

Lecture 16: Obstruents

Mark Hasegawa-Johnson

ECE 537: Speech Processing

- 1 Types of unvoiced excitation
- 2 Pole-zero cancellation of back-cavity resonances
- 3 Front cavity is a quarter-wave resonator
- 4 Voiced obstruents
- 5 Summary

Outline

- 1 Types of unvoiced excitation
- 2 Pole-zero cancellation of back-cavity resonances
- 3 Front cavity is a quarter-wave resonator
- 4 Voiced obstruents
- 5 Summary

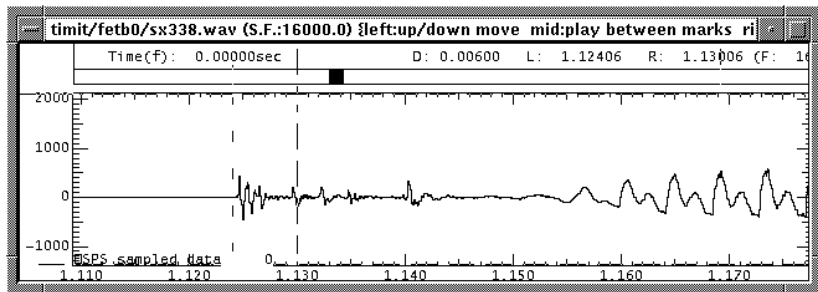
Types of unvoiced excitation

- Transient (Voiced, Unvoiced, Ejective)
- Frication
- Aspiration
- Clicks and Implosives

Sequence of events prior to release of a plosive (stop) consonant

- 1 The vocal tract is closed at a supraglottal constriction (e.g., lips, tongue tip, tongue body), but the glottis remains open
- 2 Air pressure builds up behind the constriction
- 3 The glottis is either opened fully to prevent voicing (unvoiced), closed fully to prevent voicing and aspiration (Amharic ejective stops), or fine-tuned to continue voicing as long as possible (voiced)

Sequence of events after release of a plosive (stop) consonant



- 1 Release of air pressure causes a **transient**
- 2 $\sim 10\text{ms}$ of turbulence at the constriction = **frication**
- 3 $> 25\text{ms}$ of turbulence at the glottis = **aspiration**
- 4 Periodic vocal fold vibration = **voicing**

Transient

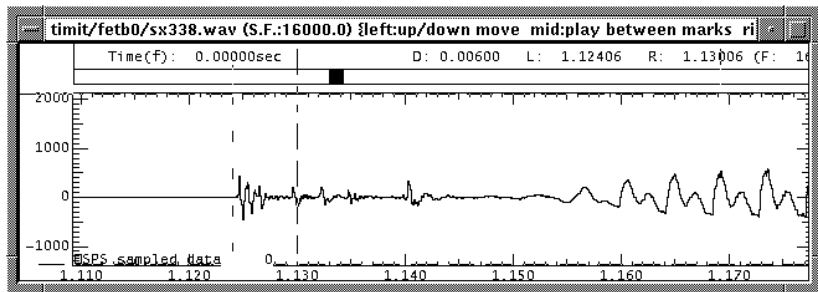
- As soon as the constriction opens, positive air pressure causes a sudden rush of air through the constriction:

$$u(t) \approx \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$$

- The pressure wave is the derivative of volume velocity:

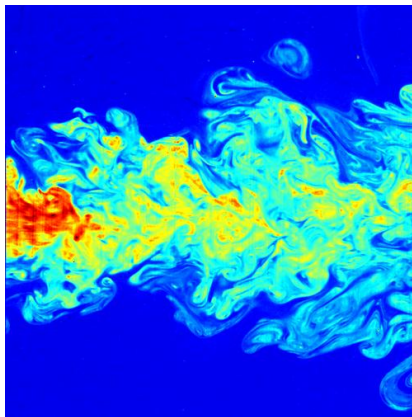
$$p(t) \propto \frac{du}{dt} = \delta(t)$$

Multiple transients



As shown in this example, we sometimes see evidence that the constriction flaps open and closed a few times before opening fully.

