

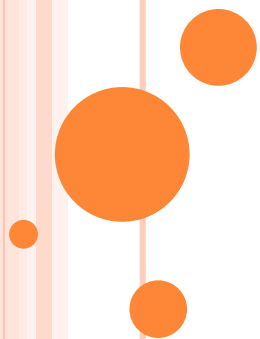
آشنایی با زبان‌شناسی رایانشی

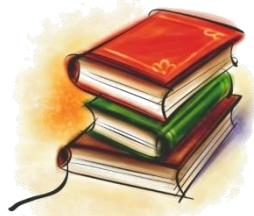
آواشناسی و پردازش سیگنال

هادی ویسی

h.veisi@ut.ac.ir

دانشگاه تهران - دانشکده علوم و فنون نوین





فهرست

○ صدا

○ تولید گفتار

- اندام‌ها
- واکداری
- طیف‌نگار و فرمنت

○ درک گفتار

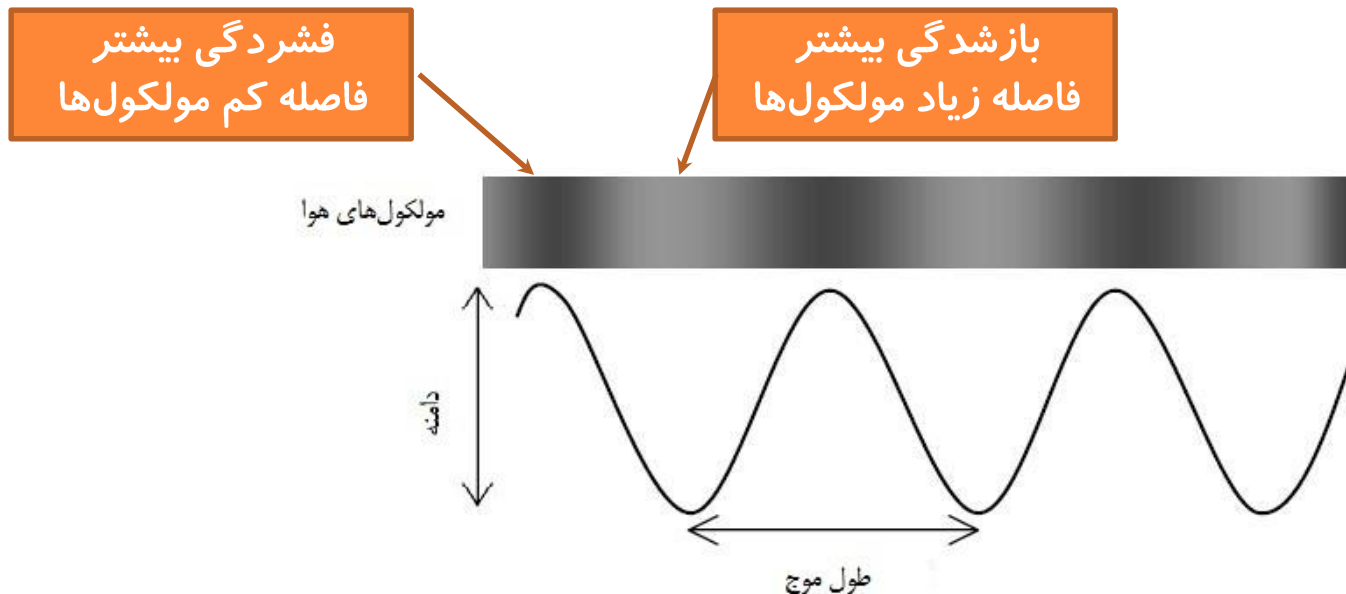
- ساختار گوش

○ آواشناسی و واج‌شناسی

- واج‌شناسی: واکه‌ها و همخوان‌ها
 - انگلیسی و فارسی
- رده‌شناسی آوایی
- واج‌گونه و هم‌تولیدی

○ تعریف صدا (Sound)

- یک موج که در اثر فشردگی‌ها و بازشدگی‌های مولکول‌های هوا ایجاد می‌شود
- تولید بر اثر ایجاد فشار و اعمال انرژی در هوا، در مسیری به موازات مسیر اعمال انرژی



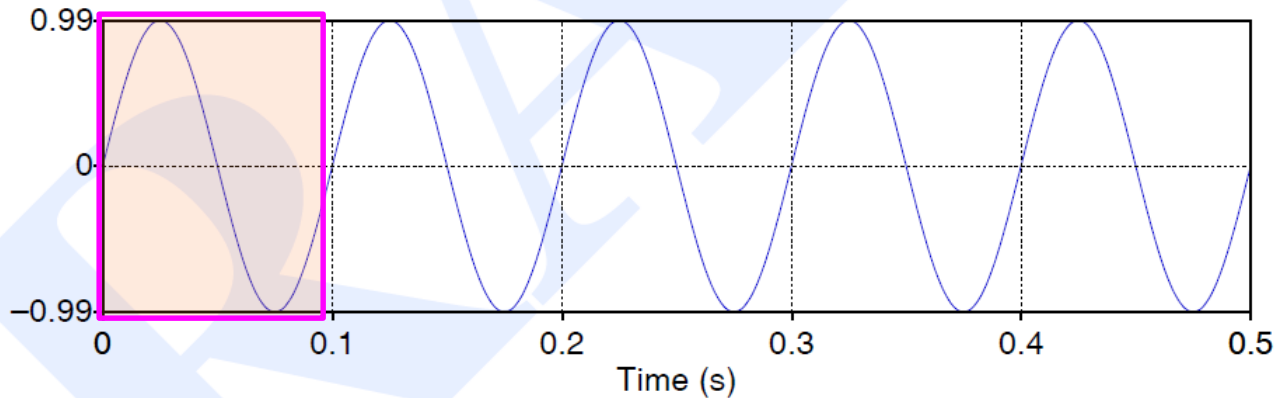
- دامنه = کار انجام شده برای تولید انرژی‌ای که مولکول‌های هوا را به حرکت درآورد



○ مفاهیم ...

• مثال: $y = A \sin(2\pi f t)$ (شکل موج با فرض $A=1$ و $f=10$)

○ برای مقادیر مختلف t



- دوره ای (متناوب، پریودیک): یک سیگنال به صورت مشابه تکرار می شود
- دوره (Period): مدت زمانی که نیاز است یک تکرار از سیگنال کامل شود $T =$
 - در اینجا $T=0.1$ Sec
- دامنه (amplitude): بیشترین مقدار محور y

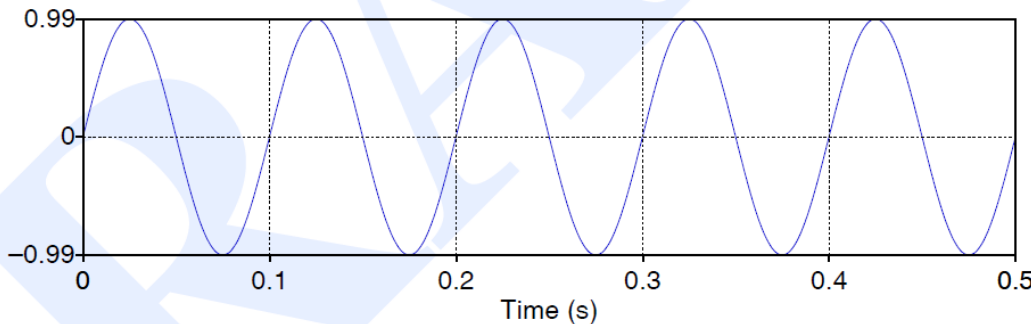


صدا ...

○ مفاهیم ...

• مثال: $y = A \sin(2\pi f t)$ (شکل موج با فرض $f=10$ و $A=1$)

○ برای مقادیر مختلف t



○ فرکانس (frequency)

• تعداد بارهایی که در یک واحد زمانی (یک ثانیه) یک سیگنال تکرار می‌شود $f =$

○ هر بار = یک دوره (تناوب یا Cycle)

• شکل: در ۰.۵ ثانیه ۵ بار تکرار شده، در یک ثانیه ۱۰ بار تکرار می‌شود ← فرکانس = ۱۰

• واحد اندازه‌گیری فرکانس = هرتز (Hz)

• داریم:

$$T = \frac{1}{f}$$

دوره تناوب

فرکانس

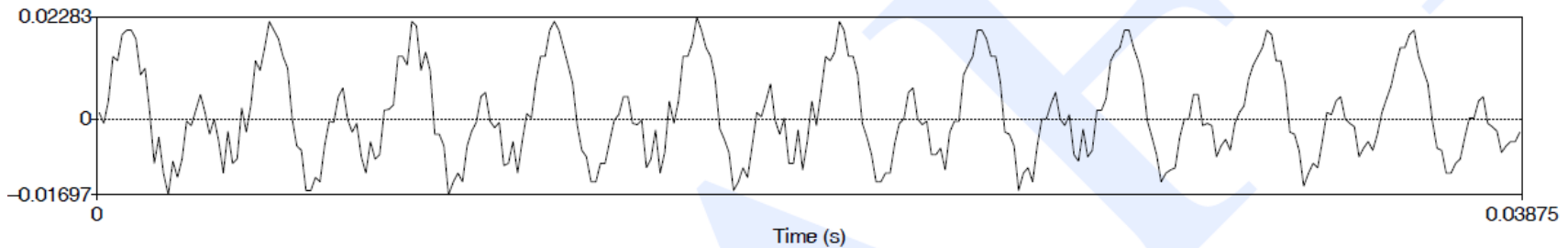
○ در شکل: $T=1/10=0.1$



صدا ...

○ مفاهیم ...

• در گفتار: سیگنال صدای /iy/ در کلمه She



• فرکانس =؟

○ تعداد ۱۰ تکرار در 0.03875 ثانیه، پس تعداد تکرارها در یک ثانیه =؟

○ $10/0.03875$



تبدیل سیگنال آنالوگ به دیجیتال

• نمونه برداری (Sampling)

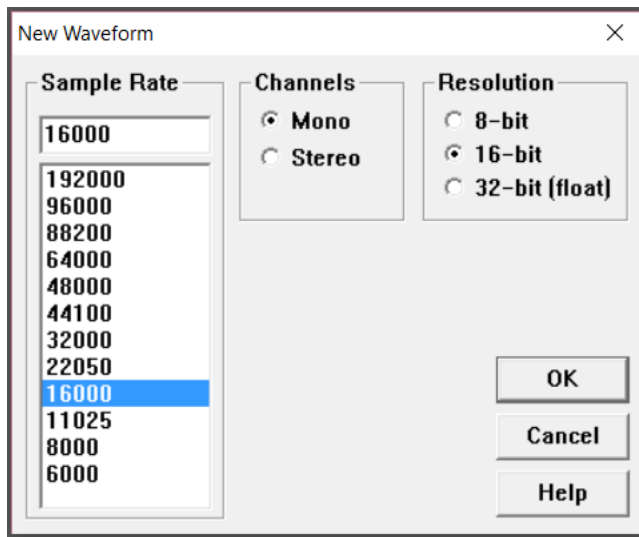
- تعداد نمونه های سیگنال در واحد زمان = نرخ نمونه برداری (Sampling Rate)
- نرخ نایکوئیست (Nyquist): نرخ نمونه برداری دیجیتال که دو برابر تعداد فرکانس واقعی سیگنال است
- نرخ نمونه برداری برای گفتار معمولا ۱۶۰۰۰ هرتز (16KHz) و ۲۲۰۵۰ هرتز (22.05 KHz) است
- نرخ نمونه برداری گفتار تلفنی 8KHz است: کانال تلفن سیگنال ها را در 4KHz فیلتر می کند

• چندی سازی (Quantization)

- تعداد بیت هایی که برای نگهداری هر نمونه لازم است
- ۸ بیت (نمونه ها از -128 تا +127 است)
- ۱۶ بیت (نمونه ها از -32768 تا +32767 است)

• کانال (Channel)

- تک کاناله (Mono)
- دو کاناله (Stereo)





صدا ...

تعداد نمونه های
سیگنال x

$$Power = \frac{1}{N} \sum_{i=1}^N x(i)^2$$

نمونه i ام سیگنال x

○ توان (Power)

- متوسط انرژی نمونه ها

○ شدت (Intensity)

- توان نرمال شده سیگنال گفتار (بر حسب دسی بل)

$$Intensity = 10 \log_{10} \frac{1}{NP_0} \sum_{i=1}^N x(i)^2$$

فشار آستانه شنوایی = $2 * 10^{-5}$



صدا ...

سطح توان

$$10 \log_{10} (P1/P2)$$

○ دسی بل (dB) = اندازه گیری دامنه صدا

- مقیاس لگاریتمی برای مقایسه دو صدا
- به دلیل گسترده بوده بازه مربوط به دامنه صداهای مختلف

○ سطح فشار صدا (SPL: Sound Pressure Level)

$$SPL(dB) = 20 \log_{10} \left(\frac{P}{P_0} \right)$$

- محاسبه میزان فشار مطلق صدای P بر حسب dB
- P_0 = فشار مطلق صدا در 0 dB
- مرجع 0 dB معادل آستانه شنوایی است
- برای یک گفتگو در فاصله حدود یک متری: SPL=60 dB
- برای یک چکش دستی مورد استفاده برای سوراخ کردن سنگ: SPL=120 dB



صدا...

○ شدت و سطح دسی بل صداهای مختلف

سطح dB	صدا	--- برابر بزرگتر از آستانه شنوایی
0	آستانه شنوایی (10^{-12}W/m^2)	10^0
10	نجوای آرام	10^1
20	اتاق نشیمن ساکت	10^2
40	گفت و گوی آرام	10^4
50	دفتر کار معمولی	10^5
60	گفت و گوی عادی	10^6
70	خیابان شلوغ شهر	10^7
80	گیتار آکوستیک در فاصله ۱ فوتی	10^8
90	عبور کامیون سنگین	10^9
100	مترو در محل ایستگاه	10^{10}
110	ابزارهای قدرتی	10^{11}
120	آستانه‌ی درد گوش	10^{12}
130	باند فرودگاه	10^{13}
140	بمب صوتی	10^{14}
150	آسیب دائمی به گوش	10^{15}
160	موتور جت (در نزدیک)	10^{16}
180	موتور راکت	10^{18}
220	دوازده فوت فاصله از دهانه توپ آتش بار (10^{10}W/m^2)	10^{22}

