

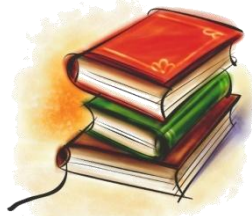
پردازش گفتار

معرفی و کاربردها

هادی ویسی

h.veisi@ut.ac.ir

دانشگاه تهران - دانشکده علوم و فنون نوین



○ معرفی

○ کاربردهای پردازش گفتار

- بازشناسی (تشخیص) گفتار
- سنتز گفتار
- دیالوگ و فهم گفتار
- شناسایی گوینده
- بهسازی گفتار
- کدینگ و فشرده‌سازی

معرفی ...

○ گفتار

اولین و ساده‌ترین راه ارتباطی



○ ارسطو:

انسان، حیوان **ناطق** است



معرفی ...

○ پردازش گفتار

- ایجاد توانایی‌های انسان در زمینه صحبت کردن، شنیدن، فهمیدن و یادگیری برای رایانه
- ساخت سیستم‌های زبان گفتاری (Spoken Language System)
- گفتار تبدیل به اولین وسیله ارتباطی بین انسان و کامپیوتر شود





معرفی

○ چرا گفتار؟

یادگیری نیاز ندارد	طبیعی
دست و چشم را درگیر نمی کند	منعطف
نرخ انتقال داده بالا دارد	بهینه
ارتباط ارزان	اقتصادی
علاوه بر کلمات، اطلاعات دیگر را نیز منتقل می کند	تاثیر گذار
انسانها آن را ترجیح می دهند	رایج
کامل تر از زبان نوشتاری است	تکامل





کاربردهای پردازش گفتار ...

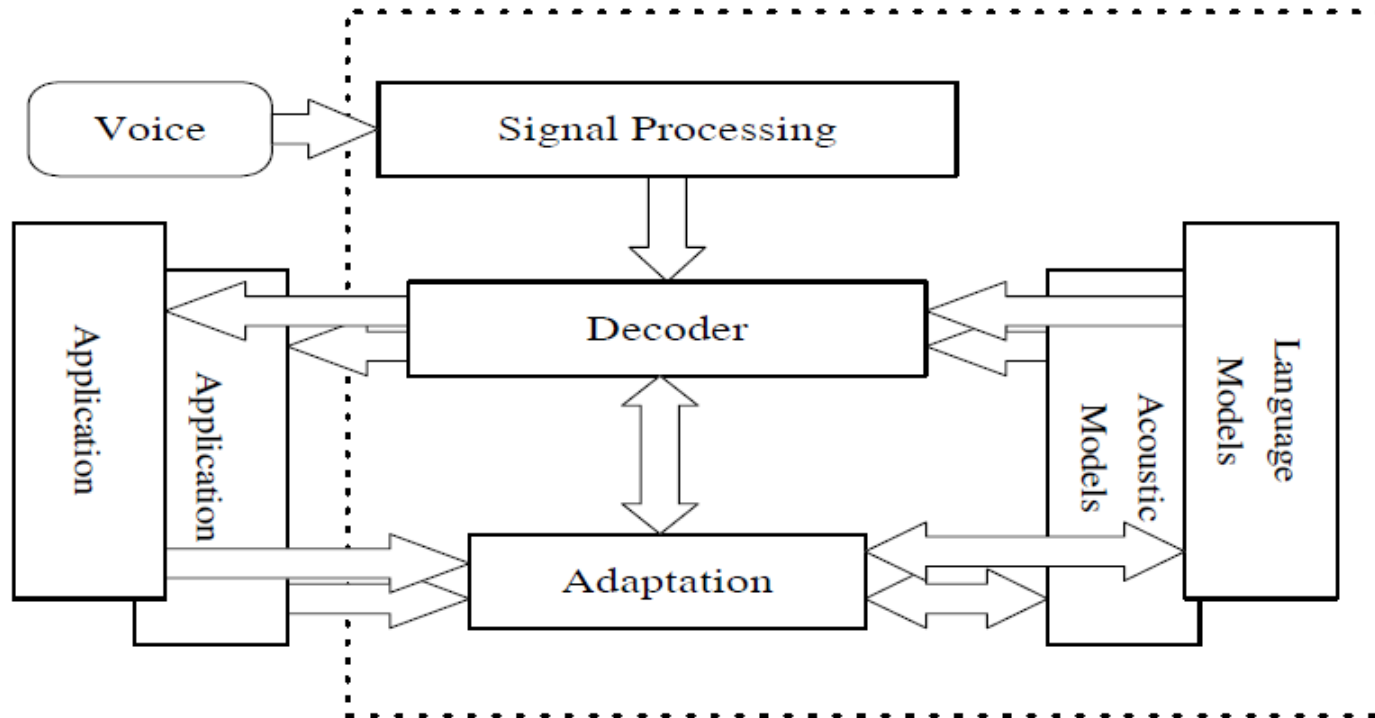
- **بازشناسی (تشخیص) گفتار (Speech Recognition)**
 - تبدیل گفتار به متن
- **سنتز گفتار (Speech Synthesis)**
 - تبدیل متن به گفتار (Text-to-Speech)
- **دیالوگ و فهم گفتار (Speech Dialogue/Understanding)**
 - درک گفتار (متفاوت با تشخیص)
- **شناسایی گوینده (Speaker Recognition)**
- **بهبودی گفتار (Speech Enhancement)**
- **کدینگ و فشرده‌سازی**



