



Digital Speech Processing

به نام خدا



گروه هوش مصنوعی، دانشکده مهندسی
کامپیوتر

گفتار پردازش رقمی

مدلسازی سیستم تولید گفتار

حمیدرضا برادران کاشانی

پاییز ۱۴۰۱



سرفصل مطالب

- ❖ دوره تناوب، فرکانس و طول موج
- ❖ موج ساکن و تشدید
- ❖ مدلسازی مجرای گفتار بصورت تیوب
- ❖ مدل‌های مجرای گفتار پیچیده تر
- ❖ مقادیر تقریبی فرکانس های فرمنت واکه ها
- ❖ مدل منبع - فیلتر

Hamidreza Baradaran Kashani



محیط انتشار و صوت

❖ محیط انتشار (medium) و صوت (sound)

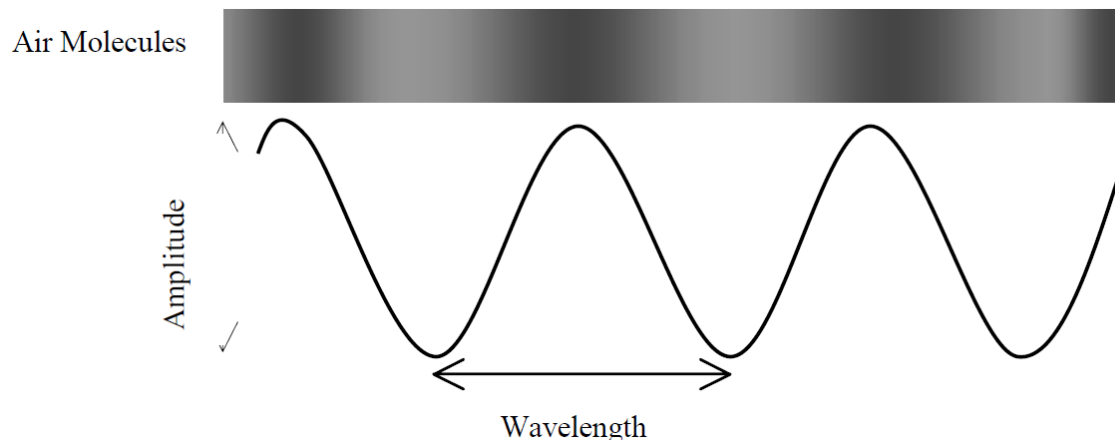
❖ محیطی که در ذرات آن آشفتگی (disturbance) پدید می آید.

❖ مولکول های هوا عامل انتشار امواج صوتی هستند.

❖ به عبارتی صوت، موجی است که در اثر حرکت مولکول های هوا (فشرده گی و بازشدگی مولکول های هوا) بوجود می آید.

❖ محیط انتشار می تواند مادی باشد مثل آب یا هوا و یا غیرمادی یعنی خلا باشد.

❖ سرعت انتشار موج صوتی در هوا ۳۵۰ متر بر ثانیه است.



Hamidreza Baradaran Kashani



دوره تناوب، فرکانس و طول موج

❖ سه تا مشخصه مهم در یک موج صوتی:

❖ دوره تناوب (Period):

❖ فاصله زمانی بین دو پیک یک موج را دوره تناوب آن گویند (بر حسب ثانیه)

❖ یا بطور کلی تر: مدت زمانی که یک سیکل موج کامل شود را پرپود گویند.

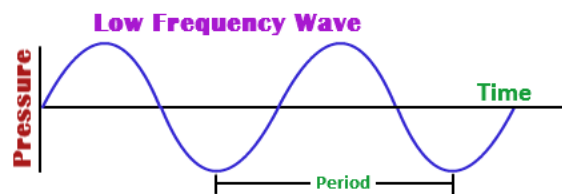
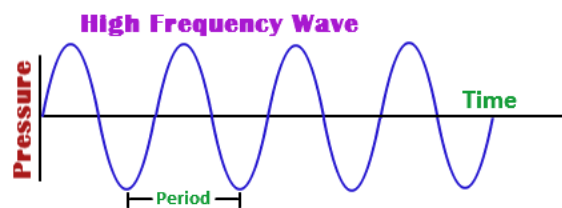
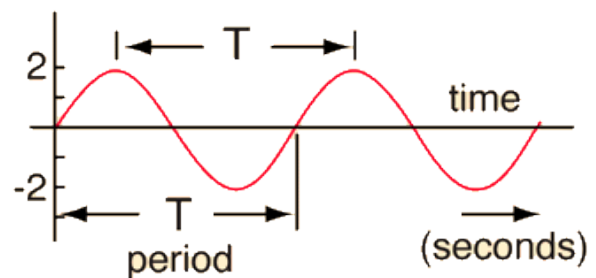
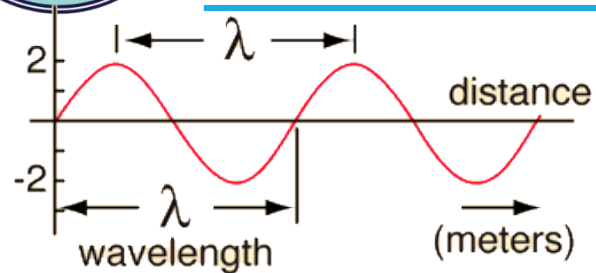
❖ فرکانس (Frequency):

❖ فرکانس عکس دوره تناوب است، به عبارتی تعداد سیکل ها در یک ثانیه را

فرکانس گویند (بر حسب هرتز)

❖ طول موج (Wavelength):

❖ به مسافتی که موج در یک پرپود طی می کند طول موج گویند (بر حسب متر)



Hamidreza Baradaran Kashani

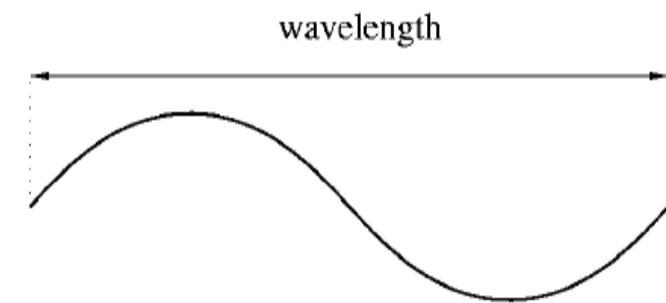
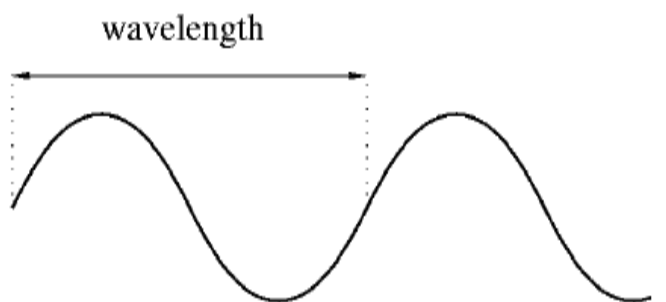


فرکانس و طول موج

❖ فرکانس بیشتر یعنی طول موج کمتر و بالعکس

❖ کم یا زیاد شدن فرکانس موج و یا طول موج

تاثیری در سرعت موج ندارد.



lower frequency
means
longer wavelength

$$f \lambda = c$$

f is frequency, λ is wavelength

c is about 350ms^{-1} (metres per second)

Hamidreza Baradaran Kashani



موج ساکن و تشدید

- ❖ ممکن است موج صوتی در حین انتشار در محیط به مانعی برخورد کند و یا به محیط دیگری برسد.
- ❖ مثلاً موج صوتی در یک لوله استوانه‌ای با انتهای بسته حرکت کند و در واقع انتهای لوله مرز بین هوای محبوس در لوله و هوای بیرون لوله است.
- ❖ اصطلاحاً موج در مرز دو محیط رفتار مرزی نشان می‌دهد که یکی از این رفتارهای مرزی مهم در ارتباط با موج صوتی بازتاب (Reflection) است.
- ❖ وقتی موج به انتهای محیط انتشار برسد و بازتاب شود با امواج دیگری که بوسیله منبع صوت تولید شده تداخل می‌کنند.



موج ساکن و تشدید

❖ امواج تابشی و بازتابشی در نقاط خاصی از محیط انتشار اثر یکدیگر را خنثی می کنند و به عبارتی ذرات محیط انتشار در آن نقاط خاص جا بجا نمی شوند و ثابتند.

