CS 224S / LINGUIST 285
Spoken Language Processing

Andrew Maas
Stanford University
Spring 2022

Lecture 2: Phonetics

Original slides by Dan Jurafsky

Homework 1

- Available on website and Canvas (we will release homeworks on website only moving forward).
  - Due on Monday 4/11 11:59pm Pacific
- Colab and written sections. Today's lecture will help with phonetic transcription!

Week 1

- Course introduction
- Course Logistics
- Course topics overview
- **Articulatory Phonetics**
- **ARPAbet transcription**

Phonetics

- ARPAbet
  - An alphabet for transcribing American English phonetic sounds.
- Articulatory Phonetics
  - How speech sounds are made by articulators (moving organs) in mouth.
- Acoustic Phonetics
  - Acoustic properties of speech sounds

Phonetics

- Modern systems are less reliant on encoding phonetic domain knowledge directly.
- Basic understanding helps with describing and debugging spoken language systems
  - E.g. how does an accent change the sound of pronunciations?
- Phonetic categories derived from *how* humans produce speech

International Phonetic Alphabet (IPA)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	’ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ʄ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ʛ Uvular	sʼ Alveolar fricative

OTHER SYMBOLS

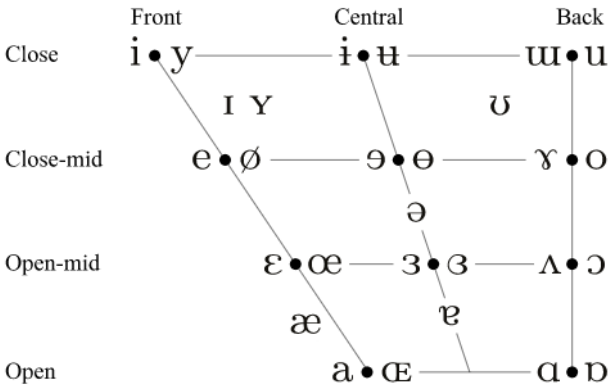
ʍ Voiceless labial-velar fricative

ɕ ʑ Alveolo-palatal fricatives

ɯ Voiced labial-velar approximant

ɺ Voiced alveolar lateral flap

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

(Wikipedia IPA with sounds)

Articulatory parameters for English consonants (in ARPAbet)

MANNER OF ARTICULATION	PLACE OF ARTICULATION													
	bilabial		labio-dental		inter-dental		alveolar		palatal		velar		glottal	
stop	p	b					t	d			k	g	q	
fric.			f	v	th	dh	s	z	sh	zh			h	
affric.									ch	jh				
nasal		m						n				ng		
approx		w						l/r		y				
flap							dx							

VOICING: voiceless | voiced

Table from Jennifer Venditti

ARPAbet Vowels

	b_d	ARPA		b_d	ARPA
1	bead	iy	9	bode	ow
2	bid	ih	10	booed	uw
3	bayed	ey	11	bud	ah
4	bed	eh	12	bird	er
5	bad	ae	13	bide	ay
6	bod(y)	aa	14	bowed	aw
7	bawd	ao	15	Boyd	oy
8	Budd(hist)	uh			

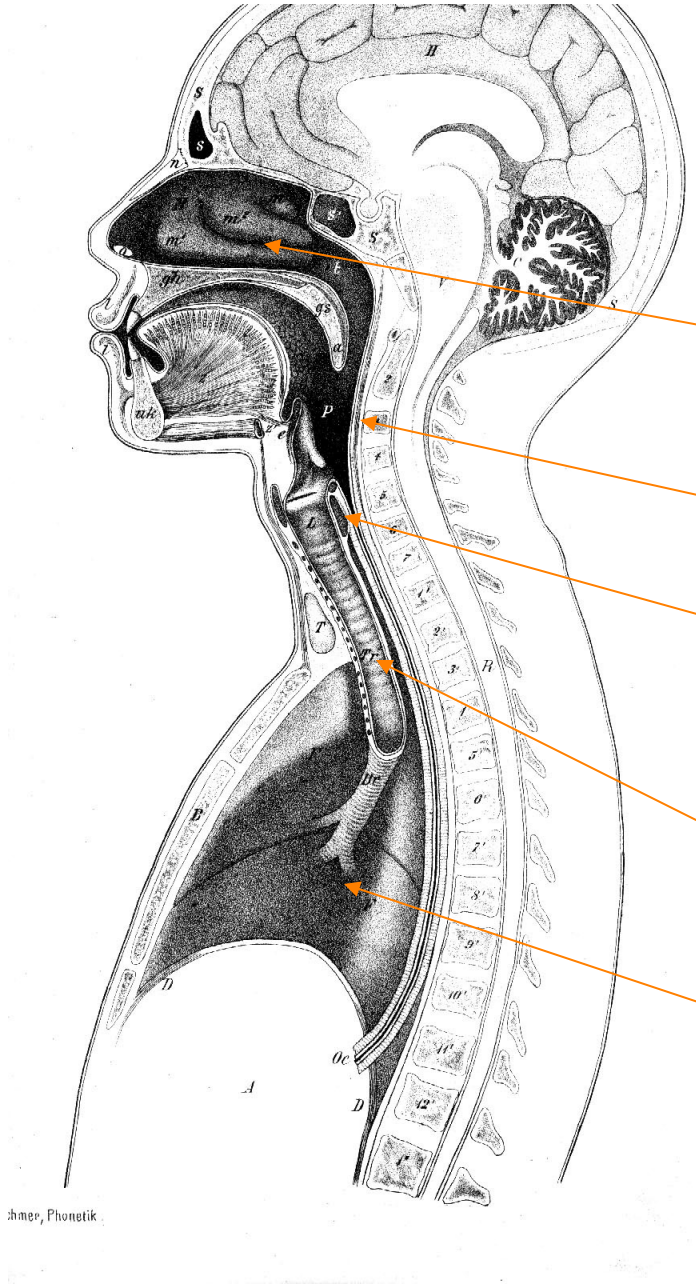
Note: Many speakers pronounce Buddhist with the vowel uw as in booed, So for them [uh] is instead the vowel in "put" or "book"