- Turbulence is generated when a jet of air is so narrow that viscous forces break it apart
- When turbulence is generated at the glottis, it is called **aspiration**
- When turbulence is generated at a supraglottal constriction, it is called **frication**



https://commons.wikimedia.org/wiki/File:False_color_image_of_the_far_field_of_a_submerged_turbulent_jet.jpg

- Turbulent vortices have uniformly distributed radii
- Spectrum of generated noise is therefore nearly white

White noise

- White noise has equal energy at all frequencies:

$$E [|V(\omega)|^2] = 1$$

- Equivalently, the sample values are uncorrelated:

$$E [v[n]] = 0$$

$$E [v[n]v[m]] = \begin{cases} 1 & n = m \\ 0 & \text{otherwise} \end{cases}$$

Strident versus Nonstrident Fricatives

- Strident fricatives (e.g., /s,ʃ/) direct the turbulent jet against the teeth, generating a loud noise source
- Nonstrident fricatives (e.g., /f,θ/) have no downstream obstacle, so quieter noise is generated only by the clash of different airstreams

Implosives and Clicks

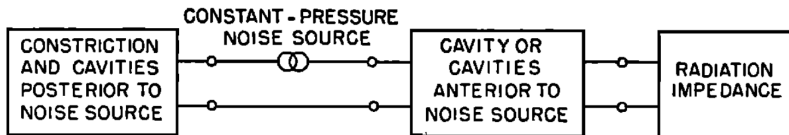
Implosives and clicks are caused by closing the vocal tract in two places, expanding the vocal tract so that pressure drops below atmospheric pressure, then opening the anterior constriction so that air rushes in.

- Implosives (as in Vietnamese /ɓ,d/) have the posterior constriction at the glottis
- Clicks (as in Zulu, e.g., <https://audiomack.com/ladysmith-black-mambazo/song/homeless-1>) have a supraglottal posterior constriction, usually the tongue body.

Outline

- 1 Types of unvoiced excitation
- 2 Pole-zero cancellation of back-cavity resonances
- 3 Front cavity is a quarter-wave resonator
- 4 Voiced obstruents
- 5 Summary

Supraglottal pressure source



Clashing airstreams cause random pressure fluctuations, which can be modeled as a supraglottal sound pressure source.

Pole-zero cancellation

The supraglottal source, $P_S(\omega)$, generates airflow equal to

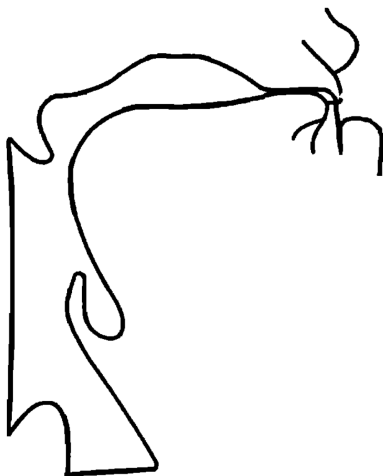
$$U_S(\omega) = \frac{P_S(\omega)}{Z_B(\omega)}$$

where $Z_B(\omega)$ is the impedance of the back cavity. Volume velocity at the lips is

$$U_L(\omega) = H(\omega)U_S(\omega) = \frac{H(\omega)P_S(\omega)}{Z_B(\omega)}$$

and therefore $U_L(\omega) = 0$ at frequencies where $Z_B(\omega) = \infty$.

- Formants are resonances of the vocal tract, from larynx to lips.
- The constriction nearly decouples back and front cavities.
- Back-cavity formants are resonances of a tube closed at the constriction, i.e., roughly the same frequencies as $Z_B(\omega) = \infty$.



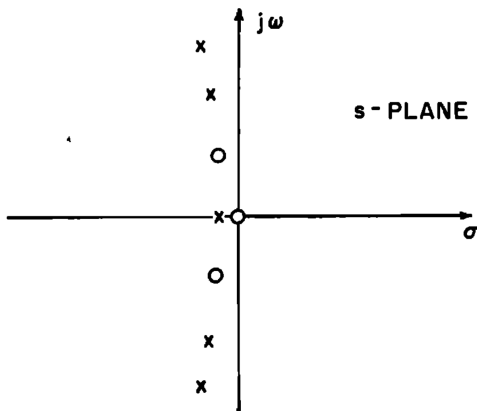
Pole-zero cancellation of back cavity resonances

The overall transfer function is:

$$\begin{aligned}\frac{U_L(z)}{P_S(z)} &= \frac{H(z)}{Z_B(z)} \\ &= \left(\prod_i \frac{z^2}{(z - p_i)(z - p_i^*)} \right) \left(\prod_j \frac{(z - z_j)(z - z_j^*)}{z^2} \right) \\ &= \prod_{i: p_i \neq z_j} \frac{z^2}{(z - p_i)(z - p_i^*)}\end{aligned}$$

- p_i include both front-cavity and back-cavity formants
- z_j are poles of $Z_B(\omega)$, which roughly cancel out the back-cavity formants

Pole-zero cancellation of the back cavity resonances



Heinz and Stevens (1961) Fig. 3(b), showing zeros that nearly cancel the back-cavity formant but not the front-cavity formant.