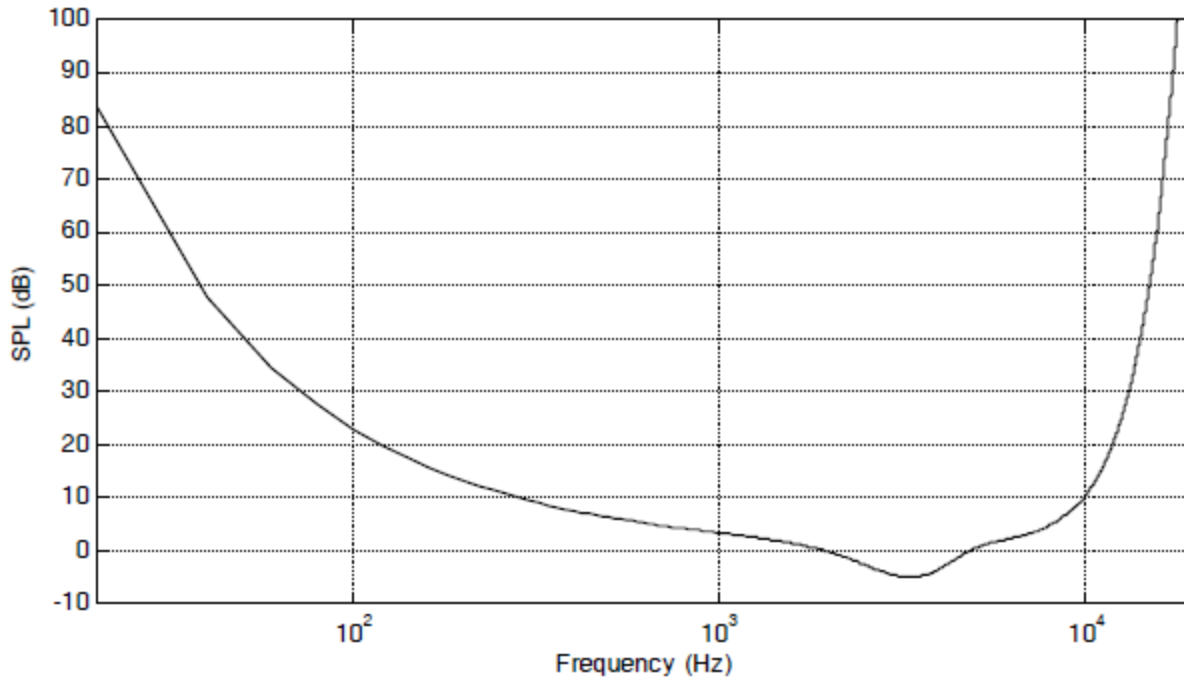
بلندترین صدای قابل تحمل توسط گوش انسان



○ آستانه مطلق شنوایی

$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \quad (dB \text{ SPL})$$

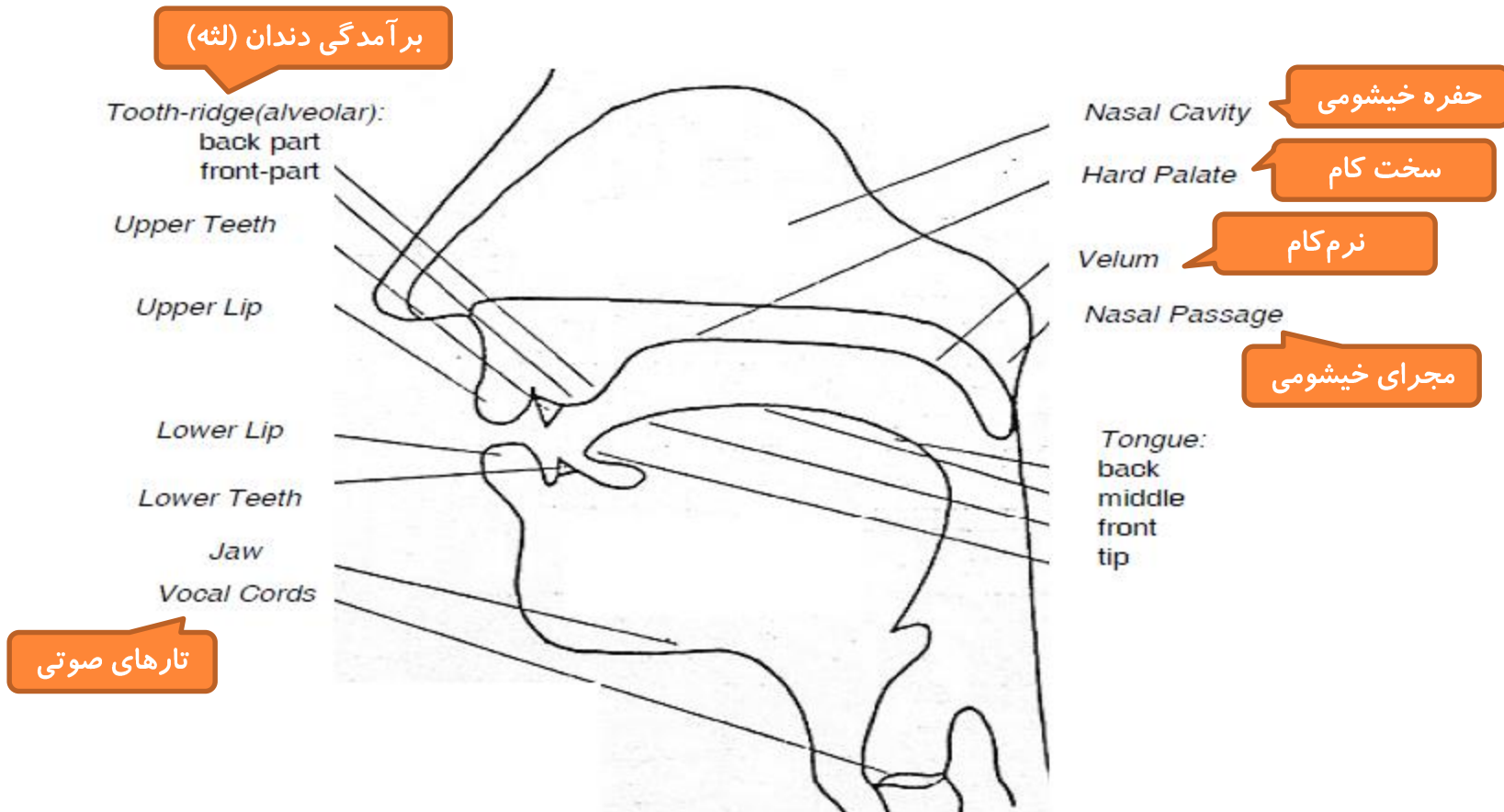
- تابعی از بسامد
- به طور تقریبی



- از ۲۰ هرتز تا ۲۰ کیلو هرتز

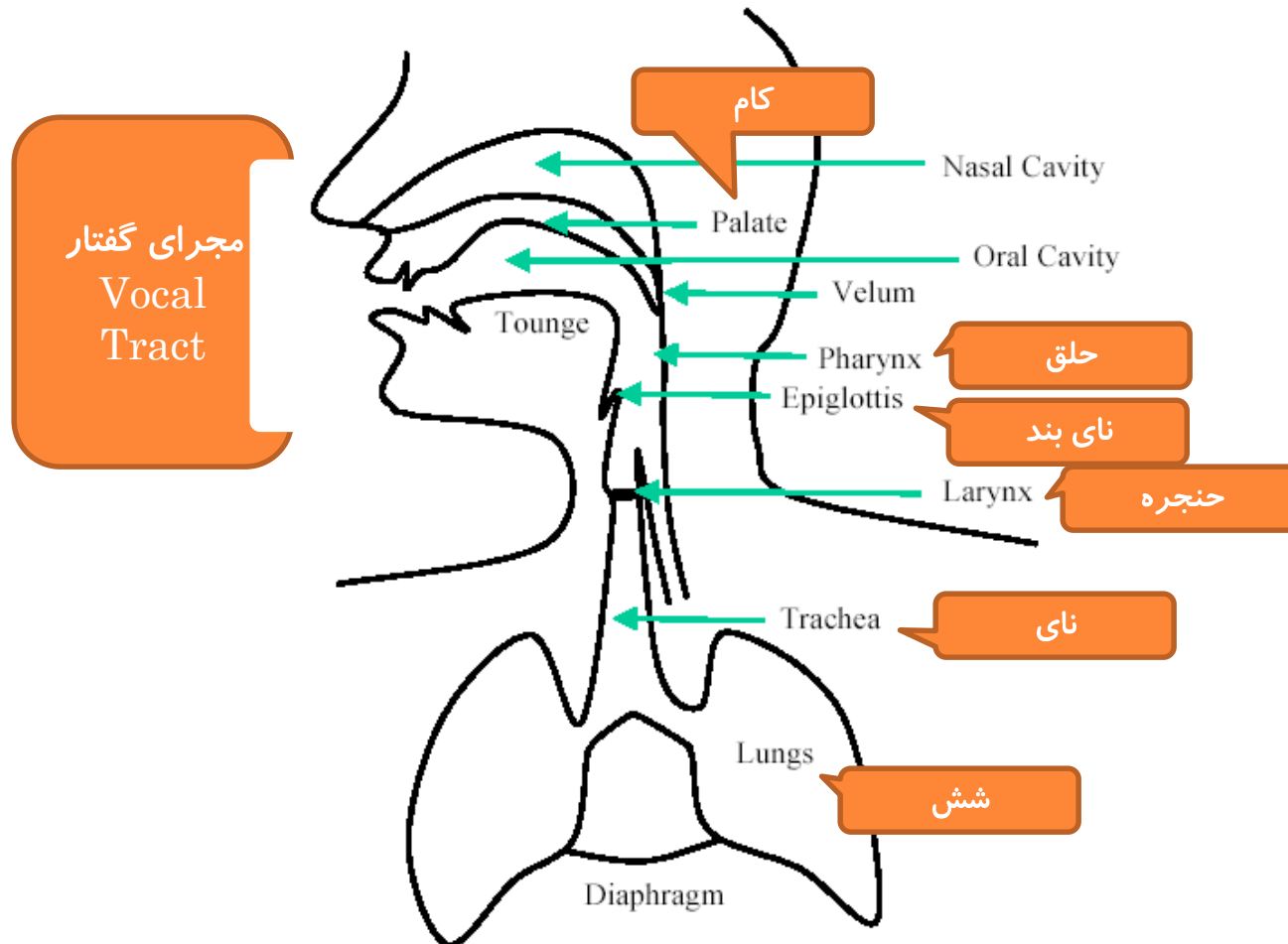
تولید گفتار: اندام‌ها ...

- تولید به وسیله امواج ناشی از فشار هوای خارج شده از دهان و بینی
- مجرای گفتار (Vocal Tract)



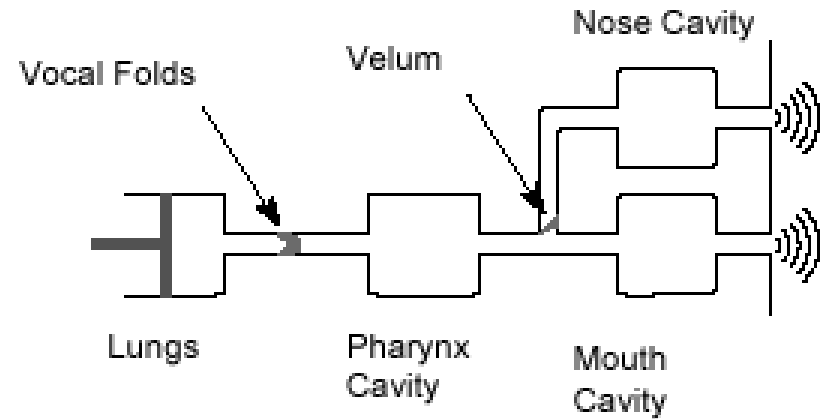
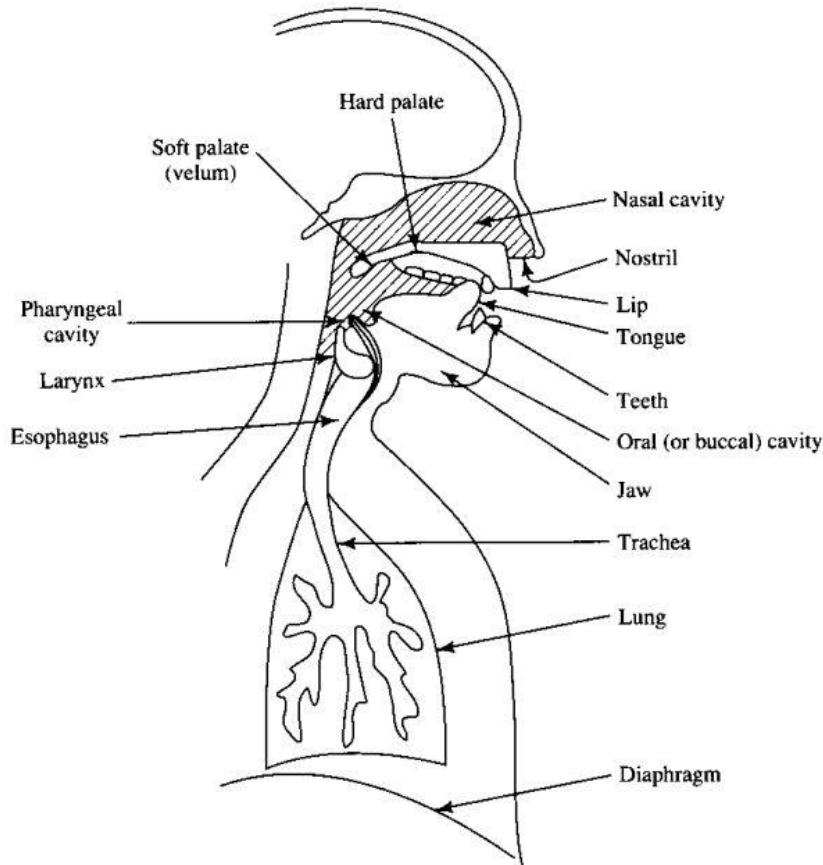


تولید گفتار: اندام‌ها ...



تولید گفتار: اندام‌ها ...

○ مدل مجرای گفتار





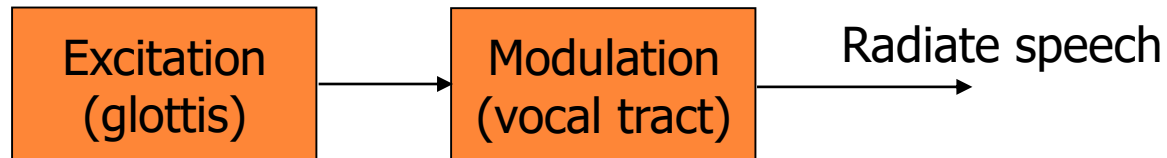
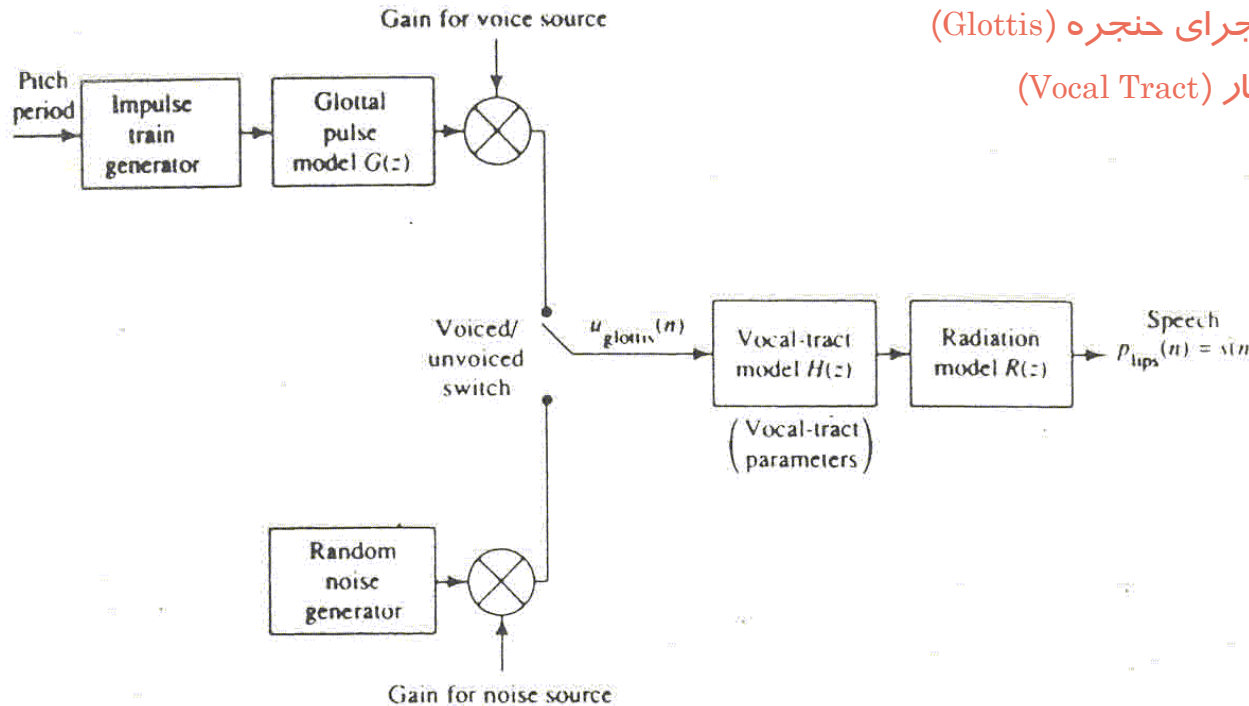
تولید گفتار: اندامها ...

○ مدل عمومی تولید گفتار در پردازش سیگنال

• مدل source-filter

○ منبع (Source) = مجرای حنجره (Glottis)

○ فیلتر = مجرای گفتار (Vocal Tract)





تولید گفتار: اندام‌ها ...

○ انواع واج‌ها

- همخوان‌ها (Consonants): تولید با انقباض در گلو یا در دهان (زبان، دندان‌ها، لب‌ها)
 - مانند ب، پ، ف، س، ...
- واکه‌ها (Vowels): تولید بدون انقباض یا انسداد قابل توجه
 - مانند آ، آو، اُ، ای، اِ

○ دستگاه تولید گفتار انسان ...

- شش‌ها: منبع هوا در حین صحبت کردن
- تارهای صوتی (Vocal Cords)(حنجره)
 - صدای واکدار (Voiced): تارهای صوتی در حین تولید یک صدای گفتار در فاصله نزدیکی از همدیگر قرار می‌گیرند و به طور تناوبی حرکت می‌کنند
 - مانند ز، ب، د، ژ
 - صدای بی‌واک (Unvoiced): تارها آنقدر از هم فاصله می‌گیرند یا محکم و سفت می‌شوند که نمی‌توانند حرکت تناوبی داشته باشند
 - مانند س، پ، ت، چ



تولید گفتار: اندام‌ها ...

○ دستگاه تولید گفتار انسان

- نرم‌کام ((Velum (Soft Palate)): مثل یک دریچه، وقتی باز می‌شود هوا از درون حفره خیشوم عبور می‌کند
 - واج‌های م، ن
- سخت‌کام (Hard Palate): یک سطح بلند و نسبتاً سفت در سقف دهان، وقتی زبان در مقابل آن قرار می‌گیرد، صدای واج همخوان تولید می‌شود
- زبان (Tongue): انعطاف‌پذیر، برای تولید واژه‌ها از سخت‌کام دور می‌شود و برای تولید همخوان‌ها در نزدیکی یا تماس با سخت‌کام یا سطوح سخت دیگر قرار می‌گیرد
- دندان‌ها (Teeth): برای تولید همخوان‌های خاص، زبان را سفت می‌کند
- لب‌ها (Lips): گرد یا گسترده برای تغییر کیفیت واژه، یا کاملاً بسته می‌شود تا برای همخوان‌های خاصی (مانند p, b, m) مانع از جریان هوا از راه دهان شود



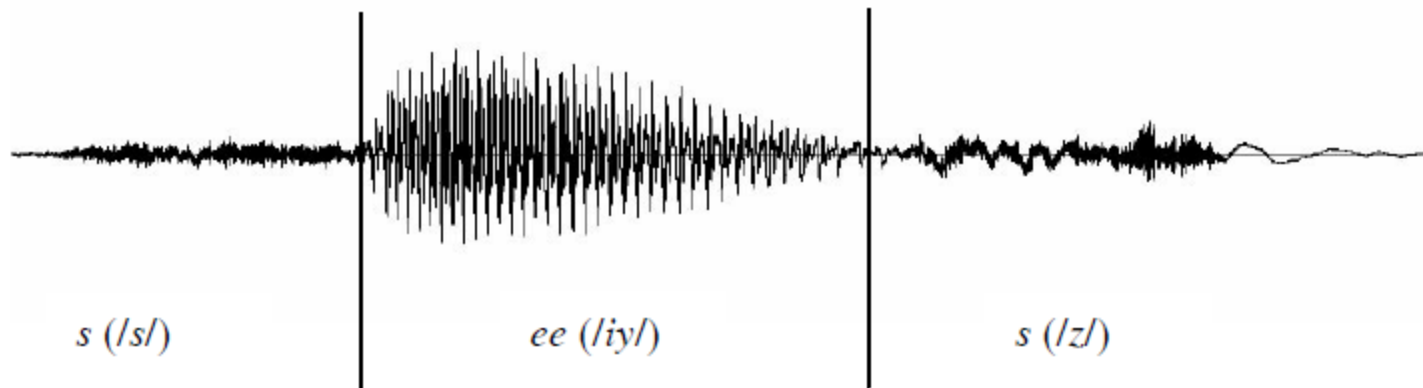
تولید گفتار: واکداری ...