https://corpus.linguistics.berkeley.edu/acip/

Speech Production: Flow, Resonance, & Articulation

- Flow
  - We (normally) speak while breathing out. Respiration provides airflow. "Pulmonic egressive airstream"
  - Airstream sets vocal folds in motion. Vibration of vocal folds produces sounds. Sound is then modulated by:
- Resonance: Shape of vocal tract causing harmonics
- Articulation: Manipulation of airflow
  - Oral tract: uvula, soft palate (velum), hard palate, tongue, lips, teeth
  - Nasal tract

Text adopted from Sharon Rose



Sagittal section of the vocal tract
(Techmer 1880)

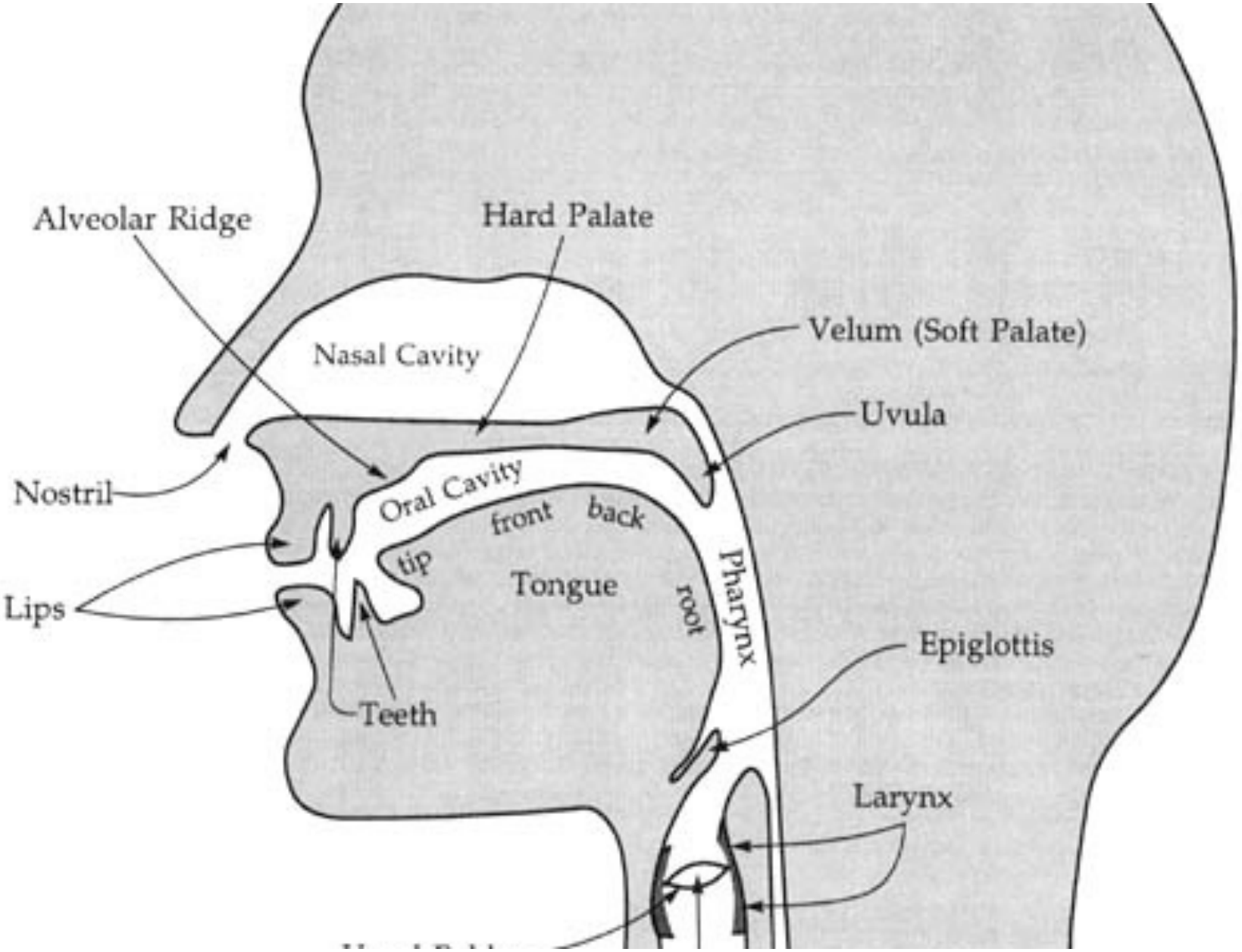
Nasal Cavity

Pharynx

Vocal Folds (within the Larynx)

