفتار: تاریخچه ...

## ○ سنتز کننده‌های مکانیکی

- بازسازی ویتستون از ماشین سخنگوی وان کمپلن (Von Kempelen)





## سنتز گفتار: تاریخچه ...

### ○ سنتزکننده‌های الکتریکی ...

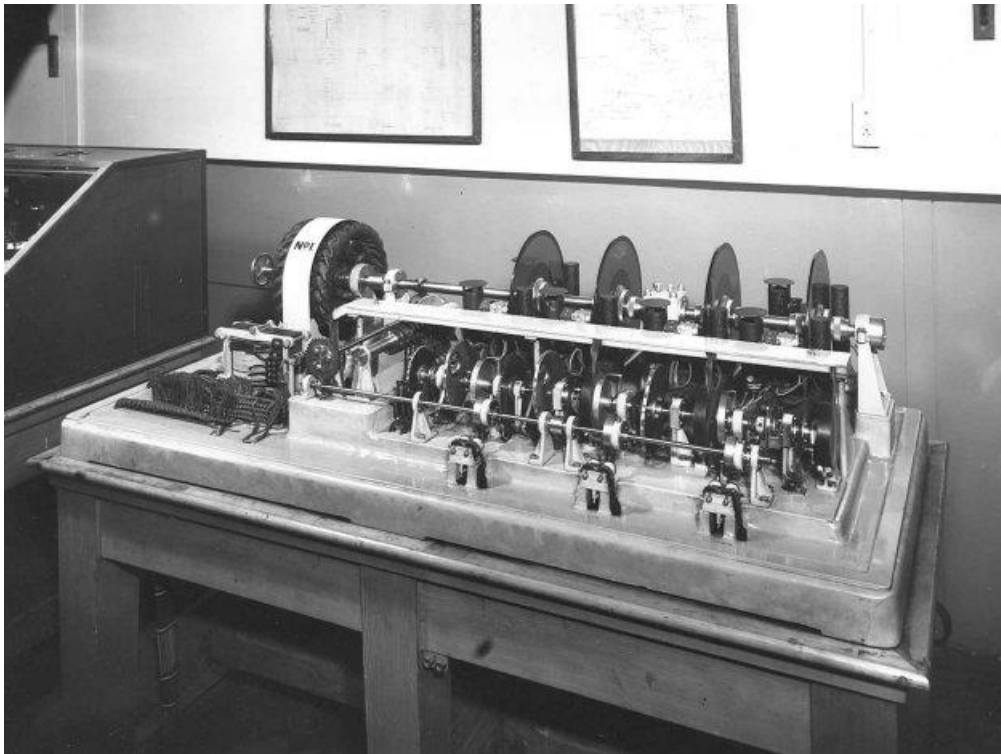
سال-اسم	توسط	روش و صدای تولیدی
1922	Stewart	دو مدار تشدیدکننده-واکه‌ها با استفاده از دو فرمنت اول
	Wagner	چهار مدل تشدیدکننده-مشابه روش Stewart
1932	Obata and Teshima	استفاده از فرمنت سوم
1939	Homer Dudlly	دستگاهی برای تحلیل گفتار و بازسازی سیگنال گفتار از روی پارامتر (مشابه مدل منبع-فیلتر)
	Voder	
1951	Franklin Cooper	تولید صدا از روی الگوهای اسپکتروگرام ضبط شده
PAT-1953	Walter Lawrence	اولین سنتز کننده فرمنت-سه تشدیدکننده الکترونیکی موازی
OVE I	Gunner Fant	اولین سنتز کننده فرمنت سری
1962	Martony	مدل کردن تابع انتقال مجرای گفتار برای واکدارها، صداهای خیشومی، و همخوان‌های بی‌واک
OVE II		
1972	John Holmes	سنتز کننده فرمنت موازی

## سنتز گفتار: تاریخچه ...

### ○ سنتزکننده‌های الکتریکی ...

#### • ساعت سخن گو (The UK Speaking Clock)

- سال ۱۹۳۶
- دو دیسک برای دقیقه،
- ۱ دیسک برای ساعت
- ۱ دیسک برای ثانیه





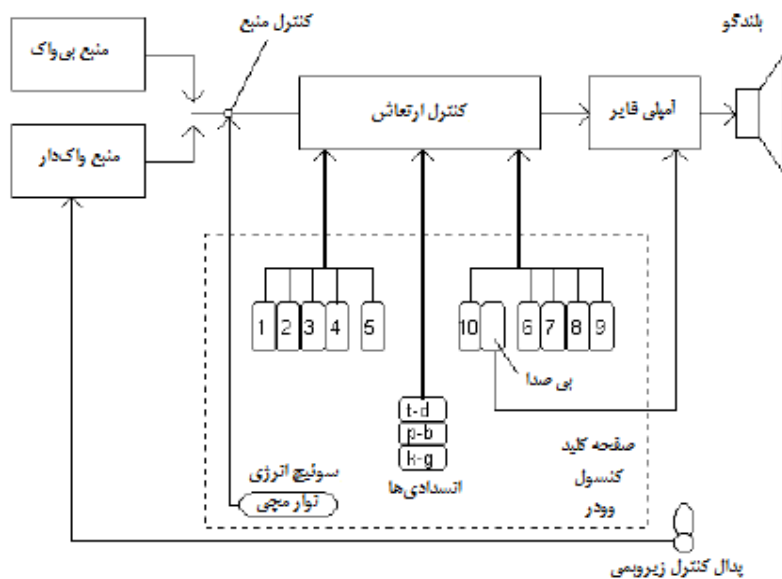
# سنتز گفتار: تاریخچه ...

## ○ سنتزکننده‌های الکتریکی ...

### • سیستم Voder توسط Homer Dudley

○ سال ۱۹۳۹ در نمایشگاه World's Fair

○ نیاز به اپراتور-کنترل با صفحه کلید



## سنتز گفتار: تاریخچه ...

○ سنتزکننده‌های الکتریکی ...