○ مکانیسم واکداری (Voiced)

- تمایز اصلی بین گونه‌های صدا در گفتار، تمایز واکداری و بی‌واکی است
- صداهای واکدار، از جمله واکه‌ها، الگوی تقریباً منظمی در ساختار زمان و بسامد دارند که صداهای بی‌واک، از قبیل همخوان‌هایی مثل s، آن را ندارند.
- صداهای واکدار انرژی بیشتری دارند

• مثال: واژه sees

○ همخوان بی‌واک /s/ واکه /iy/ همخوان واکدار /z/





تولید گفتار: واگذاری ...

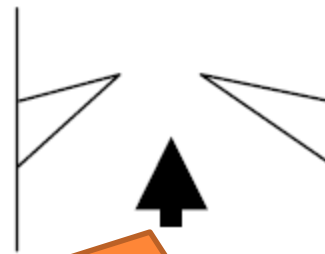
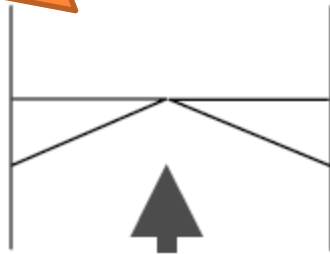
○ تمایز بین واج‌های واگ‌دار و بی‌واک

- لرزش (واگ‌دار) یا عدم لرزش (بی‌واک) تارهای صوتی در حین تولید واج
 - واکه‌ها در طول زمان تولیدشان واگ‌دار هستند و طنین‌های (timber) مجزای هر واکه با استفاده از لب‌ها و زبان ایجاد می‌شوند.
- تفاوت در سرعت لرزش تارهای صوتی برای افراد مختلف
 - برای یک مرد بالغ ۶۰ بار در ثانیه (هرتز)
 - برای یک زن جوان یا یک بچه ۳۰۰ هرتز یا بیشتر
- بسامد پایه (F_0 : Fundamental Frequency): سرعت چرخه باز و بسته شدن تارهای صوتی در حنجره در حین تولید صداهای واگ‌دار
 - مبنای تناوبی برای هارمونیک‌های با بسامد بالاتر (تولید شده توسط حفره‌های تشدید حلقی و دهانی)
 - کمک به درک زیروبمی گفتار (Pitch) (افت و خیز شبه موسیقایی آهنگ صدا)

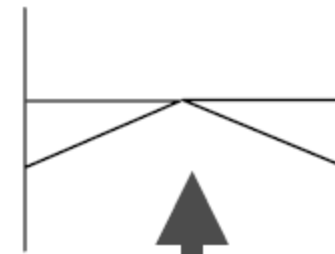
تولید گفتار: واگذاری ...

○ چرخه تارهای صوتی در حنجره ...

بسته شدن تارهای صوتی



باز شدن تارهای صوتی بر اثر فشار هوای شش‌ها



برگشتن تارهای صوتی به حالت قبل با کاهش فشار هوا



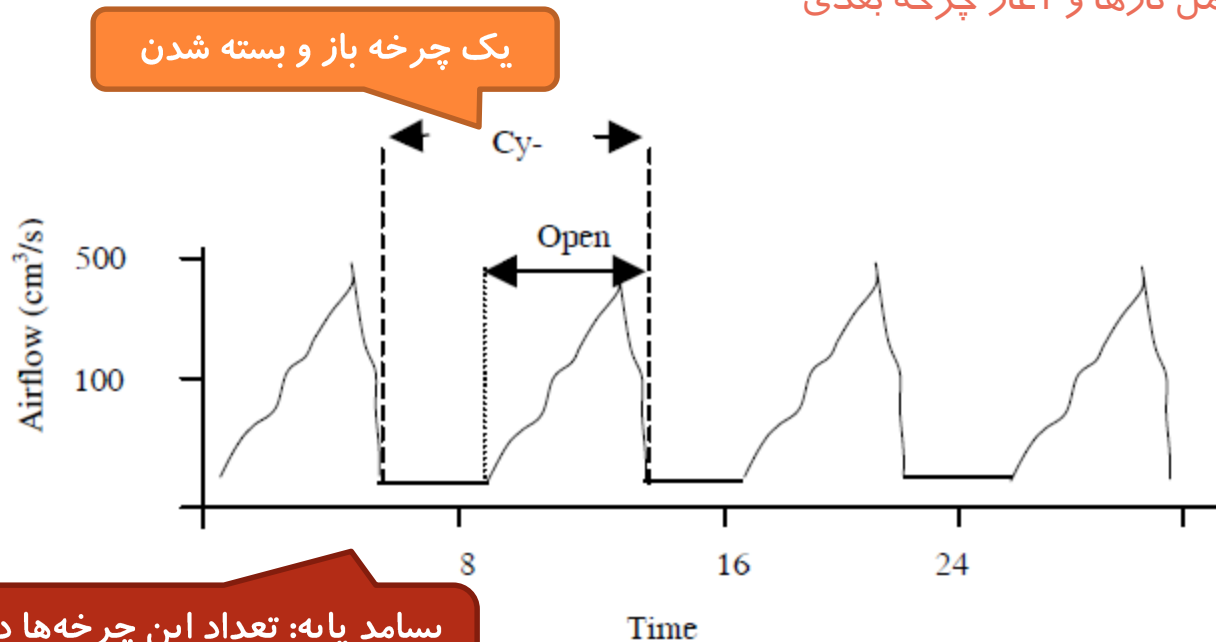
• محل به هم رسیدن تارهای صوتی = چاکنای (Glottis)



تولید گفتار: واگذاری ...

○ چرخه تارهای صوتی در حنجره

- مرحله بسته بودن: عدم جریان هوا
- مرحله باز شدن: باز شدن چاکنای و ایجاد جریان هوا
- بعد از یک افزایش کوتاه مدت، برگشت تارهای صوتی به موقعیت اصلی و کاهش جریان هوا تا بسته شدن کامل تارها و آغاز چرخه بعدی

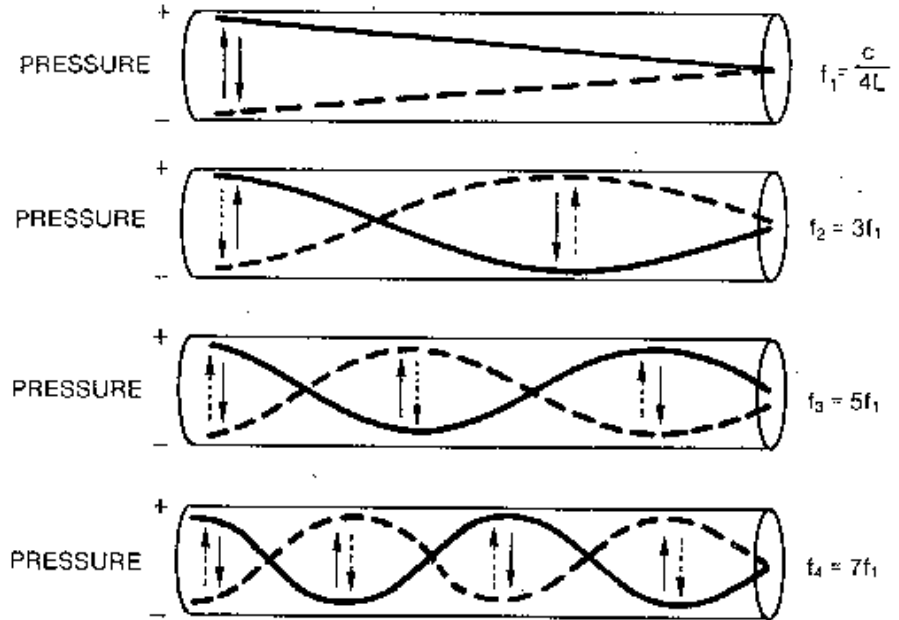
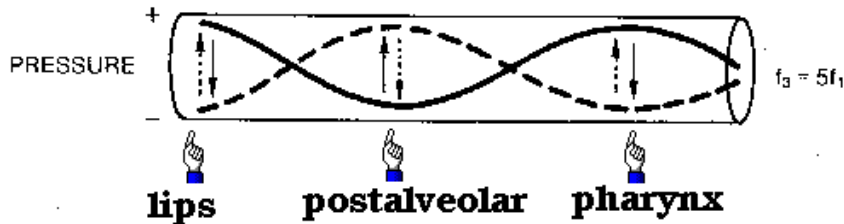
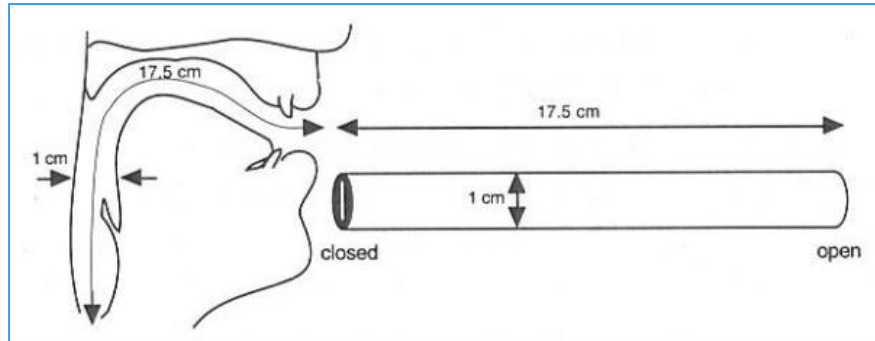


بسامد پایه: تعداد این چرخه‌ها در ثانیه
(هرتز) - در اینجا حدود 120 هرتز

تولید گفتار: طیف‌نگار و فرمنت ...

○ مجرای گفتاری

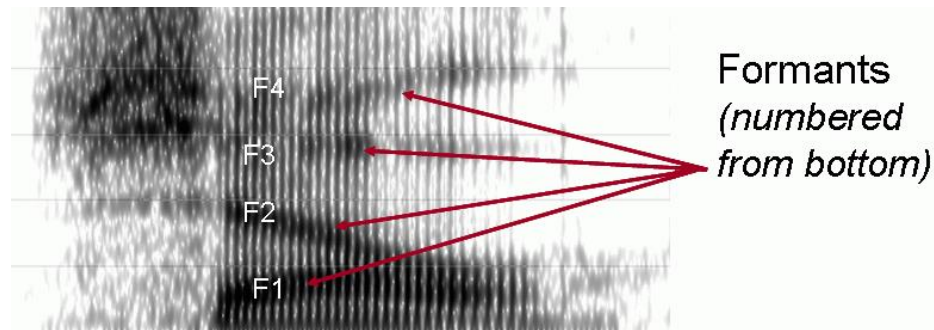
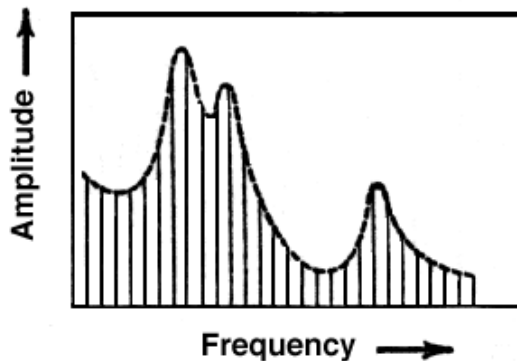
- لوله مستقیمی با سطح مقطع عرضی یکسان، انتهای آن در چاکنای بسته و سر آن در ناحیه لب‌ها باز است



تولید گفتار: طیف‌نگار و فرمنت ...

○ فرمنت

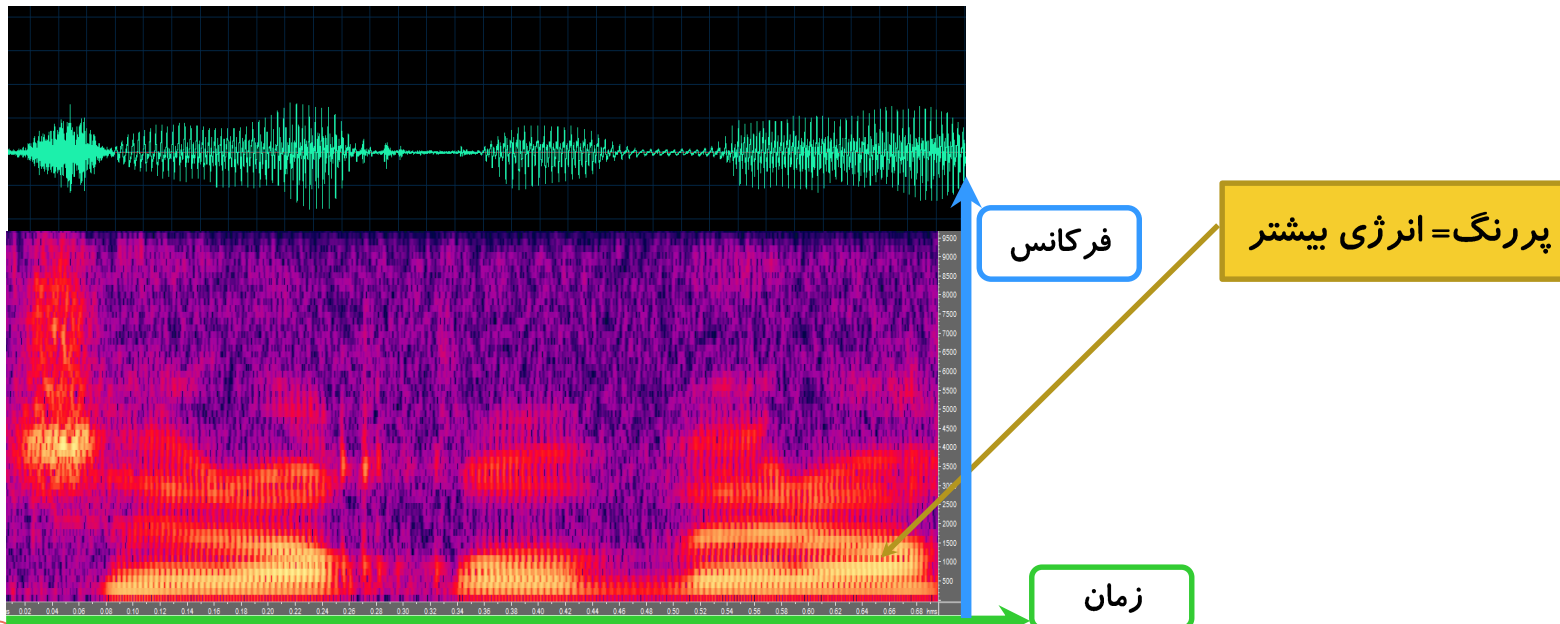
- موج چاکنای متناوب بوده و شامل بسامد پایه (F_0) و چند هارمونیک (مضرب‌های F_0) است
- تحریک مجرای گفتار در بالای چاکنای توسط انرژی چاکنایی
- تغییر تشدیدها با تغییر شکل مجرای گفتاری
- اهمیت بیشتر هارمونیک‌های نزدیک تشدیدها
- فرمنت (Formant) در گفتار، تشدیدهای حفره‌های اندام‌های گفتاری
 - بیانگر نوعی از پیکره‌بندی اندام‌های گفتاری (مثل طنین‌های مختلف یک واکه)



تولید گفتار: طیف‌نگار و فرمنت ...