❖ به علت ثابت ماندن آن نقاط، به موج تشکیل شده **موج ساکن (standing wave)** می گویند.

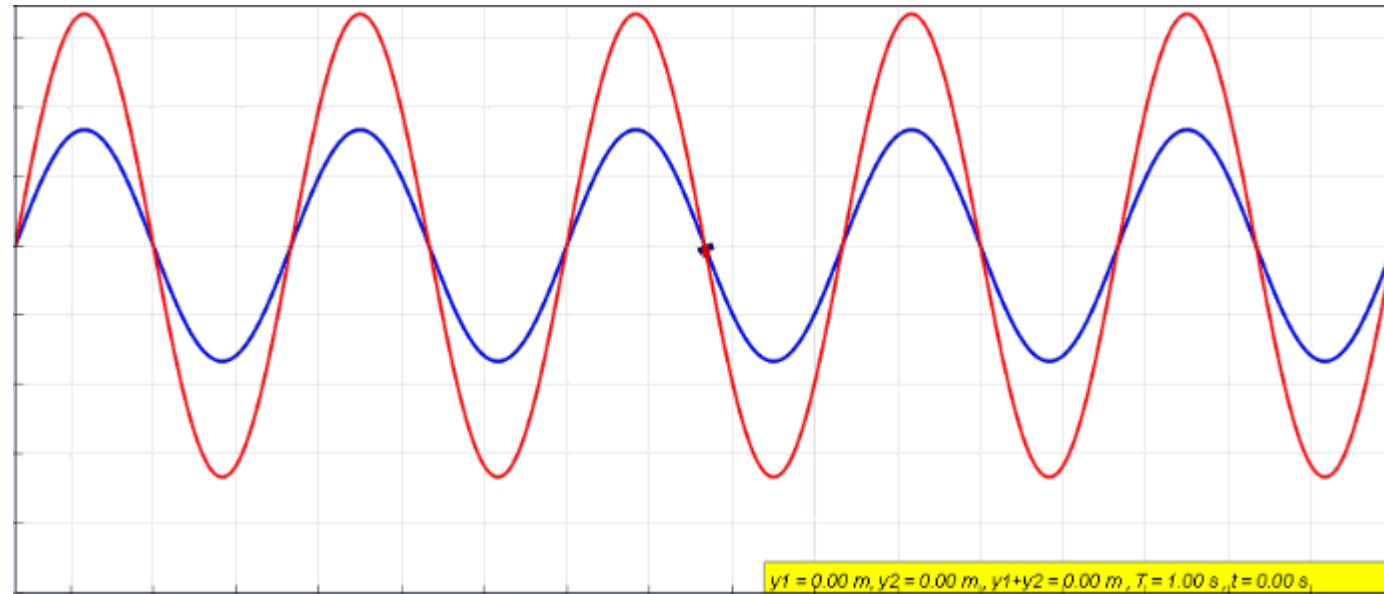
❖ به عبارتی به ازای فرکانسی خاص، تداخل این امواج منجر به تولید موج ساکن می شود.

❖ به آن فرکانس خاص فرکانس تشدید یا فرکانس طبیعی آن جسم گویند.



موج ساکن و تشدید

Node
Antinode



Animation of a standing wave (*red*) created by the superposition of a left traveling (*blue*) and right traveling (*green*) wave

Hamidreza Baradaran Kashani



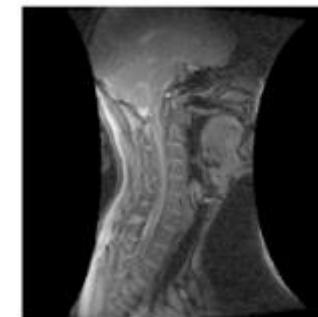
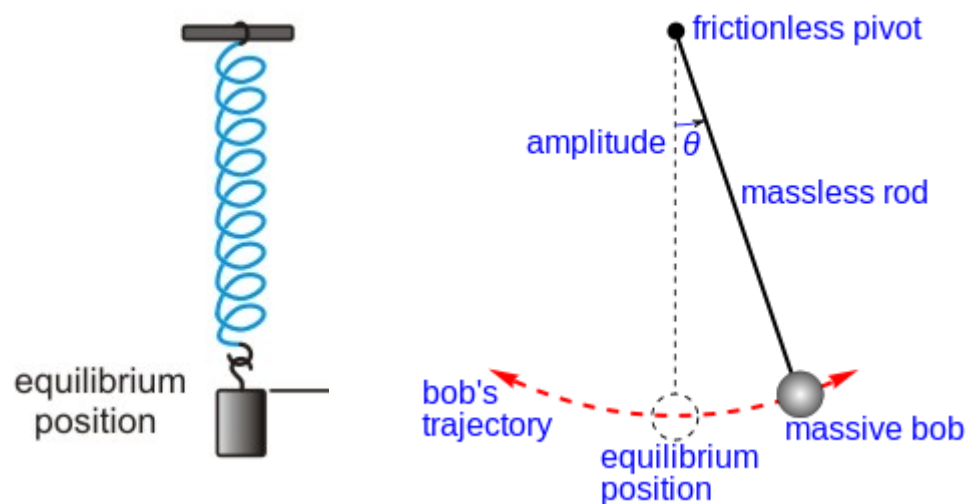
فرکانس تشدید

❖ هر سیستم مکانیکی تمایل دارد در برخی از فرکانس ها با بیشترین دامنه ممکن نوسان کند.



❖ نمونه ای از سیستم های تشدید:

❖ تاب بچه، پاندول ساعت، سیستم جرم و فنر و **مجرای گفتار**

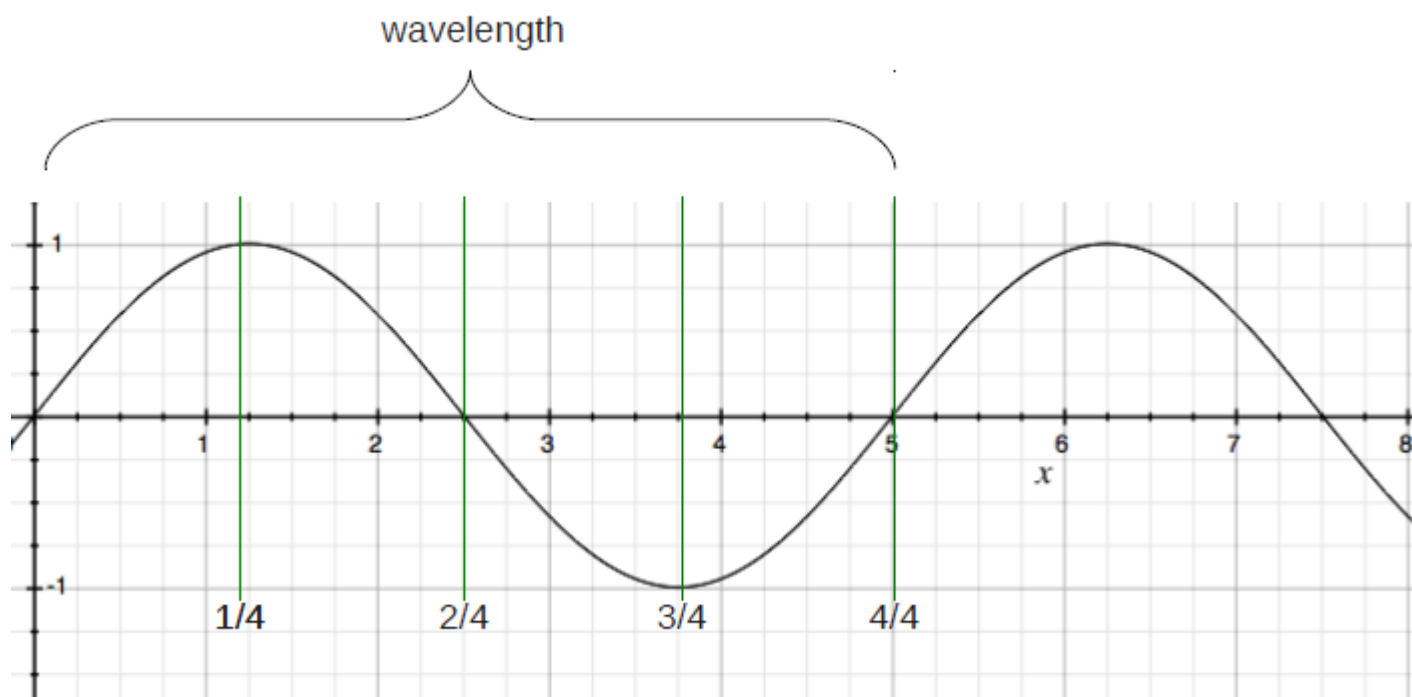


Hamidreza Baradaran Kashani



مدلسازی مجرای گفتار بصورت تیوب

❖ با توجه به شکل زیر مشاهده می شود که پیک در $1/4$ طول موج رخ می دهد.