Trachea

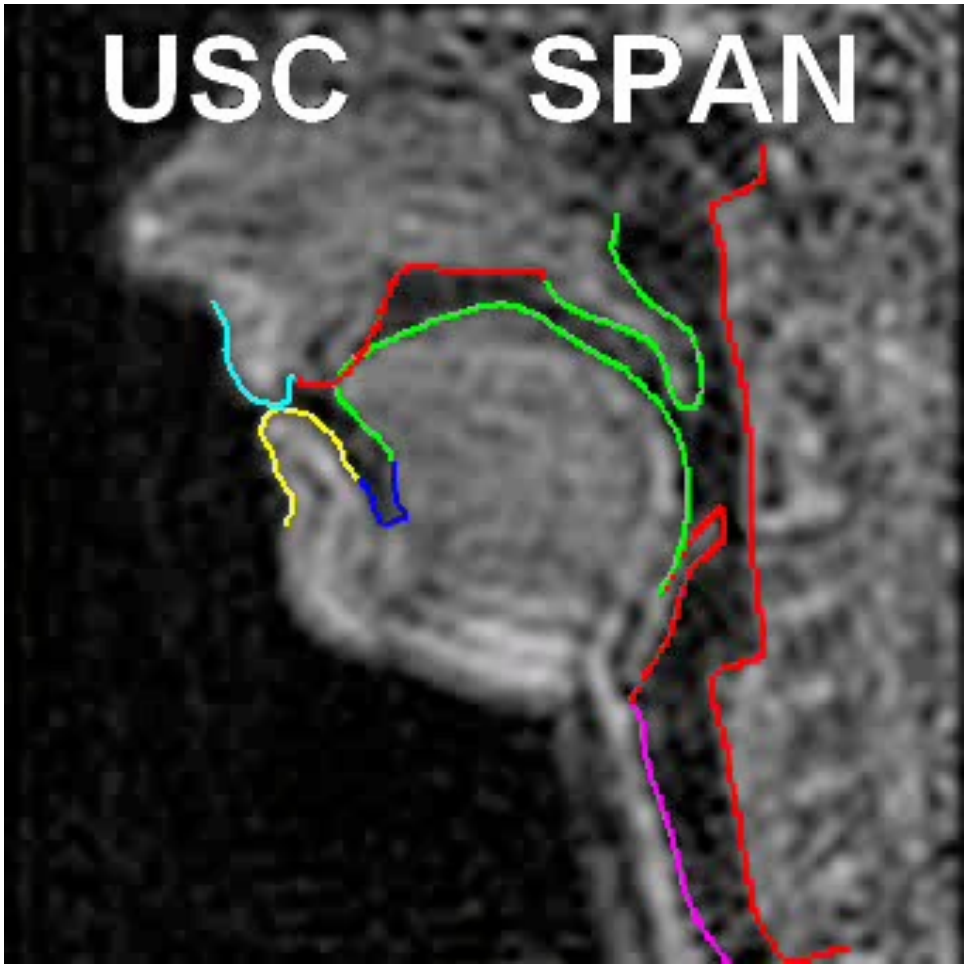
Lungs



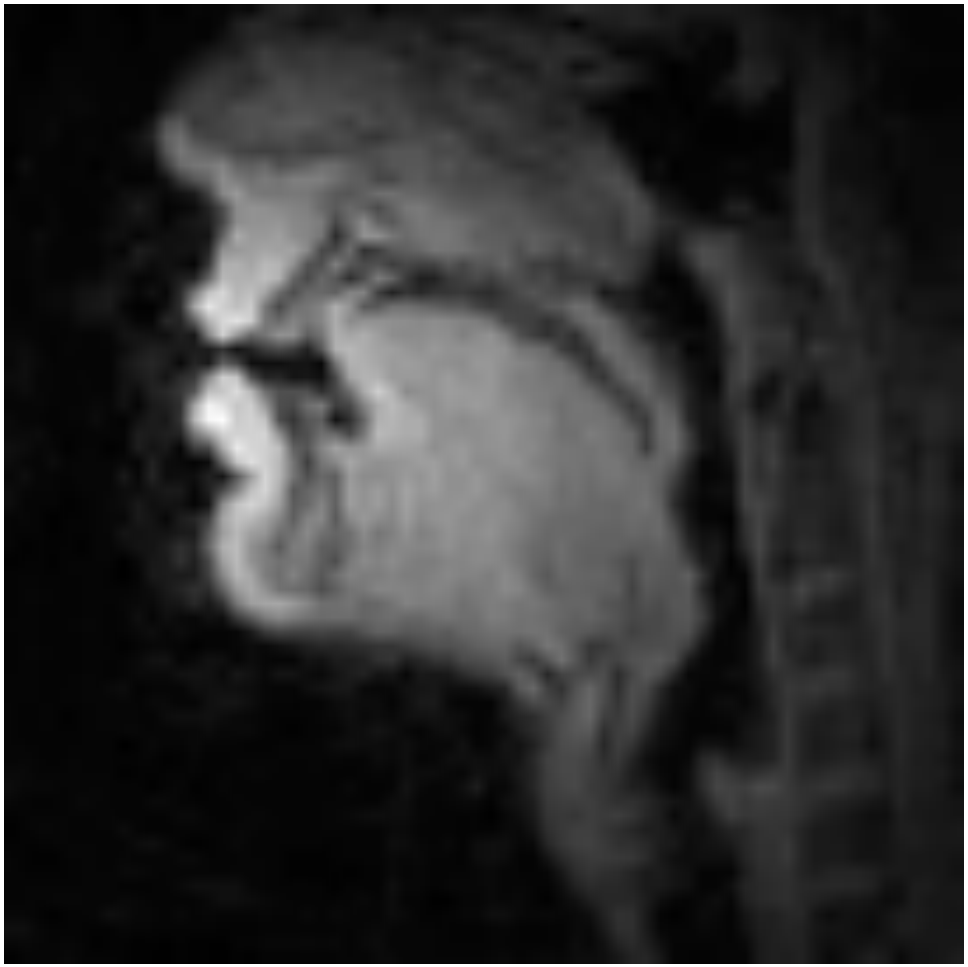
From Mark Liberman's Web Site, from Language Files (7th ed)

USC's SAIL Lab

Shri Narayanan



Tamil

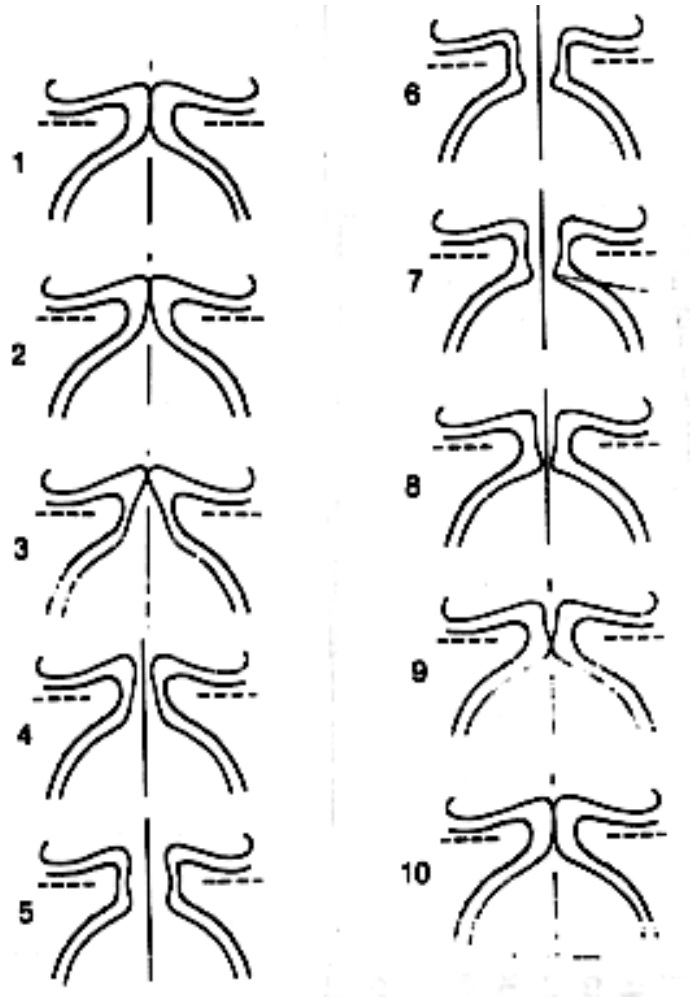


Larynx and Vocal Folds

- The Larynx (voice box)
  - A structure made of cartilage and muscle
  - Located above the trachea (windpipe) and below the pharynx (throat)
  - Contains the vocal folds
  - (adjective for larynx: laryngeal)
- Vocal Folds (older term: vocal cords)
  - Two bands of muscle and tissue in the larynx
  - Can be set in motion to produce sound (voicing)

