24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

24.910

Laboratory Phonology

The source-filter model of speech production

- Readings for next week:
  - Ladefoged 'Elements of Acoustic Phonetics (2nd edn) ch. 10 from p.160, ch. 11 to p.183.
  - Liljencrants & Lindblom (1972).
    - Note: Mel is an auditory frequency scale like the Bark scale.
- Assignments (due 3/6):
  - Write up analysis of mandarin affricates.
  - Acoustics assignment.

I. Vowel inventories

- Common vowel inventories:

i u	i u	i u
		e o
	e o	ɛ ɔ
a	a	a

Arabic,   Spanish,   Italian,
Nyangumata,   Swahili,   Yoruba,
Aleut, etc.   Cherokee, etc. Tunica, etc.

- Unattested vowel inventories:

i	i ɯ	i u
e	e ɤ	ɪ ʊ
a	a	

Why?

Spectrogram image removed due to copyright restrictions.
See: http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/course/chapter8/8.3.htm

# Source-Filter Model of Speech Production
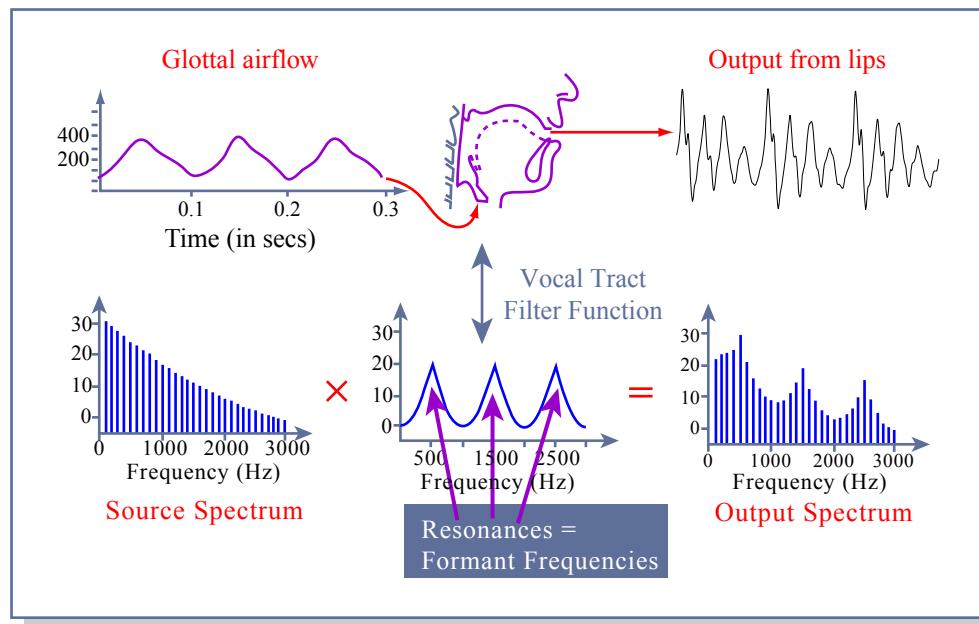
Figure by MIT OpenCourseWare.

Sources

- A source is an input of acoustic energy into the speech production system.

- There are two basic types:
  - Voicing
  - Noise

Voicing source

- Voicing is a periodic source produced by modulation of the airflow from the lungs by the vocal folds.

Epiglottis

False vocal folds

Vocal folds

Trachea

Figure by MIT OpenCourseWare.

(1) (6)

(2) (7)

(3) (8)

(4) (9)

(5) (10)
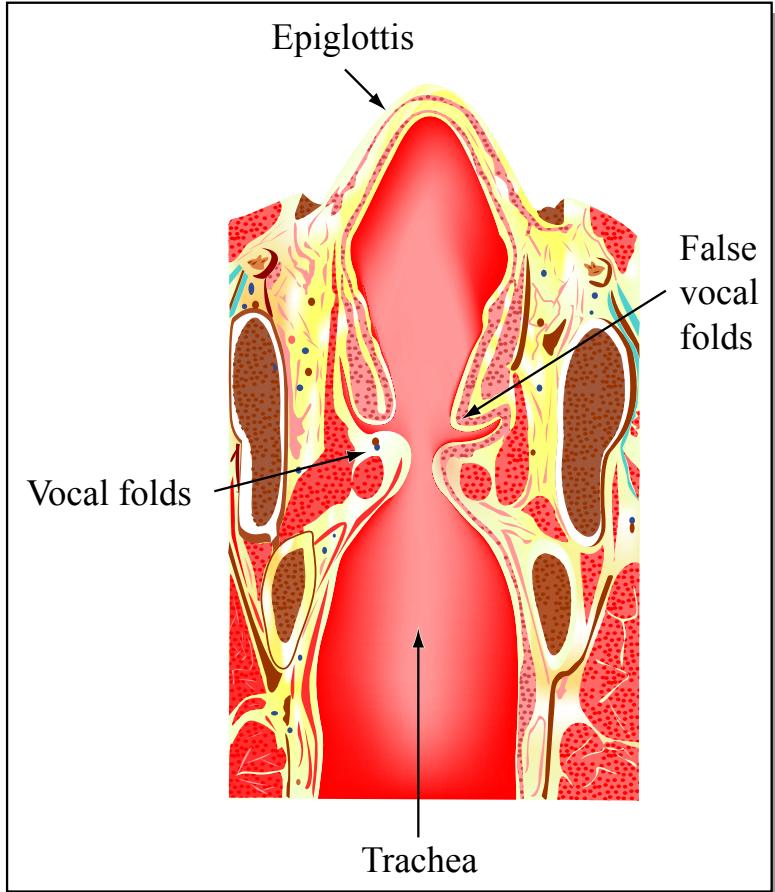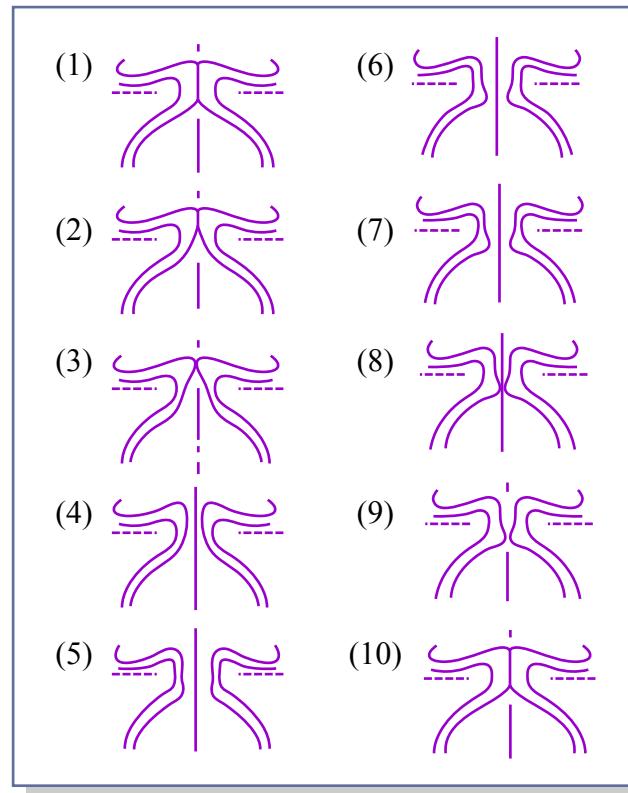
Figure by MIT OpenCourseWare.

Voicing source

Video removed due to copyright restrictions.

Please see
"The vocal tract and larynx"
to view.

Video removed due to copyright restrictions.

Please see
"The vibrating vocal folds"
in Peter Ladefoged's
Vowels and Consonants.

Voicing source

- vocal fold vibration produces a complex periodic wave, whose spectrum contains energy at the fundamental frequency of laryngeal vibration and multiples of the fundamental frequency (harmonics).

SPEAKER P.B.

	FUNDAMENTAL FREQUENCY	SUBGLOTTAL PRESSURE
	107 Hz	8.6 cm H_2O
	97 Hz	8.4 cm H_2O
	94 Hz	8.1 cm H_2O
	89 Hz	7.3 cm H_2O
	83 Hz	6.9 cm H_2O

LITERS SEC

0   msec   10
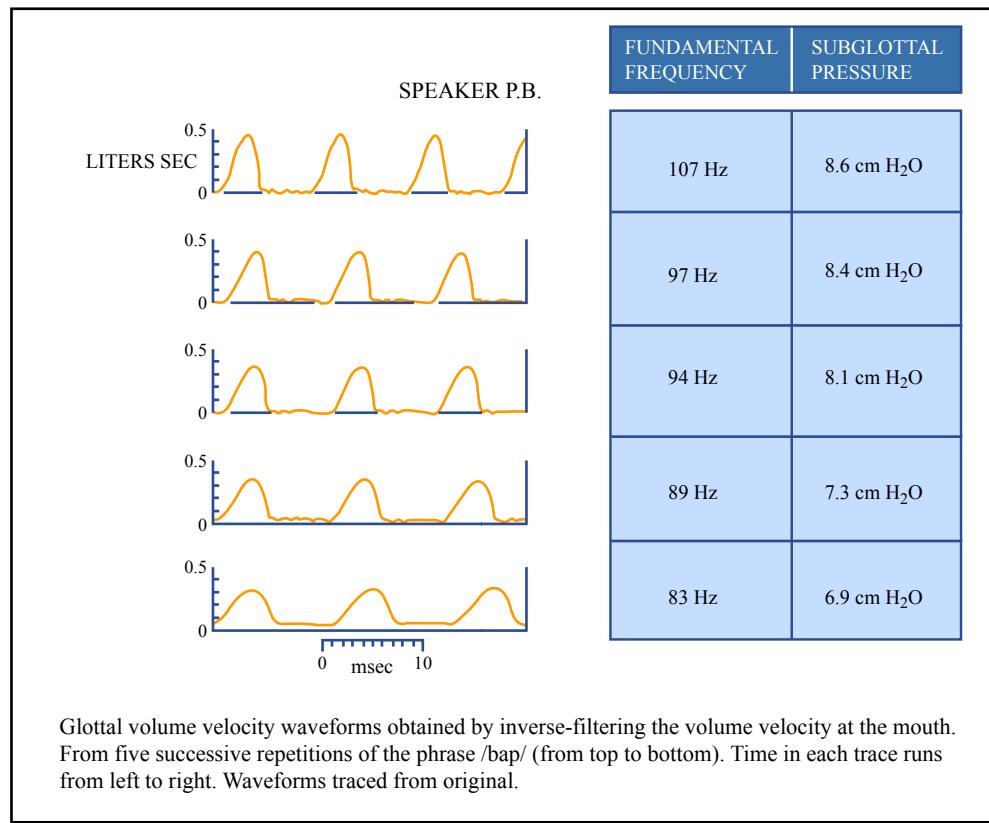
Glottal volume velocity waveforms obtained by inverse-filtering the volume velocity at the mouth. From five successive repetitions of the phrase /bap/ (from top to bottom). Time in each trace runs from left to right. Waveforms traced from original.

Noise source

- Air turbulence generated at an obstruction involves random (aperiodic) pressure fluctuations over a wide range of frequencies.