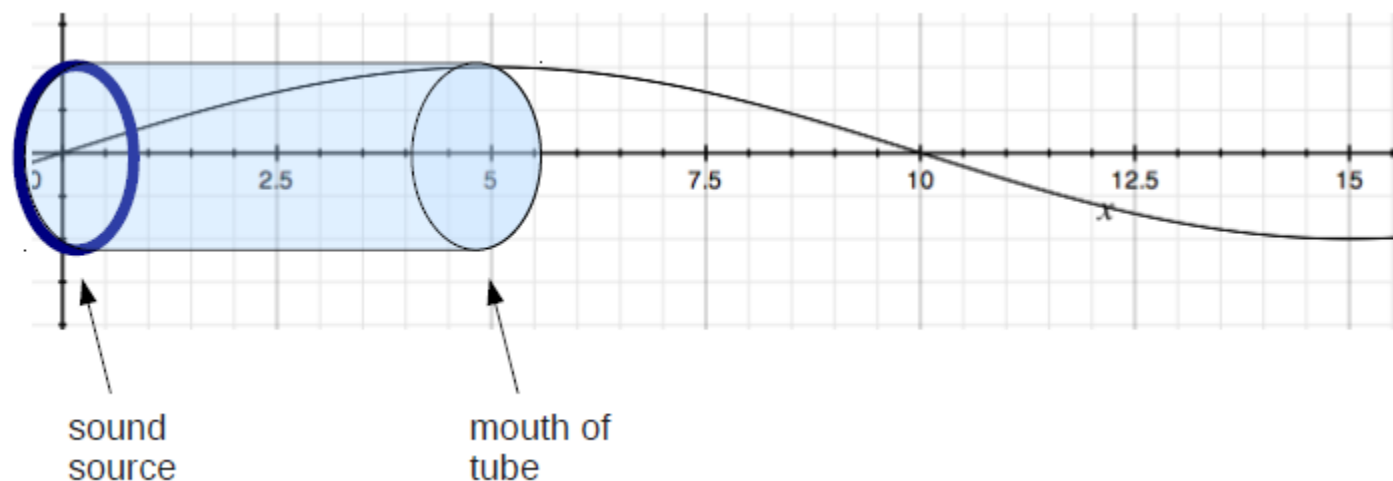
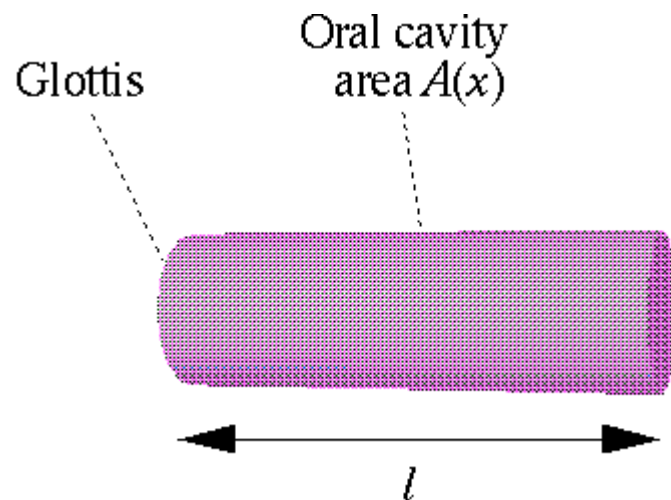


Hamidreza Baradaran Kashani



مدلسازی مجرای گفتار بصورت تیوب

❖ اگر صوت از تیوبی خارج شود که طولش $1/4$ طول موج باشد، صوت خارج شده از تیوب بیشترین دامنه و بلندی صدا را خواهد داشت.



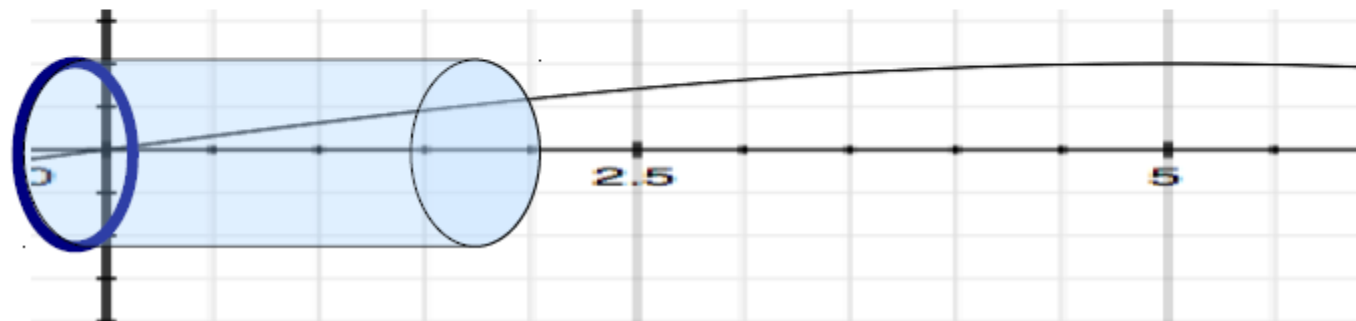
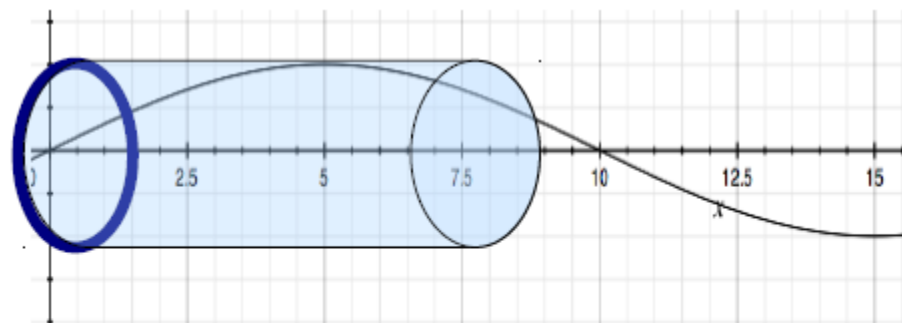
❖ در ساده ترین مدل، مجرای گفتار را می توان بصورت یک تیوب ساده با سطح مقطع عرضی یکسان که انتهای آن در سمت چاکنای بسته و در سمت لب ها باز است.

Hamidreza Baradaran Kashani



مدلسازی مجرای گفتار بصورت تیوب

❖ اگر طول تیوب بیشتر یا کمتر از طول موج باشد، آنگاه صدای خارج شده ضعیف تر است.