کاربردهای پردازش گفتار ...

○ بازشناسی (تشخیص) گفتار ...

• ASR: Automatic Speech Recognition



کاربردهای پردازش گفتار ...

○ بازشناسی (تشخیص) گفتار ...

- تایپ گفتاری و سیستم دیکته

- تشخیص فرامین و دستورات صوتی

- اجرای برنامه ها در رایانه با بیان نام آنها، کنترل لوازم خانگی با صوت
- فرمان دادن به رباتها و فرمانهای صوتی در خودرو
- کیوسکهای اطلاعات و دستگاههای خود پرداز بانکها
- استفاده در بازیهای کامپیوتری (افزایش قابلیتها و جذابیت)
- و ...



کاربردهای پردازش گفتار ...



○ بازشناسی (تشخیص) گفتار

- کاربردهای مربوط به معلولین، ناشنویان و نابینایان
 - معلولین حرکتی: صحبت کردن برای استفاده از وسایل و ابزارها
 - ناشنویان: تایپ سخنان دیگران، تبدیل گفتار به حرکات ایما و اشاره
 - نابینایان: تایپ گفتاری و ورود اطلاعات، صحبت کردن برای استفاده از وسایل و ابزارها

• سیستم‌های اطلاع‌رسانی

- تلفن گویاها

• سیستم‌های آموزش

- پرسش و پاسخ

• کاربردهای ترکیبی: ترجمه گفتار به گفتار

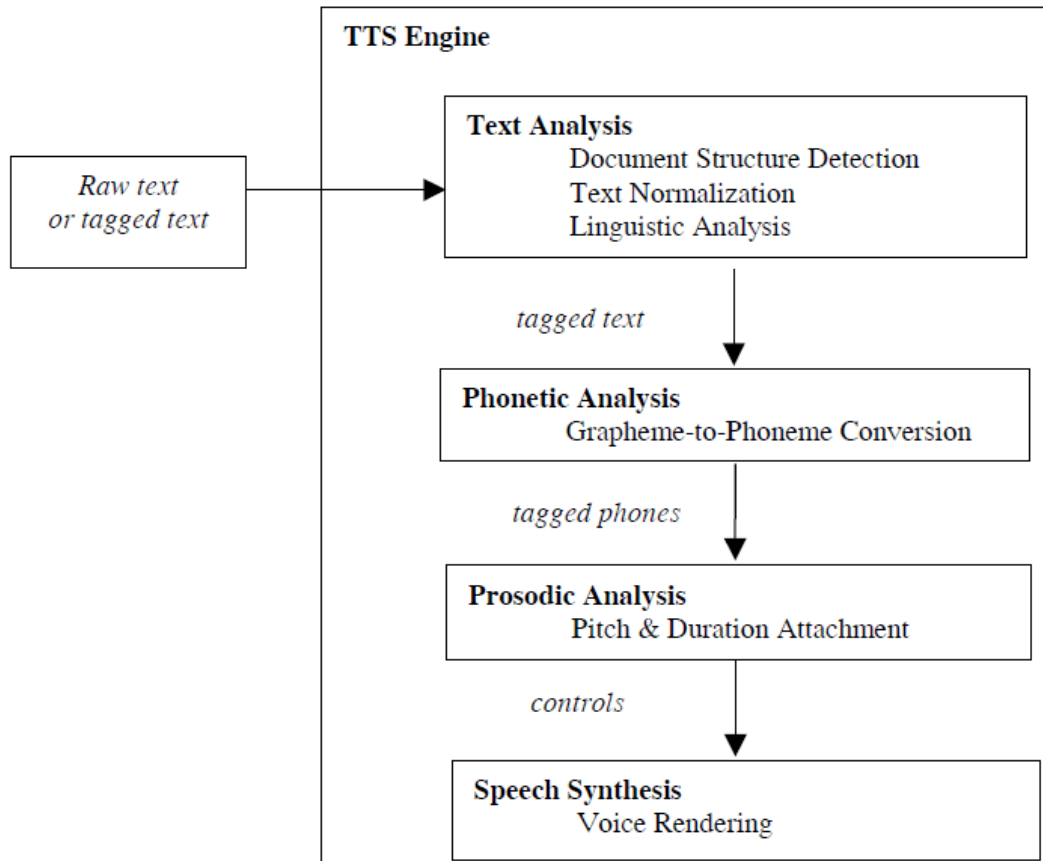




کاربردهای پردازش گفتار ...

○ سنتز گفتار ...

• TTS: Text-to-Speech



کاربردهای پردازش گفتار ...

○ سنتز گفتار ...

- سخنگو کردن کامپیوتر

- سیستم‌های آموزشی (کتاب‌های الکترونیکی، محتوای آموزشی، آموزش از راه دور)
- وب سایت‌ها (اخبار و اطلاعات، آموزش)
- تلفن‌های همراه (خواندن پیامک، نقشه، کتاب و ...)
- رفع مشکلات و محدودیت‌های افراد ناتوان جسمی در صحبت کردن و خواندن



کاربردهای پردازش گفتار ...



○ سنتز گفتار ...

- رفع مشکلات و محدودیت‌های افراد ناتوان در خواندن
 - نابینایان، افراد کم‌سواد و بی‌سواد و سالمندان
 - سیستم‌های صفحه‌خوان (Screen Reader) ویژه نابینایان
 - JAWS
 - HAL
 - Talks (موبایل)

- سامانه‌های اطلاع‌رسانی مانند سیستم‌های تلفنی (تلفن گویا)، کیوسک‌ها، نوبت‌دهی بانک‌ها
 - تغییر سریع و آسان پیغام‌های صوتی بدون نیاز به ضبط صدا

- استفاده در نرم‌افزارهای دیگر: مترجم گفتار به گفتار، OCR

- فشردن سازی گفتار (کد کردن)

کاربردهای پردازش گفتار ...

○ دیالوگ و فهم گفتار ...