- Noise generated at the glottis is called aspiration, while noise generated elsewhere is called frication.
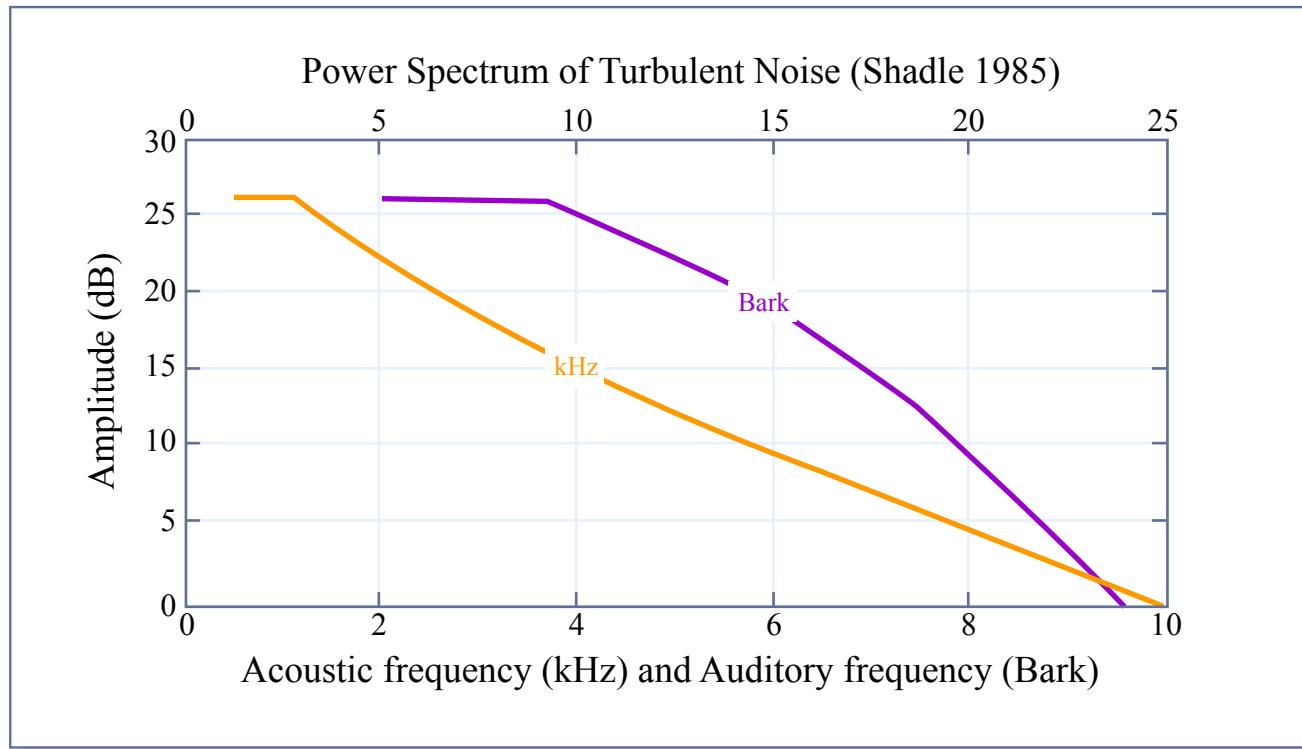

Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

Source

A sound may involve more than one source.

- E.g. a voiced fricative combines voicing and frication noise

- breathy voice can combine voicing and aspiration noise.

- voiceless fricatives can have noise generated at the glottis and at the supralaryngeal constriction.

Filters

- The vocal tract acts as a filter, modifying the source waveform.
- An acoustic filter is a device which passes or amplifies certain frequencies and attenuates others.
- An important characteristic of a filter is its **transfer function** - the ratio of the output to the input depending on frequency.
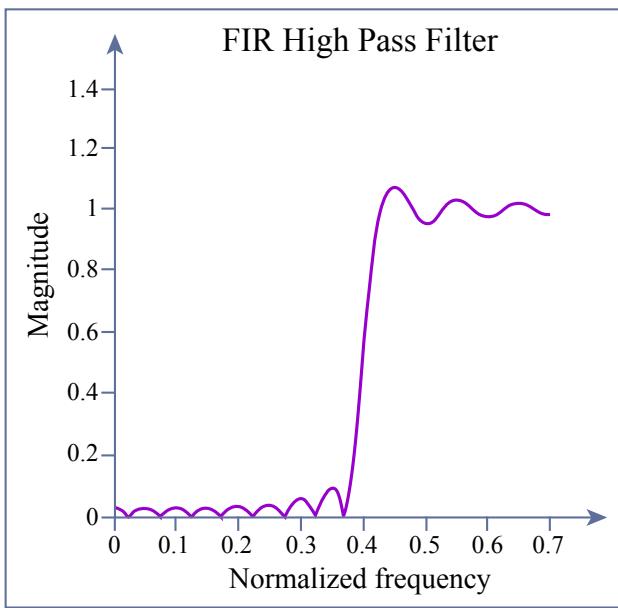

Figure by MIT OpenCourseWare.
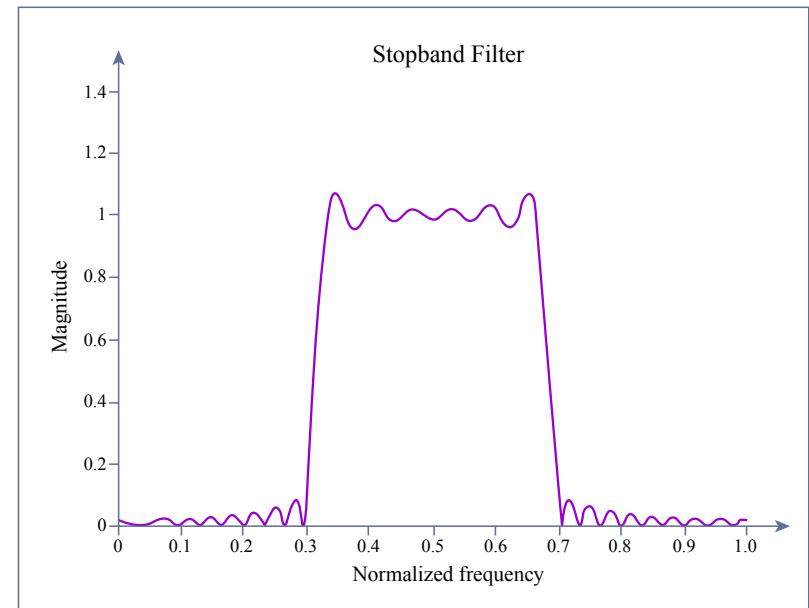
Figure by MIT OpenCourseWare.

The vocal tract as a filter

- The sound wave at some distance from the speaker is the result of filtering the source with the vocal tract filter, plus the radiation characteristics of the lips/nose.

- How does the vocal tract operate as a filter?
- What determines the characteristics of the filter?

Outline
- Resonators are filters.
- The column of air in the vocal tract is a resonator, hence a filter.
- The characteristics of the filter depend on the shape of the vocal tract - we will explore the relationship by considering simple cases.

Resonators

- A body, like a mass of air in a tube, naturally vibrates at one or more frequencies.
- If a pulse of energy is imparted to the air, it will vibrate at these natural frequencies.
- If a source (driving force) which is vibrating at a natural frequency of the body is applied to it (e.g. a tuning fork is held over the tube of air), the body will resonate with the source, i.e. vibrate strongly at the same frequency.
- If a source of a different frequency is applied to the body it will vibrate with less amplitude.

Resonators

- We can plot the response of a resonator to a range of input frequencies. Peaks mark the resonant frequencies.
- Resonators vary in the range of frequencies they will respond strongly to
  - ➢ a sharply tuned resonator only responds strongly to frequencies very close to the resonant frequency (it has a narrow bandwidth).
 - ➢ A resonator with a wide bandwidth will vibrate strongly in response to a wider range of input frequencies.
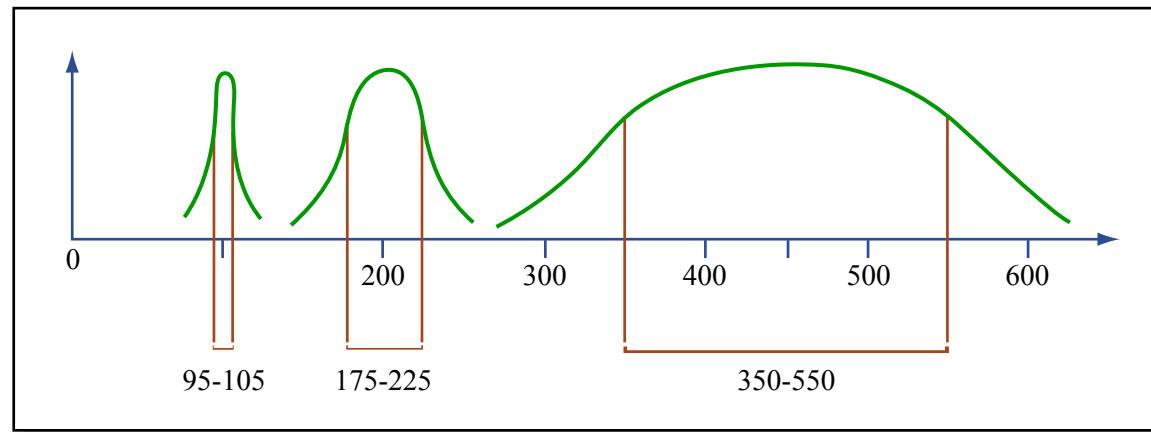
Figure by MIT OpenCourseWare. Adapted from Ladefoged, P. *Elements of Acoustic Phonetics*. 2nd ed. University of Chicago Press, 1996.

Resonators are filters

A resonator acts as an acoustic filter:
- Consider the vibration of a column of air driven by a vibrating source.
- The source is the input to the filter. This input produces vibration in the column of air.
- This vibration produces a sound wave in the external air - the output of the filter.
- The strength with which an input frequency is output from the filter will depend on the resonance characteristics of the air column - if it is near a resonant frequency, it will be passed through to the output at full strength, if it is not near a resonant frequency, it will have low amplitude in the output.
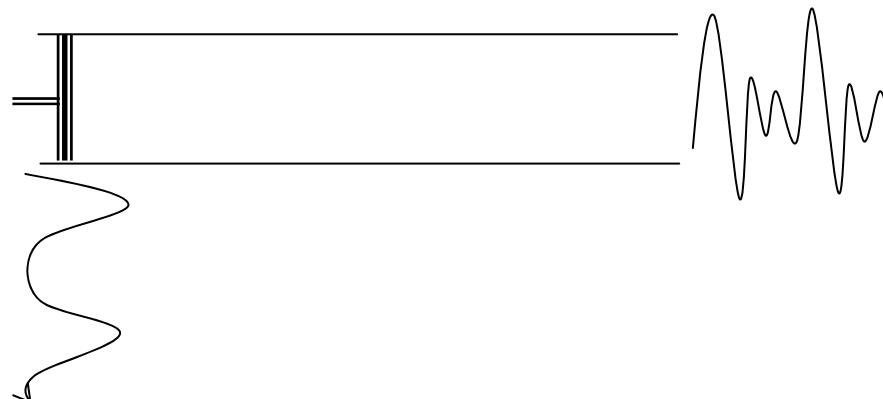
The vocal tract is a filter

- In speech the input has energy at a range of frequencies (voicing or noise, or both) and the resonator has multiple resonant frequencies of varying bandwidths.
- The amplitude in the output wave of each frequency component of the source depends on the extent to which the air in the vocal tract resonates at that frequency. Thus the vocal tract filters the source.
- The resonances of the vocal tract are called formants.