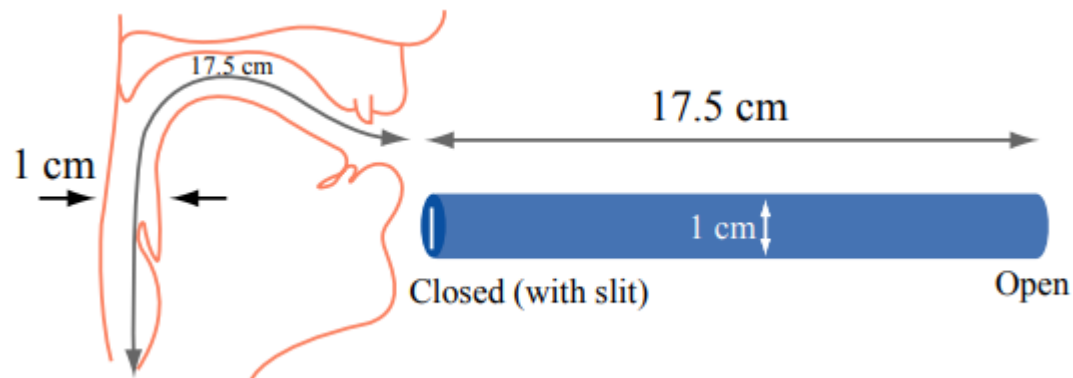


Hamidreza Baradaran Kashani



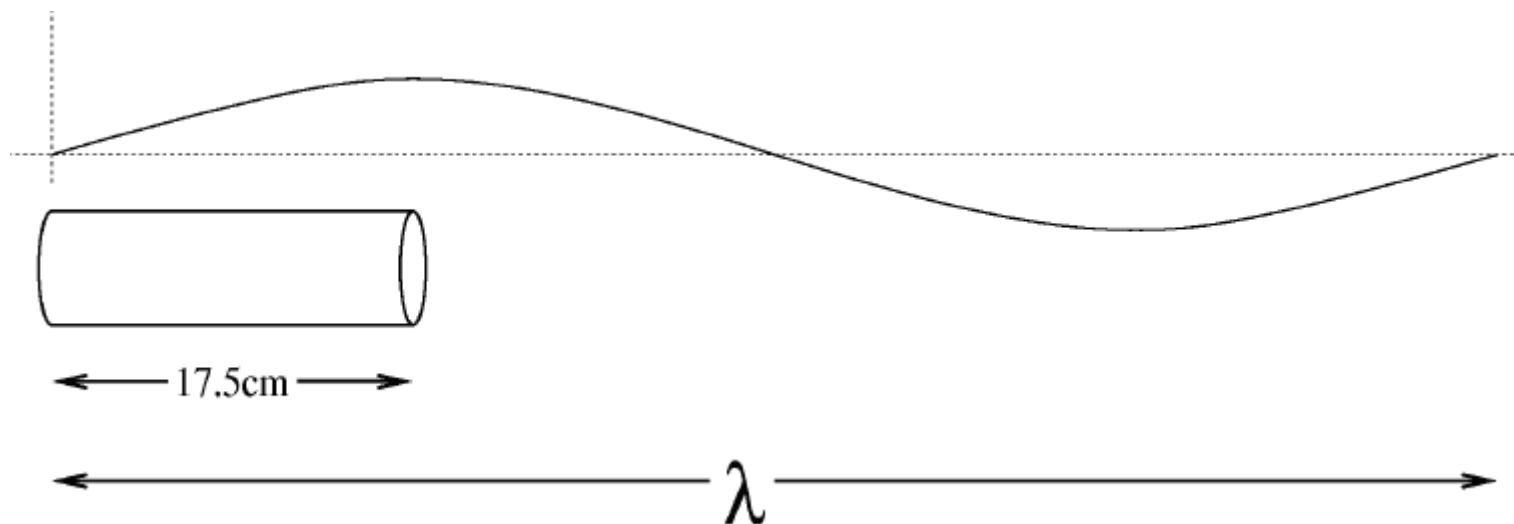
مدلسازی مجرای گفتار بصورت تیوب

❖ برای یک مرد بزرگسال، طول مجرای گفتار بطور طبیعی ۱۷/۵ سانتی متر است.



$$\lambda = 4 \times 17.5 \text{ cm} = 70 \text{ cm} = 0.7 \text{ m}$$

$$f = \frac{c}{\lambda} = \frac{350 \text{ ms}^{-1}}{0.7 \text{ m}} = 500 \text{ Hz}$$



Hamidreza Baradaran Kashani



مدلسازی مجرای گفتار بصورت تیوب

❖ البته تیوب فرکانس های تشدید دیگری هم دارد (غیر از بزرگترین طول موج که برابر با $\lambda_1 = 4L$)

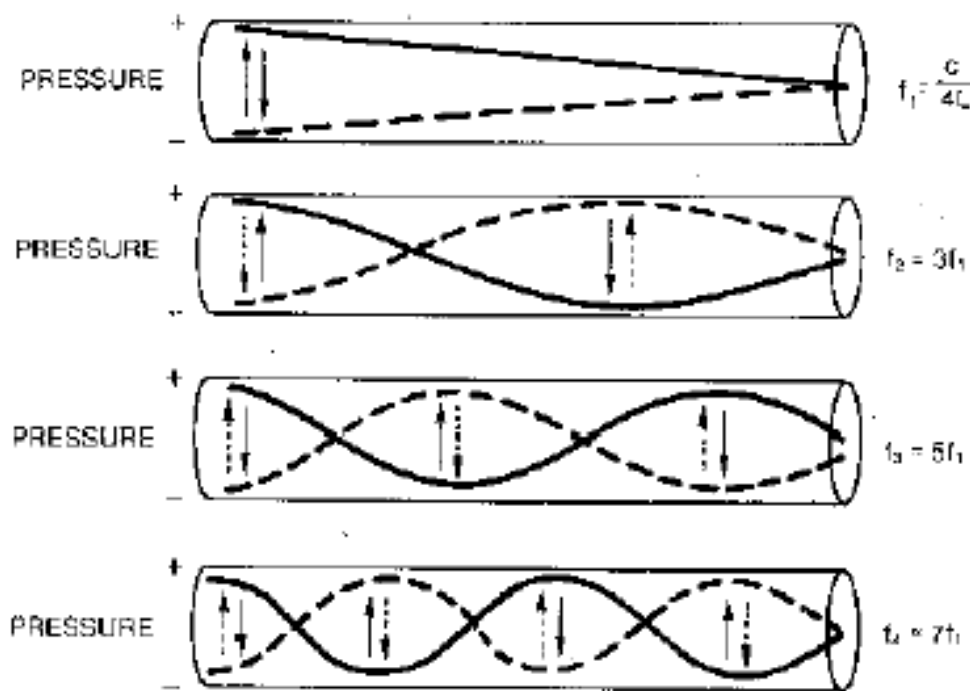
❖ به عبارتی می توان گفت که امواج با طول موج های دیگری هم داخل این تیوب ایجاد می شوند، از جمله:

$$\lambda_1 = 4L \rightarrow f_1 = \frac{c}{\lambda_1} = 500 \text{ Hz}$$

$$\lambda_2 = \frac{4}{3}L \rightarrow f_2 = \frac{c}{\lambda_2} = 1500 \text{ Hz}$$

$$\lambda_3 = \frac{5}{4}L \rightarrow f_3 = \frac{c}{\lambda_3} = 2500 \text{ Hz}$$

$$\lambda_4 = \frac{7}{4}L \rightarrow f_4 = \frac{c}{\lambda_4} = 3500 \text{ Hz}$$



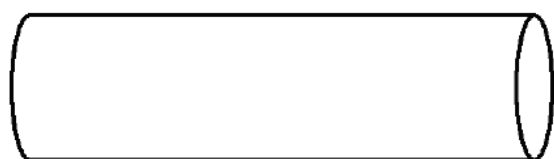
$$f_n = \frac{c}{4L} (2n-1),$$
$$n = 1, 2, 3, \dots$$

Hamidreza Baradaran Kashani

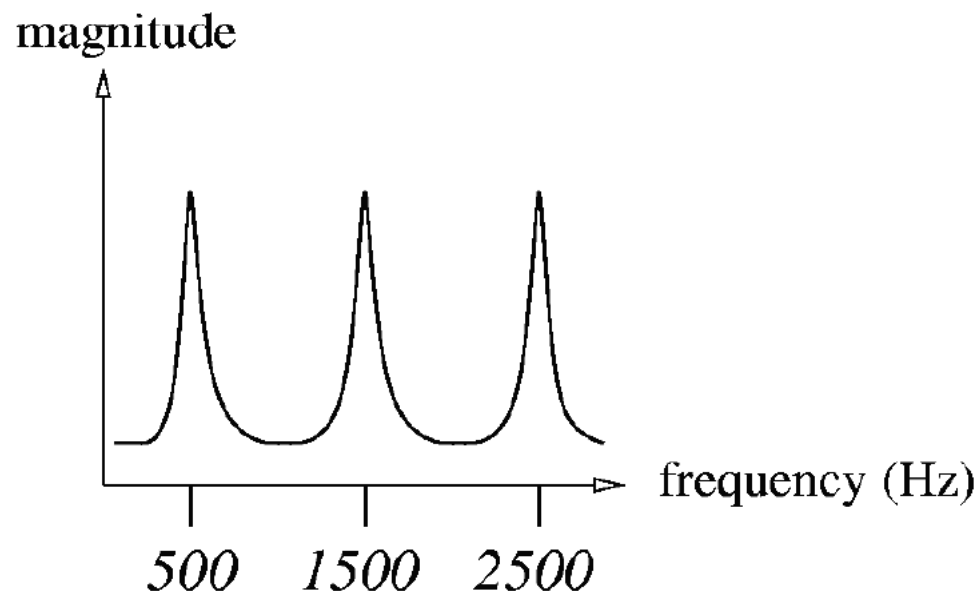


پاسخ فرکانسی مدل تیوب ساده

❖ پاسخ فرکانسی مدل تیوب ساده معروف به پاسخ فرکانسی واکه خنثی (schwa) در زبان انگلیسی است.