○ طیف‌نگار (Spectrogram)

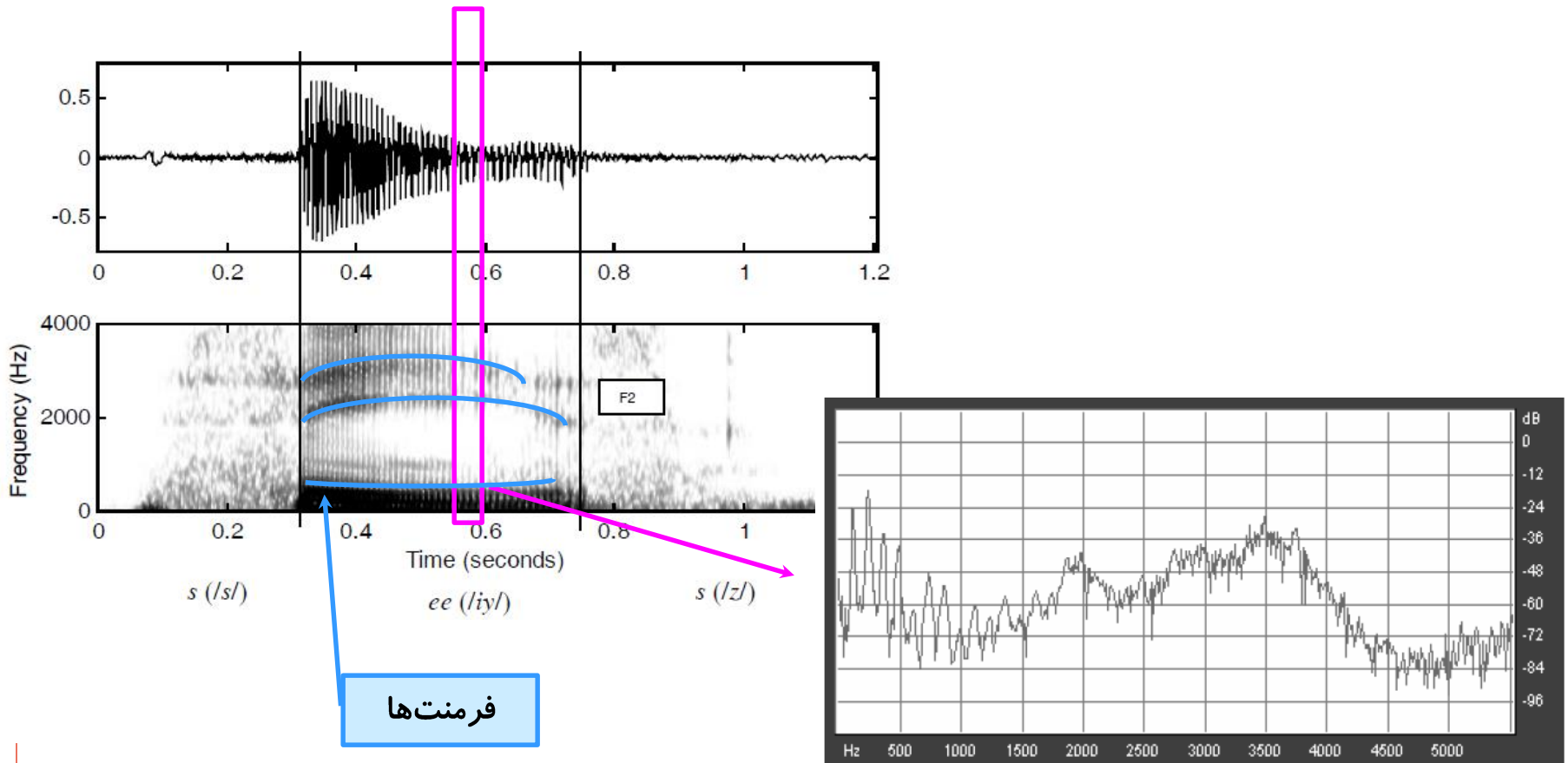
- نمایش زمان-فرکانس سیگنال گفتار
- تیرگی یا روشنی = میزان انرژی
- نوارهای افقی پرانرژی = فرمنت‌ها
- توزیع غیرهمسان انرژی در تشدیدهای واکه
- هارمونیک‌های بسامد پایه در تشدیدهای طبیعی وضعیت حفره اندام گفتاری





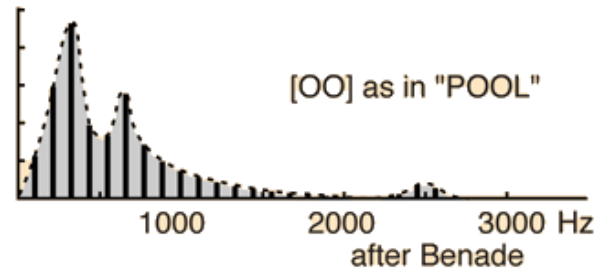
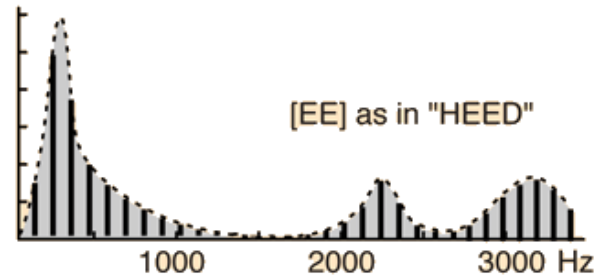
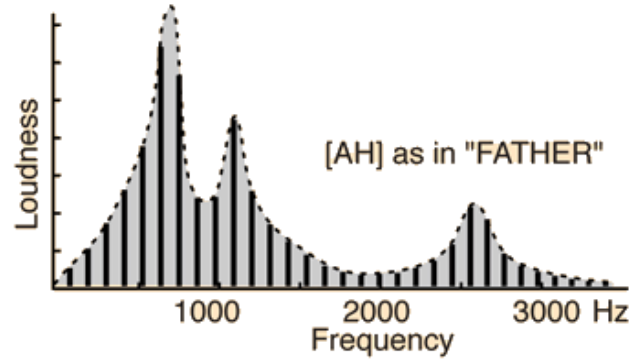
تولید گفتار: طیف‌نگار و فرمنت ...

○ مثال: واژه /iy/ در واژه sees



تولید گفتار: طیف‌نگار و فرمنت ...

○ مثال





تولید گفتار: طیف‌نگار و فرمنت ...

○ مقادیر نمونه فرکانس فرمنت‌ها

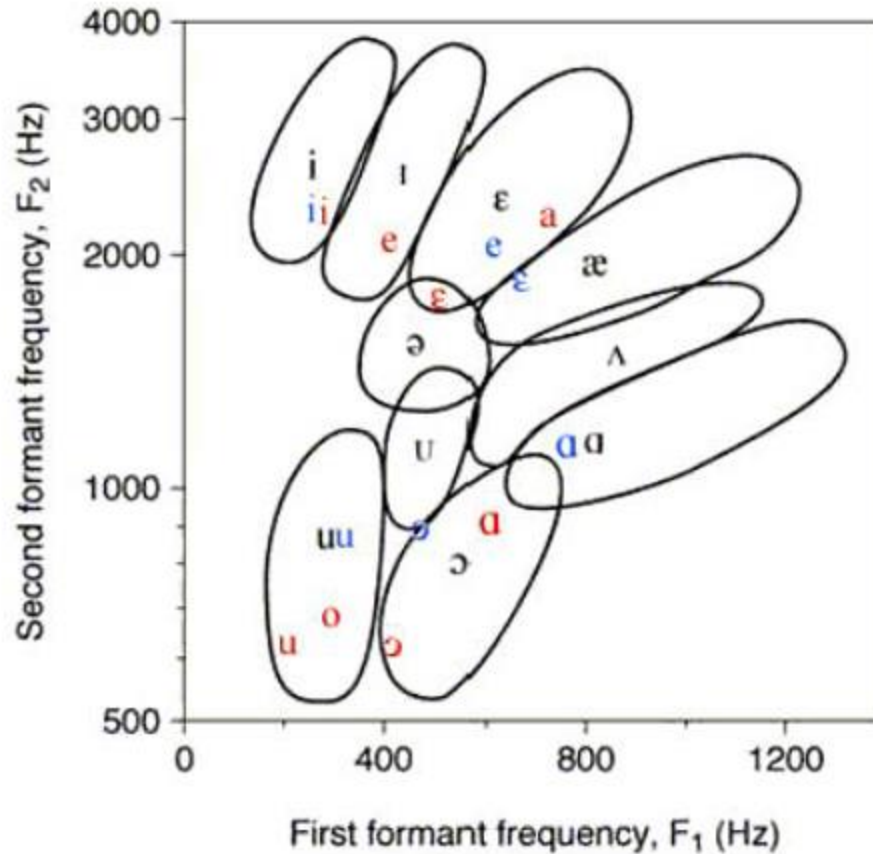
• یکی از ویژگی‌ها برای تشخیص واج‌ها

Phonetic Symbol	Example Word	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
/ow/	bought	570	840	2410
/oo/	boot	300	870	2240
/u/	foot	440	1020	2240
/a/	hot	730	1090	2440
/uh/	but	520	1190	2390
/er/	bird	490	1350	1690
/ae/	bat	660	1720	2410
/e/	bet	530	1840	2480
/i/	bit	390	1990	2550
/iy/	beet	270	2290	3010



تولید گفتار: طیف‌نگار و فرمنت

○ دسته‌بندی واج‌های انگلیسی بر اساس دو فرمنت اول





درک گفتار ...

○ دو مؤلفه اصلی شنیداری در سیستم درک

- اندام‌های شنیداری جانبی (گوش‌ها)
- سیستم عصبی شنیداری (مغز)

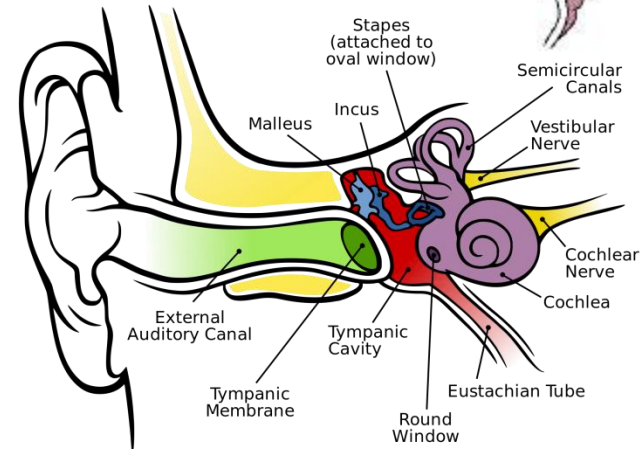
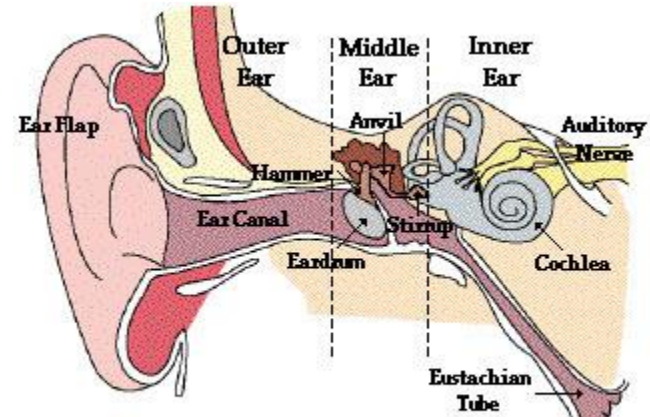
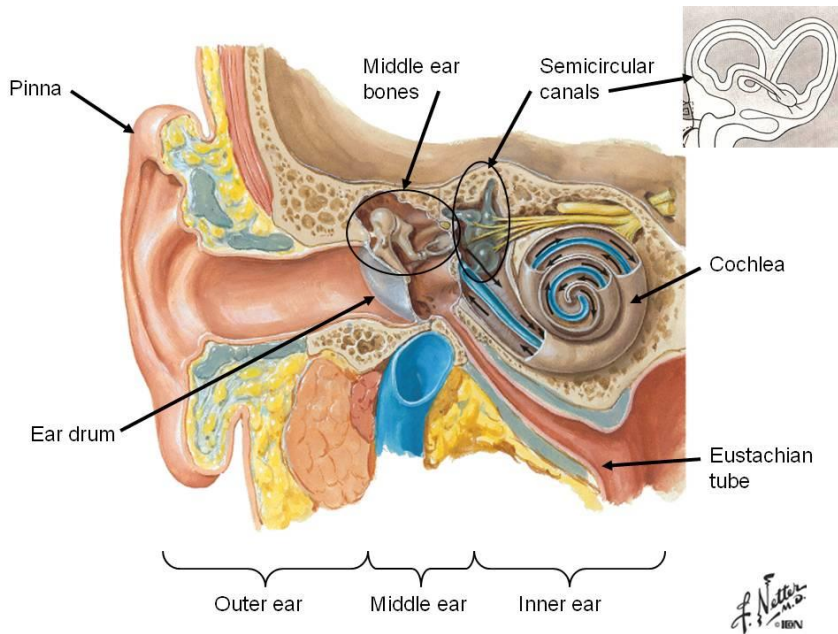
○ مراحل

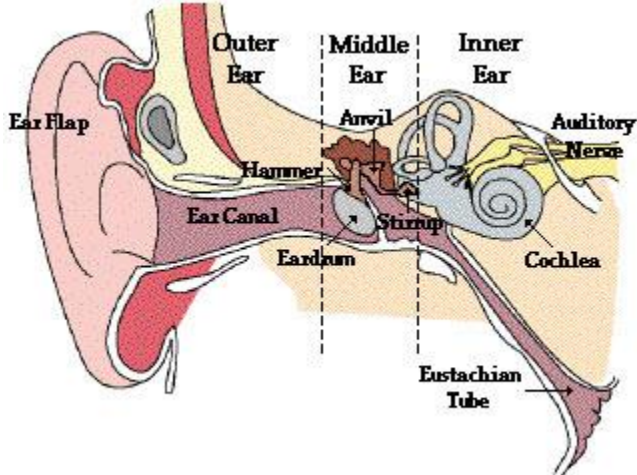
- دریافت سیگنال صوتی توسط گوش
- تبدیل به ارتعاش مکانیکی بر روی غشای پایه (Basilar Membrane)
- ارائه به صورت یک سری پالس و انتقال به وسیله عصب‌های شنوایی
- استخراج اطلاعات ادراکی در مراحل مختلف سیستم عصبی شنیداری

درک گفتار ...

○ ساختار گوش ...

• سه بخش: گوش بیرونی، گوش میانی و گوش درونی



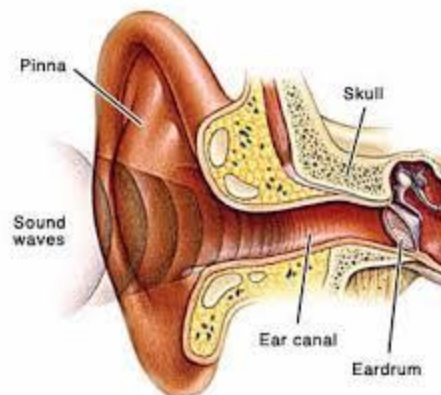


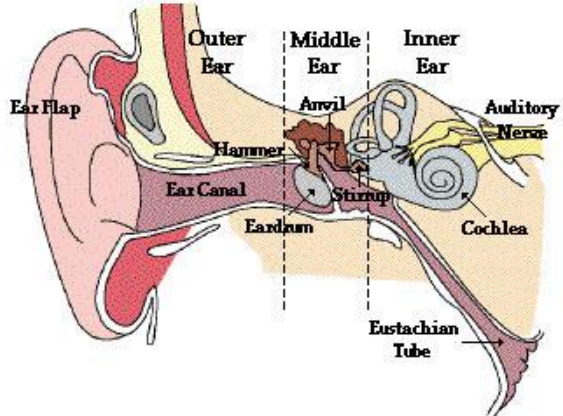
درک گفتار: ساختار گوش ...

○ گوش بیرونی

• شامل

- (۱) لاله گوش
- (۲) کانال شنیداری خارجی (یک لوله انتقال صدا با طول حدود ۲.۵ سانتی‌متر)
- (۳) پرده گوش (Eardrum)
- با رسیدن نوسانات فشار هوا از بیرون به پرده گوش، پرده مرتعش شده و ارتعاشات را به استخوان‌هایی که در سمت مقابلش قرار دارند می‌فرستد
- بسامد (فشرده‌گی و بازشدگی متناوب) ارتعاش پرده گوش با موج فشار صدای ورودی یکسان است





درک گفتار: ساختار گوش ...

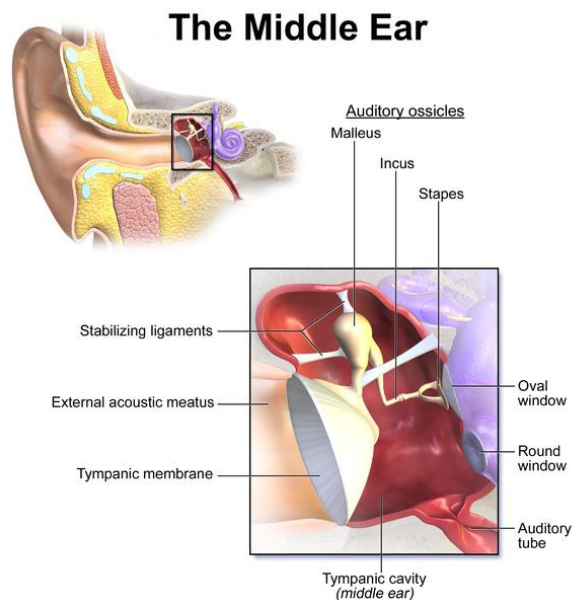
○ گوش میانی

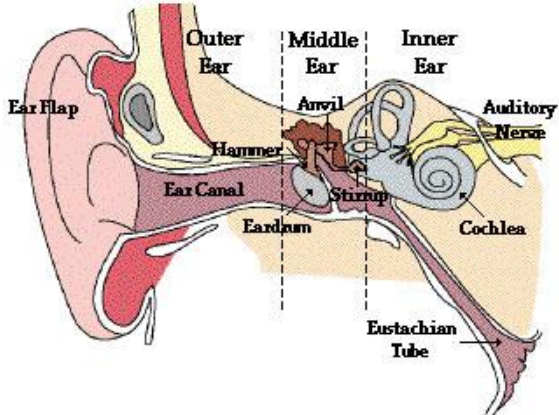
- حفره‌ای پر از هوا با عرض ۱.۳ سانتی‌متر و حجم حدود ۶ سانتی‌متر مکعب
- هوا از لوله‌ای که این حفره را به بینی و حلق متصل می‌کند، به گوش میانی می‌رسد
- لوله/شیپور استاش (Eustachian tube)

- شامل سه استخوان کوچک (ossicles) با نام‌های Incus، Malleus (hammer) و Stapes (stirrup) و (anvil)

○ انتقال لرزش پرده

○ محدود کردن دامنه





درک گفتار: ساختار گوش ...