Source Spectrum

Resonances =
Formant Frequencies

Output Spectrum
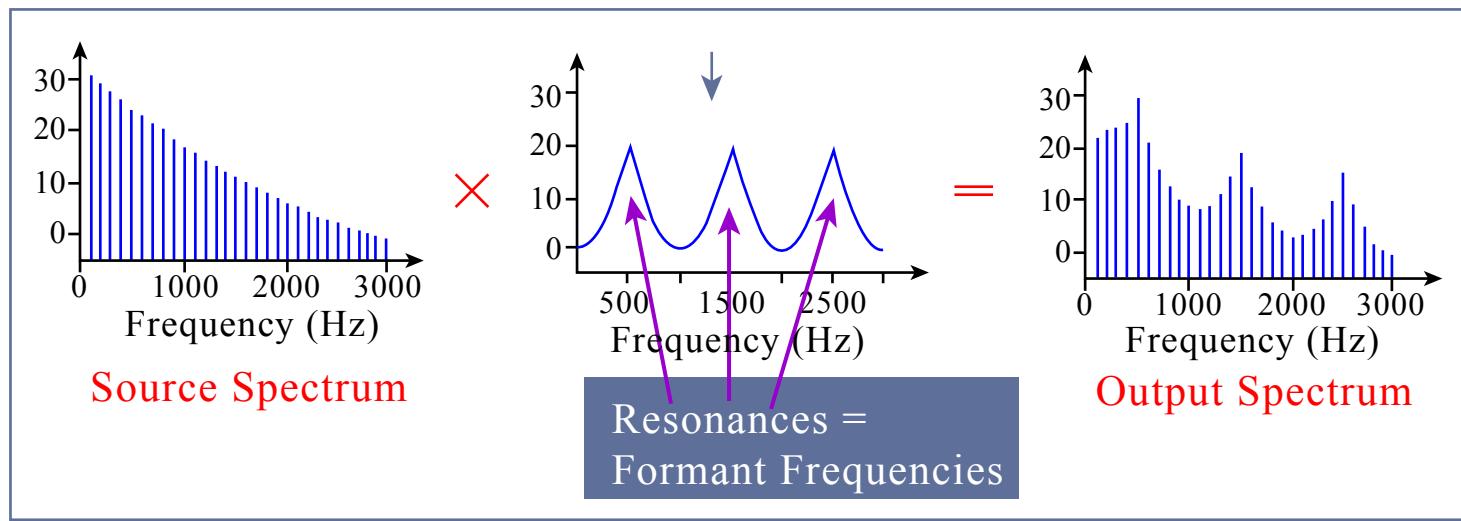
Image by MIT OpenCourseWare.

Resonant frequencies of air in a tube

What determines the resonant frequencies of a mass of air in a tube?

- The resonant frequencies (formants) of air in the vocal tract depend on how the cross-sectional area of the vocal tract varies over its length.
- The relationship between vocal tract shape and transfer function is complex - we will consider the simple case of a uniform tube.

How air in a uniform tube vibrates

- A compression or rarefaction travelling down a tube is reflected off the ends of the tube.
- There is reflection even at the open end of a tube, but with a change of phase - a compression is reflected as a rarefaction, and vice versa.
- The vocal tract in a vowel can be approximated by a tube which is closed at one end (the glottis) and open at the other (the lips).
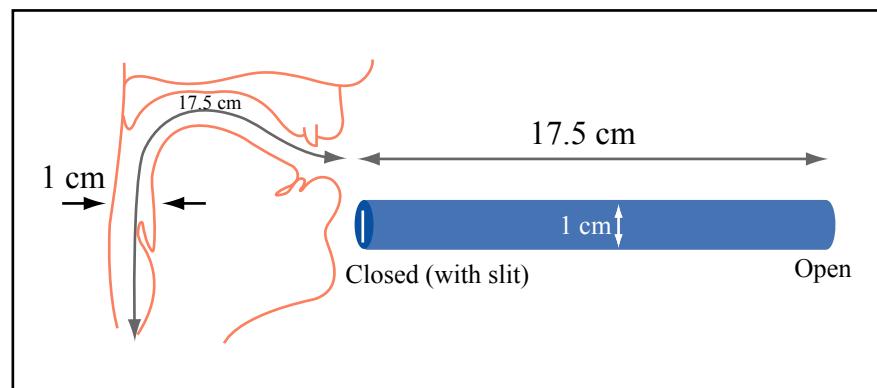
Image by MIT OpenCourseWare. Adapted from Ladefoged, P. *Elements of Acoustic Phonetics*. 2nd ed. Chicago, IL: University of Chicago Press, 1996.

How air in a tube vibrates
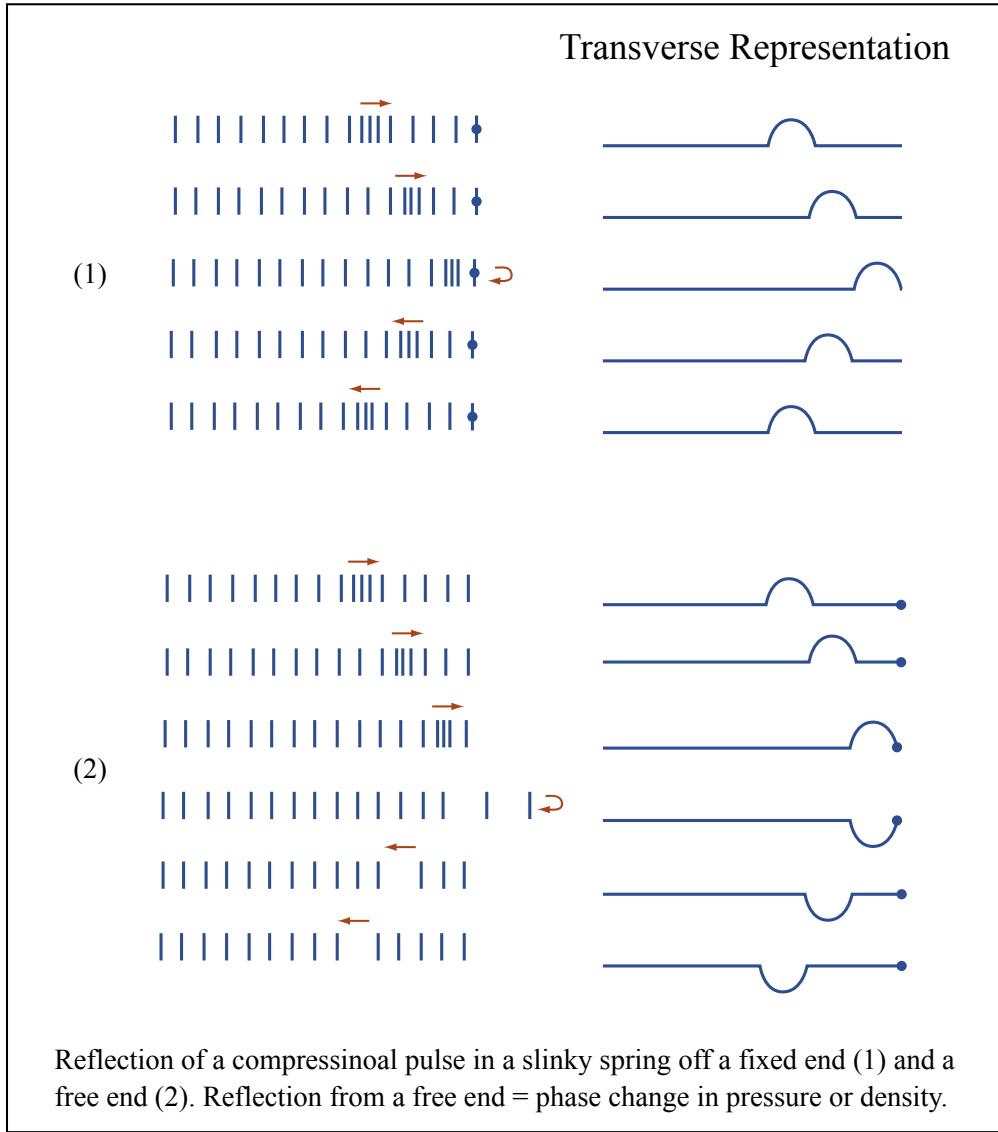
Transverse Representation

(1)

(2)

Reflection of a compressinoal pulse in a slinky spring off a fixed end (1) and a free end (2). Reflection from a free end = phase change in pressure or density.

Resonance in a uniform tube, open at one end

- For resonance to occur, compressions must reach the closed end at the same time as the excitation source produces another compression peak, and rarefactions must reach the closed end at the same time as the source produces a rarefaction.
- In a tube which is open at one end:
  i.   a compression is reflected from the open end as a rarefaction.
  ii. This rarefaction is reflected back off the closed end.
  iii. The rarefaction is then reflected from the open end as a compression.
  ➢   rarefactions do the opposite.
- So source and reflected wave will interfere constructively if, e.g., the time it takes for the wave to travel 2 times up and down the tube is equal to the half the period of the wave.

Resonance in a tube open at one end
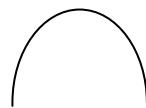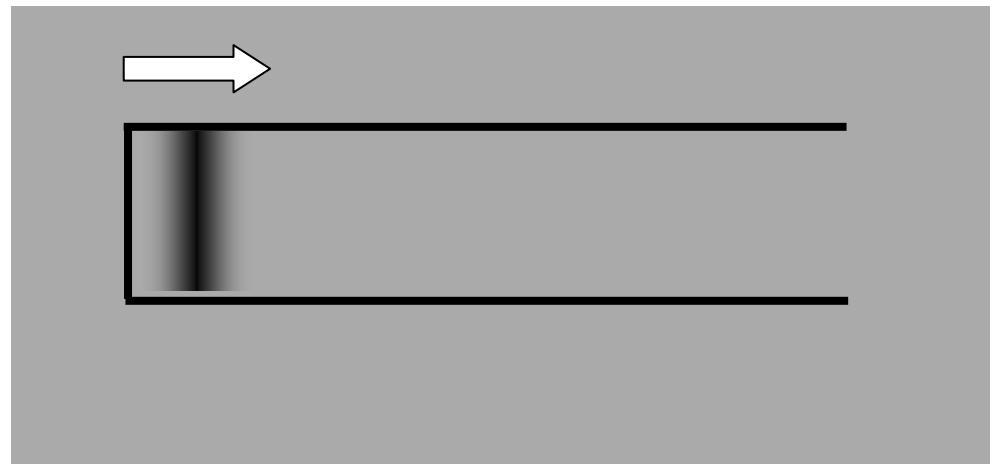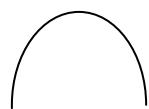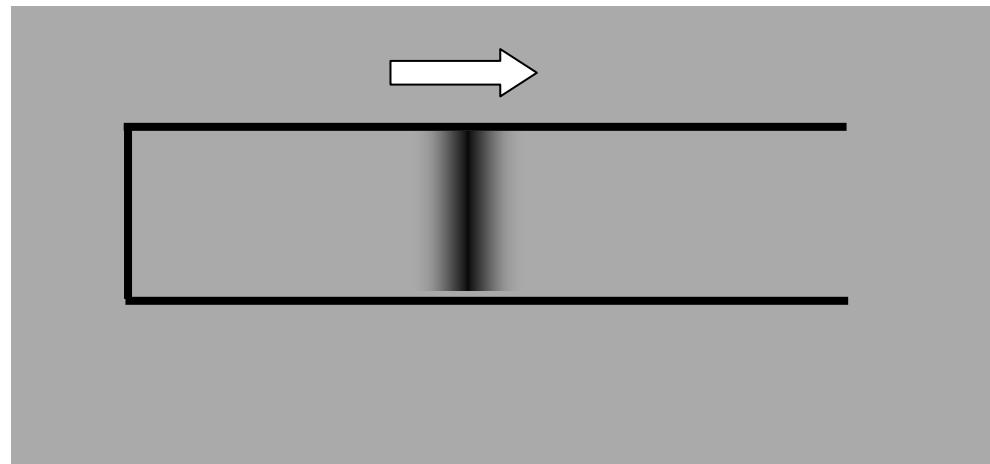
Resonance in a tube open at one end