17.5cm



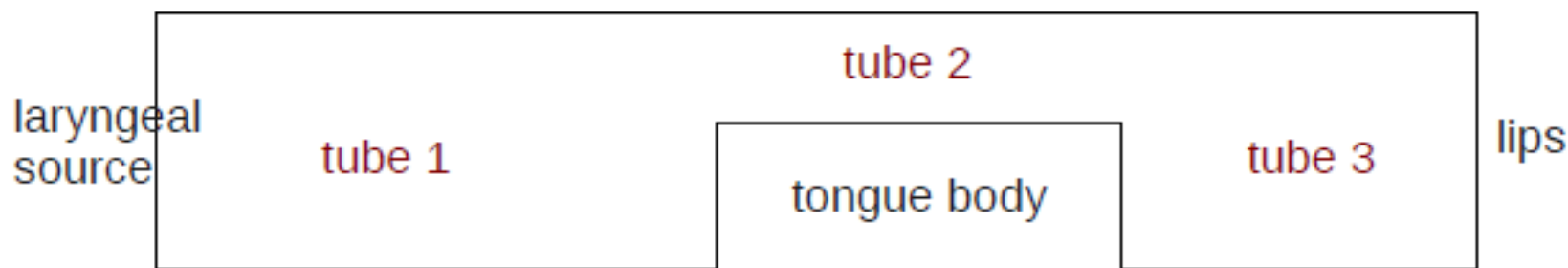
❖ در واقع می توان گفت که هارمونیک های منبع تحریک که فرکانس های نزدیک به فرکانس های فرمنت دارند با بیشترین دامنه خارج شده و سایر هارمونیک تضعیف (damp) می شوند.

Hamidreza Baradaran Kashani



مدل های مجرای گفتار پیچیده تر

- ❖ در بیشتر مواقع مجرای گفتار را نمی توان بصورت یک مدل تیوب ساده در نظر گرفت.
- ❖ با حرکت لبها و بخصوص **زبان**، شکل مجرای گفتار تغییر پیدا کرده و بدنبال آن فرکانس های تشدید مجرای گفتار نیز تغییر می کند.
- ❖ مثلاً مدل ۳ تیوبی که در آن اثر زبان نیز مدل می شود:



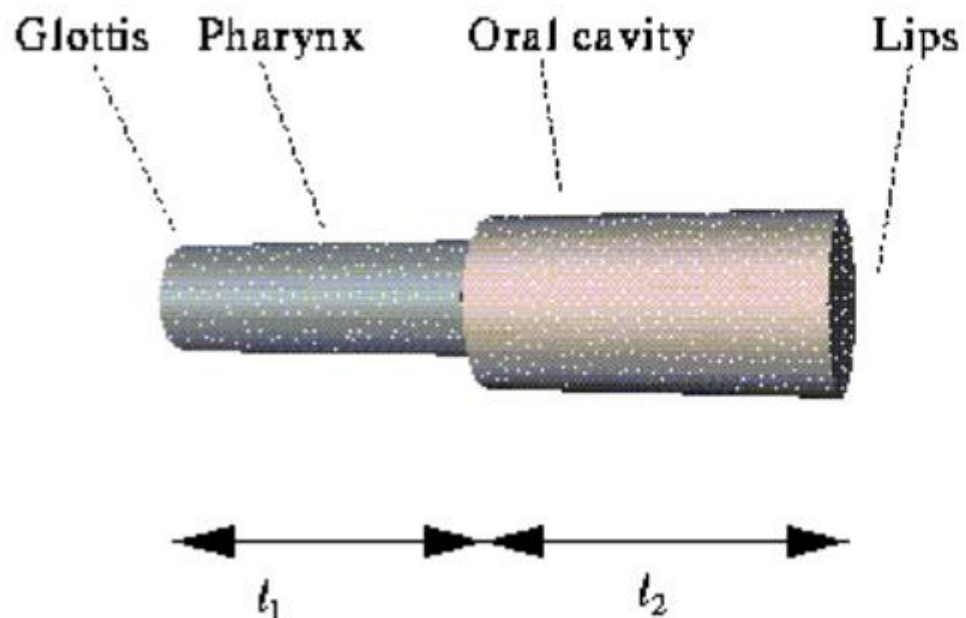
- ❖ فرکانس فرمتهای به طول تیوب ها و سطح مقطع آنها بستگی دارد.

Hamidreza Baradaran Kashani

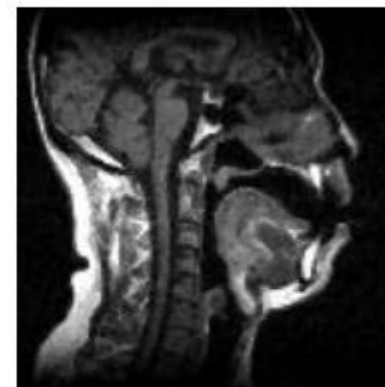


مدل های مجرای گفتار پیچیده تر

Approximation to the vowel [ɑ]:



Vocal tract profile of natural (Japanese) [ɑ]

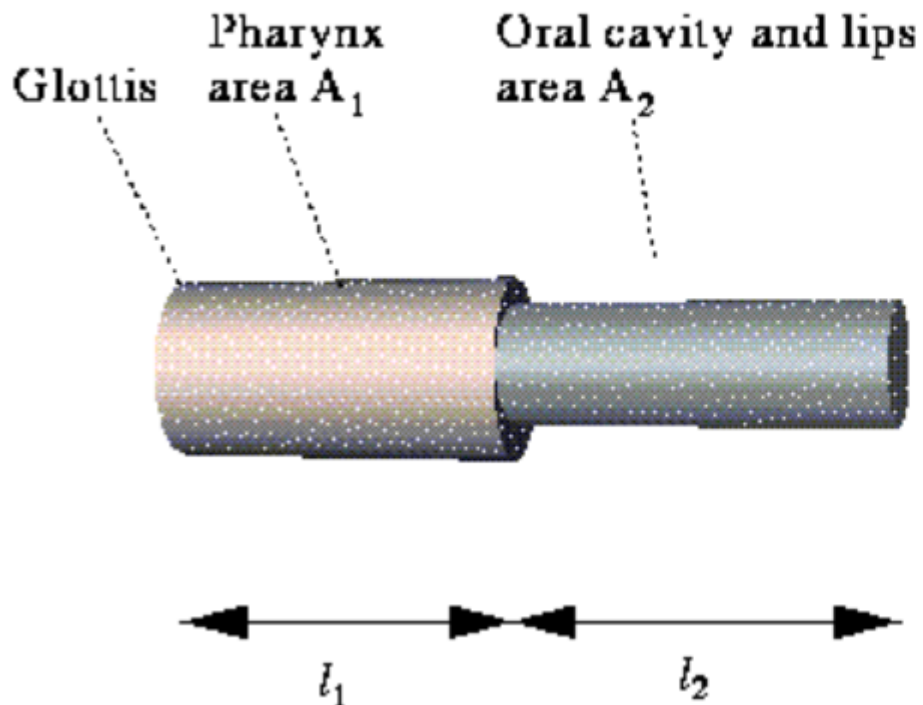


Hamidreza Baradaran Kashani



مدل های مجرای گفتار پیچیده تر

Approximation for vowel [i]:



Vocal tract profile of (Japanese) [i]