A spectrum with two poles and one zero tends to be flat between the frequencies of the two poles:

FIG. 4. Comparison of spectra measured by Hughes and Halle⁴ for the fricative /s/ as spoken by one subject in different contexts (dashed curves), and spectra obtained as the output of appropriate electric networks (solid curves).

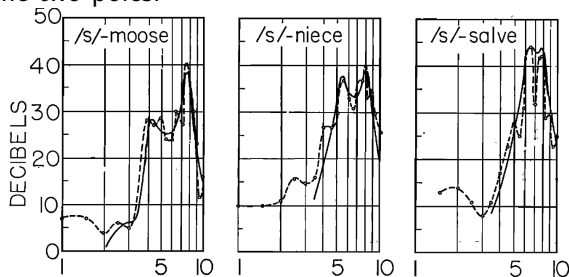
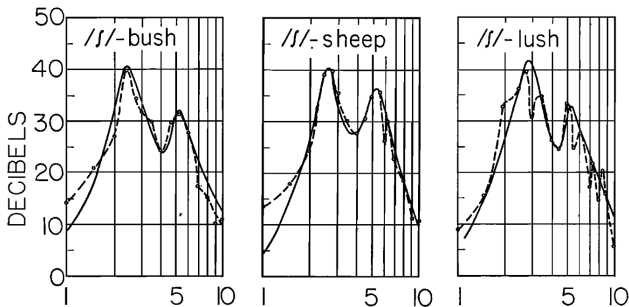


FIG. 5. Same as Fig. 4, for fricative /ʃ/.



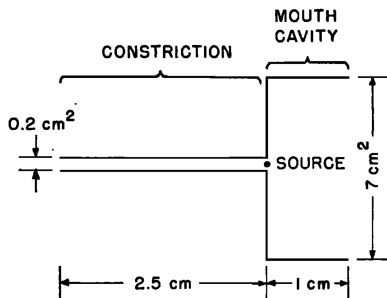
Outline

- 1 Types of unvoiced excitation
- 2 Pole-zero cancellation of back-cavity resonances
- 3 Front cavity is a quarter-wave resonator
- 4 Voiced obstruents
- 5 Summary

The cavity in front of a plosive or fricative constriction is much wider than the constriction, so it can be modeled as a quarter-wave resonator, with front-cavity resonances at

$$r_k = \frac{c}{4l_f} + (k-1)\frac{c}{2l_f}$$

where l_f is the length of the front-cavity constriction.



Front cavity is a quarter-wave resonator

Here are some rough estimates of front-cavity length, based on $c = 354\text{m/s}$ and $r_1 = c/4l_f$:

Phonemes	l_f (cm)	r_1 (Hz)
/f/	0.5	17700
/θ/	1.0	8850
/s,t/	2.0	4425
/ʃ/	4.5	2000

Perceptual identification as a function of r_1

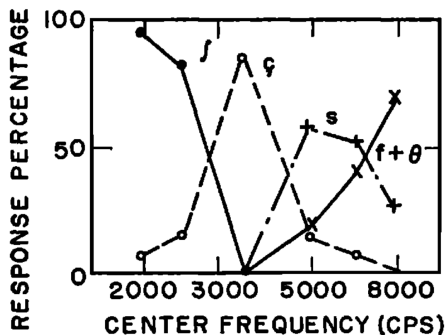
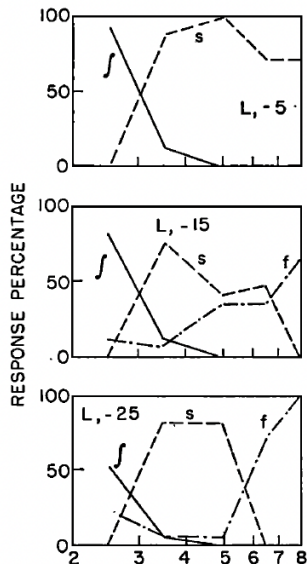


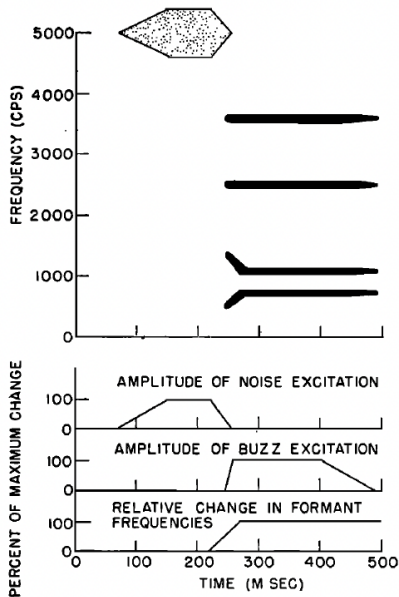
FIG. 8. Results of identification test for isolated synthetic fricative consonants. The abscissa represents the resonant frequency of the electric circuit used to generate the stimuli. Responses of /f/ and /θ/ are combined.