Resonance in a tube open at one end
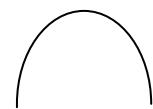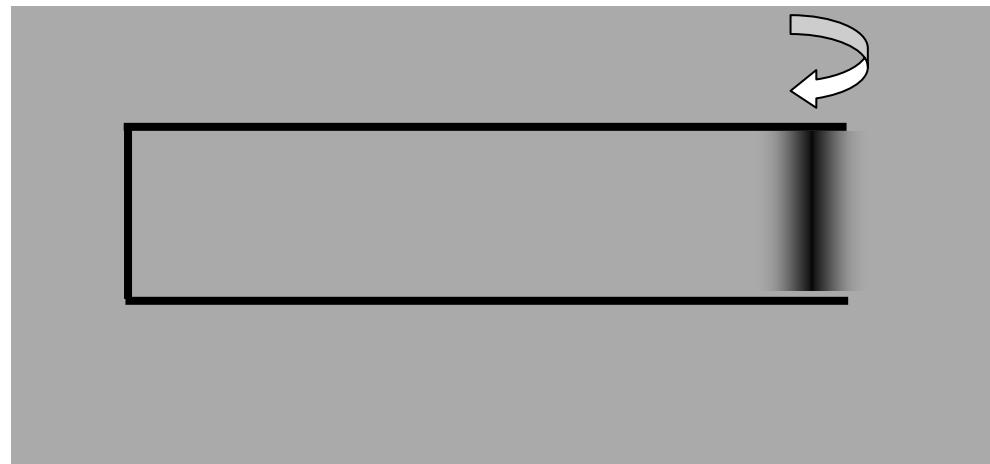
Resonance in a tube open at one end
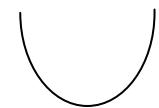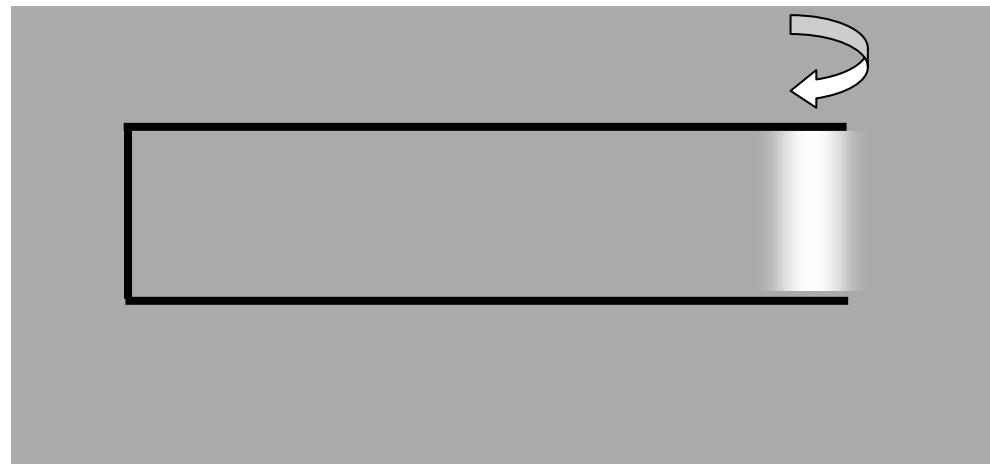
# Resonance in a tube open at one end
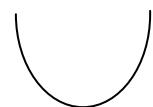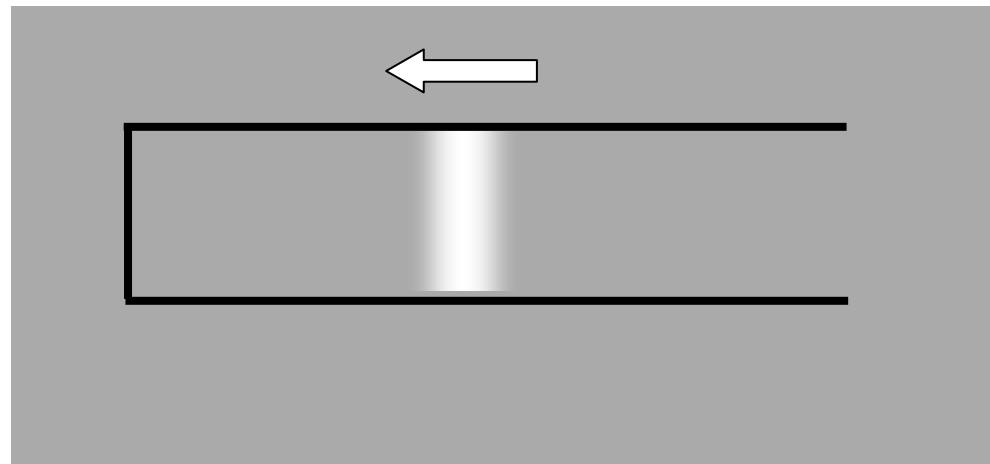
# Resonance in a tube open at one end
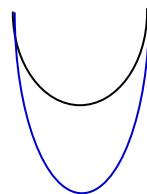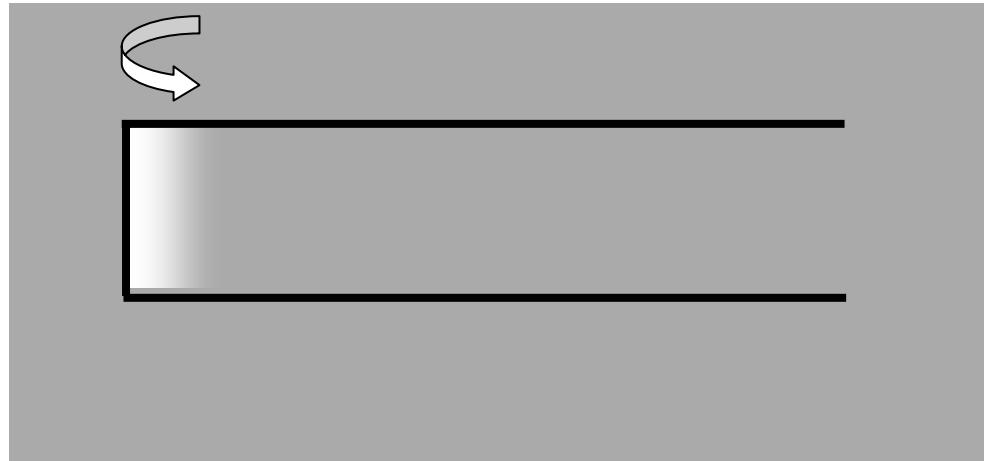
source emits a rarefaction

# Resonance in a tube open at one end
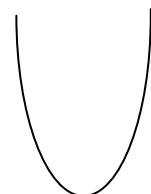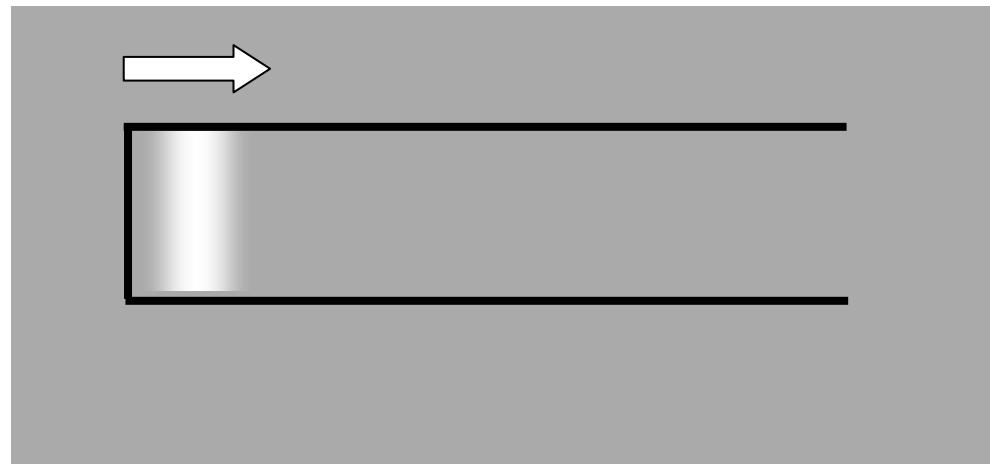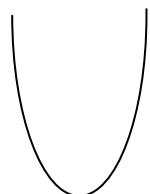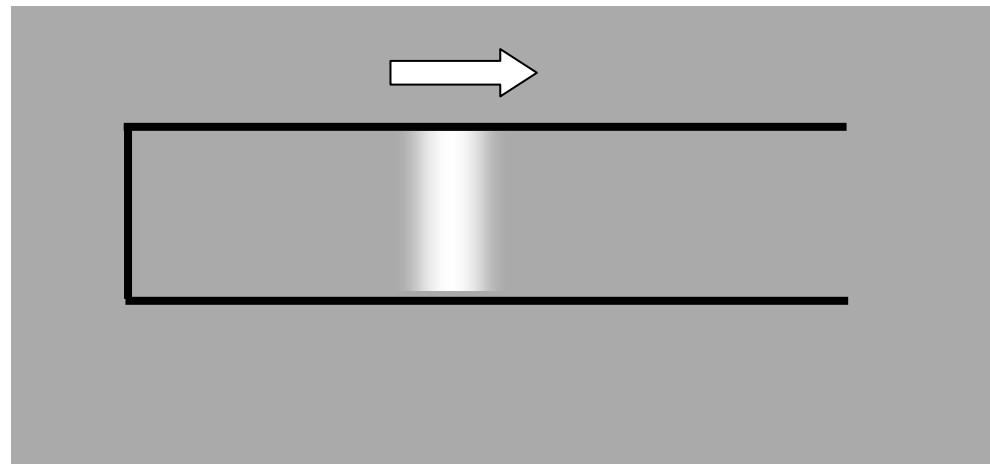
Resonance in a tube open at one end