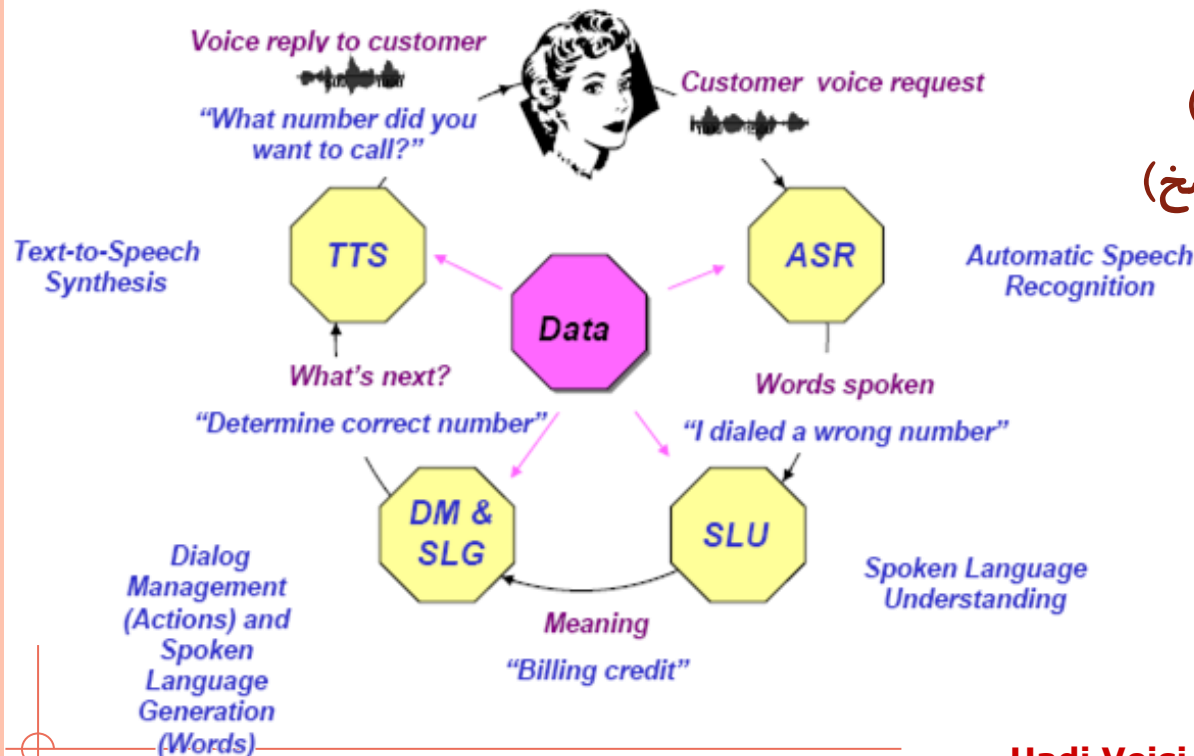
• Spoken Dialogue System

• تشخیص گفتار (ورود اطلاعات)

• درک (فهم) گفتار