• سیستم OVE

○ استفاده از فرمت‌های اول و دوم



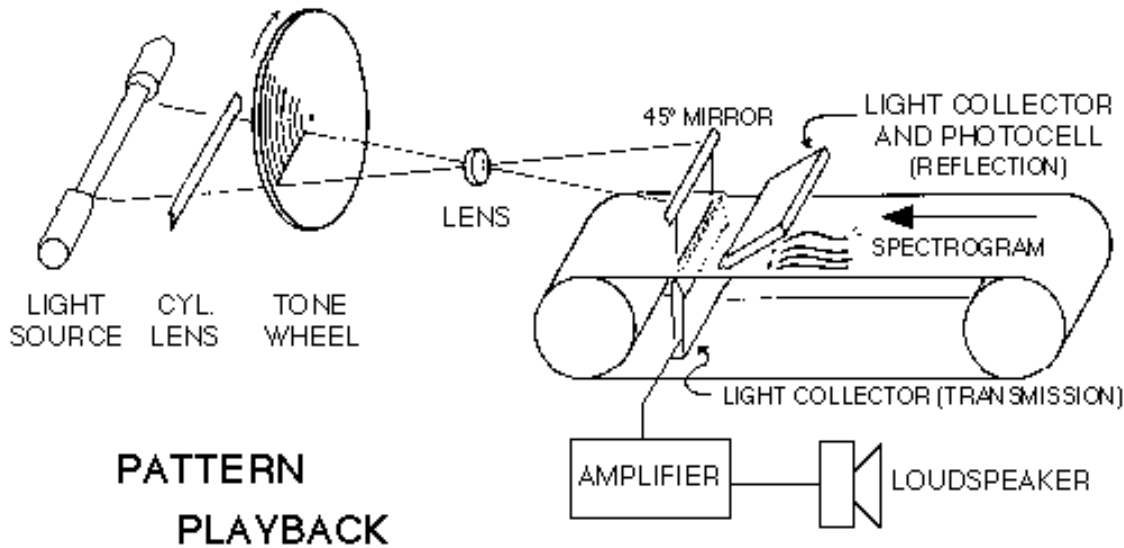


# سنتز گفتار: تاریخچه ...

## ○ سنتزکننده‌های الکتریکی

### • سیستم Cooper

- عبور نور از درون یک دیسک
- نور ارسالی در هر باند فرکانسی متناسب با انرژی صدا





## سنتز گفتار: تاریخچه ...

### ○ سنتز کننده‌های مدرن

روش و صدای تولیدی	توسط	سال
اولین سنتز کننده تولیدی DAVO (articulatory) با نام	George Rosen در دانشگاه MIT	1958
اولین برنامه کامپیوتری سنتز واجی مبتنی بر قاعده	John Kelly and Louis Gerstman	1961
اولین سیستم کامل تبدیل متن به گفتار با روش سنتز تولیدی	Noriko Umeda در آزمایشگاه الکتروتکنیکی ژاپن	1968
اولین سیستم سنتز پیوندی مبتنی بر دوواجی با بسامدهای فورمنت پارامتری	Rex Dixon and David Maxey	1968
سنتز کننده تولیدی	Paul Mermelstein, James Flanagan از آزمایشگاه‌های بل	1976
ماشین متن‌خوان برای افراد نابینا	Raymond Kurzweil	1976
سیستم سنتز پیوندی مبتنی بر دوواجی با پیش‌بینی خطی	Joe Olive و همکارانش در آزمایشگاه‌های بل	1977



## سنتز گفتار: تاریخچه ...

### ○ سنتز کننده‌های مدرن

روش و صدای تولیدی	توسط	سال
سیستم MITalk مبتنی بر فرمنت	Jonathan Allen و Dennis Klatt و Sheri Hunnicut در MIT	1979
سیستم Klattalk (مبنای سیستم‌های تجاری قرن بیستم)	Dennis Klatt در MIT	1981
Votrax اولین IC سنتز گفتار مبتنی بر فرمنت	Votrax	1980
سیستم Echo سنتزگر پیوندی دوواجی	شرکت Street Electronics	1982
سیستم تجاری Prose-2000 مبتنی بر روش کلات	شرکت Telesensory Systems	1982
سیستم Infovox - نخستین سنتز کننده فرمنت چندزبانه	Rolf Carlson, Bjorn Granstrom, Sheri Hunnicutt در سوئد	1982
سیستم تجاری DECTalk مبتنی بر روش کلات	شرکت Digital Equipment Corporation	1983



## سنتز گفتار: تاریخچه

### ○ سنتز کننده‌های مدرن

روش و صدای تولیدی	توسط	سال
توسعه روش اصلاح نوای PSOLA و استفاده در سنتزگرهای پیوندی	France از Charpentier and Moulines Telecom	1985
سنتز پیوندی مبتنی بر پیکره	محققان در ATR ژاپن	1990
سنتز در سیستم عامل ویندوز	مایکروسافت	2000



## سنتز گفتار: روش

### ○ تبدیل متن به گفتار شامل دو بخش اصلی

- پردازش متن

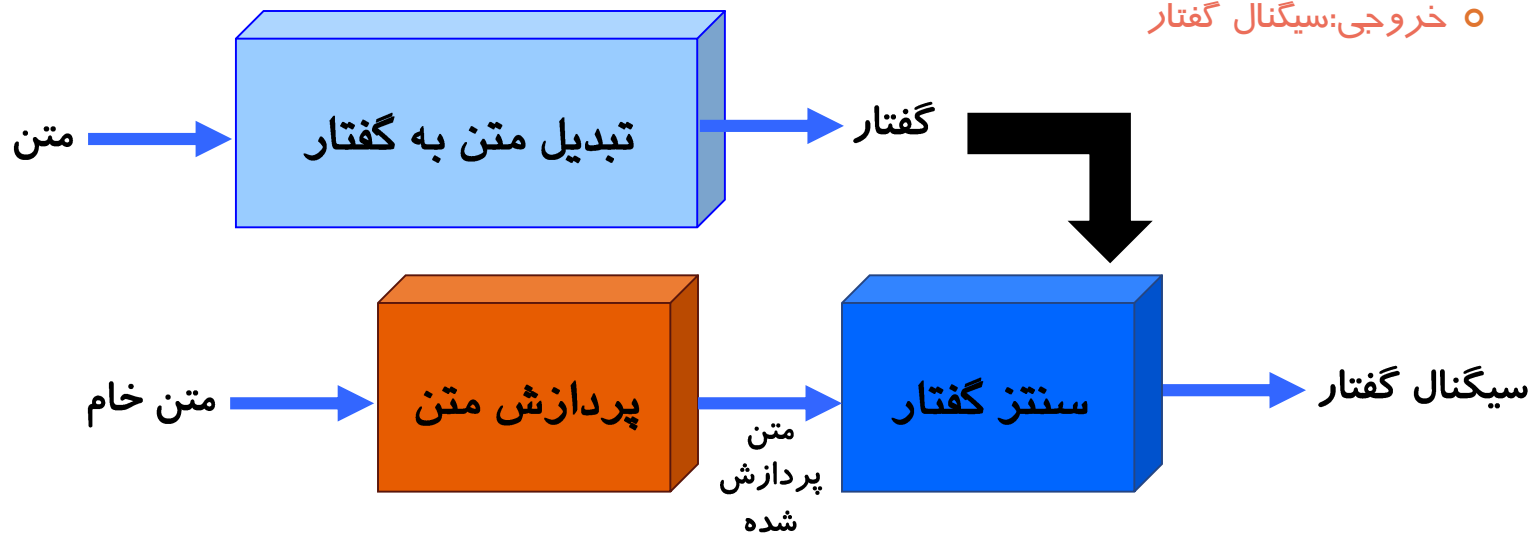
- ورودی متن خام

- خروجی: صورت تلفظی متن و اطلاعات زبانی دیگر

- تولید گفتار (سنتز گفتار)