Text from slides by Sharon Rose UCSD LING 111 handout

Voicing:



- Air comes up from lungs
- Forces its way through vocal cords, pushing open (2,3,4)
- This causes air pressure in glottis to fall, since:
  - when gas runs through constricted passage, its velocity increases (**Venturi tube effect**)
  - this increase in velocity results in a drop in pressure (**Bernoulli principle**)
- Because of drop in pressure, vocal cords snap together again (6-10)
- Single cycle: ~1/100 of a second.

Figure & text from John Coleman's web site

Vocal Fold Vibration

UCLA Phonetics Lab Demo

Voicelessness

- When vocal cords are open, air passes through unobstructed
- Voiceless sounds: p/t/k/s/f/sh/th/ch
- If the air moves very quickly, the turbulence causes a different kind of phonation: whisper

Consonants and Vowels

- Consonants: phonetically, sounds with audible noise produced by a constriction
- Vowels: phonetically, sounds with no audible noise produced by a constriction

- (it's more complicated than this, since we have to consider syllabic function, but this will do for now)

Text adapted from John Coleman


USC
SPAN

Place of Articulation

- Consonants are classified according to the location where the airflow is most constricted.
- This is called place of articulation
- Three major kinds of place articulation:
  - **Labial** (with lips)
  - **Coronal** (using tip or blade of tongue)
  - **Dorsal** (using back of tongue)