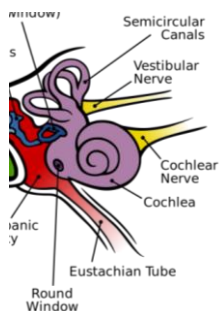
○ گوش درونی

- پنجره بیضی شکل (Oval window): غشایی بر روی واسطه استخوانی گوش درونی (حلزونی گوش)

○ از آنجا که دیواره‌های حلزونی گوش استخوانی هستند، انرژی به وسیله فعالیت مکانیکی اندام‌های درون گوش و ایجاد فشار بر غشای موجود در اطراف دریچه بیضی شکل، منتقل می‌شود.

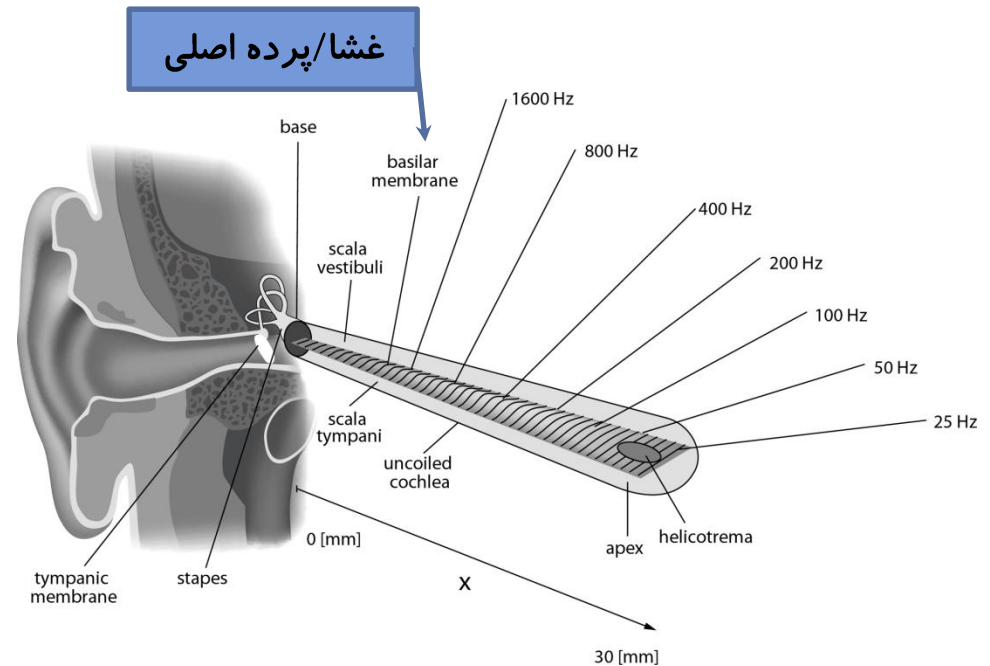
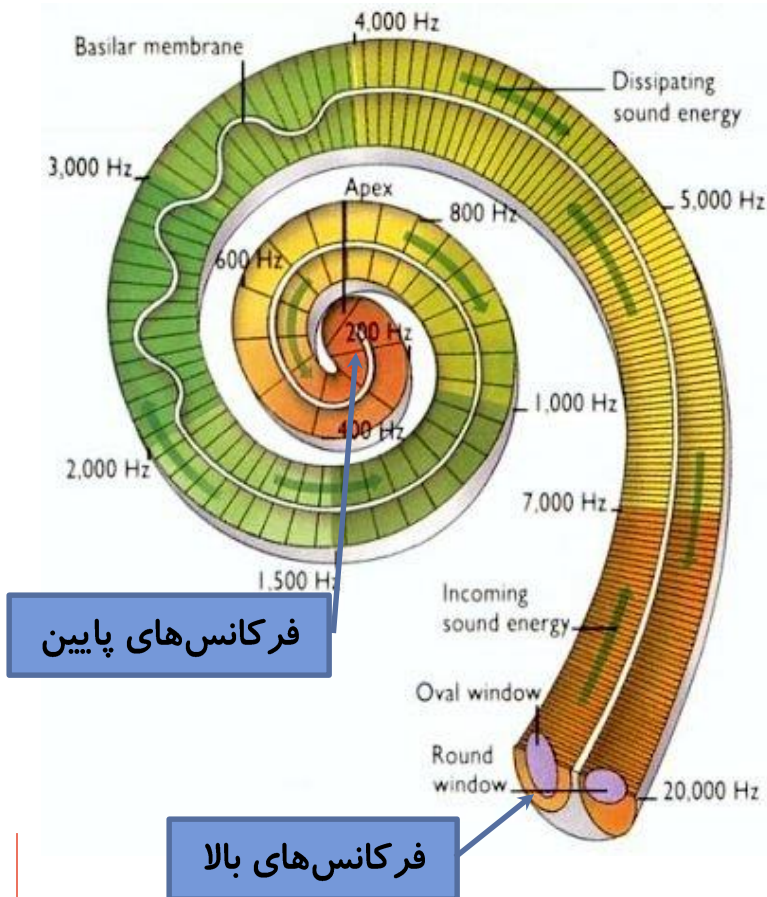
• حلزونی گوش (Cochlea)

- قسمت مربوط به درک گفتار در گوش درونی که به طور مستقیم با عصب شنوایی در ارتباط است و صدا را به مغز می‌رساند
- یک لوله مارپیچ با طول ۳.۵ سانتی‌متر و حدود ۲.۶ بار به دور خود پیچیده است.
- حلزونی گوش یک بانک فیلتر است که خروجی‌های آن به ترتیب موقعیت منظم می‌شوند.
- تبدیل بسامد به مکان: فیلترهایی که کمترین فاصله را از پایه حلزونی دارند به بسامدهای بالاتر و فیلترهایی که به بالای حلزون نزدیک‌ترند به بسامدهای پایین‌تر پاسخ می‌دهند.
- تبدیل لرزش‌های صوتی به پالس‌های عصبی



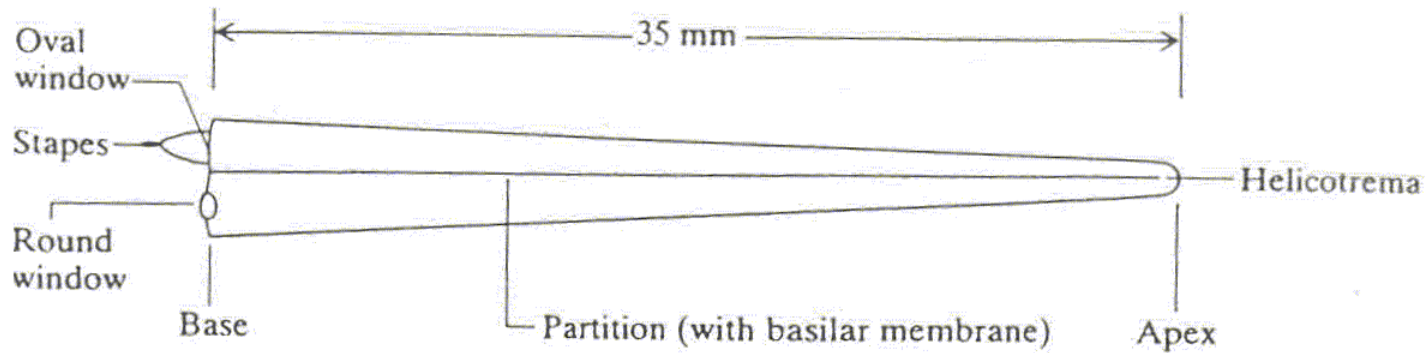
درک گفتار: ساختار گوش ...

○ گوش درونی: حلزونی گوش (Cochlea) ...



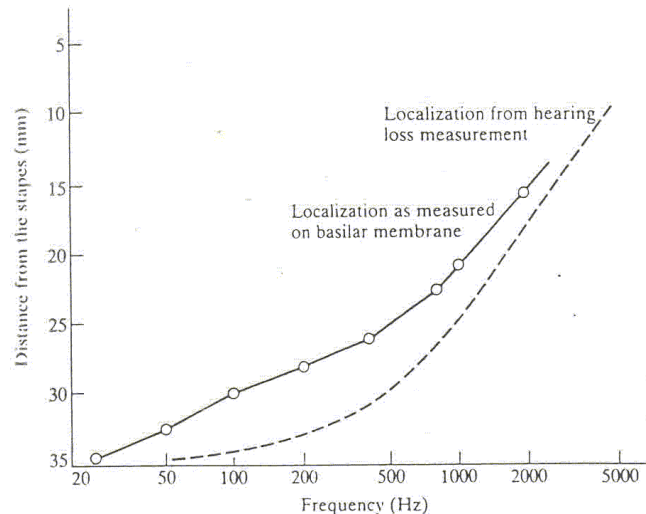
درک گفتار: ساختار گوش ...

○ گوش درونی: حلزونی گوش (Cochlea) به صورت باز شده



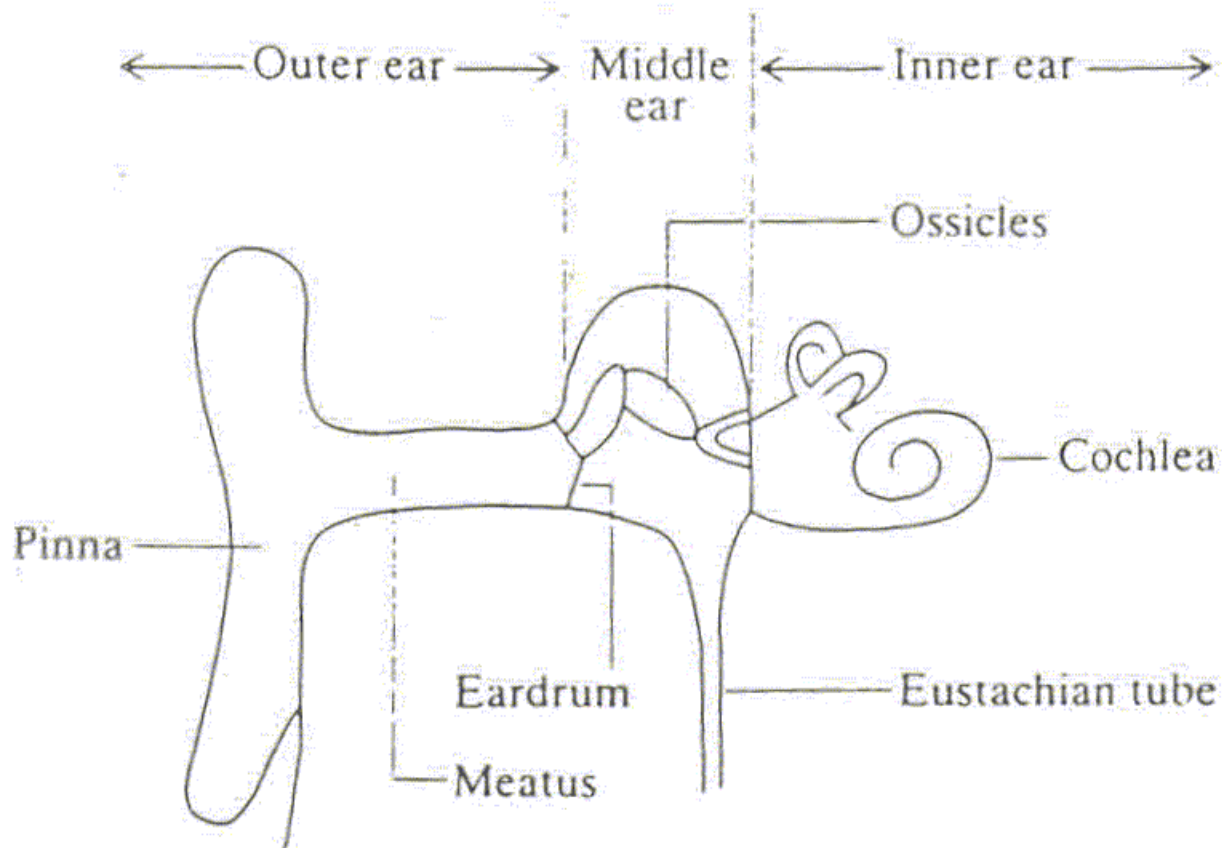
○ گوش درونی: حلزونی گوش (Cochlea)

• دامنه در طول پرده اصلی



درک گفتار ...

○ ساختار گوش





درک گفتار ...

○ مشخصه‌های ادراکی و فیزیکی صدا ...

کیفیت ادراکی	کمیت فیزیکی
بلندی (Loudness)	شدت (Intensity)
زیروبمی (Pitch)	بسامد پایه (Fundamental frequency)
طنین (Timbre)	شکل طیفی (Spectral shape)
زمان‌بندی (Timing)	زمان آغاز/پایان (Onset/offset time)
موقعیت (Location)	اختلاف فاز در شنیدن با هر دو گوش

- مشخصه‌های ادراکی دارای هم‌بستگی با یک مشخصه فیزیکی است، اما این ارتباط پیچیده است؛ زیرا سایر مشخصه‌های فیزیکی آن صدا ممکن است به نحوه پیچیده‌ای بر ادراک آن تأثیر بگذارند.

○ صداهایی که سطح شدت بیشتری دارند معمولاً بلندتر هستند، اما حساسیت گوش با تغییر بسامد و کیفیت صدا تغییر می‌کند.

○ فرق اساسی بین کیفیت‌های فیزیکی و ادراکی، پدیده درک کردن غیرهمسان صدای نواخت‌هایی با بلندی یکسان اما با بسامدهای متغیر

○ نواخت‌هایی که زیروبمی متغیری دارند، ذاتاً بلندی دریافت‌شده متفاوتی نیز دارند



درک گفتار ...

○ مشخصه‌های ادراکی و فیزیکی صدا ...

• زیروبمی صدا \sim بسامد پایه

- هرچه بسامد پایه بیشتر باشد، زیروبمی دریافتی بیشتر است
- اما، با افزایش شدت و ثابت ماندن بسامد، زیروبمی دریافتی تغییر خواهد کرد.

• پوشاندن برخی نواخت‌ها توسط نواخت‌های دیگر

- وقتی گوش در معرض دو یا چند نواخت مختلف قرار گیرد، ممکن است یکی از نواخت‌ها سایر نواخت‌ها را پوشاند
- نواخت‌های خالص، صداهای مرکب، نوفه‌های دارای پهنای باند کم و دارای پهنای باند زیاد، دارای توانایی در پوشاندن صداهای دیگر هستند

• توانایی تشخیص جهت صدا توسط انسان

- توانایی ما را در تشخیص جهت منبع صدا وابسته به شنیدن با دو گوش است.
- موقعیت‌یابی صدا توسط انسان، زمان و شدت صدا به ترتیب دارای تأثیرات متفاوتی بر صداهای با بسامد پایین و بسامد بالا هستند
- صداهایی که بسامد پایین دارند، عمدتاً بر اساس تفاوت زمانی بین دو گوش تمایز داده می‌شوند در حالی که صداهایی با بسامد بالا بیشتر بر اساس تفاوت‌های شدت دو گوش موقعیت‌یابی می‌شوند

درک گفتار ...

○ مشخصه‌های ادراکی و فیزیکی صدا

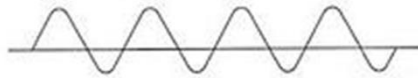
• طنین (Timbre)

- مشخصه‌ای از حس شنوایی که با آن می‌توان بین دو صدای شبیه به هم که بلندی و زیربومی یکسان دارند، تمایز قائل شد.
- وقتی تمام تفاوت‌هایی از دو صدا که به راحتی اندازه‌گیری می‌شوند، کنترل شوند تا یکسان باشند، مابقی تفاوت‌های درک شده به طنین اختصاص داده می‌شود.
- این مسئله در موسیقی راحت‌تر شنیده می‌شود، یک نت واحد با اکتاو و مدت زمان استمرار یکسان، وقتی با ویولن نواخته می‌شود با نواخته شدن آن با فلوت فرق می‌کند

○ ویژگی مهمی در تحلیل موسیقی: تشخیص نوع موسیقی و نوع ساز



Tuning fork



Flute



Voice



Violin

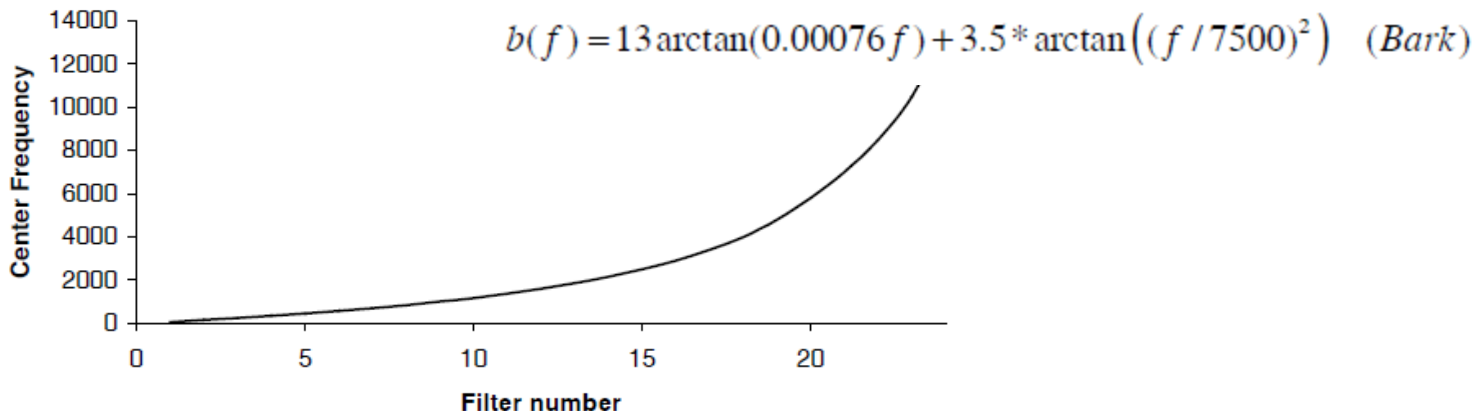




درک گفتار ...