- ورودی: صورت تلفظی متن و اطلاعات زبانی دیگر

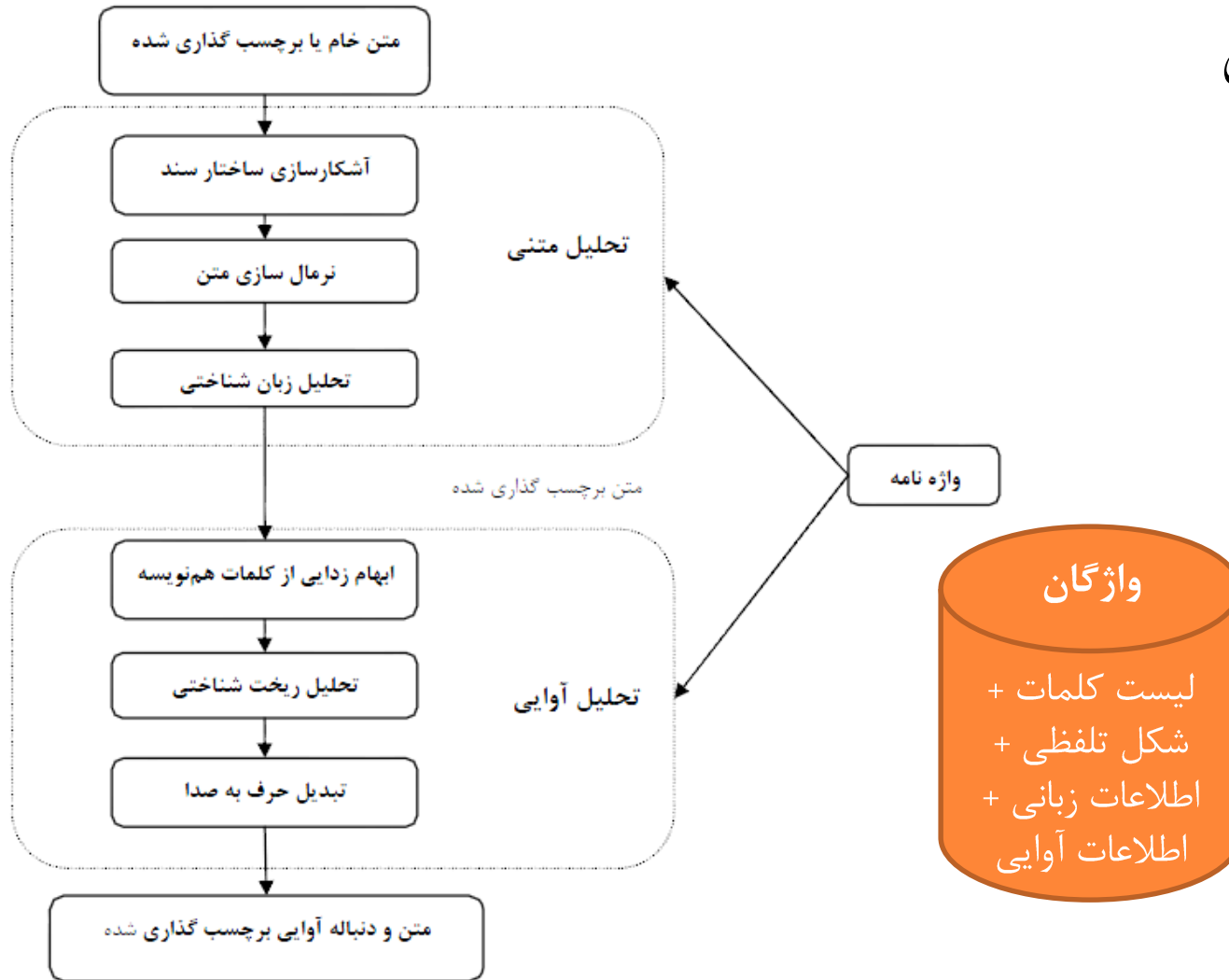
- خروجی: سیگنال گفتار





# سنتز گفتار: پردازش متن ...

## پردازش متن







## سنتز گفتار: پردازش متن ...

### پردازش متن

#### تشخیص ساختار متن = واحدیابی (Tokenization)

- جداکردن جمله‌ها، پاراگراف‌ها و بخش‌ها
- جمله‌ها: اغلب با یکی از علامت‌های پایانی {!?.} - همیشه ساده نیست
- مثال (انگلیسی): نقطه پایان جمله است یا بخشی از یک کلمه مخفف (مانند Dr.) است

Mr. Smith came by. He knows that it costs \$1.99, but I don't know when he'll be back (he didn't ask, "when should I return?")... His website is www.mrsmithhhhhh.com. The car is 72.5 in. long (we don't know which parking space he'll put his car in.) but he said "...and the truth shall set you free," an interesting quote.

- پاراگراف: شناسایی با نشانه‌های <CR> (carriage return) یا <NL> (new line)

- نوع اطلاعات و تشخیص ساختار وابسته به نوع متن است
- ایمیل: نیاز به اطلاعات فرستنده و گیرنده، بخش امضا
- وب سایت: آدرس، لینک، رنگ متن، ...

متن خام یا برچسب گذاری شده

آشکارسازی ساختار سند

نرمال سازی متن

تحلیل زبان شناختی

تحلیل متنی



# سنتز گفتار: پردازش متن ...

## پردازش متن

### • نرمال‌سازی متن: یکدست و استاندارد کردن متن

- اصلاح املایی و تبدیل کردن موارد خاص به صورت قابل بیان
- تصحیح کاراکترهای خاص و جدا کردن کلمات زبان‌های دیگر
- علائم نگارشی مانند . ! ? \* &

اعداد، شماره حساب‌ها، اعداد کسری مانند ۲۰٪، ۲/۵، ۰۹۱۲۴۳۳۱۲۵۴، 550-342 (234) +44

مبلغ و واحدهای پول مانند ، ۲۰۰۰۰۰ تومان

ساعت و تاریخ مانند 5:20 am و ۱۳۹۳/۱۰/۰۱

اختصارات مانند Dr، (ص)، خخخخخ

واحدها مانند mm

عبارات ریاضی و فرمول‌ها (شیمی)

شکل‌ها

و ...

:-) or :) SMILEY FACE (humor, laughter, friendliness, sarcasm)

:( or :( FROWNING FACE (sadness, anger, or disapproval)

;-) or :) WINKING SMILEY FACE (naughty)

:-D OPEN-MOUTHED SMILEY FACE (laughing out loud)

متن خام یا برچسب گذاری شده

آشکارسازی ساختار سند

نرمال سازی متن

تحلیل زبان شناختی

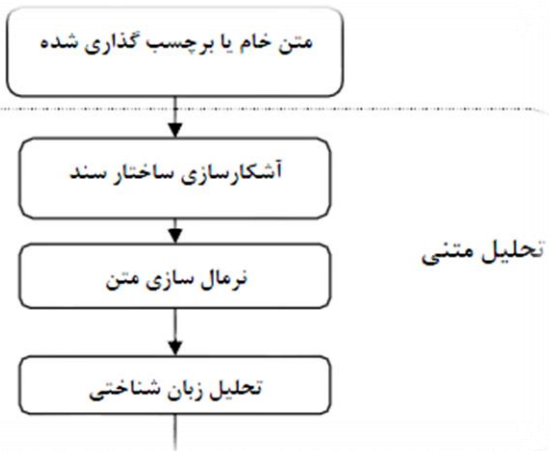
تحلیل متنی

12/19/94 (US)	December nineteenth ninety four
19/12/94 (European)	December nineteenth ninety four
04/27/1992	April twenty seventh nineteen ninety two
May 27, 1995	May twenty seventh nineteen ninety five
July 4, 94	July fourth ninety four
1,994	one thousand nine hundred and ninety four
1994	nineteen ninety four

CO	Colorado	commanding officer
	conscientious objector	carbon monoxide
IRA	Individual Retirement Account	Irish Republican Army
MD	Maryland	doctor of medicine
	muscular dystrophy	



## سنتز گفتار: پردازش متن ...



### ○ پردازش متن

#### • تحلیل زبانی: استخراج اطلاعات معنایی و نحوی

- تشخیص پاره گفتار (POS) یا نوع واژه، مثلاً، اسم خاص یا فعل
- معنای واژه، مثل river bank در مقابل money bank
- همبستگی واژه‌ها در عبارات، مثل ضرب‌المثل‌ها، عبارات نحوی، بندها، جمله‌ها
- ارجاع و مترادف بین واژه‌ها و عبارات
- تعیین گونه نحوی، مثل پرسش، نقل قول، امر، و غیره
- تعیین تمرکز معنایی (تأکید)
- تعیین گونه معنایی و کنش گفتاری، مثل درخواست، اطلاع‌رسانی، روایتگری، و غیره
- تحلیل ژانر و سبک
- کاربرد در ایجاد نوا و تلفظ‌نویسی
- تشخیص تلفظ کلمات هم نویسه و کسره اضافه
  - مثال (فارسی): مرد (اسم) = مَرْد؛ مرد(فعل)=مُرد
  - مثال (انگلیسی): "You record your voice" در مقابل "I'd like to buy that record"



# سنتز گفتار: پردازش متن ...

تحلیل آوایی

ابهام زدایی از کلمات هم‌نویسه

تحلیل ریخت‌شناختی

تبدیل حرف به صدا

متن و دنباله آوایی برچسب گذاری شده

## ○ پردازش متن

### • رفع ابهام هم‌نویسه‌ها

○ هم‌نویسه (Homograph): صورت نوشتاری یکسان اما تلفظ متفاوت

○ مثال (فارسی): مرد (اسم) = مَرَد؛ مرد (فعل) = مُرَد

○ مثال (انگلیسی): object [ah b jh eh k t] (verb) or [aa b jh eh k t] (noun)