• تولید جمله (تولید پاسخ)

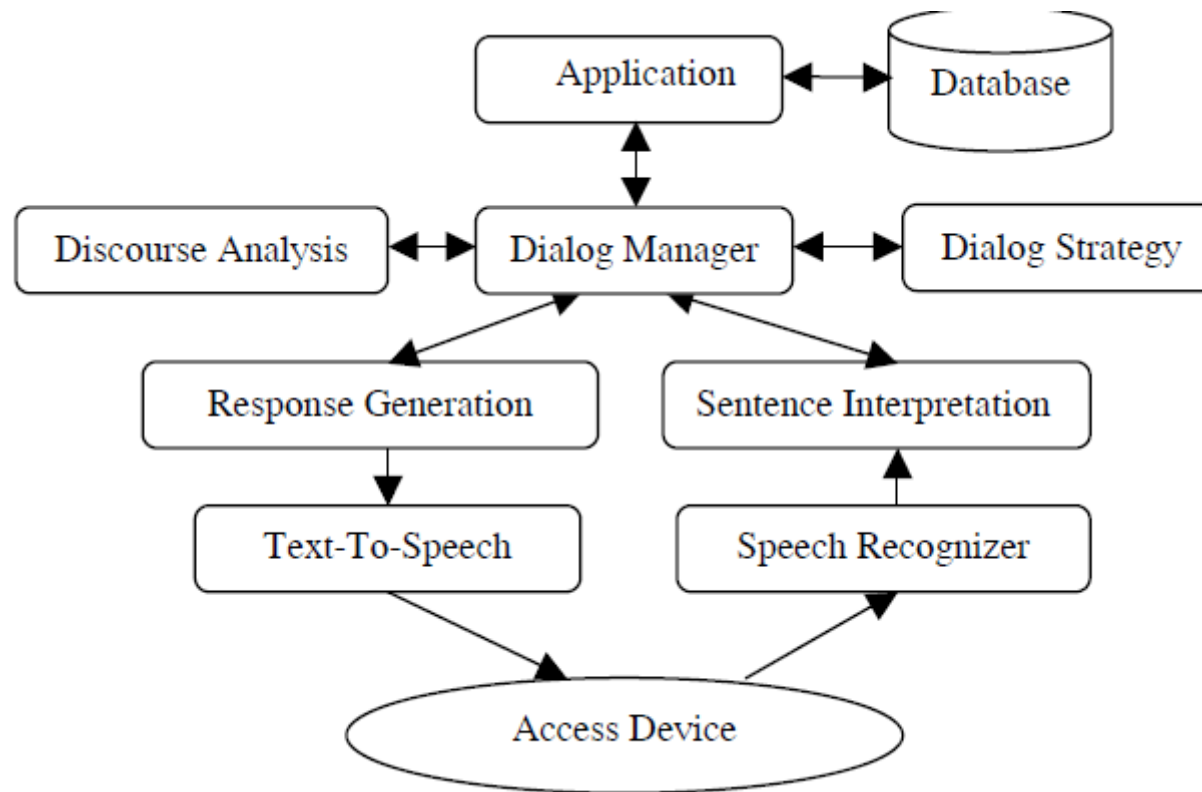
• تولید گفتار (خواندن پاسخ)