Places of articulation

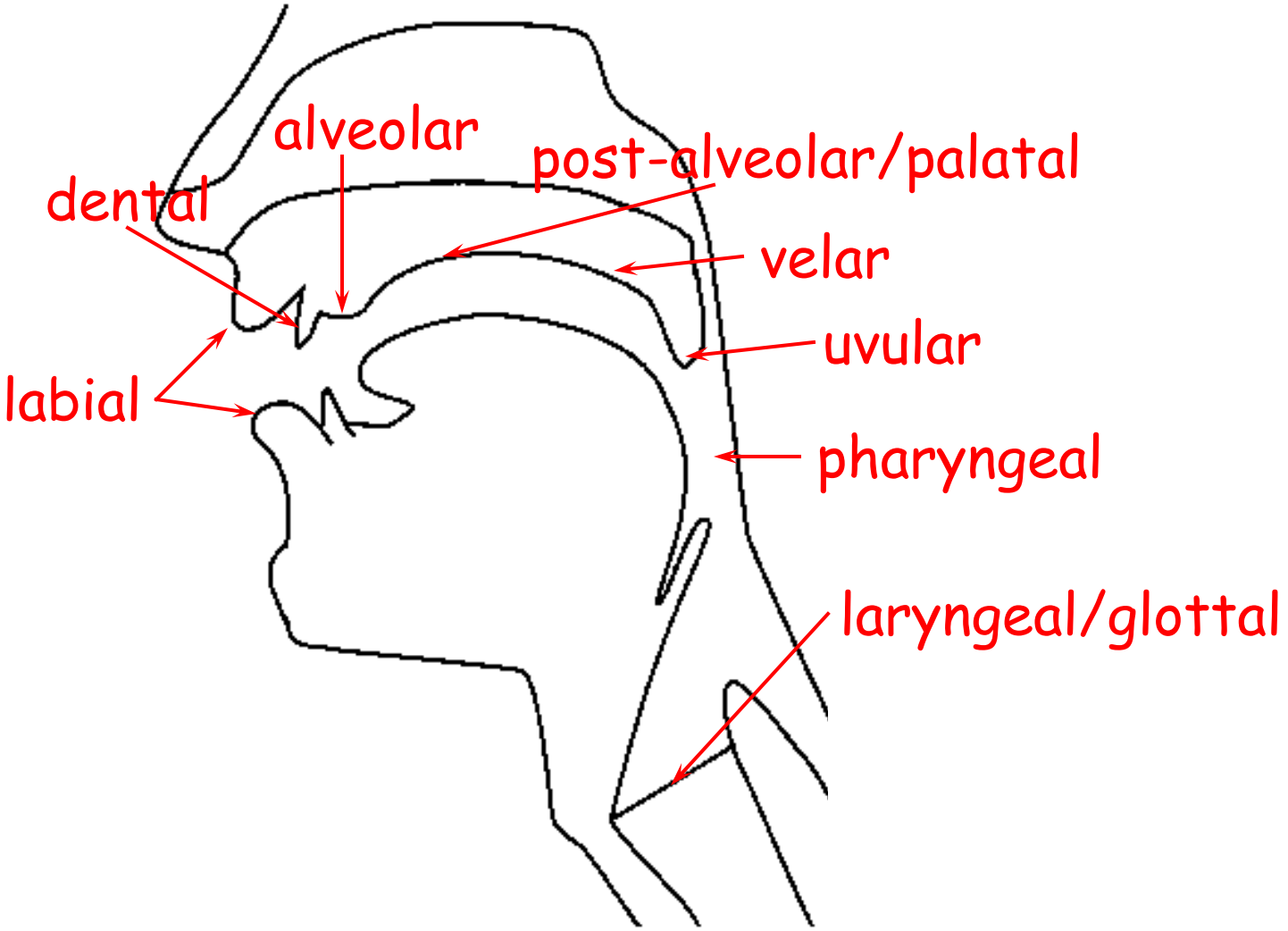
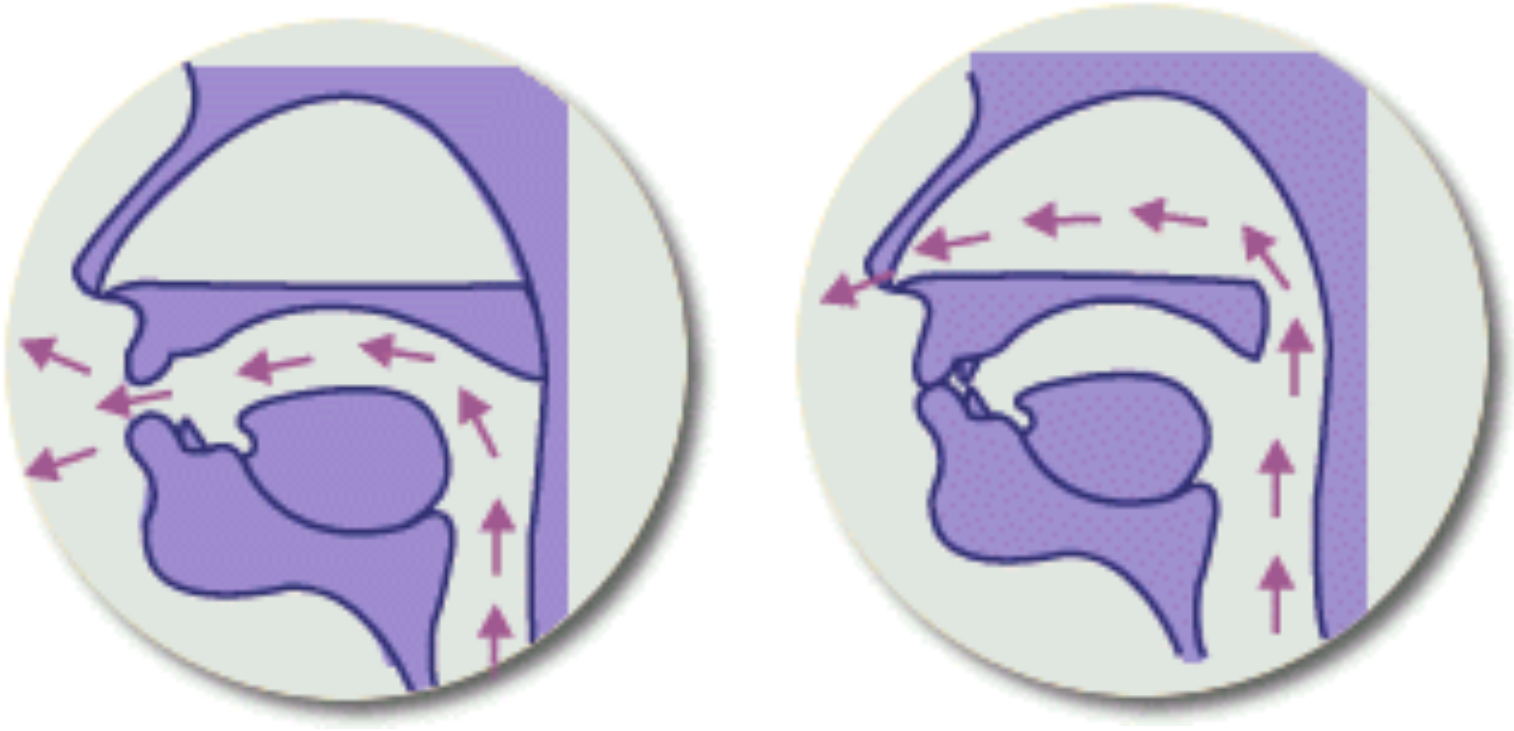


Figure thanks to Jennifer Venditti

Manner of Articulation

- Stop: complete closure of articulators, so no air escapes through mouth
- Oral stop: palate is raised, no air escapes through nose. Air pressure builds up behind closure, explodes when released
  - p, t, k, b, d, g
- Nasal stop: oral closure, but palate is lowered, air escapes through nose.
  - m, n, ng