### • تحلیل ریخت‌شناسی (مورفولوژی)

○ جداکردن پسوندها و پیشوندها

○ پسوند (انگلیسی): -s, 's, -ing, -ed, -est, -ment و غیره

○ پیشوند (انگلیسی): in-, un-, non-, pre-, sub- و غیره

○ تجزیه واژه‌های مرکب

○ یافتن سرنخ‌هایی از ریشه کلمات

○ کمک به یافتن تلفظ درست کلمات تصریفی و اشتقاقی (کلماتی که در واژگان نیستند)



## سنتز گفتار: پردازش متن ...

تحلیل آوایی

ابهام زدایی از کلمات هم‌نویسه

تحلیل ریخت‌شناختی

تبدیل حرف به صدا

متن و دنباله آوایی برچسب‌گذاری شده

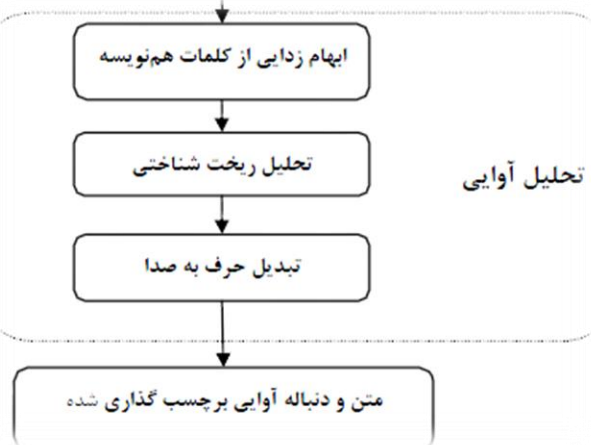
### ○ پردازش متن

#### • تبدیل حروف به صدا (تولید صورت تلفظی)

- استفاده از اطلاعات دریافتی در پردازش‌های مراحل قبل
- استفاده از واژگان (دیکشنری)
- حاوی لیست کلمات، شکل تلفظی آنها و اطلاعات زبانی و آوایی مرتبط
- باید قادر به تلفظ آوایی برای هر توالی از کلمات باشد
  - برخی از کلمات در واژگان نیستند
  - نیاز به روش‌های خودکار برای تلفظ نویسی
- روش‌های خودکار تلفظ نویسی
  - مبتنی بر قاعده: اگر  $k$  در اول کلمه باشد و بعد از آن  $n$  بیاید خوانده نمی‌شود (مانند *knife*)
  - یادگیری خودکار مبتنی بر داده: درخت *CART* یا روش آماری *HMM* یا شبکه عصبی



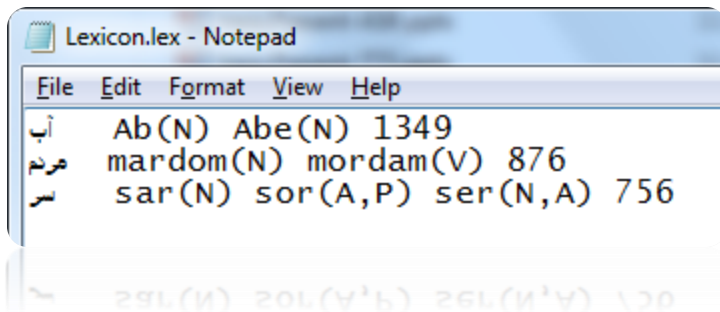
# سنتز گفتار: پردازش متن ...



## ○ ساخت واژگان

### ● به دو صورت

- شامل خود کلمه به طور کامل
- شامل ریشه‌ی کلمات و وندهای آنها
- مثال: کتاب‌ها، کتابی، کتاب‌هایی، برادرهایشانند، کتاب‌های...



### ● شامل

- مقوله‌های نحوی مختلف یک واژه
- تلفظ‌های مختلف هر کلمه و هر مقوله نحوی
- الگوی تکیه هر تلفظ
- آمارها و اطلاعات کمکی

### ● کاربرد در تحلیل آوایی



## سنتز گفتار: مشکلات پردازش متن ...

- ابهام ناشی از عدم نمایش واژه‌های کوتاه در نظام نوشتاری
  - دَسْت، دَسْتِ، دَسْتِ یا ...

- عدم وجود مرزهای دقیق بین کلمات

آنها وی را از بیمارستان مرخص نمی‌کردند

- مشکل کسره اضافه (که خاص زبان فارسی است)

سامانه متن خوان فارسی آریانا به دست متخصصین ایرانی (از دانشگاه صنعتی شریف) در سال ۱۳۸۹ طراحی و ساخته شد.

sAmAneye matnxAne fArsie AriAna be daste motaxasesine irani  
parantez baz az dane\$gahe sanatie \$arif parantez baste dar sAle  
hezaro sisado ha\$tado noh tarAhi va sAxte \$od noGte



## سنتز گفتار: مشکلات پردازش متن ...

### ○ مشکل کسره اضافه

- مهمترین عامل در وجود/عدم وجود کسره: ساختار نحوی جمله
  - کلمه‌هایی با نقش دستوری فعل، حرف اضافه و ضمیر، کسره اضافه نمی‌گیرند
- عامل دیگر نحوی نگارش و تلفظ کلمات است
  - کلماتی که با واکه‌ی بلند /A/ تمام می‌شوند، کسره به صورت /ye/ در تلفظ و به صورت /ی/ در نوشته ظاهر می‌شود. مثال: بقا ⇐ بقای
  - کلماتی که تلفظشان با واکه مرکب /w/ تمام می‌شود، کسره اضافه به صورت /ve/ تلفظ می‌شود. مثال: خسرو ⇐ خسرو
- عامل دیگر، عامل معنایی است (بسیار پیچیده)
  - مثال "اسرائیل بمب ساخت آمریکا را به هوا پرتاب کرد". از نظر نحوی، تلفظ کلمه‌ی ساخت به هر دو صورت با کسره و بدون کسره امکان‌پذیر است.
- عدم قطعیت در وجود کسره اضافه در برخی از جمله‌های فارسی
  - مثال: "آمریکا بمب ساخت عراق را نابود کرد".
    - با کسره: آمریکا بمب ساختِ عراق را نابود کرد
    - بدون کسره: آمریکا بمب ساخت، عراق را نابود کرد
    - هر دو تلفظ چه از نظر نحوی و چه از نظر معنایی صحیح هستند.





## سنتز گفتار: مشکلات پردازش متن ...

○ مشکل هم نویسه‌ها (در زبان فارسی بسیار شدید است)

یکی فعل و یکی اسم

- گُرد و کُرد
- مَرَد و مُرد

هر دو اسم

- شکر و شُکر

هر دو فعل

- می شوید (شدن) و می شوید (شستن)
- می روید (رفتن) و می روید (رویدن)





## سنتز گفتار: مشکلات پردازش متن ...

### ○ شکل‌های مختلف نوشتار کلمات (مشکل کردن نرمال‌سازی متن)

- "می‌شود، میشود، می‌شود" یکی هستند
- "مسئولیت، مسوولیت، مسؤلیت" یکی هستند

### ○ نوشتن از راست به چپ

- متفاوت از زبان انگلیسی ⇐ ترکیب دو زبان پیچیده می‌شود

### ○ استثناهای زیاد

- مشکل بودن پشتیبانی از چندین زبان (فارسی، انگلیسی، چینی، ...)
- متفاوت بودن مرز کلمه‌ها و واج‌ها
- تفاوت در نوشتن اعداد
- ...