Resonance in a tube open at one end
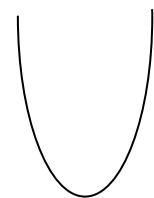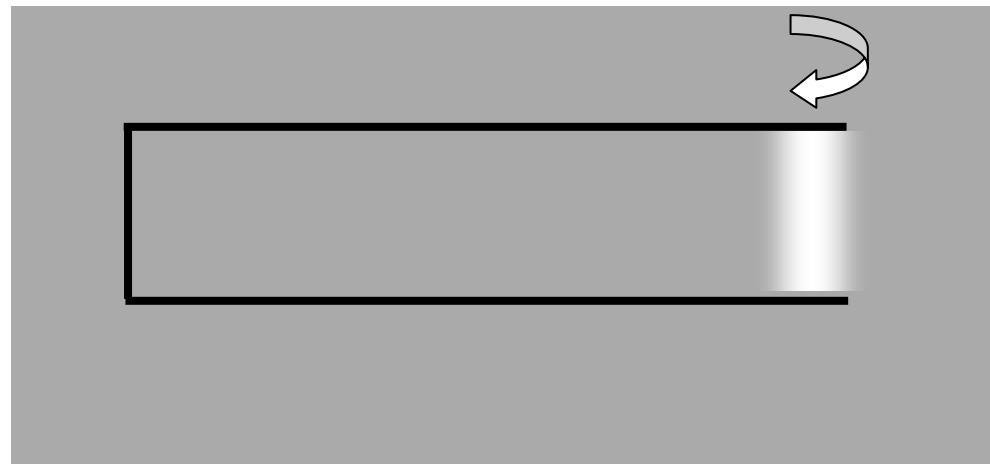
Resonance in a tube open at one end
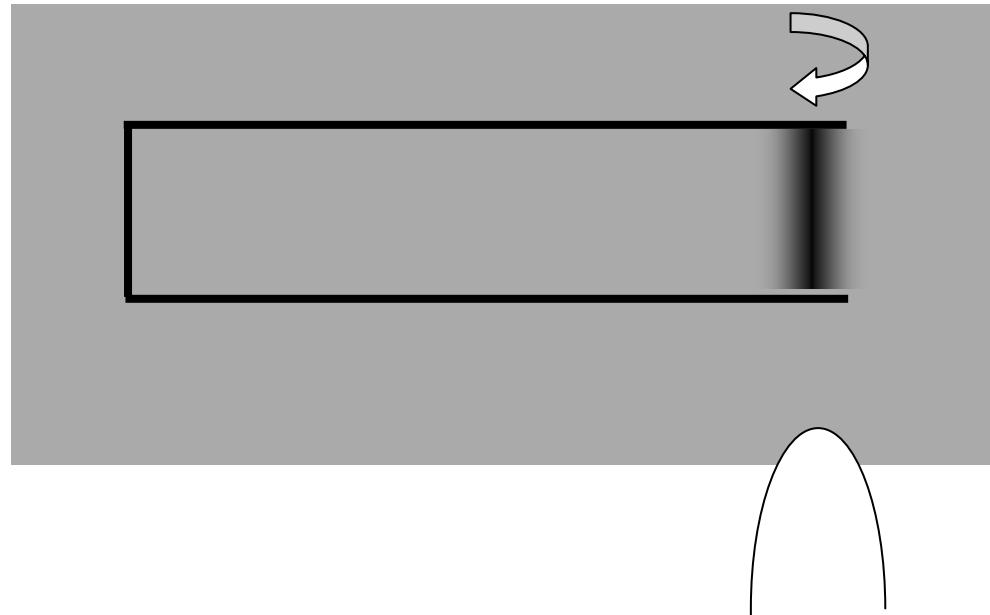
Resonance in a tube open at one end
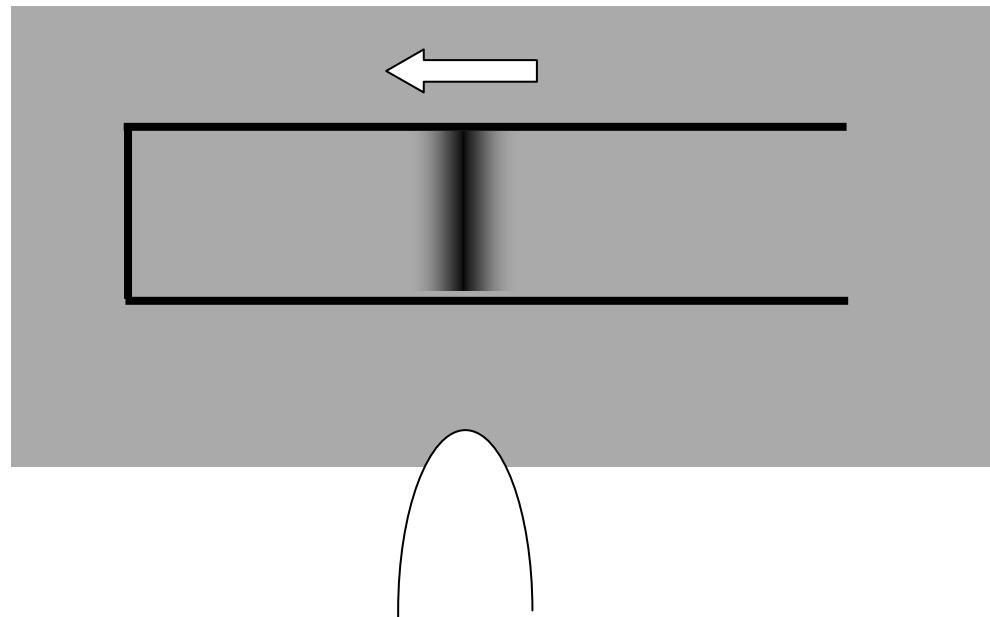
Resonance in a tube open at one end
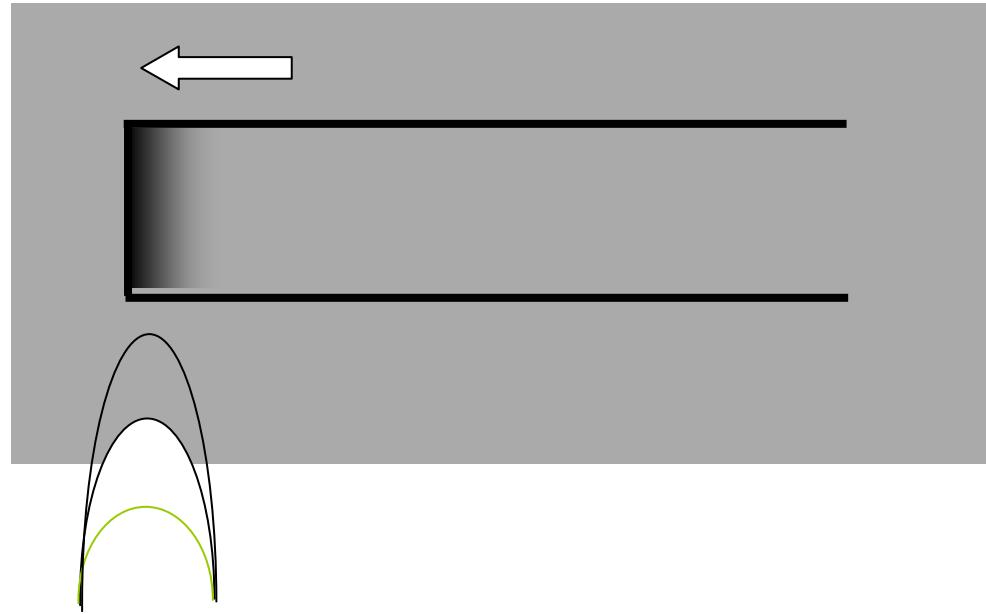
source emits a compression

Resonance in a tube open at one end

- Source and reflected wave will interfere constructively if, e.g., the time it takes for the wave to travel 2 times up and down the tube is equal to the half the period of the wave.
- For a tube of length L, for a wave with period T, travelling with speed c,

$$\frac{T}{2} = \frac{2L}{c}$$

- $f = 1/T$ so this resonant frequency $F = c/4L$

- 17.5cm is a reasonable figure for length of a (large) vocal tract, speed of sound in air is approximately 35000cm/s.

Resonance in a tube open at one end

- This is only one resonant frequency (the lowest, or first, resonance)
- The general requirement is that compressions emitted from the source must coincide with reflected compressions (likewise for rarefactions) - they don't have to coincide with every reflected compression.
- So the time it takes the wave to travel 2L must be equal to 1/2, 3/2, 5/2 .... i.e. (2n-1)/2 periods of the source wave.

$$\frac{(2n-1)T}{2} = \frac{2L}{c}$$

$$T = \frac{4L}{(2n-1)c} \qquad \text{so:} \qquad F_n = \frac{(2n-1)c}{4L}$$

Standing waves

- Under the conditions for resonance, the air in a tube vibrates in a standing wave.
- Standing waves arise when identical waves are travelling in opposite directions - e.g. a wave and its reflection from a surface. standing wave demo
- A standing wave contains fixed points at which air particles do not move (velocity nodes). Movement is maximal between the nodes (at the velocity anti-nodes).
- Velocity anti-nodes correspond to pressure nodes - pressure does not change - and velocity nodes correspond to pressure anti-nodes.

Standing waves and resonance in tubes

- The air in a tube which is open at one end naturally vibrates in standing waves.
- There must be a velocity node at the closed end since the fixed end prevents particle movement.
- There must be a pressure node at the open end - pressure is always equal to external pressure.
- The simplest wave that meets these conditions is a standing wave that goes from maximum pressure at the glottis to zero (atmospheric pressure) at the lips.

velocity

pressure

Standing waves and resonance in tubes

What is the frequency of this wave?

velocity
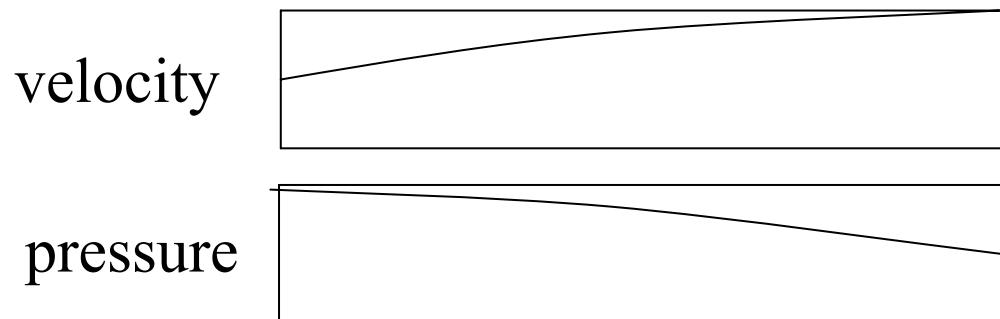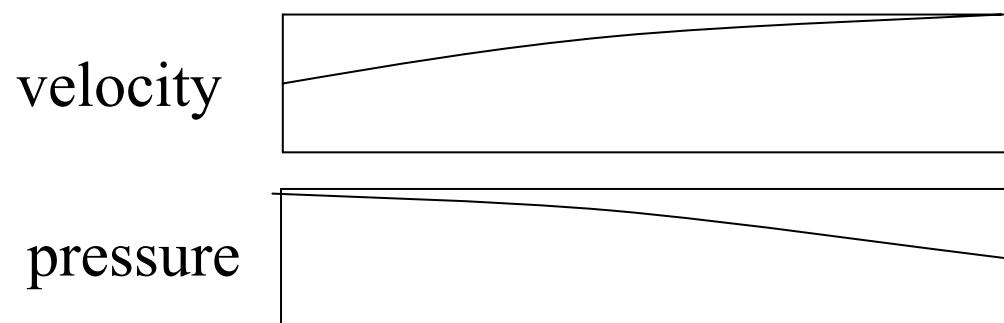
pressure

Relating frequency to wavelength

- The distribution of pressure fluctuations in space due to a sound wave depends on the frequency of the wave and the speed at which the wavefront travels.
- Consider a wave with frequency f Hz and speed c cm/s.
- In one second it travels c cm.
- In one second it goes through f cycles.
- So there are f cycles in c cm.
- So the length of one cycle (the wavelength) is c/f.