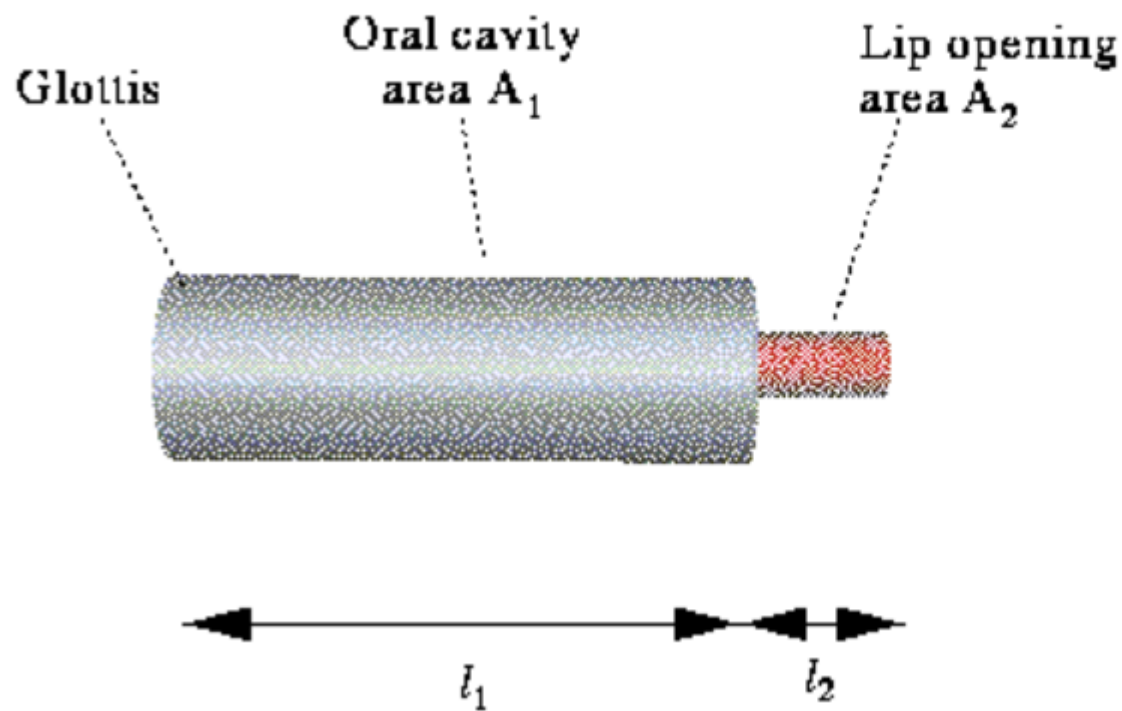


Hamidreza Baradaran Kashani



مدل های مجرای گفتار پیچیده تر

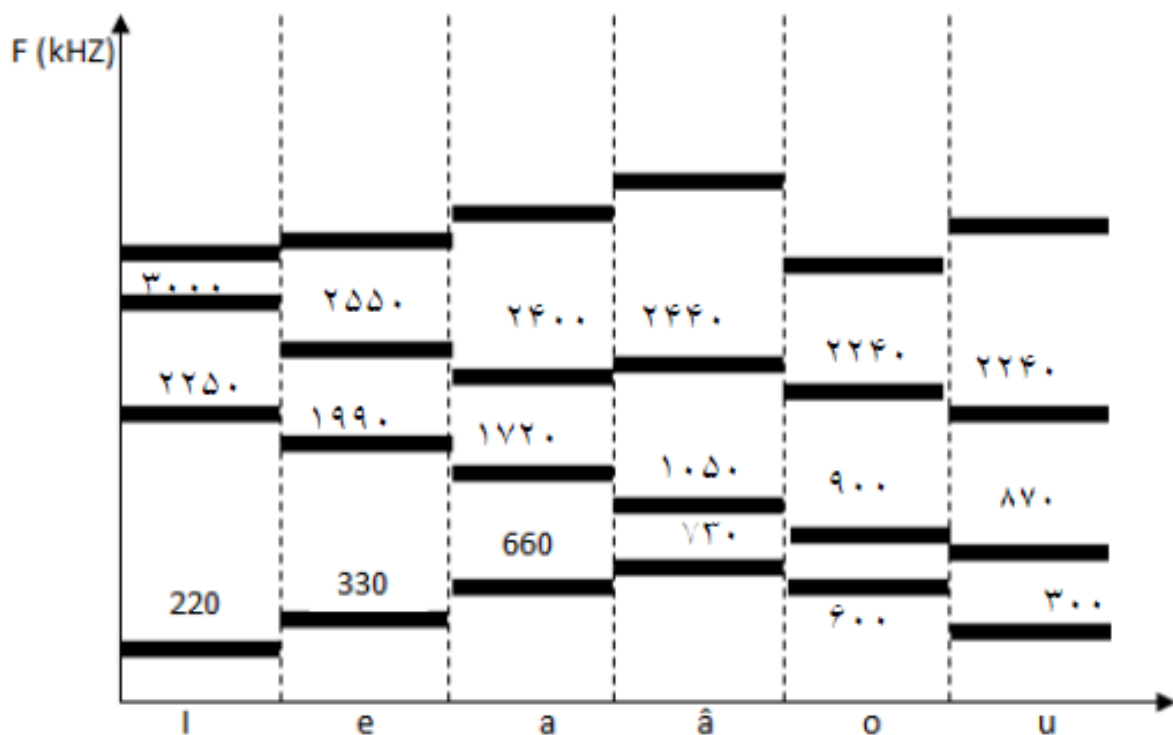
Uniform vocal-tract shape with lip-rounding (a very rough approximation for vowel [u])



Hamidreza Baradaran Kashani



مقادیر تقریبی فرمنت ها برای واکه های فارسی



❖ فرمنت اول

❖ معرف میزان باز یا بسته بودن دهان است

❖ فرمنت دوم

❖ متناظر با عقب یا جلو بودن زبان در حین ادای واکه است.

❖ هر چه زبان جلوتر باشد مقدار فرمنت دوم واکه بیشتر است.

❖ فرمنت سوم

❖ متناسب با گردی دهان است. هر چه دهان گردتر مقدار

فرمنت سوم کوچکتر است.

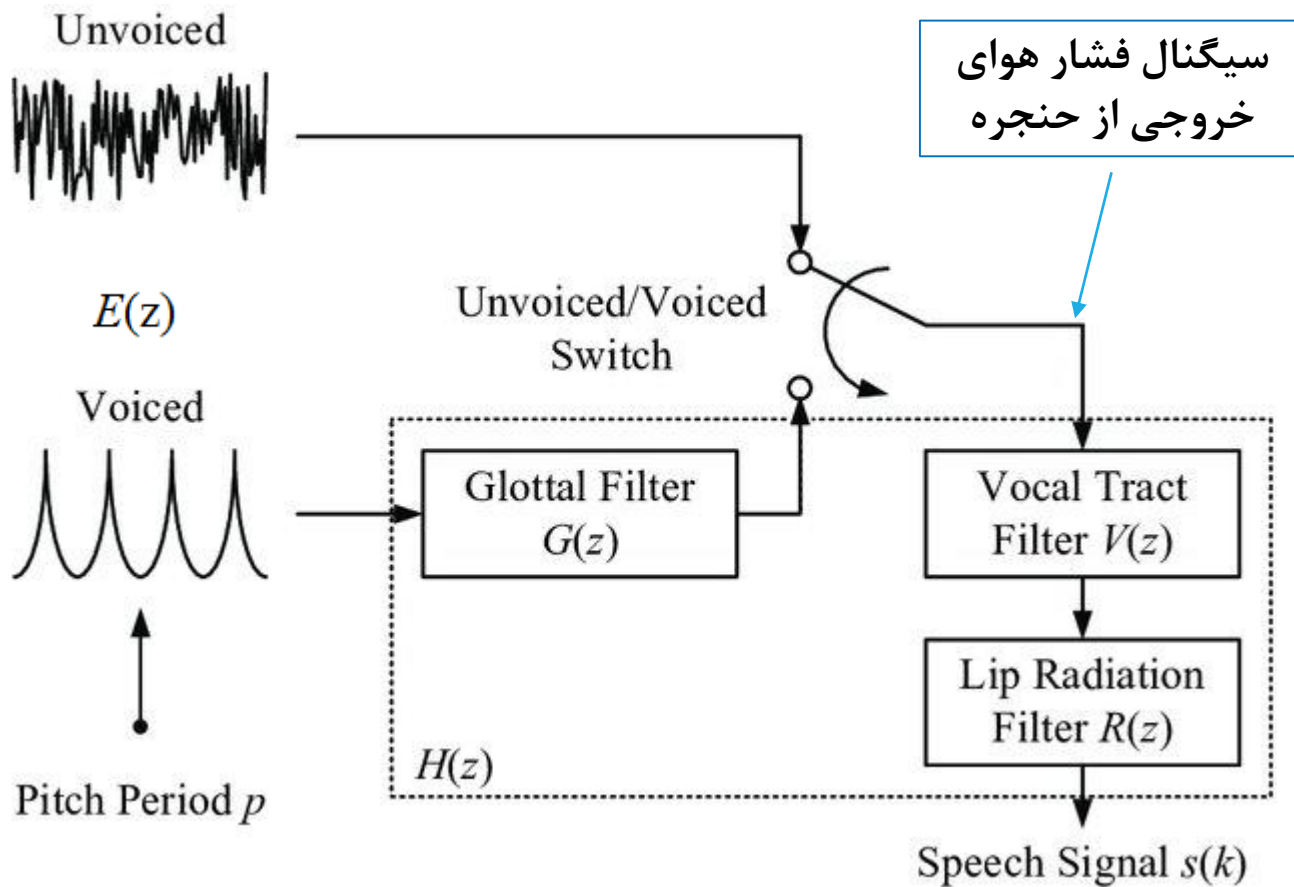
$$F_1, F_2, F_3, F_4 \propto \frac{1}{L}$$