## سنتز گفتار: مشکلات پردازش متن

### ○ زبان‌های آوایی (phonetic languages)

- زبان‌هایی که فرم نوشتار و خواندن کلمات یکسان است
  - اسپانیایی، فنلاندی، کردی

- کم بودن مشکلات پردازش متن

دهرگه که تم لی بکه وه

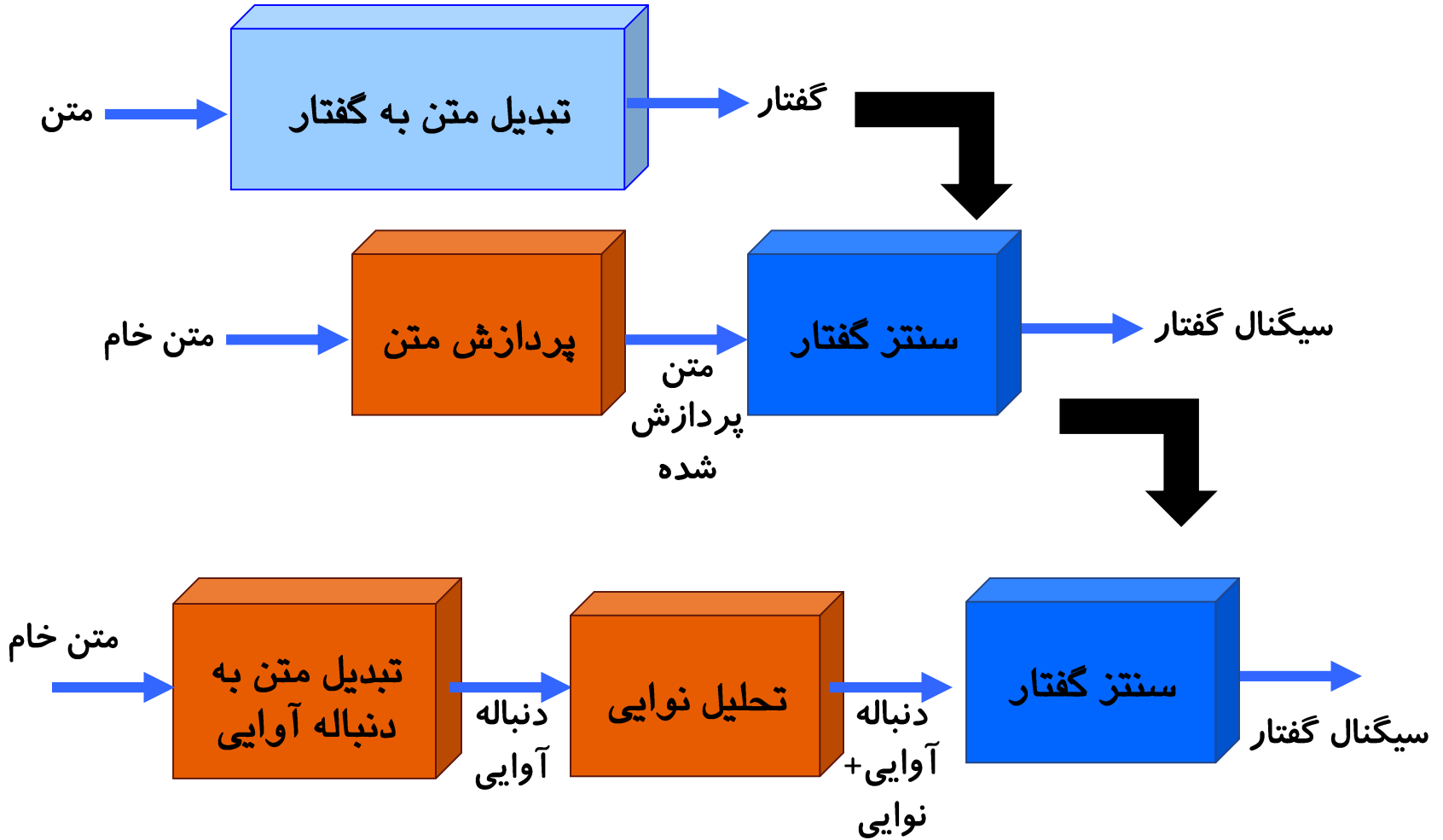
کفت و ماندووی ریگای دوورم؛

وه کوو جاران،

که میک له لات ده مینمه وه و راده بوورم



# سنتز گفتار: تحلیل نوایی ...





## سنتز گفتار: تحلیل نوایی ...

### ○ تحلیل نوایی (Prosody) ...

- پیام گفتاری یا نوشتاری = معنای صریح (Denotation)
- نوا = اثرات احساسی و هدف‌دار اعمال شده توسط گوینده یا استنباط شده توسط شنونده = بخشی از معنای ضمنی (Connotation)
- بررسی آهنگ و لحن بیان جملات
- بافت پیچیده‌ای از اثرات فیزیکی و آوایی
- نقش حمایتی در هدایت بازیابی پیام‌های اصلی (معنای صریح) توسط شنونده
- نقش اساسی در نشان دادن معنای ضمنی، یا مقصود گوینده نسبت به پیام، نسبت به شنونده و نسبت به کل رویداد ارتباط

متن خوان فارسی آریانا در سال ۱۳۸۹ ساخته شد.

متن خوان فارسی آریانا در سال ۱۳۸۹ ساخته شد؟

متن خوان فارسی آریانا در سال ۱۳۸۹ ساخته شد؟



## سنتز گفتار: تحلیل نوایی ...

### ○ تحلیل نوایی (Prosody)

- مکث (Pause): مشخص کردن عبارات و نفس گرفتن
- زیروبمی (Pitch): سرعت چرخه تارهای صوتی (بسامد پایه یا F0) - تابعی از زمان
- سرعت/دیرش نسبی (Rate/Relative Duration): زمان‌بندی و ریتم واج
- بلندی (Loudness): دامنه/حجم صدای نسبی
- پیش‌بینی نوای گفتار کار بسیار مشکلی است
- نوای خوب به درک گوینده یا خواننده از معنای متن یا پیام بستگی دارد
- قانون طلایی سخنران رومی Quintilian می‌گوید "برای اینکه کسی بتواند خوب صحبت کند، و کلمات را با تأکید صحیح تلفظ کند، باید معنی آنچه می‌گوید را بفهمد، کاملاً از آن راضی باشد و تأثیری را که می‌خواهد بر دیگران ایجاد کند در خود داشته باشد»
- سیستم‌های امروزی حالت ساده‌ای از نوا دارند



## سنتز گفتار: تحلیل نوایی ...

### ○ تکیه (Stress) و تأکید (Accent) ...

#### • دو نوع تکیه

○ تکیه واژگانی (Lexical Stress): در سطح واژگان - آنچه در لغت‌نامه‌ها آورده می‌شود

○ مثال: **employer**

○ تکیه/تأکید زیروبمی (Pitch Accent): بر اساس معنی (در سطح پاره گفتار)

○ مثال: I didn't say **employ**er, I said employ**ee**

#### • مثال: «هر کس باید تنها با زمان کشتی بگیرد»

○ تکیه در هجای اول یعنی «تن»: هر کس باید «به تنهایی» و بدون یآوری با زمانه دست و پنجه نرم کند

○ هجای دوم آن یعنی «ها»: هر کس باید فقط با زمانه کشتی بگیرد، نه با چیز دیگری

#### • مثال: «منصور، پدر احمد رفت.»

○ تکیه بر روی هجای اول: «منصور» مخاطب است و به او اطلاع داده می‌شود که پدر احمد رفته است.