Oral vs. Nasal Sounds



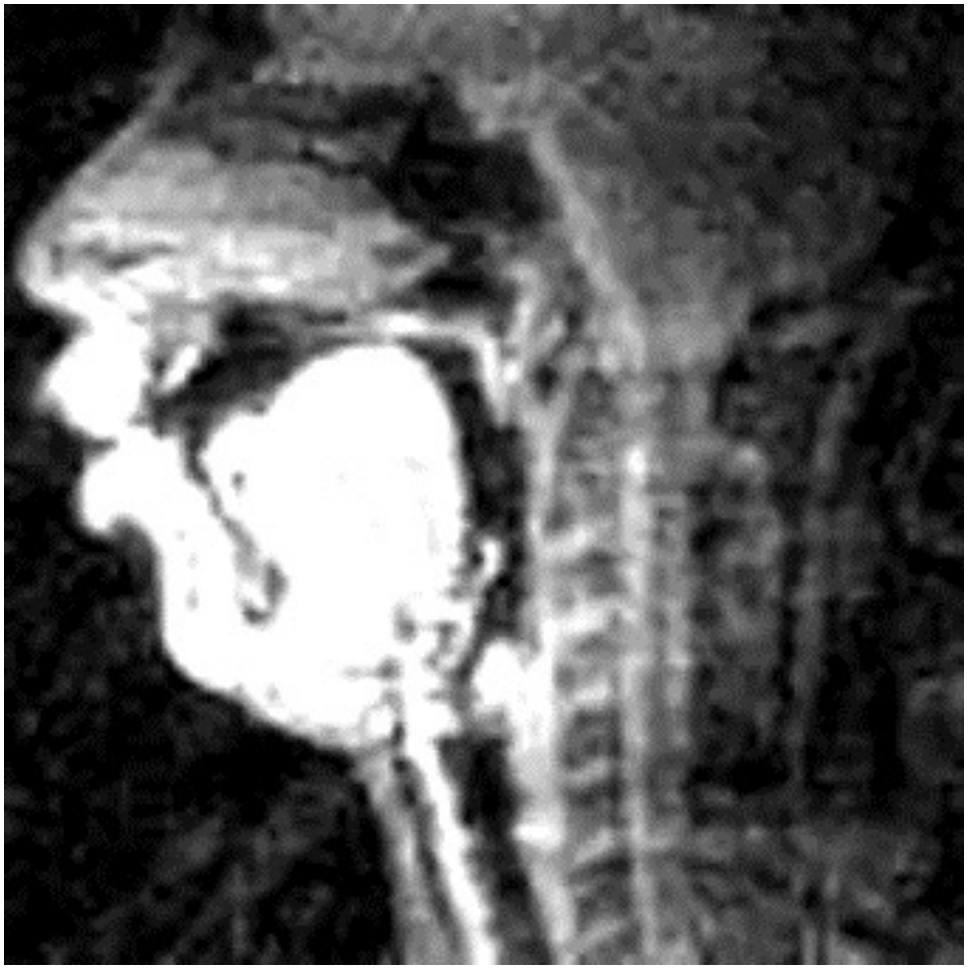
Thanks to Jong-bok Kim for this figure!

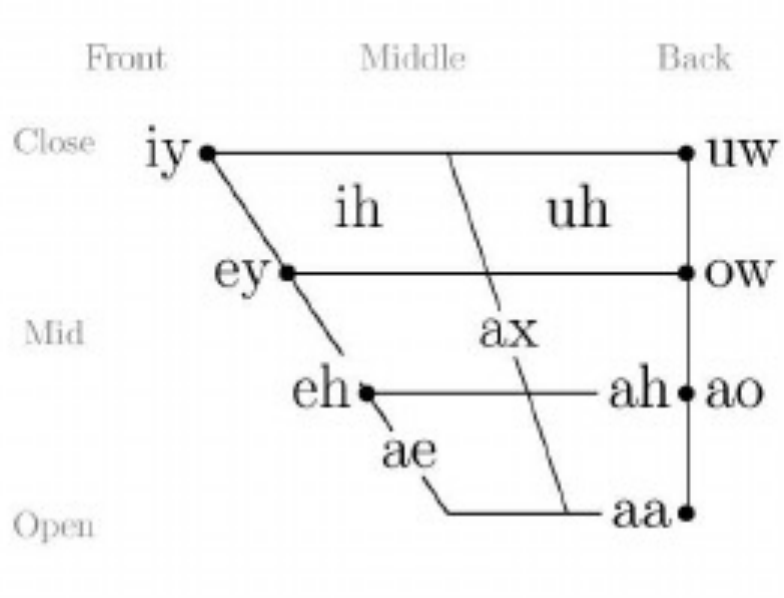
More on Manner of articulation of consonants