- Usual formulation: $c = f\lambda$ Where λ is wavelength
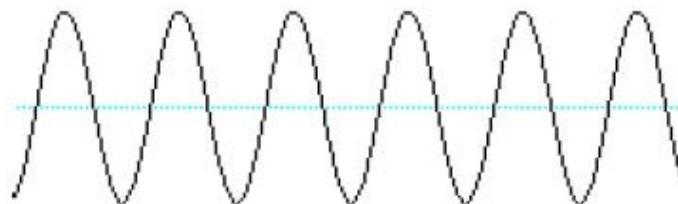
Standing waves and resonance in tubes

What is the frequency of this wave?
* The transition from maximum to zero pressure is only one quarter of a cycle.
* For a tube of length L, $L = \lambda/4$
* $\lambda = 4L$
* $f = c/\lambda = c/4L$

velocity

pressure
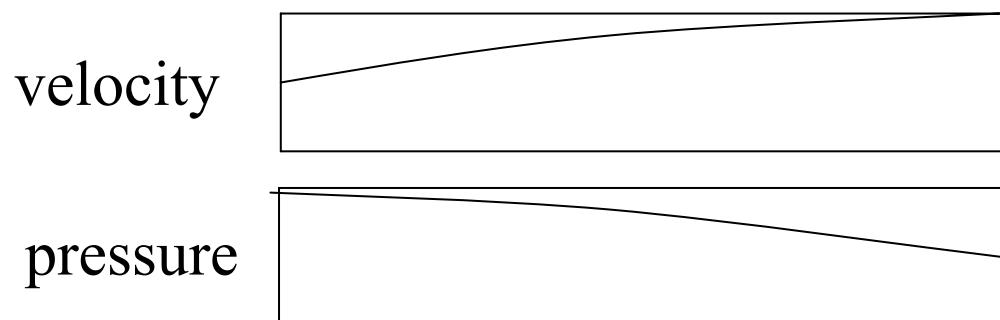
Standing waves and resonance in tubes

- Other standing waves that meet the boundary conditions:
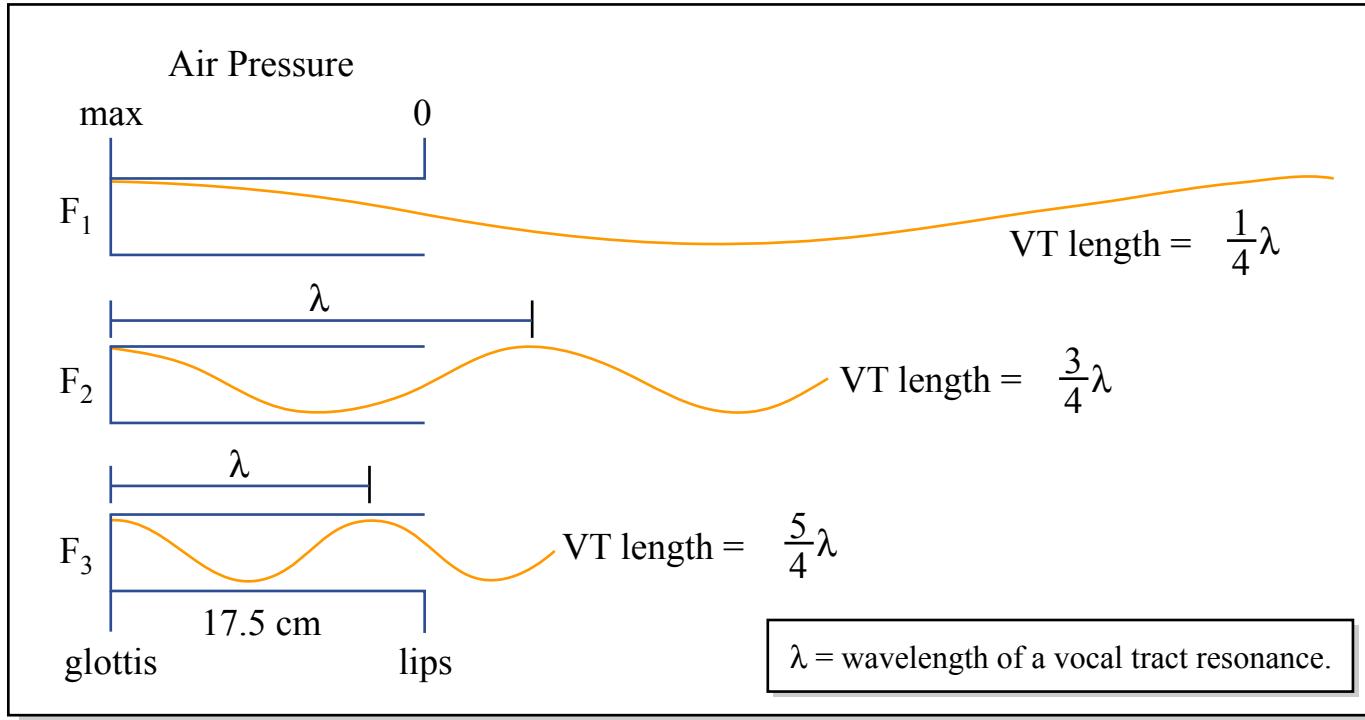
Air Pressure

max   0

F_1 VT length = $\frac{1}{4}\lambda$

λ

F_2 VT length = $\frac{3}{4}\lambda$

λ

F_3 VT length = $\frac{5}{4}\lambda$

17.5 cm

glottis   lips

λ = wavelength of a vocal tract resonance.

Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

- In general, air in a tube of length L, closed at one end, will have resonant frequencies (formants) at: $$F_n = \frac{(2n-1)c}{4L}$$

Tubes models of the vocal tract

- The vocal tract in vowels can be approximated by a tube that is closed at one end and open at the other.
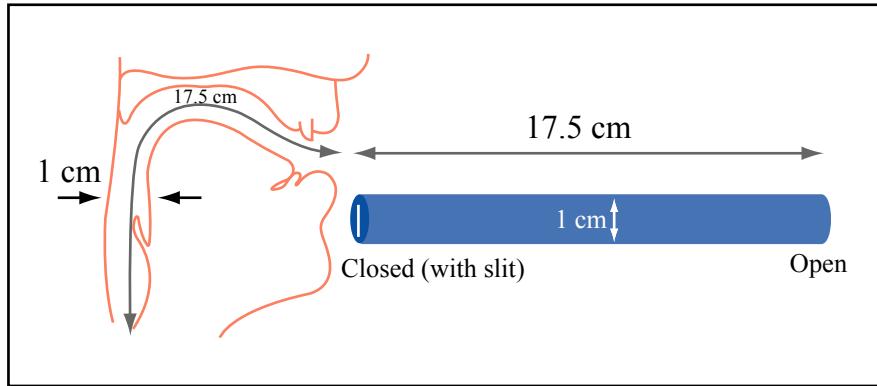

$$F_n = \frac{(2n-1)c}{4L}$$

- The cross-sectional area of the vocal tract is closest to being uniform in a mid central vowel [ɜ] as in RP English *bird* [bɜd], or at the end of *sofa* [soʊfə].
- Caution: Most things that are transcribed as schwa do not have a uniform vocal tract.

Tubes models of the vocal tract

- Another simple tube model that approximates a vocal tract shape is a tube (almost) closed at both ends.
- This is similar to a sound made with a narrow constriction at the lips, or the cavity behind an interior constriction.
- Air in a tube closed at both ends naturally vibrates in standing waves with nodes at both ends.
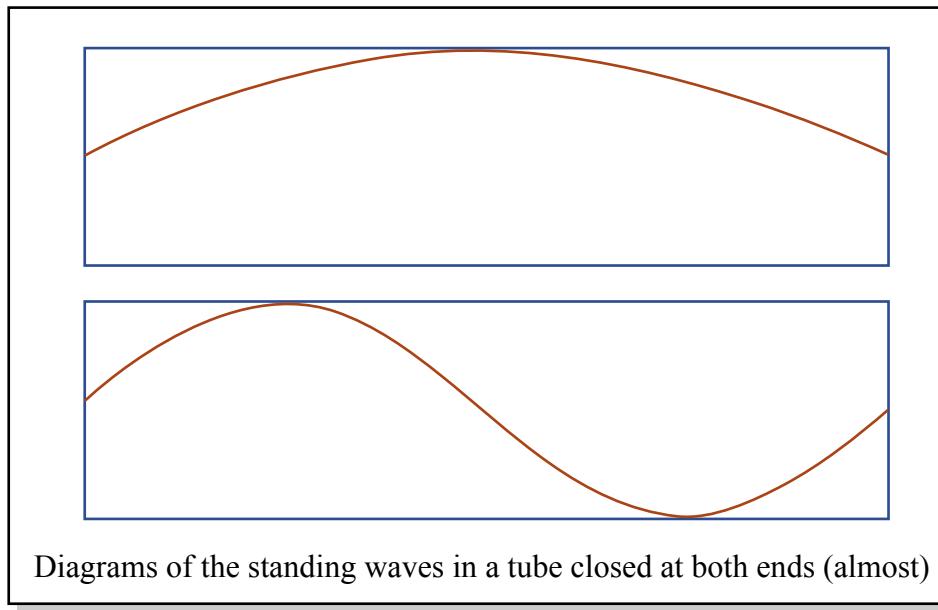
Diagrams of the standing waves in a tube closed at both ends (almost)

Tubes models of the vocal tract

- $L = n\lambda/2$
- $\lambda = 2L/n$
- $F_n = c/\lambda$

$$F_n = \frac{nc}{2L}$$
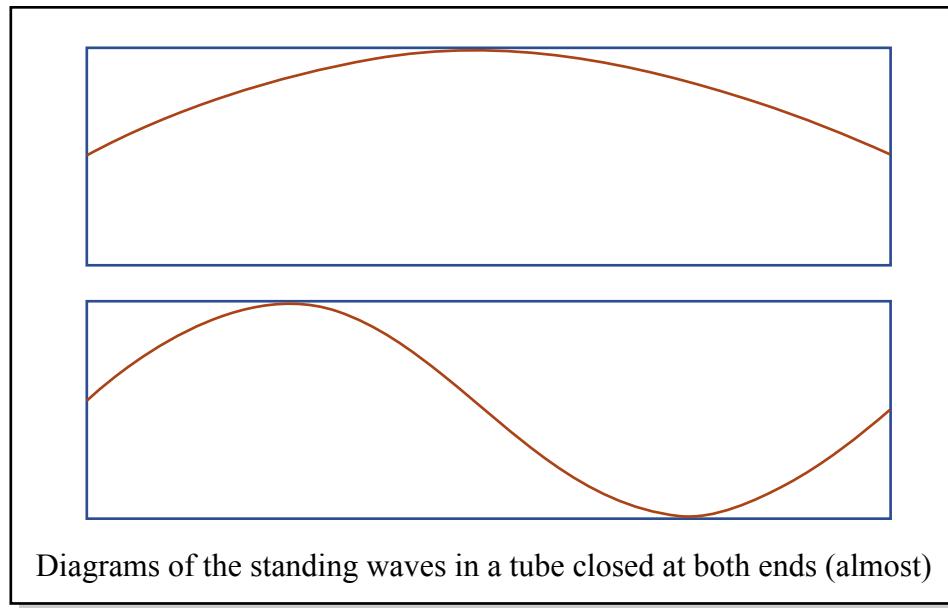
Diagrams of the standing waves in a tube closed at both ends (almost)

Source-Filter Model of Speech Production
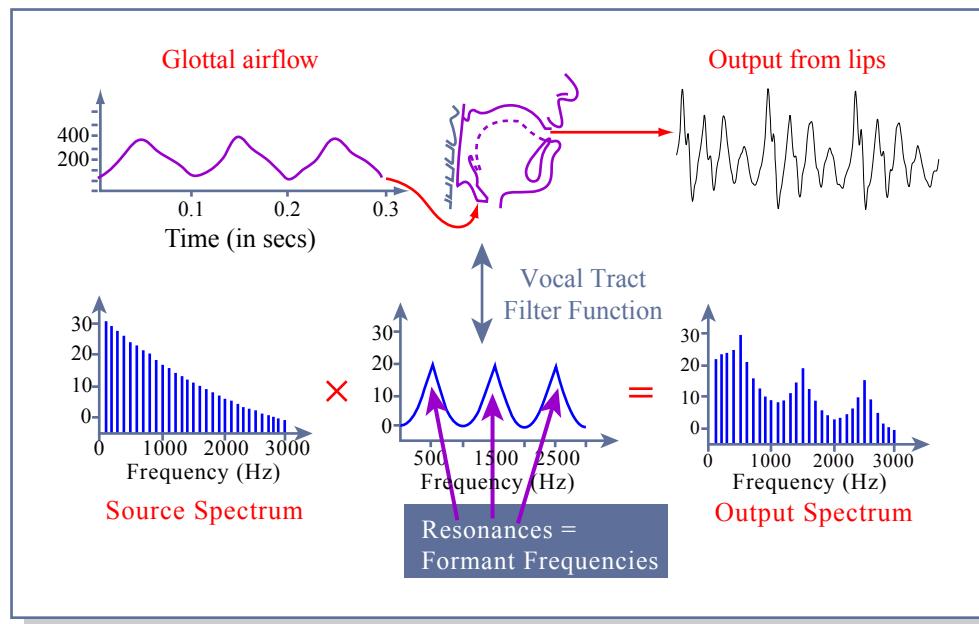
Figure by MIT OpenCourseWare.