❖ فرکانس های تشدید با عکس طول لوله صوتی شخص متناسب است

Hamidreza Baradaran Kashani



مدل منبع - فیلتر



❖ هدف:

❖ تجزیه سیگنال گفتار به دو بخش:

❖ منبع: جریان هوا در تارهای صوتی

❖ فیلتر: مجرای گفتار

❖ در حالت بیواک:

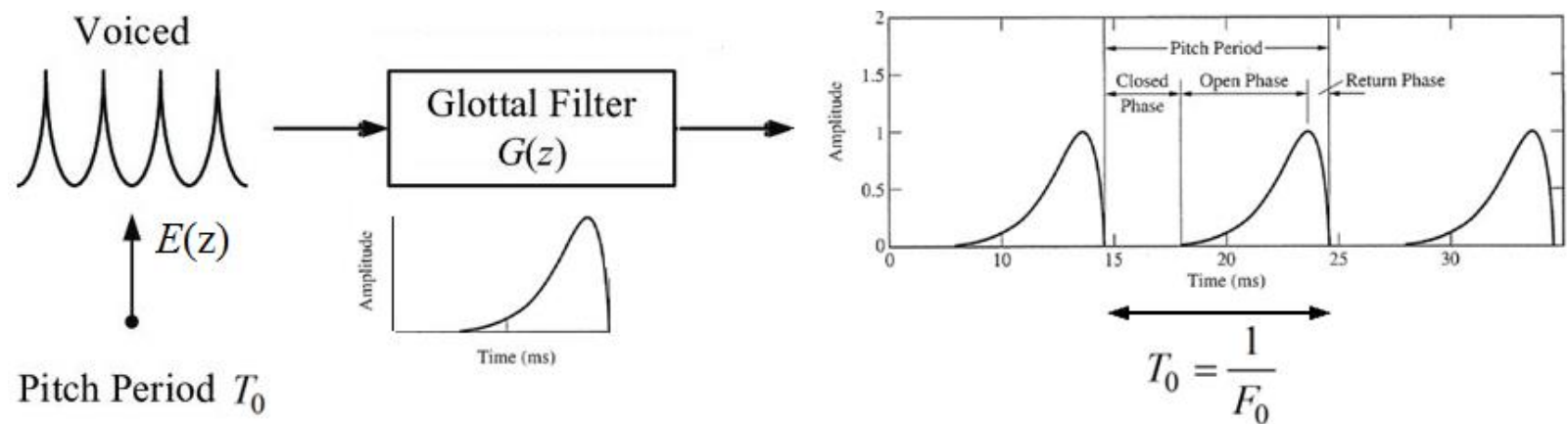
❖ تارهای صوتی نوسان ندارند و یک سیگنال نویزی داریم.

Hamidreza Baradaran Kashani



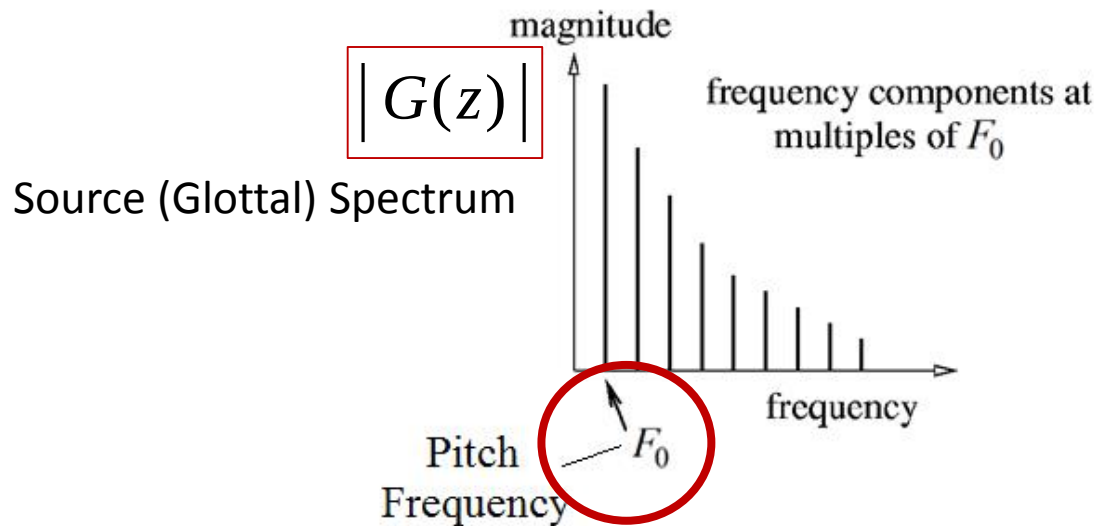
مدل منبع - فیلتر

❖ تولید سیگنال تحریر
برای گفتار واکدار



❖ در حالت واکدار، پالس های کنترل حنجره از طرف مغز تولید شده و وارد تابع تبدیل حنجره می شوند.