○ تحلیل بسامدی ...

- مطالعه مکانیزم گوش درونی و عصب شنوایی نشان می‌دهند که مشخصه‌های ادراکی صداهایی که بسامدهای مختلفی دارند، ساده یا خطی نیستند
- عملکرد حلزونی گوش: بانک فیلتری با پهنای باندهای مختلف و با هم‌پوشانی
- مقیاس بسامد بارک (Bark Frequency Scale): مقیاسی برای تعیین بانک فیلترها
 - تناسب پردازش انرژی طیفی صدا در مقیاس بارک با اطلاعات طیفی در حال پردازش توسط گوش
 - مقیاس بارک دارای بازه‌ای از ۱ تا ۲۴ بارک است معادل با ۲۴ نوار بحرانی شنوایی انسان
 - تفکیک ادراکی در بسامدهای پایین‌تر دقیق‌تر است = اهمیت دادن بیشتر به بسامدهای پایین‌تر

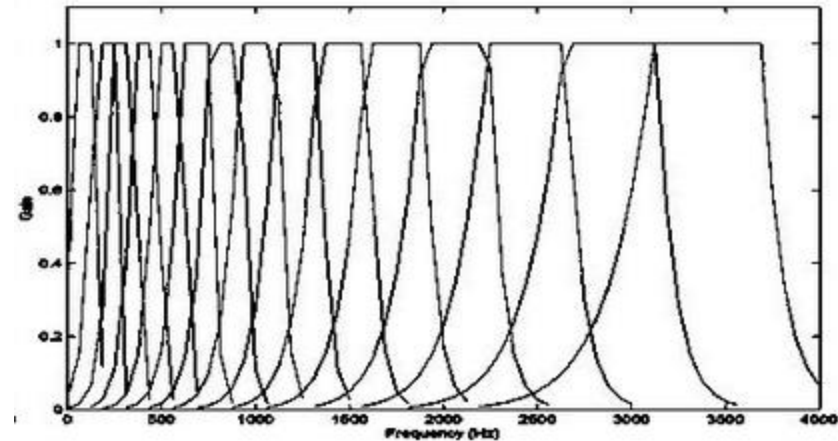




درک گفتار ...

○ تحلیل بسامدی: مقیاس بارک ...

Bark Band #	Edge (Hz)	Center (Hz)
1	100	50
2	200	150
3	300	250
4	400	350
5	510	450
6	630	570
7	770	700
8	920	840
9	1080	1000
10	1270	1170
11	1480	1370
12	1720	1600
13	2000	1850
14	2320	2150
15	2700	2500
16	3150	2900
17	3700	3400
18	4400	4000
19	5300	4800
20	6400	5800
21	7700	7000
22	9500	8500
23	12000	10500
24	15500	13500



تعداد بیشتر فیلترها
در بسامد پایین



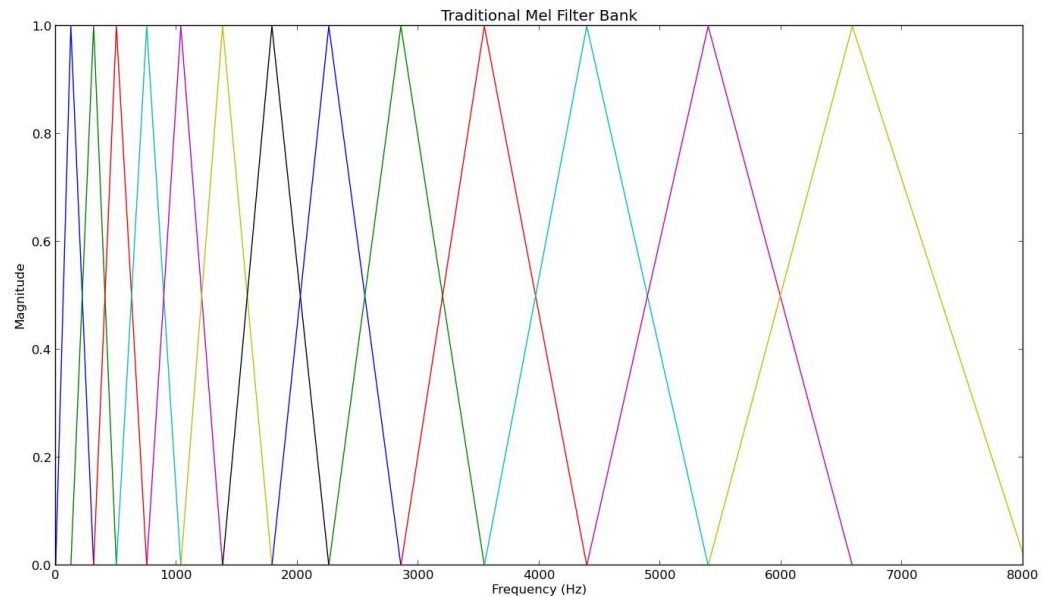
درک گفتار ...

○ تحلیل بسامدی: مقیاس مل (Mel)

- برای کمتر از یک کیلوهرتز خطی، و برای بیشتر از آن لگاریتمی $B(f) = 1125 \ln(1 + f / 700)$

- استفاده در استخراج ویژگی (MFCC)

Index	Bark Scale		Mel Scale	
	Center Freq. (Hz)	BW (Hz)	Center Freq. (Hz)	BW (Hz)
1	50	100	100	100
2	150	100	200	100
3	250	100	300	100
4	350	100	400	100
5	450	110	500	100
6	570	120	600	100
7	700	140	700	100
8	840	150	800	100
9	1000	160	900	100
10	1170	190	1000	124
11	1370	210	1149	160
12	1600	240	1320	184
13	1850	280	1516	211
14	2150	320	1741	242
15	2500	380	2000	278
16	2900	450	2297	320
17	3400	550	2639	367
18	4000	700	3031	422
19	4800	900	3482	484
20	5800	1100	4000	556
21	7000	1300	4595	639
22	8500	1800	5278	734
23	10500	2500	6063	843
24	13500	3500	6964	969





آواشناسی و واج‌شناسی ...

○ واج‌شناسی (Phonemics)

- توصیف صداهای گفتاری در یک زبان مشخص
- واج (Phoneme)
- کوچک‌ترین واحد صدای گفتار در زبان که باعث تمایز د و واژه از همدیگر می‌شود

○ آواشناسی (Phonetics)

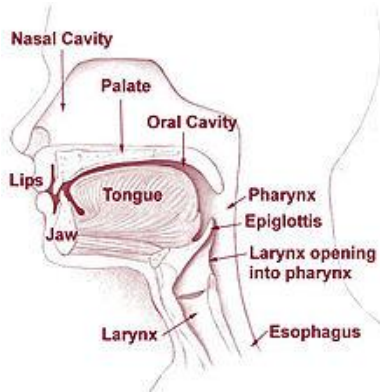
- توصیف دقیق صداهای گفتاری
- مستقل از یک زبان خاص
- آوا (Phone): یک صدای مشخص (نمود خارجی آکوستیک واج)
 - مثال: واج /t/ در انگلیسی دو نمود آکوستیک متفاوت در دو واژه sat و meter دارد
 - برای تشخیص گفتار بهتر است دو نوع t را دو آوای مختلف مدل کرد

- یکسان گرفتن آوا و واج برای بیان واحدهای مستقل از گوینده و مستقل از بافت صداهای معنادار

آواشناسی ...

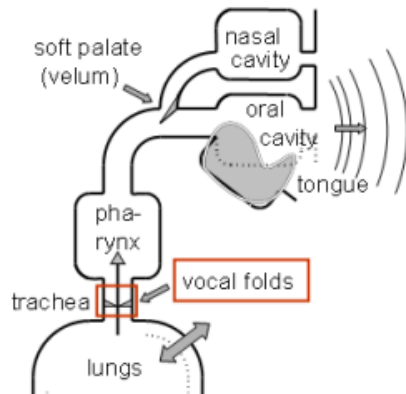
○ آواشناسی تولیدی (Articulatory Phonetics)

- نحوه تولید صداها بر اساس وضعیت اندام‌های تولید صدا
- محل و نحوه قرار گرفتن اندام‌ها



○ آواشناسی آکوستیک (Acoustic Phonetics)

- تاکید بر نحوه معیارهای قابل مشاهده و قابل اندازه‌گیری در شکل موج سیگنال گفتار
- ایجاد دانش (نظری و عملی) برای بازشناسی گفتار و تبدیل متن به گفتار



آواشناسی ...

○ آواشناسی تولیدی (Articulatory Phonetics)

• دسته‌بندی همخوان (Consonant)

○ محل تولید (Point of Articulation): محل اصلی انقباض در مجرای صوتی

○ دلبلی (Bilabial) مانند ب، لبی‌دندانی (Labiodental) مانند ف،

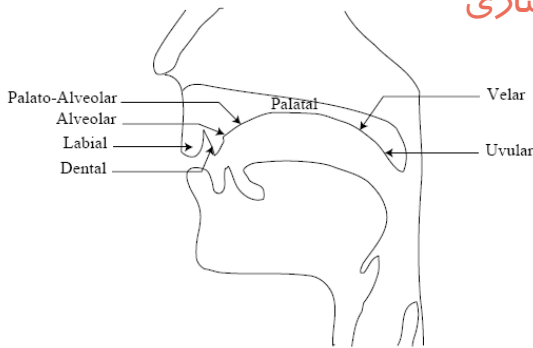
○ چاکنایی (Glottal) مانند ه، ...

○ نحوه تولید (Manner of Articulation): میزان انقباض و نحوه آزادسازی

○ انفجاری (Plosive) مانند پ، سایشی (Fricative) مانند س،

○ خیشومی (Nasal) مانند م، ...

○ صداداری (Voicing): واگذار بودن یا بی‌واکی



• دسته‌بندی واکه (Vowel)

○ ساختاردهی واکه‌ها مشکل‌تر است: زبان آزاد است و با اندام‌های دیگر تماس ندارد

○ دسته‌بندی بر اساس

○ بالا یا پایین بودن زبان، جلو باشد یا عقب، لب گرد باشد یا غیرگرد، خیشومی با غیرخیشومی



واچ‌شناسی ...

○ واچ‌شناسی (Phonemics)

- انگلیسی: ۳۱ تا ۶۴ واج
- فارسی: ۲۹ تا ۴۵ واج
 - بسته به نوع تحلیل
- بیشترین تعداد واج = ۴۵، متعلق به زبان Chipewyan (شمال کانادا)
 - ۳۹ همخوان و ۶ واکه
- کمترین تعداد واج = ۱۳، متعلق به زبان Hawaiian



واچ‌شناسی ...

○ واچ‌های انگلیسی

• ۴۱ واچ

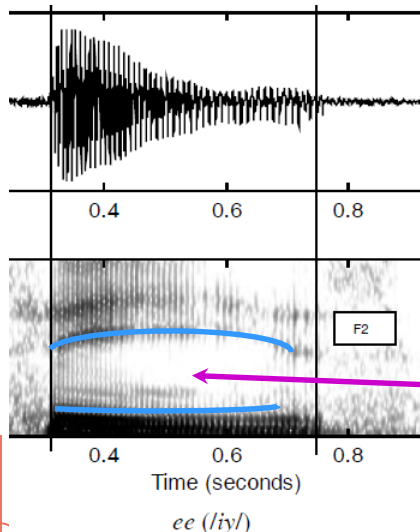
توصیف	مثال	واچ
پیشین، بسته، غیر گرد (نرم)	fill, hit, lid	ih
پیشین، باز، غیر گرد (سخت)	at, carry, gas	ae
پسین، باز، غیر گرد	father, ah, car	aa
نیمه‌باز، پسین، غیر گرد	cut, bud, up	ah
نیمه‌باز، پسین، گرد	dog, lawn, caught	ao
واکه مرکب به صورت: aa+ih	tie, ice, bite	ay
مرکزی، نیمه‌بسته (شوا)	ago, comply	ax
پیشین، نیمه‌بسته، غیر گرد (سخت)	ate, day, tape	ey
پیشین، نیمه‌باز، غیر گرد	pet, berry, ten	eh
مرکزی، نیمه‌باز، غیر گرد	turn, fur, meter	er
پسین، نیمه‌بسته، گرد	go, own, tone	ow
واکه مرکب به صورت: aa+uh	foul, how, our	aw
واکه مرکب به صورت: ao+ih	toy, coin, oil	oy
پسین، نیمه‌بسته، غیر گرد (نرم)	book, pull, good	uh
پسین، بسته، گرد	tool, crew, moo	uw
انفجاری، دولبی، واگذار	big, able, tab	b
انفجاری، دولبی، بی‌واک	put, open, tap	p
انفجاری، لثوی، واگذار	dig, idea, wad	d
انفجاری، لثوی، بی‌واک	talk, sat	t
ضربه‌ای، لثوی	meter	t
انفجاری، نرم کامی، واگذار	gut, angle, tag	g
انفجاری، نرم کامی، بی‌واک	cut, ken, take	k
سایشی، لبی‌دندانی، بی‌واک	fork, after, if	f
سایشی، لبی‌دندانی، واگذار	vat, over, have	v
سایشی، لثوی، بی‌واک	sit, cast, toss	s
سایشی، لثوی، واگذار	zap, lazy, haze	z
سایشی، دندانی، بی‌واک	thin, nothing, truth	th
سایشی، دندانی، واگذار	then, father, scythe	dh
سایشی، پس‌لثوی، بی‌واک	she, cushion, wash	sh
سایشی، پس‌لثوی، واگذار	genre, azure	zh
ناسوده، کناری، لثوی	Lid	l
ناسوده، کناری، نرم کامی	elbow, sail	l
ناسوده، برگشتی	red, part, far	r
غلطان، رسا، کامی	yacht, yard	Y
غلطان، رسا، لبی - نرم کامی	with, away	W
سایشی، چاکتایی، بی‌واک	help, ahead, hotel	hh
خیشومی، دولبی	mat, amid, aim	m
خیشومی، لثوی	no, end, pan	n
خیشومی، نرم کامی	sing, anger	ng
انفجاری - سایشی، لثوی، بی‌واک: t+sh	chin, archer, march	ch
انفجاری - سایشی، لثوی، واگذار: d+zh	joy, agile, edge	jh



واحدشناسی: واکه‌ها (انگلیسی) ...

○ واکه‌ها (Vowels) ...

- شکل و موقعیت زبان در حفره دهانی، فشردگی زیادی در جریان هوا به وجود نمی‌آورد
- تغییر در موقعیت مکانی زبان با تغییر تشدید به هر واکه ویژگی خاص خودش را می‌دهد
 - مثل بطری‌هایی با اندازه و شکل مختلف
 - انرژی اولیه وارد شده به حفره‌های حلقی و دهانی در بسامد پایه مرتعش می‌شود
 - تشدیدهای عمده حفره‌های دهانی و حلقی برای واکه‌ها در فرمنت اول (F1) و فرمنت دوم (F2)
- تعیین فرمنت‌ها با موقعیت زبان و شکل تارهای صوتی



- مقدار F1 متناسب با بخش پسین یا حلقی حفره صوتی (از چاکنای تا زبان)
 - حفره طولانی‌تر = تشدید کمتر
- مقدار F2 متناسب با اندازه و شکل قسمت دهانی و قسمت جلویی زبان
- مثال: واکه واژه see
- زبان در دهان به جلو حرکت می‌کند و حفره‌ای طولانی ایجاد می‌کند = F1 کمتر
- بخش جلویی حفره دهانی بسیار کوتاه = F2 بیشتر

فاصله زیاد دو فرمنت



واحدشناسی: واژه‌ها (انگلیسی) ...

Vowel Labels Mean F1 (Hz) Mean F2 (Hz)

<i>iy (feel)</i>	300	2300
<i>ih (fill)</i>	360	2100
<i>ae (gas)</i>	750	1750
<i>aa (father)</i>	680	1100
<i>ah (cut)</i>	720	1240
<i>ao (dog)</i>	600	900
<i>ax (comply)</i>	720	1240
<i>eh (pet)</i>	570	1970
<i>er (turn)</i>	580	1380
<i>ow (tone)</i>	600	900
<i>uh (good)</i>	380	950
<i>uw (tool)</i>	300	940

○ واژه‌ها (Vowels) - انگلیسی ...