○ تکیه بر روی هجای دوم: «منصور» نام پدر احمد است که گوینده رفتن وی را اطلاع می‌دهد



## سنتز گفتار: تحلیل نوایی ...

### ○ تکیه (Stress) و تأکید (Accent) ...

- در یک واژه هجاهایی را که با تکیه ادا می‌شوند «هجای تکیه‌بر» می‌خوانند = تغییر معنی

#### ○ مثال فارسی

- «کریم» == «بخشنده» یا «کر هستیم»
- «ولی» == «اما» یا «سرپرست»
- «کتابی» == «یک کتاب» یا «به شکل کتاب=رسمی»

#### ○ مثال انگلیسی

- «address» == «آدرس» و «مخاطب ساختن»
- «permit» == «جواز» و «اجازه دادن»
- «content» == «محتوا» و «راضی و خشنود»

- تعیین محل تکیه با واژه‌نامه یا مدل‌های آماری و درخت CART





## سنتز گفتار: تحلیل نوایی ...

### ○ تکیه (Stress) و تأکید (Accent) در فارسی

- تکیه اسامی، صفات، و بیشتر قیود در هجای پایانی است
  - سر: (اسم) هجای اول
  - کتاب: (اسم) هجای دوم
  - خوب: (صفت) هجای اول
  - سنجیده: (صفت) هجای سوم
- تکیه افعال مثبت، بر آخرین هجای نخستین سازه (ریشه) واقع می‌شود
  - در "xar-id-am"، تکیه روی "id" است = هجای پایانی ستاک ماضی (xarid)
- تکیه افعال منفی، بر تنها هجای تکواژ نفی واقع می‌شود



## سنتز گفتار: تحلیل نوایی ...

### ○ نواخت (Tone) ...

- تغییرات زیروبمی در یک واژه که باعث تغییر معنا در آن واژه می‌شود
- بیانگر سطوح یا تغییرات مهم  $F_0$  بر روی هجاها

### • زبان‌های نواختی (Tone Languages): دارای نواخت هستند

○ تغییر نواخت منجر به تغییر معنی کلمه می‌شود

○ چینی، ژاپنی، تایلندی و ...

○ مثال: در زبان‌های چینی ماندارین (Mandarian Chinese)

- تکواژ «li» بر حسب اینکه «زیر» تلفظ شود یا «بم» به دو معنی «کلابی» و «شاه بلوط» است
- تکواژ «ma» اگر «یکنواخت» تلفظ شود یا «زیر» یا «بم» در مفاهیم مختلف «مادر»، «اسب» و «سرزنش» به کار می‌رود

### • زبان‌های آهنگی: زبان فارسی

○ تغییر سطح و جهت زیروبمی، عدم تغییر معنای واژگانی، تغییر معنای بافتی پاره گفتارها

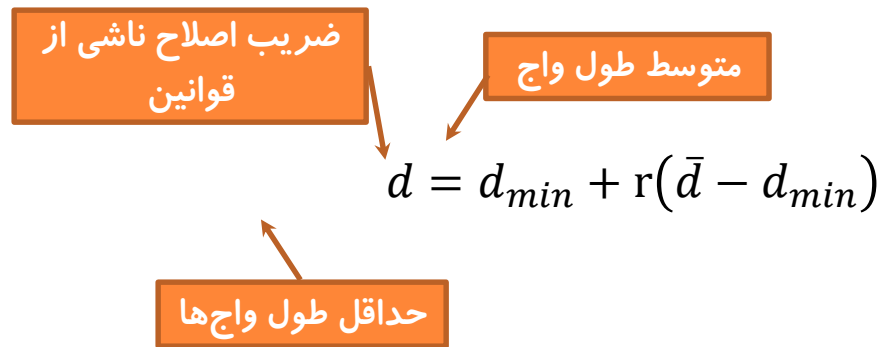


# سنتز گفتار: تحلیل نوایی ...

## ○ دیرش (Duration)

- زیروبمی و دیرش به طور کامل مستقل نیستند (دارای عوامل تاثیر گذار مشترک)
- کشش واکه‌ها در زبان عربی و انگلیسی ممیز معنی است در حالی که در زبان فارسی چنین نیست
- Sheep و Ship یا Beat و Bit

### • روش‌های مبتنی بر قاعده



### • روش‌های مبتنی بر CART

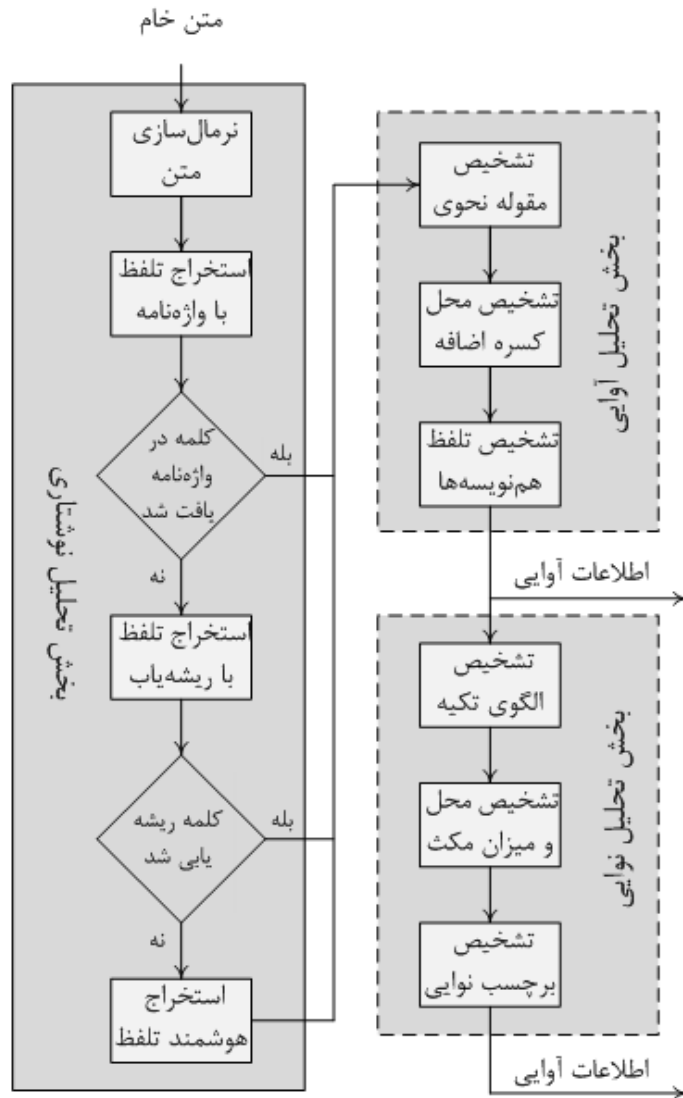
- استفاده از ویژگی‌های خود آوا، تکیه واژگانی اولیه (دارد/ندارد)، بافت سمت چپ و راست (هر کدام یک آوا)

طولانی شدن واکه پایانی و همخوان‌های بعدی در هجاهای پیش‌سکوتی
کوتاه شدن تمام بخش‌های هجایی در جایگاه غیر پیش‌سکوتی
کوتاه شدن بخش‌های هجایی اگر در هجای پایانی واژه نباشند
همخوان‌ها در جایگاه غیر از ابتدای واژه کوتاه می‌شوند
آواهای بدون تکیه و دارای تکیه‌ی ثانویه کوتاه می‌شوند
واکه‌های مورد تأکید طولانی می‌شوند
واکه‌ها ممکن است طبق مشخصه‌های آوایی بافت خود کوتاه یا بلند شوند
همخوان‌ها ممکن است در خوشه‌ها کوتاه شوند