❖ مهمترین پارامتر در منبع: F_0

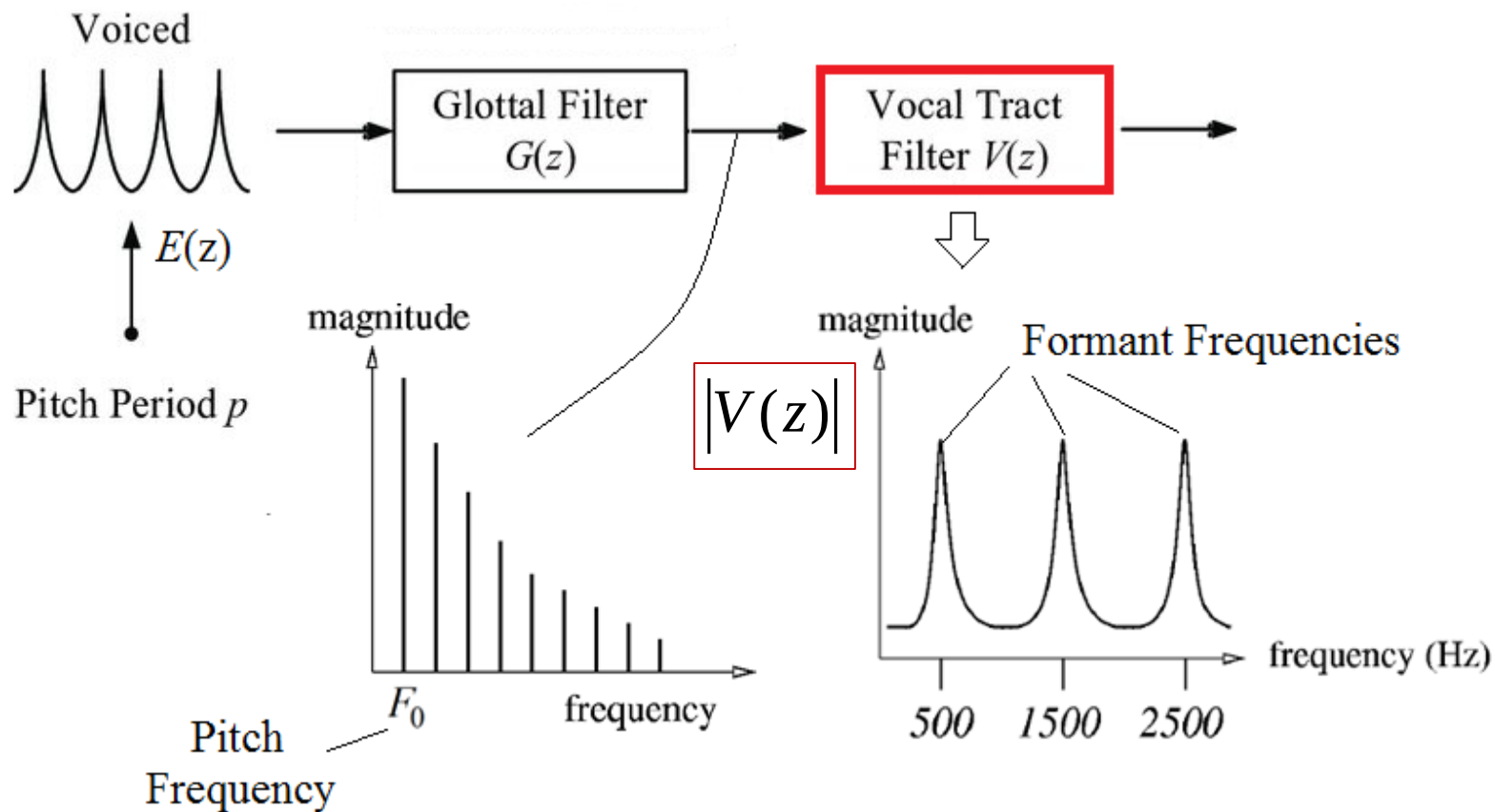


Hamidreza Baradaran Kashani



مدل منبع - فیلتر

❖ فیلتر مجرای گفتار



❖ یک مدل تمام قطب

❖ مهمترین پارامتر مجرای گفتار:
فرکانس های فرمنت

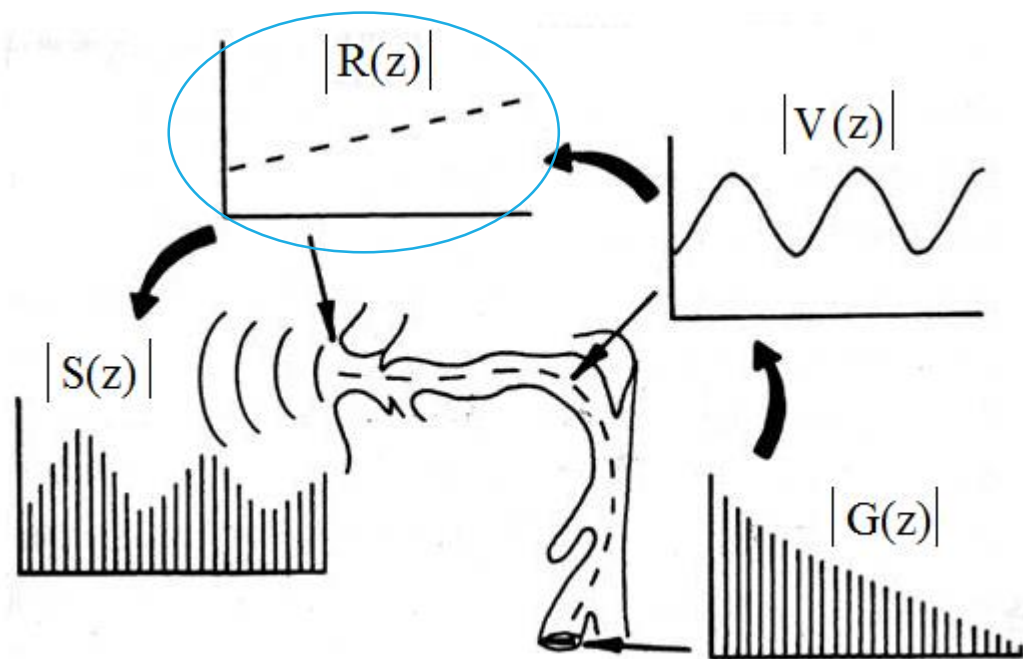
❖ فیلتر شکل روبرو: متناظر با واژه
schwa



مدل منبع - فیلتر

❖ تابع تشعشع لب ها $(R(z))$

❖ به دلیل تفاوت دو محیط، موج فیلتر شده در گذر از مجرای صوتی به هوای بیرون، رفتار مرزی از خود نشان داده و تشعشع (radiation) در لب ها رخ می دهد.

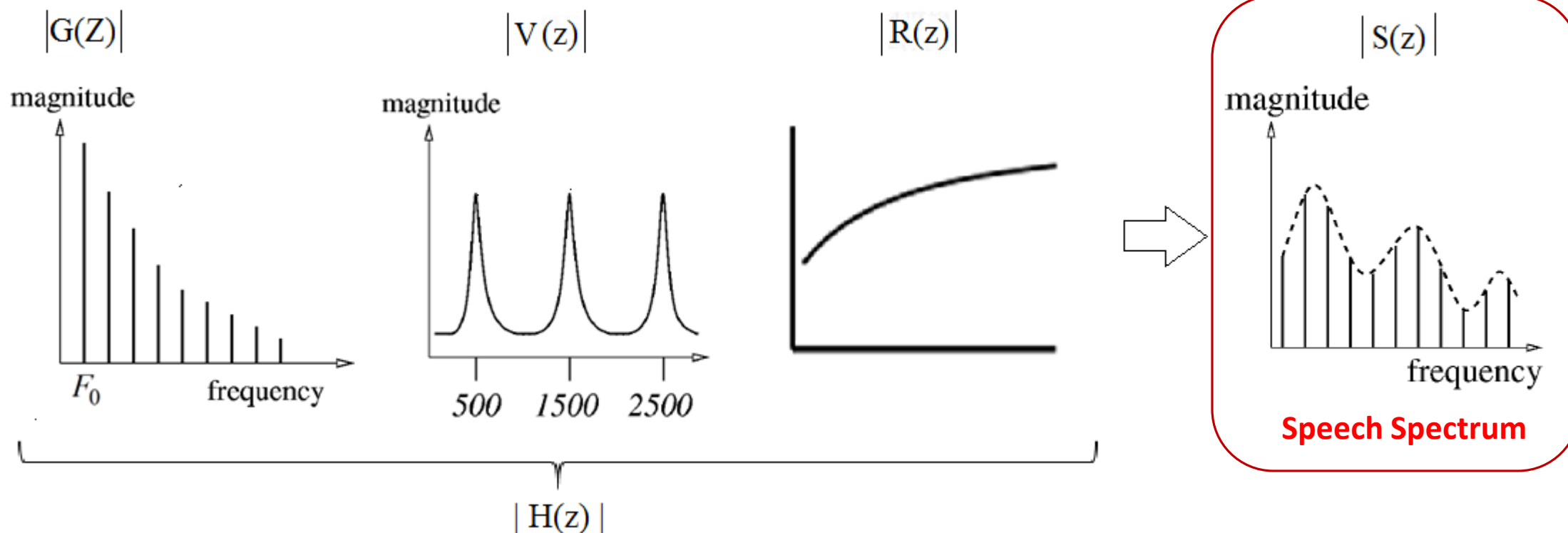


$$R(z) = 1 - \alpha z^{-1}$$

Hamidreza Baradaran Kashani



مدل منبع - فیلتر



Hamidreza Baradaran Kashani



با تشکر از اساتید و همکاران گرامی:
آقای دکتر همایونیپور
آقای دکتر کبودیان
آقای دکتر ویسی
خانم دکتر علی نژاد