• ۱۲ واج

○ واژه مرکب (Diphthongs)

• ترکیب دو واژه با هم با حرکت موقعیت زبان از جایی به جای دیگر

• ۴ واج

Diphthong Labels

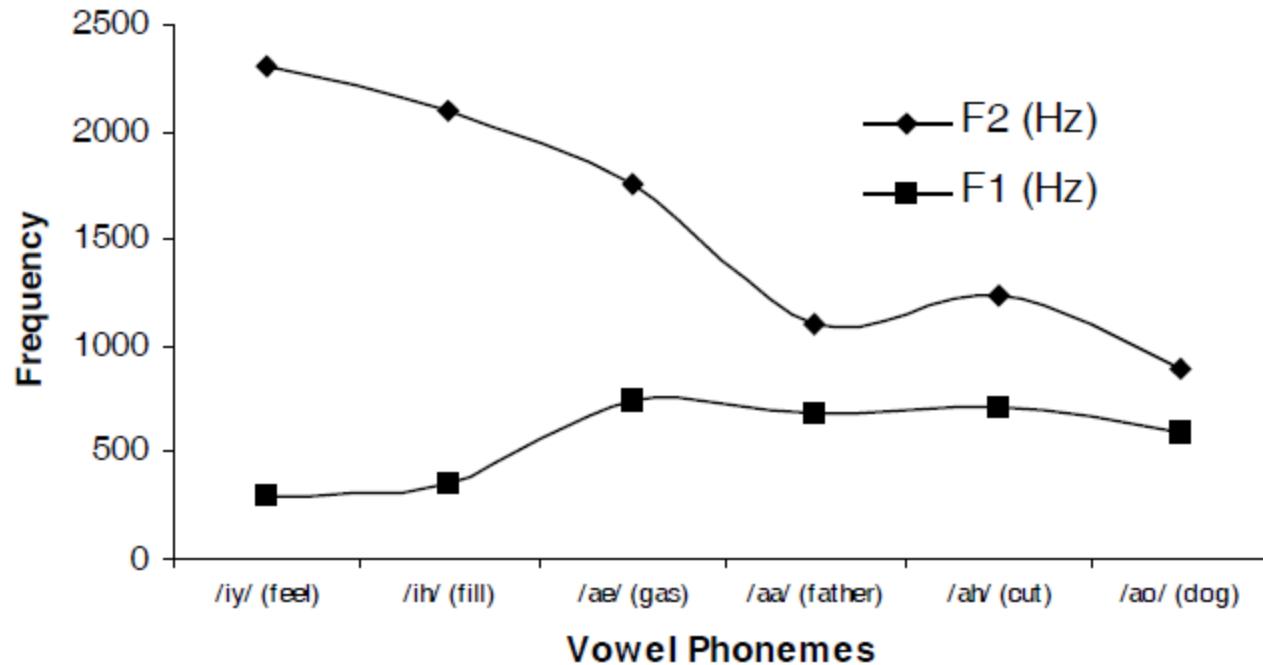
Components

ay (tie)	<i>/aɪ/</i> → <i>/iɪ/</i>
ey (ate)	<i>/eh/</i> → <i>/iɪ/</i>
oy (coin)	<i>/aɔ/</i> → <i>/iɪ/</i>
aw (foul)	<i>/aɪ/</i> → <i>/uɪ/</i>



واحدشناسی: واکه‌ها (انگلیسی) ...

○ مقادیر F1 و F2 برای تولید برخی واکه‌های انگلیسی

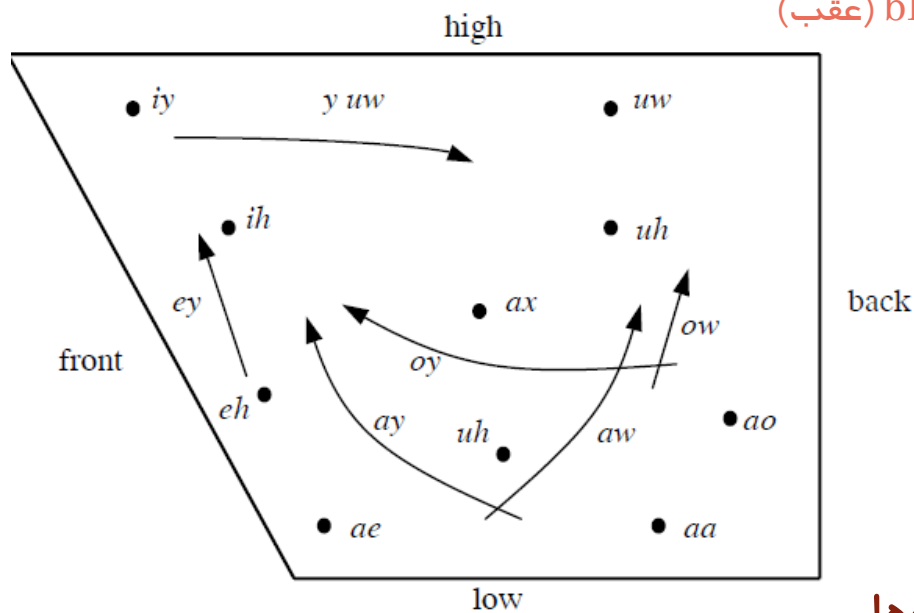




واحدشناسی: واژه‌ها (انگلیسی) ...

○ نحوه تولید برای انگلیسی

- اندام اصلی: قسمت میانه تا عقب زبان
- دامنه حرکت [جلو↔عقب] و [بالا↔پایین]
- صدای /iy/ در see (بالا) به /aa/ در father (پایین)
- صدای /uw/ در see (جلو) به /uw/ در blue (عقب)



- اندام ثانویه تولید واژه در انگلیسی = لب‌ها

○ حرکت از /iy/ (مسطح و کمی باز و تا حدودی گسترده) به /uw/ (گرد)



واج‌شناسی: واکه‌ها (انگلیسی) ...

ویژگی‌های واج‌شناختی واکه‌های اصلی زبان انگلیسی

واکه	افراشته	افتاده	پیشین	پسین	گرد	شدت
ly	+	-	+	-	-	+
lh	+	-	+	-	-	-
ae	-	+	+	-	-	+
aa	-	+	-	-	-	+
ah	-	-	-	-	-	+
ao	-	+	-	+	+	+
ax	-	-	-	-	-	-
eh	-	-	+	-	-	-
ow	-	-	-	+	+	+
uh	+	-	-	+	-	-
uw	+	-	-	+	-	+



واج‌شناسی: همخوان‌ها (انگلیسی) ...

○ همخوان‌ها (Consonants)

- تولید با فشردگی یا گرفتگی‌های عمده در حفره‌های حلقی و یا دهانی
- برخی همخوان‌ها واک‌دار هستند و برخی بی‌واک
- بسیاری از همخوان‌ها به صورت جفت هستند

○ از نظر تولیدی یکسان، اما یکی از جفت‌ها صدایی اضافه‌تر دارد

○ مثال: جفت‌های با وجه تمایز واک‌داری - /b, d, g, z, zh, v/ و /p, t, k, s, sh, f/

شیوه تولید	مثال واجی	مثال واژه‌ای	چگونگی تولید
انفجاری (Plosive)	/p/	tat, tap	گرفتگی حفره دهانی
خیشومی (Nasal)	/m/	team, meet	گرفتگی حفره خیشومی
سایشی (Fricative)	/s/	sick, kiss	نوفه‌ی جریان هوای متلاطم
روان برگشتی (Retroflex liquid)	/r/	rat, tar	شبه واکه، زبان افراشته و به سمت عقب گرد شده
روان کناری (Lateral liquid)	/l/	lean, kneel	شبه واکه، زبان در مرکز و جریان هوا در کناره‌های آن
غلطان (Glide)	/y/, /w/	yes, well	شبه واکه



واحدشناسی: همخوان‌ها (انگلیسی) ...

○ شبه‌واکه (Semivowel)

- آوای واک‌داری که بدون انسداد کامل یا کشیدگی تارهای صوتی، ایجاد می‌شوند
- روان‌ها (Liquid): /l, r/
- غلتان‌ها (Glide): /y, w/
- معادل واکه‌های /iy, uw/ - موقعیت آغازین آن‌ها کمی کوتاه‌تر است

○ انفجاری-سایشی (Affricate)

- ترکیب یک انسدادی (مثل /t/) و یک سایشی (مثل /sh/) = (t+sh)=ch در church
- در انگلیسی انفجاری-سایشی‌ها جفت‌های واک‌دار/بی‌واک هستند
- /ch/ (t+sh) و /j/ (d+zh)

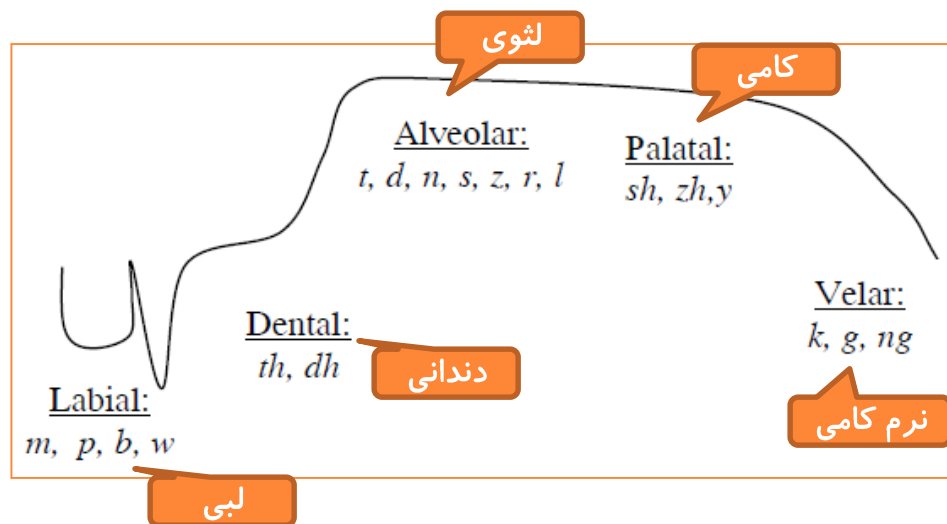


واج‌شناسی: همخوان‌ها (انگلیسی) ...

همخوان	مثال	واکداری؟	شیوه تولید
b	big, able, tab	+	انفجاری
p	put, open, tap	-	انفجاری
d	dig, idea, wad	+	انفجاری
t	talk, sat	-	انفجاری
g	gut, angle, tag	+	انفجاری
k	cut, oaken, take	-	انفجاری
v	vat, over, have	+	سایشی
f	fork, after, if	-	سایشی
z	zap, lazy, haze	+	سایشی
s	sit, cast, toss	-	سایشی
dh	then, father, scythe	+	سایشی
th	thin, nothing, truth	-	سایشی
zh	genre, azure, beige	+	سایشی
sh	she, cushion, wash	-	سایشی
jh	joy, agile, edge	+	انفجاری - سایشی
ch	chin, archer, march	-	انفجاری - سایشی
l	lid, elbow, sail	+	کناری
r	red, part, far	+	برگشتی
y	yacht, onion, yard	+	غلطان
w	with, away	+	غلطان
hh	help, ahead, hotel	+	سایشی
M	mat, amid, aim	+	خیشومی
n	no, end, pan	+	خیشومی
ng	sing, anger, drink	+	خیشومی

تولید همخوان‌های انگلیسی

• ۲۴ واج





واج‌شناسی: همخوان‌ها (انگلیسی) ...

○ همخوان‌های انگلیسی براساس جایگاه (ستون) و شیوه (ردیف) تولید

چاکنایی	نرم‌کامی	کامی	لثوی	دندانی	لبی - دندانی	لبی	
?	kg		t d			p b	انفجاری
	ng		N			m	خیشومی
h		sh zh	s z	th dh	f v		سایشی
			r				برگشتی رسا
			l				کناری رسا
		y				w	غلطان



واج‌شناسی: همخوان‌ها (انگلیسی) ...

الفبای آوانگاری بین‌المللی (IPA: International Phonetic Alphabet)

- سامانه نوشتاری و آوانگاری استاندارد شده توسط انجمن آواشناسی بین‌المللی برای نشان دادن صداهای ایجاد شده در زبان‌های گوناگون

vowels

IPA	ASCII	examples
ʌ	^	cup, luck
ɑ:	a:	arm, father
æ	@	cat, black
ə	..	away, cinema
e	e	met, bed
ɜ:ʳ	e:(r)	turn, learn
ɪ	i	hit, sitting
i:	i:	see, heat
ɒ	o	hot, rock
ɔ:	o:	call, four
ʊ	u	put, could
u:	u:	blue, food
aɪ	ai	five, eye
aʊ	au	now, out
oʊ/əʊ	Ou	go, home
eəʳ	e..(r)	where, air
eɪ	ei	say, eight
ɪəʳ	i..(r)	near, here
ɔɪ	oi	boy, join
ʊəʳ	u..(r)	pure, tourist

consonants

IPA	ASCII	examples
b	b	bad, lab
d	d	did, lady
f	f	find, if
g	g	give, flag
h	h	how, hello
j	j	yes, yellow
k	k	cat, back
l	l	leg, little
m	m	man, lemon
n	n	no, ten
ŋ	N	sing, finger
p	p	pet, map
r	r	red, try
s	s	sun, miss
ʃ	S	she, crash
t	t	tea, getting
tʃ	tS	check, church
θ	th	think, both
ð	TH	this, mother
v	v	voice, five
w	w	wet, window
z	z	zoo, lazy
ʒ	Z	pleasure, vision
dʒ	dZ	just, large

- بر پایه الفبای لاتینی

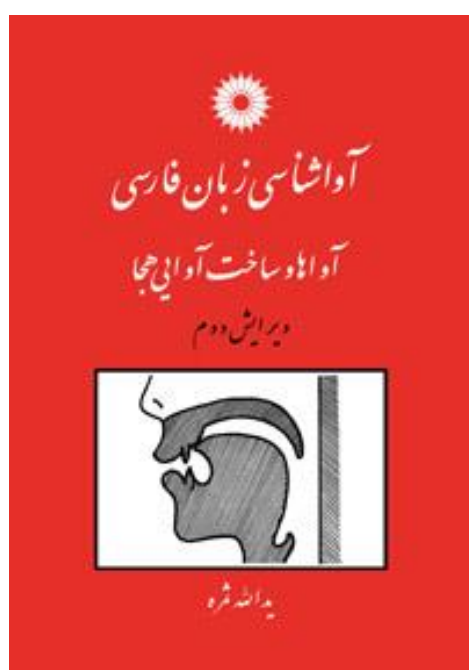
روش آوانگاری ARPabet
مبتنی بر نمادهای ASCII
است

واج‌شناسی: واکه‌ها (فارسی) ...

○ واکه‌های فارسی

• ۶ واج (+۲ واکه مرکب)

مثال		
علی	ی	i
انسان	ا	e
اکبر	آ	a
بورس	و	u
شد	أ	o
آب	آ	σ
حوله	أو	ou
میل	ای	ei





واج‌شناسی: همخوان‌ها (فارسی) ...

○ همخوان‌های فارسی

• ۲۵ واج (۲۳ واج)

واج روان / و / در زبان‌های عربی و گُردی استفاده می‌شود ولی در فارسی تقریباً فراموش شده است و فقط در بعضی از لهجه‌های فارسی استفاده می‌شود.

شبه واکه	
ل	l
ر	r
م	m
ن	n
ی، یه	j

انفجاری-سایشی	
ج	dʒ
چ	tʃ

انفجاری		
	ب	b
	پ	p
	ت، ط	t
	د	d
تاکید	ک (پیشین)	c
کار	ک (پسین)	k
اگر	گ (پیشین)	ɣ
سیگار	گ (پسین)	g
	ق، غ	G
	ع، ع	ɛ

سایشی	
و	v
ف	f
ث، س، ص	s
ز، ذ، ض، ظ	z
ه، ح	h
ش	ʃ
ژ	ʒ
خ	χ



واج‌شناسی: همخوان‌ها (فارسی) ...