# سنتز گفتار: پردازش متن و تحلیل نوایی

## تبدیل متن به دنباله آوایی و نوایی

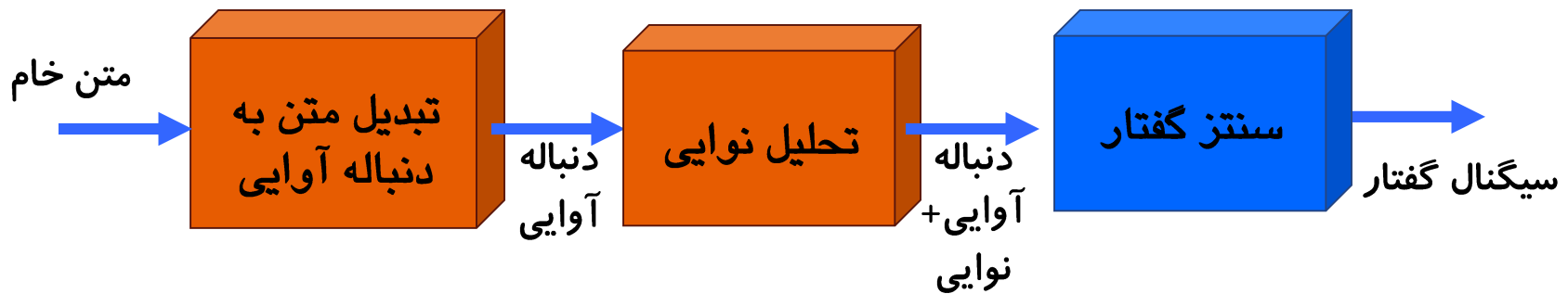




## سنتز گفتار: تولید سیگنال گفتار ...

### ○ تولید سیگنال گفتار = شکل موج

- ورودی: متن تجزیه شده به رشته آوایی + اطلاعات نوایی



### ○ تولید صدای طبیعی

- خوشایند بودن
- قابل فهم بودن

### ○ انعطاف در صدای تولیدی

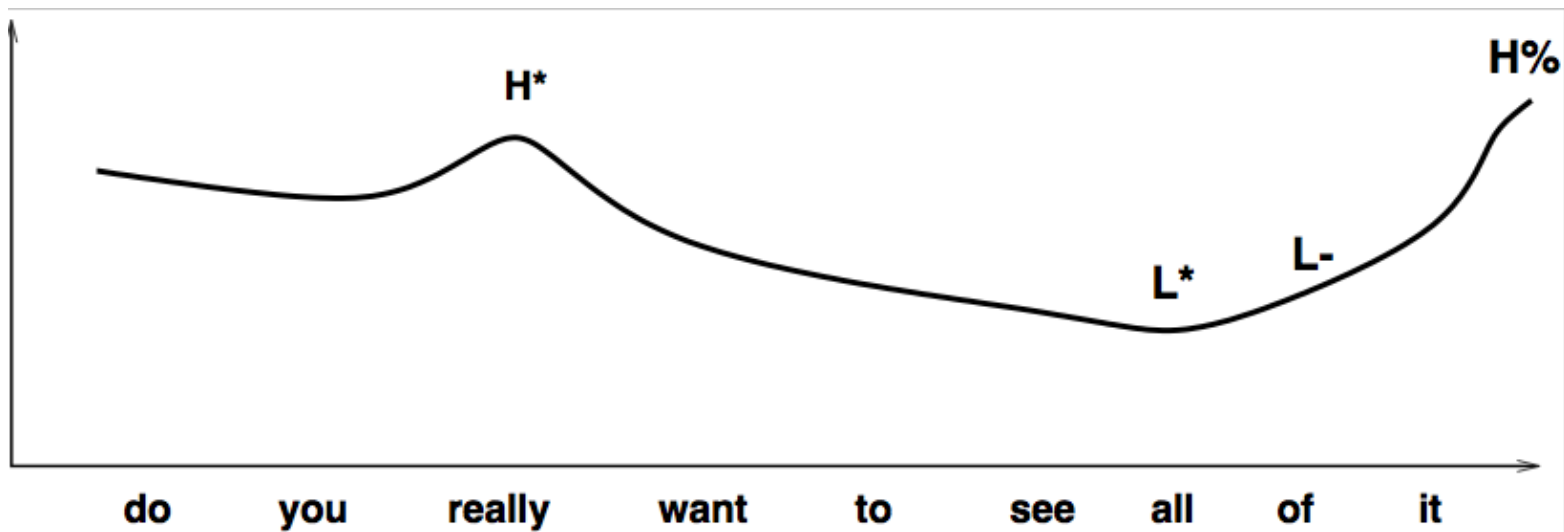
- تغییر آسان جنسیت، زیری و بمی، دامنه‌ی زیری و بمی و امثال آن



# سنتز گفتار: تولید سیگنال گفتار ...

## ورودی

		H*												L*		L- H%					
do		you		really						want			to		see		all		of	it	
d	uw	y	uw	r	ih	l	iy	w	aa	n	t	t	aɪ	s	iy	ao	l	ah	v	ih	t
110	110	50	50	75	64	57	82	57	50	72	41	43	47	54	130	76	90	44	62	46	220





## سنتز گفتار: تولید سیگنال گفتار ...

### ○ روش‌ها ...

#### • سنتز تولیدی (Articulatory)

- مدل کردن حرکات مکانیکی اندام‌های تولید و توزیع‌های سرعت و فشار صدا در شش‌ها، حنجره، و دستگاه صوتی و خیشومی
- هر آوا توسط قوانینی که حرکات اندام گویایی را در زمان و مکان مدل می‌نماید، ارائه می‌گردد
- پیاده‌سازی آن نسبت به روش‌های دیگر مشکل است
- اولین سنتز کننده مفصلی در سال ۱۹۵۸ توسط George Rosen در دانشگاه MIT

#### • سنتز فرمندی (Formant)

- مدل کردن مجرای گفتار به صورت ترکیب چند لوله صوتی (هر کدام بسامد تشدید مخصوص دارند)
- سه تا چهار فرمنت اول را به عنوان ویژگی‌های یک واحد صوتی در نظر می‌گیرند
- بسامد، پهنای باند، دامنه فرمنت‌ها، مشخصات سیگنال تحریک شامل واکدار یا بی‌واک بودن، دوره زیروبمی در صورت واکدار بودن و مقدار بهره
- استفاده از مدل منبع-فیلتر

○ سنتز گر کلات Klattalk در ۱۹۸۱



## سنتز گفتار: تولید سیگنال گفتار ...

### ○ روش‌ها

#### • سنتز پیوندی (Concatenative)

- مبتنی بر پیکره
- قطعات صدای ذخیره شده در کنار یکدیگر قرار می‌گیرند تا صدای اصلی را تولید کنند.
- قطعه: آوا، دو آوایی، هجا، واژه یا حتی جمله
- مشکل اصلی: گسستگی‌های ایجاد شده در محل پیوند و کمبود نوا در گفتار تولید شده
- تحقیقات: طبیعی‌سازی
- روش انتخاب واحد (Unit Selection)

#### • سنتز آماری پارامتری (Statistical Parametric)

- استخراج پارامترهایی از صداهای مختلف = ایجاد مدل آماری صدا (آموزش)
- تولید صدا بر اساس مدل‌سازی آماری پارامترها
- مشکل: صدای مصنوعی
- روش: آماری مبتنی بر مدل مخفی مارکوف (به صورت‌های مختلف)
- شروع توسط یوشیمورا و توکودا در آزمایشگاه ناگویا در سال ۲۰۰۱ (سیستم Straight)





## سنتز گفتار: تولید سیگنال گفتار ...

### روش‌ها (دسته‌بندی از نظر دیگر)

#### • مبتنی بر قانون (synthesis by rule)

- استفاده از مجموعه‌ای از قوانینی که به طور دستی معین شده‌اند
- مانند سنتز گره‌های فرمنت و سنتز تولیدی

#### • مبتنی بر داده (data-driven synthesis)

- پارامترهای سنتز کننده به طور خودکار از داده‌های گفتار واقعی به دست می‌آیند
- مانند سنتز گره‌های پیوندی





## سنتز گفتار: تولید سیگنال گفتار ...

### ○ مهم‌ترین ویژگی‌های روش‌های تولید سیگنال گفتار

- کیفیت (و متغیر بودن کیفیت)

- کیفیت باید در جملات و متون مختلف بالا باشد

- محاسبات و تأخیر: زمانی که طول می‌کشد تا سنتز گر شروع به صحبت کند

- منابع حافظه: سنتز گرهای مبتنی بر قانون حافظه کمتر نیاز دارند

- منابع پردازشی: سنتز گرهای پیوندی در جستجوی توالی بهینه پردازش زیاد نیاز دارند

- انعطاف در صدای تولیدی

- سرعت متغیر: تولید گفتار سریع توسط افراد نابینا و مرور اجمالی مطالب مفید است - سنتز گرهای پیوندی که شکل موج را اصلاح نمی‌کنند در این ویژگی مشکل دارند