کاربردهای پردازش گفتار ...

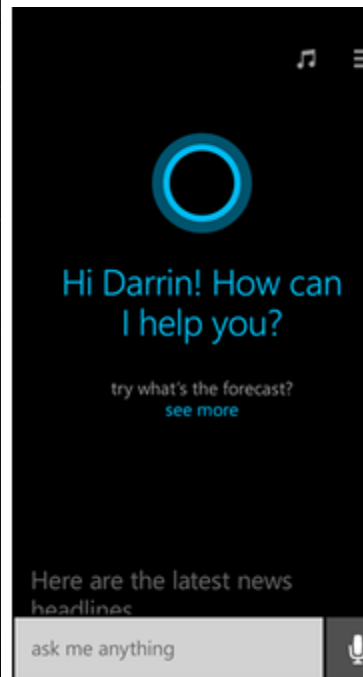
○ دیالوگ و فهم گفتار ...



کاربردهای پردازش گفتار ...

○ دیالوگ و فهم گفتار ...

- iPhone روی Apple Siri
- Microsoft Cortana روی ویندوز موبایل ۸.۱
- Google Now روی اندروید ۴.۱ به بعد (و iOS)





کاربردهای پردازش گفتار ...

○ دیالوگ و فهم گفتار (سیستم‌های تلفنی)

- Microsoft Tellme
- Jupiter (اطلاعات آب و هوای ۵۰۰ شهر دنیا)

Jupiter

A conversational interface for on-line weather information over the phone.

1-888-573-8255

(outside the USA: 1-617-258-0300)

<http://www.sls.lcs.mit.edu/jupiter>

Spoken Language Systems Group,
MIT Laboratory for Computer Science



Microsoft
Tellme.

Say it. Get it.

or call Tellme at
1-800-555-TELL

+1-800-555-8355

کاربردهای پردازش گفتار ...

○ شناسایی گوینده ...

● تشخیص گوینده (Speaker Identification)

○ کسی که صحبت می کند، کیست؟

○ زمینه های مشابه

○ تشخیص جنسیت (Gender Identification)

○ تشخیص زبان (Language Identification)

● تایید گوینده (Speaker Verification)

○ آیا او واقعاً حسین است؟

● جداسازی گوینده (Speaker Diarization)

○ جداسازی بخش های یک مکالمه (دو یا چند نفره) به تفکیک گوینده

○ چه کسی، چه زمانی صحبت کرده است؟



کاربردهای پردازش گفتار ...

○ شناسایی گوینده

- تشخیص گوینده (Speaker Identification)
- کاربردهای امنیتی و جنایی



• تایید گوینده (Speaker Verification)

- رمز صوتی (کنترل دسترسی)
- در تلفن بانکها
- دسترسی به رایانه یا دستگاه خاص

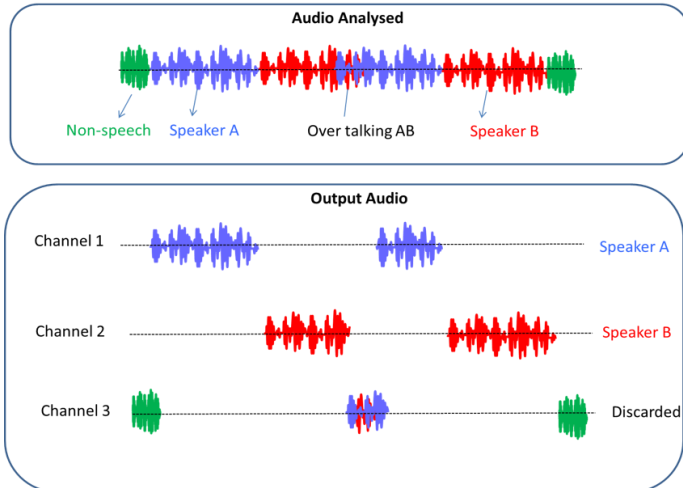
• جداسازی گوینده (Speaker Diarization)