○ همخوان‌های فارسی براساس جایگاه (ستون) و شیوه (ردیف) تولید

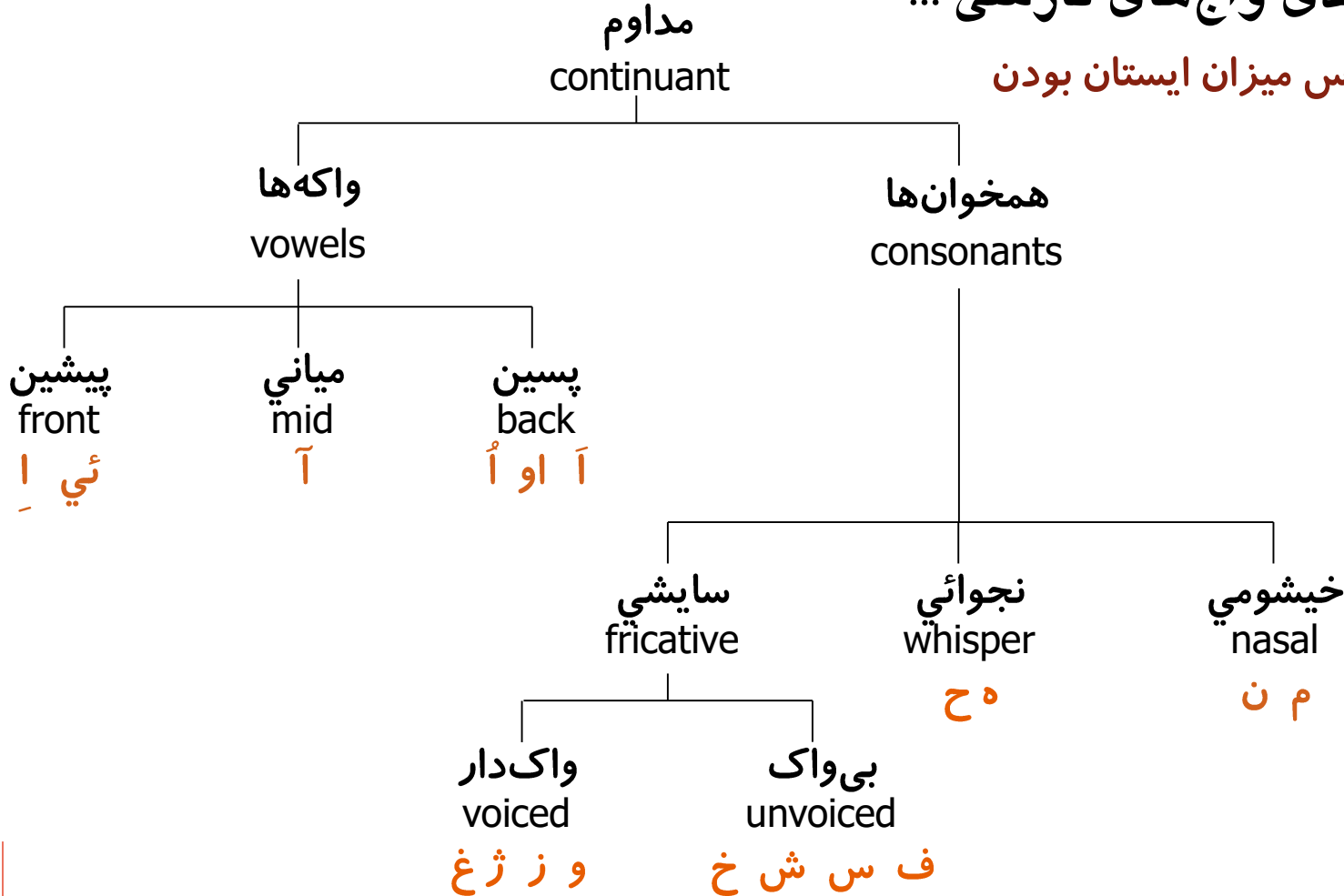
چاکنایی	نرم‌کامی	کامی	لثوی-کامی	لثوی	دندانی	لبی-دندانی	لبی	
ء، ع	ق	ک، گ			ت، د		ب، پ	انفجاری
				ن			م	خیشومی
ه	خ		ش، ژ	س، ز		ف، و		سایشی
			چ، ج					انفجاری-سایشی
		ای		ل				روان
				ر				لرزشی



واج‌شناسی: فارسی ...

○ تقسیم‌بندی واج‌های فارسی ...

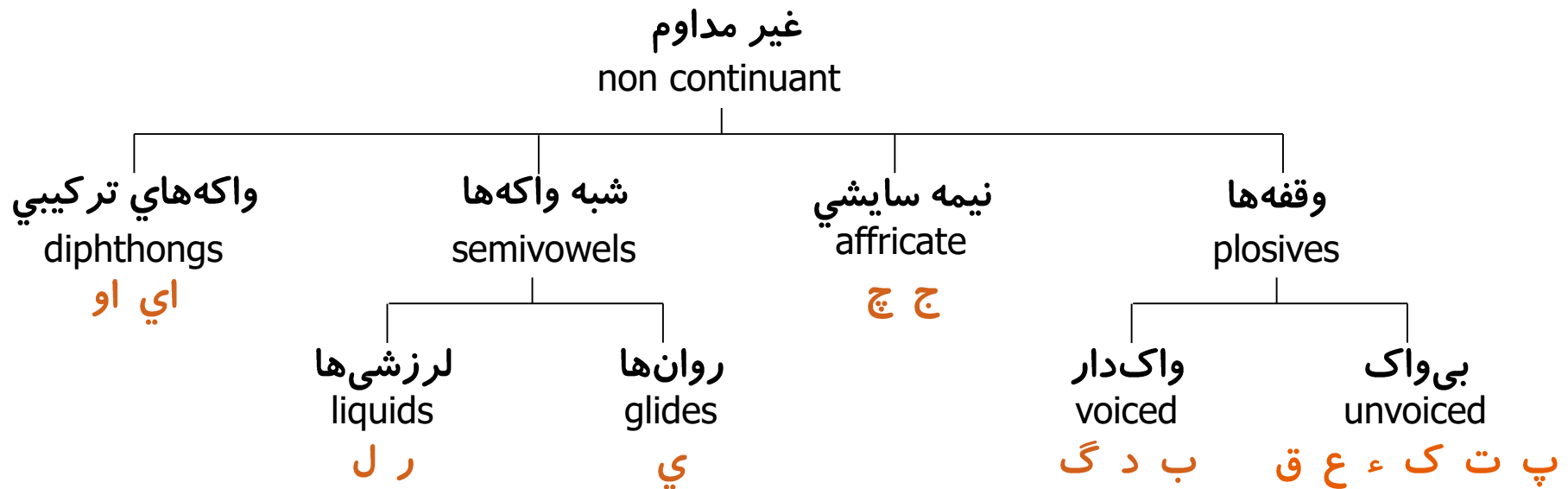
- بر اساس میزان ایستان بودن





واج‌شناسی: فارسی

○ تقسیم‌بندی واج‌های فارسی





واحدشناسی: فارسی

◉ علائم اختصاری و IPA

IPA	Char	Code	Farsi Letter	Phonetic Description
I	i	105	ای	high front unrounded
e	e	101	ا	mid front unrounded
a	a	97	آ	low front unrounded
u	u	117	او	high back rounded
o	o	111	اُ	mid back rounded
/	/	47	اَ	low back rounded
\	\	92	پ،	unvoiced bilabial plosive closure
p	p	112	پ،	unvoiced bilabial plosive
`	`	96	پ،	voiced bilabial plosive closure
b	b	98	ب،	voiced bilabial plosive
-	-	45	ت، ط	unvoiced alveolar plosive closure
t	t	116	ت، ط	unvoiced dental plosive
=	=	61	د	voiced dental plosive closure
d	d	100	د	voiced dental plosive
@	@	64	ک	unvoiced palatal plosive closure
c	c	99	ک	unvoiced palatal plosive
*	*	42	ک	unvoiced velar plosive closure
k	k	107	ک	unvoiced velar plosive
!	!	33	گ	voiced palatal plosive closure
;	;	59	گ	voiced palatal plosive
&	&	38	گ	voiced velar plosive closure
g	g	103	گ	voiced velar plosive
^	^	94	ق، غ	voiced uvular plosive closure
q	q	113	ق، غ	voiced uvular plosive
((40	ا، و، ع	glottal stop closure
]]	93	ا، و، ع	glottal stop
\$	\$	36	چ	unvoiced alveopalatal affricate closure
'	'	39	چ	unvoiced alveopalatal affricate
#	#	35	ج	voiced alveopalatal affricate closure
,	,	44	ج	voiced alveopalatal affricate
f	f	102	ف	unvoiced labiodental fricative
v	v	118	و	voiced labiodental fricative
s	s	115	س، ش، ص	unvoiced alveolar fricative
z	z	122	ز، ذ، ظ، ض	voiced alveolar fricative
.	.	46	ش	unvoiced alveopalatal fricative
[[91	ژ	voiced alveopalatal fricative
x	x	120	خ	unvoiced uvular fricative
h	h	104	ه، ح	unvoiced glottal fricative
l	l	108	ل	lateral alveolar
r	r	114	ر	trill alveolar
m	m	109	م	nasal bilabial
n	n	110	ن	nasal alveolar
y	y	121	ی	approximant palatal



رده‌شناسی آوایی

○ رده‌شناسی آوایی (Phonetic Typology)

- علاوه بر تاثیر مکانیسم‌های دهانی، خیشومی، حلقی، و چاکنایی در تولید آواها، برخی اثرات صوتی نیز در این موضوع موثرند
- در واژه‌های زبان ژاپنی، کشش (Length) یک ویژگی متمایز کننده است
 - کلمات kado (گوشه) و kaado (کارت) به لحاظ طیفی کاملاً یکسان بوده و تنها تفاوت آن‌ها در این است که kado در تمام بافت‌ها بسیار کوتاه‌تر ادا می‌شود.
- صدای r لرزشی و انسدادی درون‌سو در زبان اسپانیولی
 - کلماتی مانند pero (اما) و perro (سگ) را از هم متمایز می‌کند
- تغییر زیروبمی (Tone) در زبان‌های آسیا و آفریقا (مانند تمام گونه‌های چینی)
 - زبان‌های نواختی (Tonal): زیروبمی در آن تقابل معنایی ایجاد کند
 - مثال در زبان چینی ماندارین (چهار نواخت اصلی)

معنی	چینی	مثال	شکل	نواخت
مادر	妈	ma	مسطح افراشته	۱
بی‌حس	麻	ma	خیزان افراشته	۲
اسب	马	ma	خیزان افتاده	۳
ناسزا گفتن	骂	ma	افتان افراشته	۴

○ مهم در تبدیل متن به گفتار



واج‌گونه

○ واج‌گونه (Allophone)

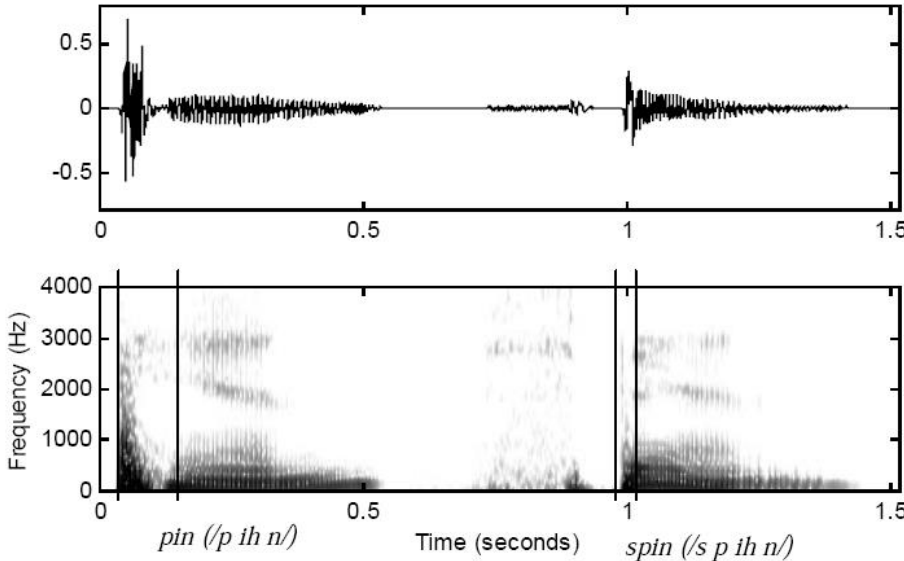
• حالت‌های مختلف تلفظی برای یک واج

• مثال ۱: صدای /p/ در

○ واژه pin

○ دم‌ش یا Aspiration جریان مشهود هوا

○ واژه spin



• مثال ۲: واج /l/ در like (رقیق) و kill (غلیظ)

• ایجاد مشکل در تشخیص گفتار

○ مدل کردن همه واج‌گونه‌ها



هم‌تولیدی ...

○ هم‌تولیدی (Coarticulation)

- فرایندی که در آن صداهای مجاور بر هم تأثیر می‌گذارند

- مثال: جمله Did you hit it to Tom?

○ تلفظ اصلی: /d ih d y uw h ih t ih t t uw t aa m/

○ تلفظی که شنیده می‌شود: /d ih jh ax hh ih dx ih t ix t aa m/

○ کامی‌شدگی /d/ قبل از /y/ در did you

○ کاهش /u/ بدون تکیه به شوا در you

○ زنبشی شدن /t/ بین دو واکه در hit it

○ کاهش شوا و واک‌رفتگی /u/ در to

○ کاهش همخوان مشدد (دوتایی) /t/ در it to

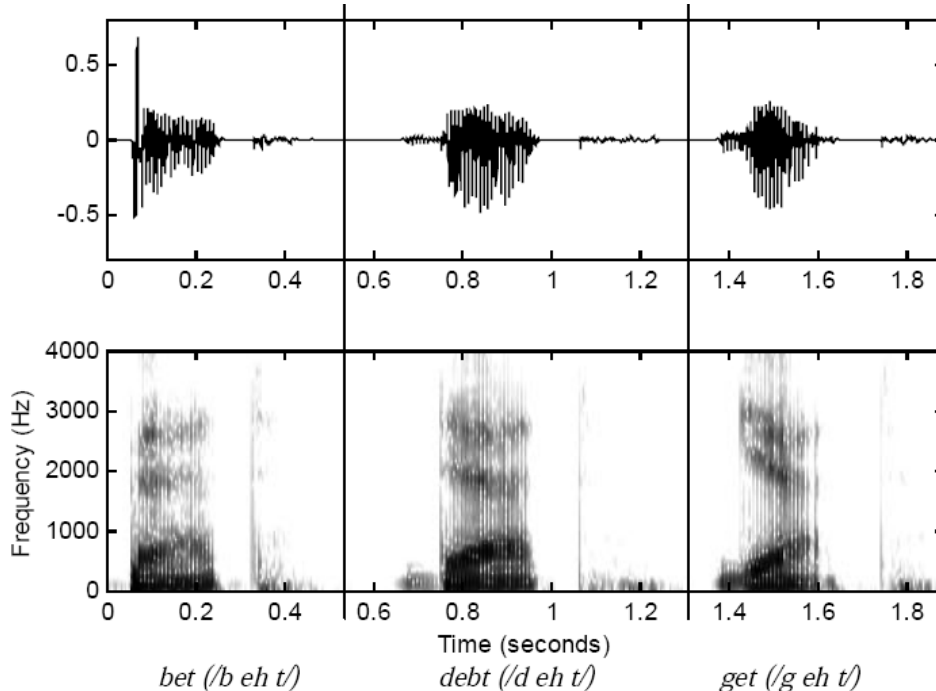
- مثال فارسی: آب پاش، متشکرم (مچکرم)



هم‌تولیدی ...

○ تأثیر هم‌تولیدی در نمایش طیفی ...

- تأثیر زیاد همخوان‌ها بر مسیر فرمنت واکه‌های مجاور
- مثال: تفاوت $F1$ و $F2$ را در واکه‌ی /eh/ در سه واژه bet، debt، و get
 - تفاوت در واج آغازین
 - سه واج آغازی نمایانگر سه جایگاه تولید اصلی (لبی، لثوی و نرم‌کامی) هستند



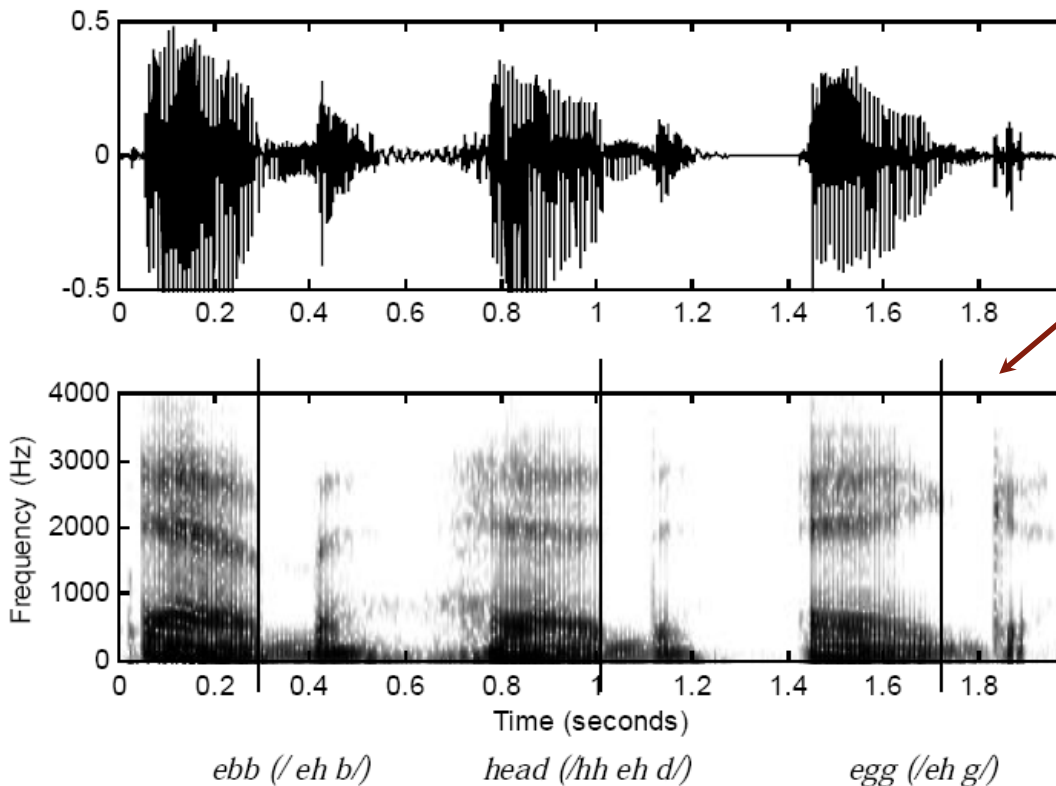
اثر هم‌تولیدی استقامت (Perseverance)
ابتدا همخوان سپس واکه



هم‌تولیدی

○ تأثیر هم‌تولیدی در نمایش طیفی

- مثال: همخوان‌های متفاوت بعد از یک واکه در *ebb* و *head* و *egg*
- افزایش گستردگی نسبی $F1$ و $F2$ را در تبدیل واکه پایانی-همخوان



اثر هم‌تولیدی تقدم (Anticipation)
ابتدا واکه سپس همخوان