- کنترل زیروبمی: مشکل سنتز گرهای پیوندی با این موضوع

- ویژگی‌های خاص صدا: تولید صداها خاص (مانند رباطی) در سایر سیستم‌های زبان گفتاری

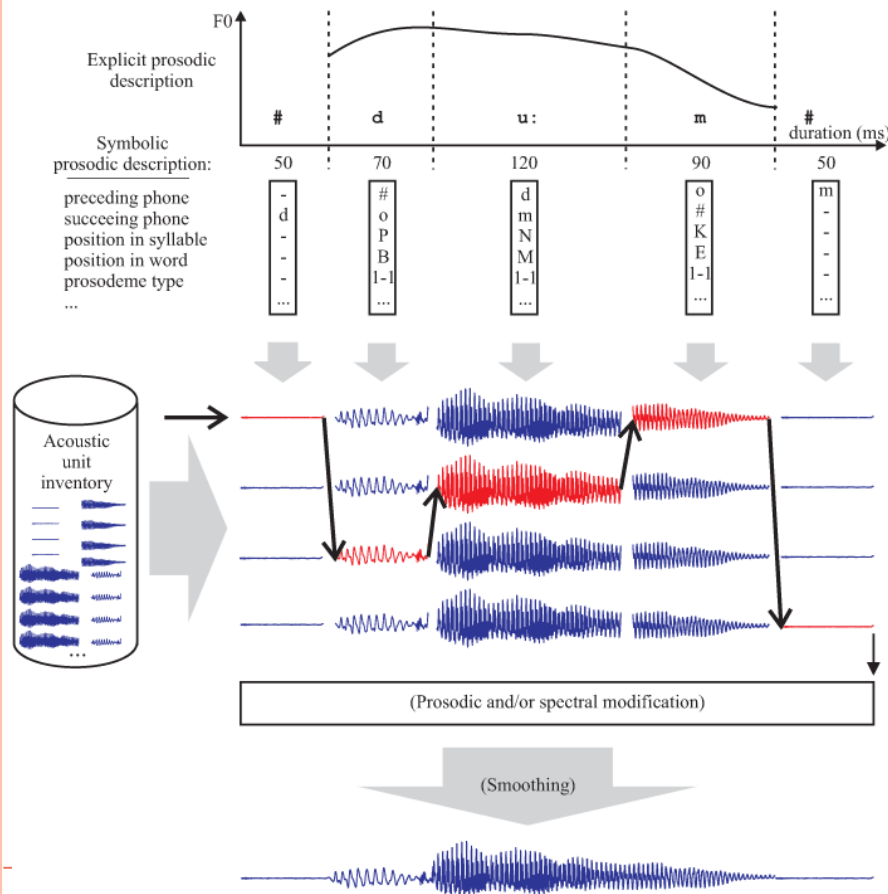
- به صورت عمومی روش‌های مبتنی بر قانون (فرمنت) و پارامتری انعطاف بالایی دارند



# سنتز گفتار: تولید سیگنال گفتار (سنتز پیوندی) ...

ایده

• الحاق قطعه گفتارهای از قبل ضبط شده متناسب با رشته واجی متن



• روش: انتخاب واحد (Unit Selection)



## سنتز گفتار: تولید سیگنال گفتار (سنتز پیوندی) ...

### مزیت

- نه به قانون نیاز دارد نه به تنظیم دستی
- هر قطعه کاملاً طبیعی است، بنابراین خروجی بسیار طبیعی است (?)
- روش رایج در سیستم‌های تجاری امروزی

### عیب

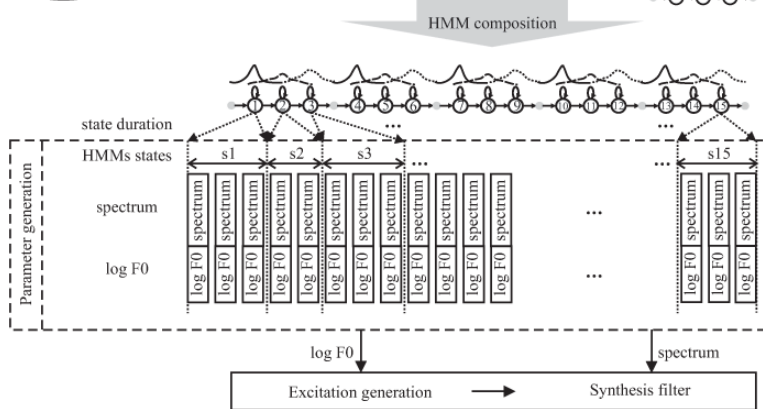
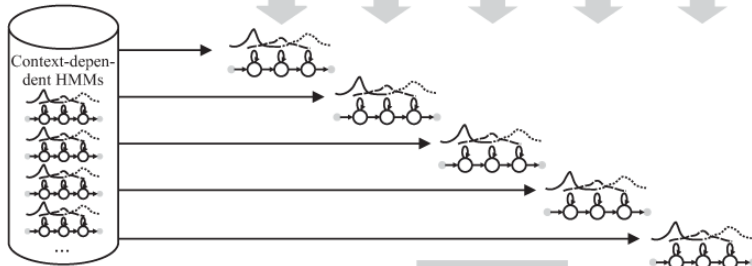
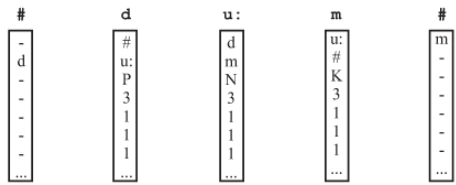
- انعطاف پایین (در تغییر نوا، گوینده و ...)
- نیاز به داده آموزش (زیاد)
- ممکن است کیفیت پایین باشد (طبیعی نباشد!!)
- اگر دو قطعه گفتار را که مجاور هم نبودند با هم الحاق کنیم، ممکن است ناپیوستگی طیفی یا نوایی وجود داشته باشد. برای شنونده، گفتار سنتز شده حاوی ناپیوستگی‌های بزرگ، کیفیت پایینی دارد.
- عیب اصلی: کیفیت متغیر (یک جمله کیفیت عالی ارائه می‌کند و در جمله بعد کیفیت بسیار بد)



# سنتز گفتار: تولید سیگنال گفتار (سنتز آماری پارامتری) ...

Symbolic prosodic description:

preceding phone  
succeeding phone  
position in syllable  
number of phones in syllable  
syllable stress  
position of syllable in word  
position of word in phrase  
...



## ایده

- استخراج ویژگی از واحدهای گفتار
- مدل کردن ویژگی‌های واحدهای مختلف
- استفاده از مدل‌ها در تولید صدا

## روش‌ها

- مدل مخفی مارکوف (HMM)



## سنتز گفتار: تولید سیگنال گفتار (سنتز آماری پارامتری)

### ○ مزیت

- انعطاف بسیار بالا
  - تطبیق آسان به صدای یک گوینده خاص
  - تغییر صدا
- کیفیت قابل قبول
- حجم پایین
  - استفاده در تلفن‌های همراه

### ○ عیب

- مشکل بودن تولید صدای طبیعی

## سنتز گفتار: ارزیابی کارایی

### بخش پردازش متن

- تشخیص درست تلفظ (شامل کسره، هم نویسه و ...)

$$\text{دقت} = \frac{\text{تعداد کلماتی که درست تلفظ شده‌اند}}{\text{تعداد کل کلمات}}$$

### بخش تولید سیگنال

- قابلیت فهم (Intelligibility)

- آزمون DRT: Diagnostic Rhyme Tests
- آزمون MRT: Modified Rhyme Test

- طبیعی بودن (Naturalness)

- معیار MOS: Mean Opinion Score