○ تایپ گفتاری صورت جلسه‌ها

○ مثال: صحن علنی مجلس

○ زمینه مشابه

○ جداسازی گفتار از غیرگفتار (VAD)





کاربردهای پردازش گفتار ...

○ بهسازی گفتار

- بهبود کیفیت و قابلیت فهم در کاربردهای عمومی
 - کاربرد در سمعک‌ها و وسایل کمک شنوایی
 - حذف نویز در کانال‌های مخابراتی (بی‌سیم‌ها و تلفن‌های همراه)
 - سیستم‌های صوتی موجود در اماکن نویزی مانند فرودگاه‌ها و کابین خلبان
 - حذف صداهای مزاحم در فایل‌های سخنرانی و کاست‌های قدیمی
 - و ...

• استفاده در سیستم‌های پردازش گفتار دیگر جهت مقاوم‌سازی

- مقاوم‌سازی سیستم‌های کدینگ و انتقال صوت با کاربرد در تلفن‌های همراه، کنفرانس‌های ویدئویی و صوتی، انتقال صوت روی اینترنت مثل تلفن اینترنتی و ...
- مقاوم‌سازی سیستم‌های بازشناسی خودکار گفتار (ASR) و تشخیص گوینده به نویز



کاربردهای پردازش گفتار ...

○ کدینگ و فشرده‌سازی

- کاهش حجم (اندازه) سیگنال

- ذخیره‌سازی (mp3)

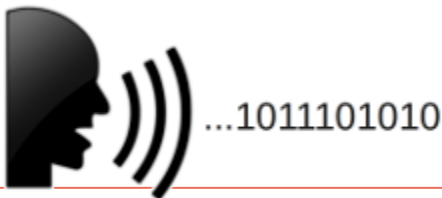
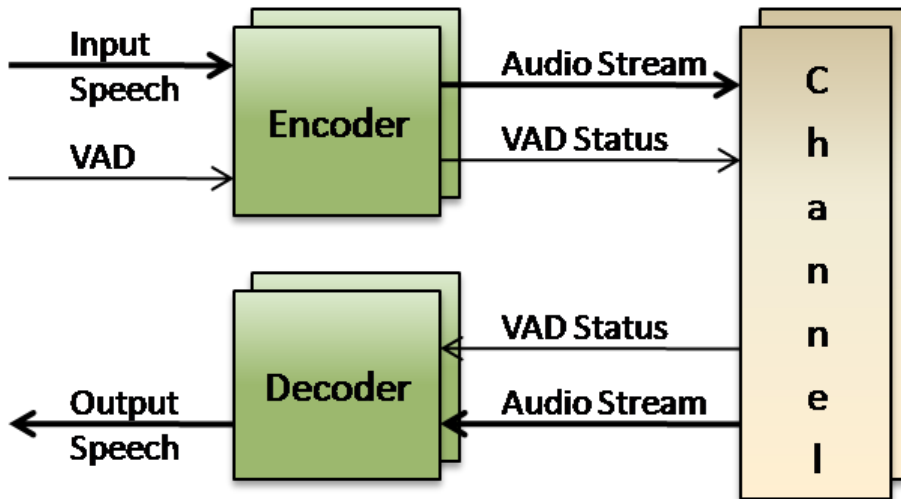
- انتقال (روی کانال‌های مخابراتی)

- تلفن همراه

- روی شبکه (اینترنت) - G.729

- بی‌سیم‌های نظامی

- افزایش امنیت انتقال



کاربردهای پردازش گفتار

○ کامپیوترهای پوشیدنی (Wearable Computers)

- کاربردهای نظامی
- عینک گوگل