Cues to vowel quality

- The main cues to vowel quality are related to the frequencies of the first two or three formants.

Evidence:

- Synthesis experiments: Varying formant frequencies in synthetic vowels can change the vowel percept (Delattre, Liberman, Cooper, and Gerstman 1952, and many others).
  - http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html

Cues to vowel quality

- The main cues to vowel quality are related to the frequencies of the first two or three formants.

Evidence:

- Multi-Dimensional Scaling studies: Shepard (1972) presents a three-dimensional MDS analysis of confusion data on American English vowels from Peterson and Barney (1952). Two nearly orthogonal axes correlate well (but non-linearly) with F1 and F2. It is possible to identify an axis which correlates well with F3, but it is also correlated with the F2 axis (not orthogonal).
  - Many other MDS studies of vowel confusion/similarity have found that the first two dimensions identified correspond well to F1, F2.
- It seems that a third dimension is required to accommodate rhotic vowels, but MDS analyses do not provide clear evidence that this dimension corresponds to F3.

Shepard (1972): MDS analysis of vowel confusions
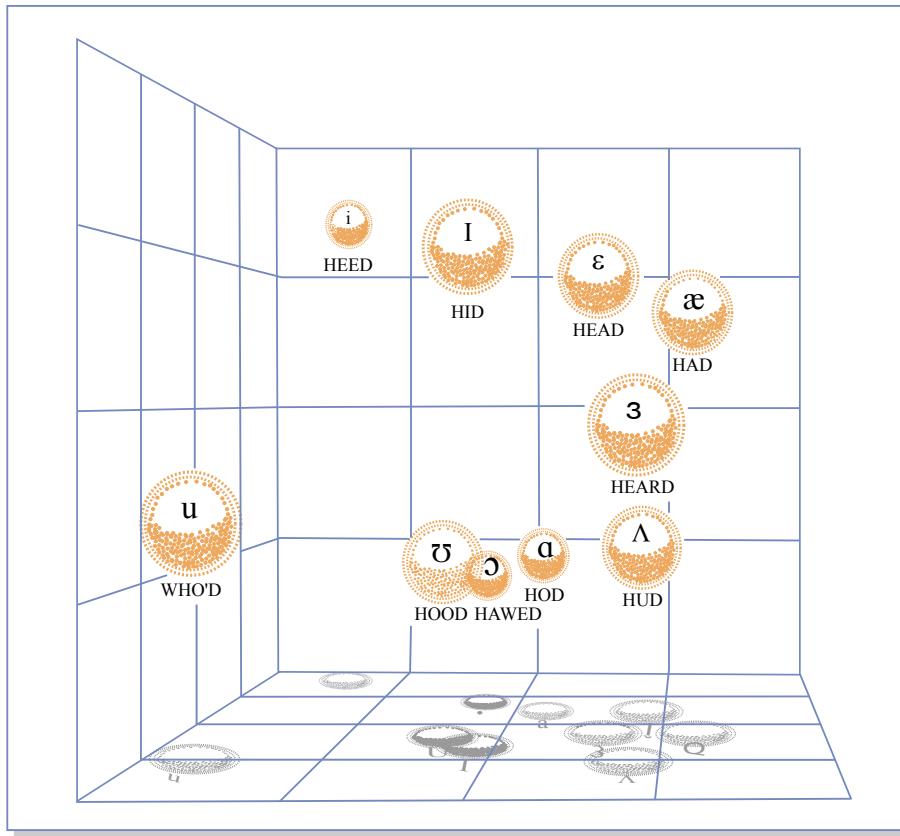
Figure by MIT OpenCourseWare. Adapted from Roger Shepard. "Psychological Representation of Speech Sounds."
In *Human Communication: A Unified View*. Edited by E. E. David and P. B. Denes. New York,
NY: McGraw-Hill, 1972, pp. 67-113.

Vowel quality

- The main cues to vowel quality are related to the frequencies of the first two or three formants.

- So in trying to understand vowel acoustics, we will focus on the frequencies of formants.

Questions to keep in mind:

- Why are the first 2-3 formant frequencies the primary cues to vowel quality?

- What is the range of formant patterns that we can produce?

- We can get a basic understanding of vowel acoustics using very simple tube models of the vocal tract.

Low vowels [ɑ, a, æ]

- ## Pharyngeal constriction

The shape of the vocal tract in the vowel [ɑ] as in father schematized as two tubes.
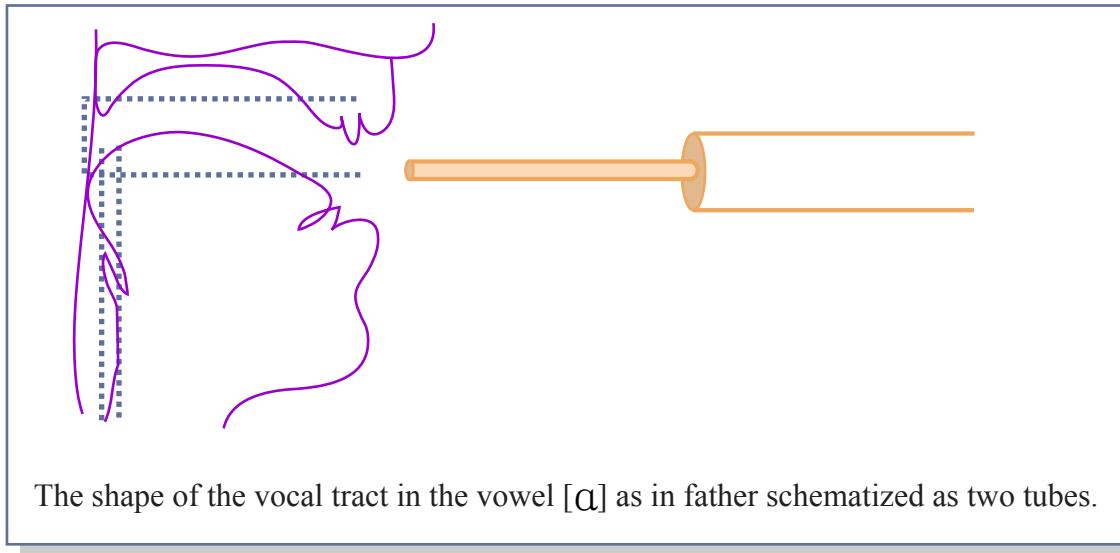
Figure by MIT OpenCourseWare.

- Since the back tube is much narrower than the front tube, each can reasonably be approximated by a tube closed at one end and open at the other.
- The resonances of the combined tubes deviate from the values we would calculate for these configurations in isolation because the resonators are acoustically coupled.
- The degree of coupling depends on the difference in cross-sectional areas.

Low vowels [ɑ, a, æ]

$$F_n = \frac{(2n-1)c}{4L}$$

Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.
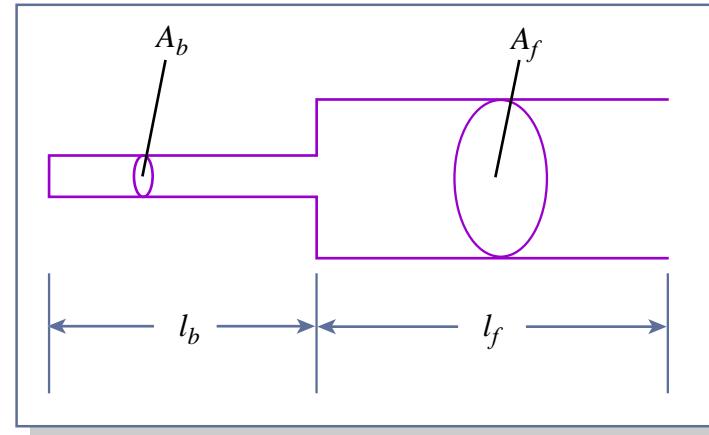
nomogram
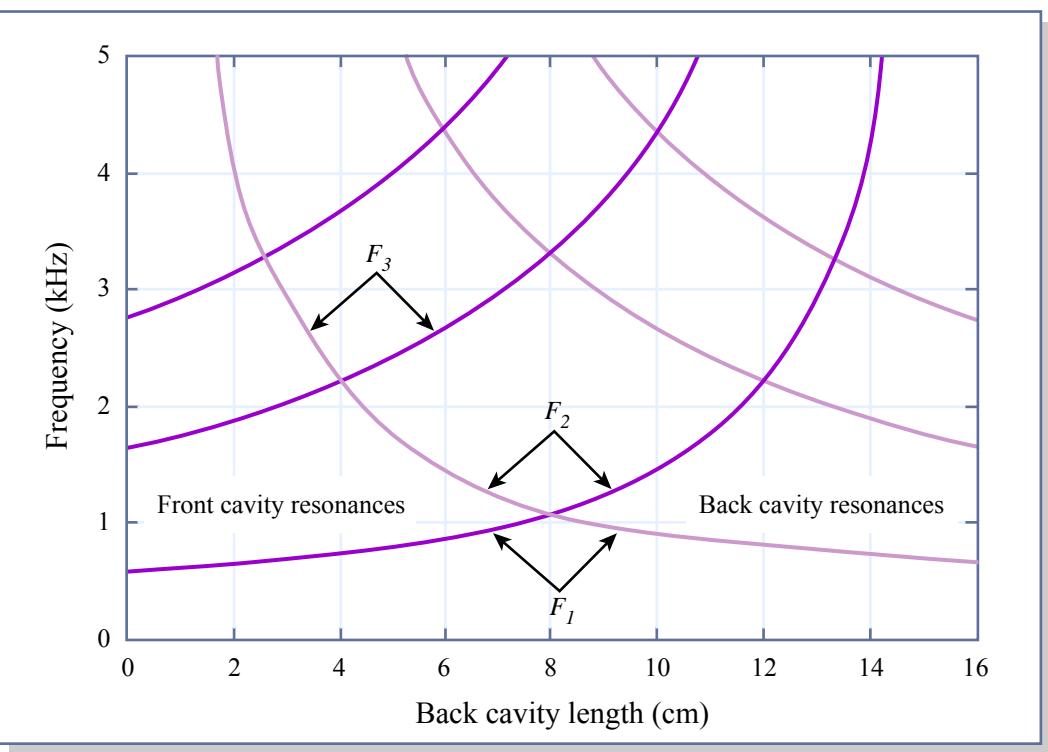
Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

# Non-low vowels (e.g. [i, e])

- ## Short constriction in the mouth

- The back cavity can be approximated by a tube closed at both ends.

$$F_n = \frac{nc}{2L}$$

- The front cavity is approximated by a tube closed at one end.

$$F_n = \frac{(2n-1)c}{4L}$$

- Neglects coupling. The degree of coupling depends on the cross-sectional area of the constriction.

- How do we account for the F1 of high vowels?

Helmholtz resonators

- The back cavity and the constriction together form a resonant system called a Helmholtz resonator.