Loud fricatives are harder to perceive as /f/. The figure shows classification probability as a function of resonant frequency when fricative level is -5dB relative to vowel (top), -15dB (middle), and -25dB (bottom).

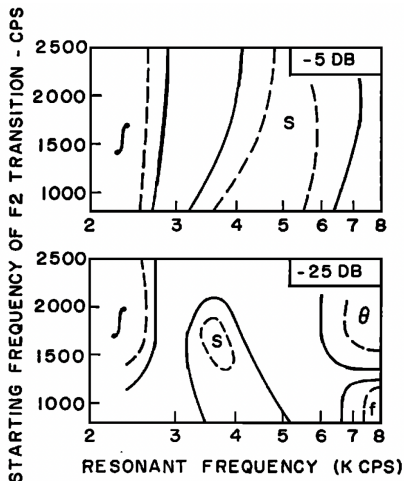


Formant transitions distinguish /f/ from /θ/.

Tongue tip consonants (/θ,s,d,f/) transition into the vowel from an /i/-like tongue shape, with $F_2 \approx 2000\text{Hz}$. Labial consonants (/b,f/) transition into the vowel from an /u/-like tongue shape, with $F_2 \approx 1000\text{Hz}$.



Combining all three cues: At high amplitudes (top) or for $r_1 < 5000\text{Hz}$, only /f,s/ are heard. At low amplitudes (bottom), the distinction between /f/ and /θ/ depends on the formant transition into the vowel.



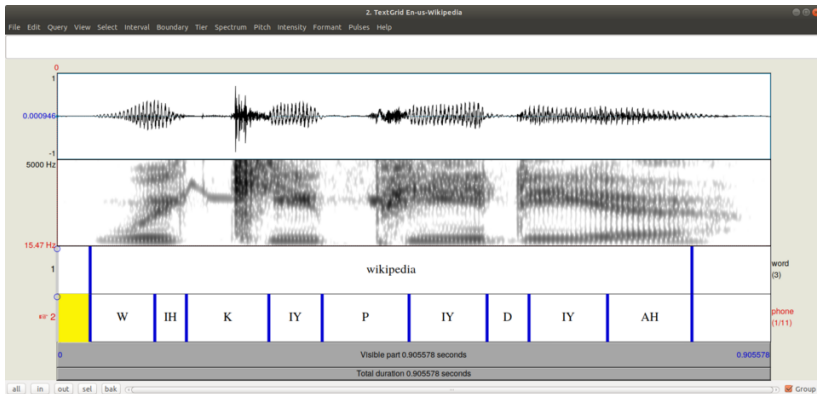
Outline

- 1 Types of unvoiced excitation
- 2 Pole-zero cancellation of back-cavity resonances
- 3 Front cavity is a quarter-wave resonator
- 4 Voiced obstruents**
- 5 Summary

Voiced Obstruents

- Voicing requires that subglottal pressure $>$ supraglottal pressure.
- To keep supraglottal pressure low, the obstruent (stop or fricative) is often shorter.
- Because supraglottal pressure is low, the transient and frication may be quieter.
- In English, unvoiced stops are also aspirated (/h/ produced after the stop release).

Unvoiced vs. Voiced Obstruents example: “Wikipedia”



Notice /d/ is shorter than /k/ or /p/ and has no aspiration after release. Nevertheless, voicing only lasts <halfway through the closure. [https://commons.wikimedia.org/wiki/File:](https://commons.wikimedia.org/wiki/File:Waveform_spectrogram_and_transcription_of_wikipedia_in_praat.png)

Waveform_spectrogram_and_transcription_of_wikipedia_in_praat.png

Outline

- 1 Types of unvoiced excitation
- 2 Pole-zero cancellation of back-cavity resonances
- 3 Front cavity is a quarter-wave resonator
- 4 Voiced obstruents
- 5 Summary**

Summary

- Transients and frication are pressure sources at the constriction.
- Zeros of the back cavity ($Z_B(\omega) = \infty$) occur at frequencies close to formants of the back cavity, and tend to cancel them.
- Front cavity resonances are at $r_k = \frac{c}{4l_f} + (k-1)\frac{c}{2l_f}$.
- Fricatives louder than -10dB are perceived to be strident (/s,ʃ/).
- /θ/ and /f/ have similar resonant frequencies (above 8kHz) and weak frication, so listeners use the formant transition (/i/-like vs /u/-like) to distinguish them.
- Voicing requires lower supraglottal pressure.