- If the length of the constriction is short, the air in it vibrates as a mass on the 'spring' formed by the air in the back cavity.

- Resonant frequency, $\quad f = \dfrac{c}{2\pi}\sqrt{\dfrac{A_c}{V l_c}} = \dfrac{c}{2\pi}\sqrt{\dfrac{A_c}{A_b l_b l_c}}$

Non-low vowels - nomogram

Resonant frequencies of the back tube (light lines), front tube (heavy lines) and Helmholtz resonance (dashed line) in the tube model. Frequency is plotted as function of different back tube lengths (l_b), with the length of the constriction fixed at 2 cm and the total length of the model fixed at 16 cm.

front cavity $\qquad F_n = \dfrac{(2n-1)c}{4L}$

back cavity $\qquad F_n = \dfrac{nc}{2L}$

back cavity + constriction

$$f = \frac{c}{2\pi}\sqrt{\frac{A_c}{A_b l_b l_c}}$$

- How would you model a mid vowel?

Perturbation Theory (Chiba and Kajiyama 1941)

- Constriction near a point of maximum velocity (V_n) lowers the associated formant frequency.

- Constriction near a point of maximum pressure raises the associated formant frequency.
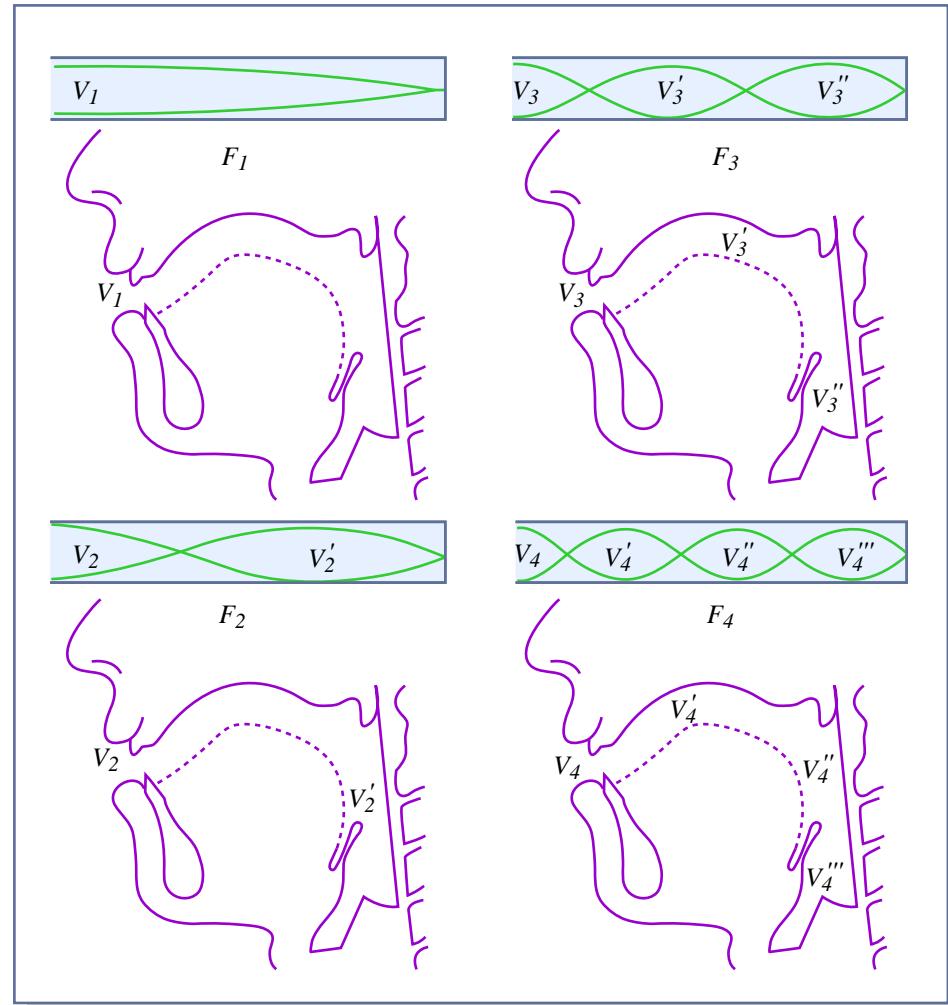

Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. Based on Chiba and Kajiyama 1941.

Perturbation Theory (Chiba and Kajiyama 1941)

- What is the effect of a pharyngeal constriction?
- Does this correspond to the tube model above?
- How do you raise F2 maximally?
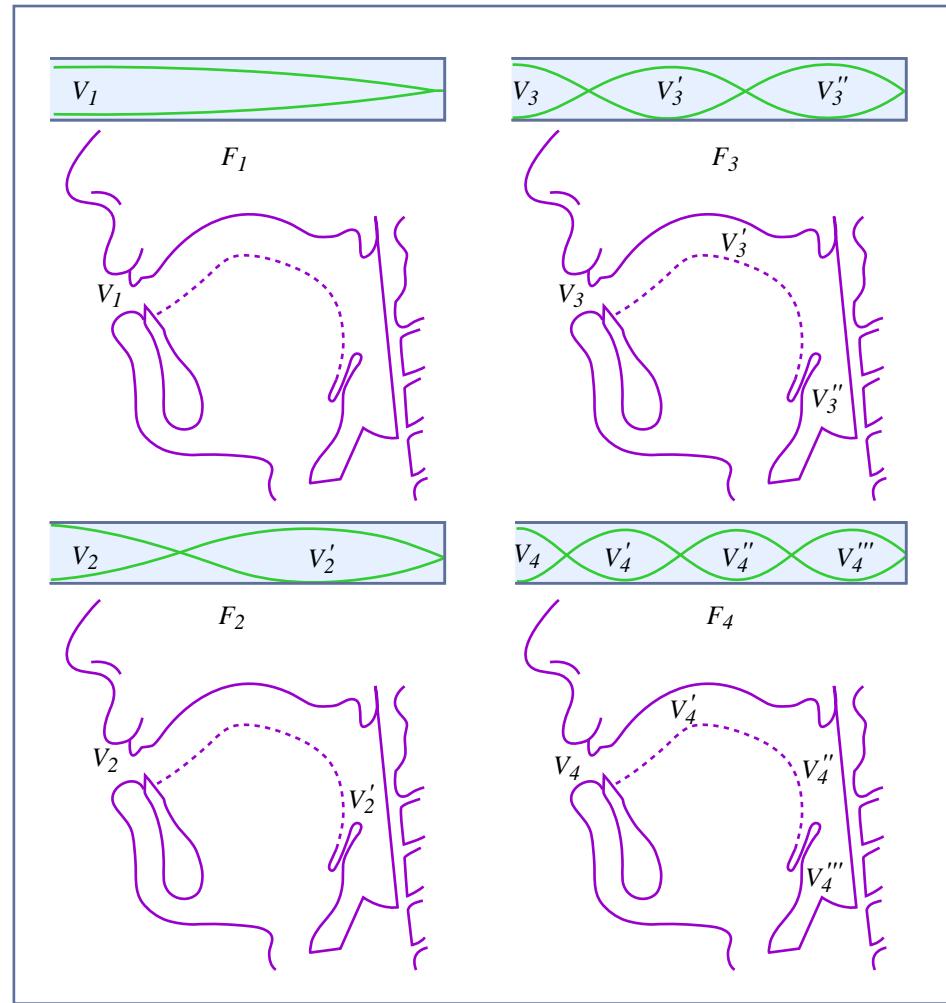
Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. Based on Chiba and Kajiyama 1941.

Perturbation Theory (Chiba and Kajiyama 1941)

A nice story about American [ɹ]:

- Three constrictions: labial (lip protrusion/rounding), palatal (bunching or retroflexion), and pharyngeal.

- All 3 are near velocity maxima for F3, hence very low F3.

- But see Espy-Wilson et al (2000).
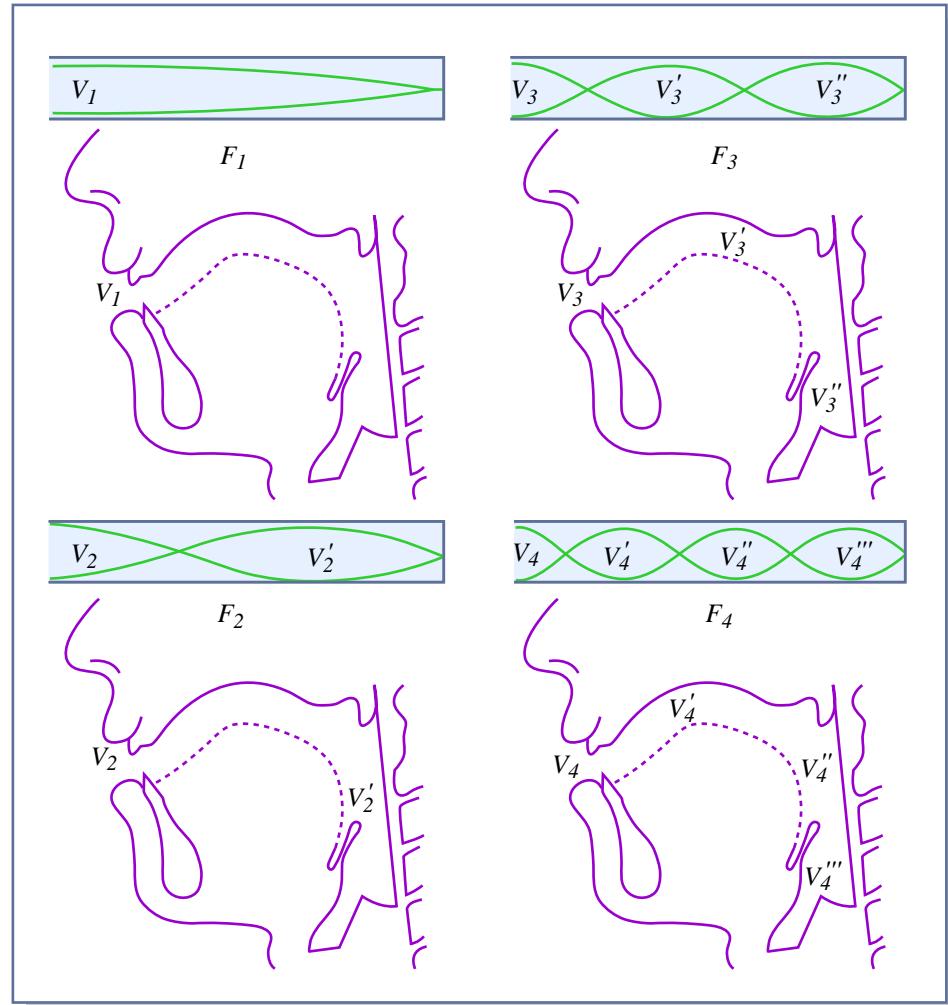
Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. Based on Chiba and Kajiyama 1941.

Perturbation Theory vs. two-tube models

- Our simple tube models ignore acoustic coupling and are therefore most valid where constrictions are narrow.

- Perturbation theory accounts for the effects of small perturbations of a uniform tube, and thus is most accurate for open constrictions.

Lip rounding

- Lip-rounding also involves lip protrusion so it both lengthens the vocal tract and introduces a constriction at the lips.

- Perturbation theory: All formants have a velocity maximum at the lips, so a constriction at the lips should lower all formants.

- Lengthening the vocal tract also lowers formants.

- Tube models + perturbation theory: A constriction at the lips lowers the resonant frequencies of the front cavity.

  - In most vowels the lowest front cavity resonance is F2, but in a front vowel it may be F3.