- Fricatives
  - Close approximation of two articulators, resulting in turbulent airflow between them, producing a hissing sound.
    - f, v, s, z, th, dh
- Approximant
  - Not quite-so-close approximation of two articulators, so no turbulence
    - y, r
- Lateral approximant
  - Obstruction of airstream along center of oral tract, with opening around sides of tongue.
    - l

Text from Ladefoged "A Course in Phonetics"

Tongue position for vowels





[iy] (bead) vs. [uw] (booed)

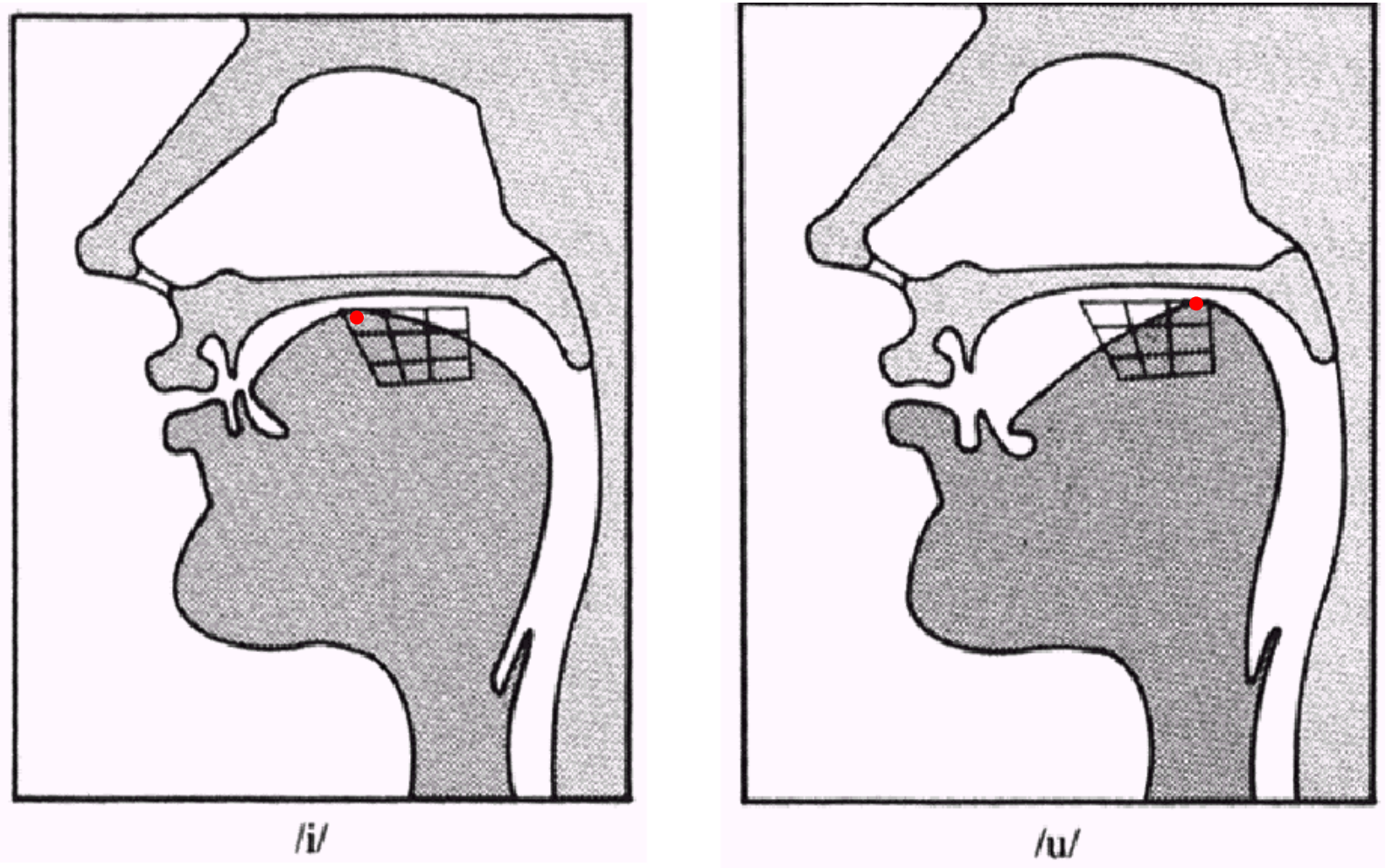


Figure from Jennifer Venditti, from a lecture given by Rochelle Newman

Articulatory parameters for English consonants (in ARPAbet)

MANNER OF ARTICULATION	PLACE OF ARTICULATION													
	bilabial		labio-dental		inter-dental		alveolar		palatal		velar		glottal	
stop	p	b					t	d			k	g	q	
fric.			f	v	th	dh	s	z	sh	zh			h	
affric.									ch	jh				
nasal		m						n				ng		
approx		w						l/r		y				
flap							dx							

VOICING: voiceless | voiced

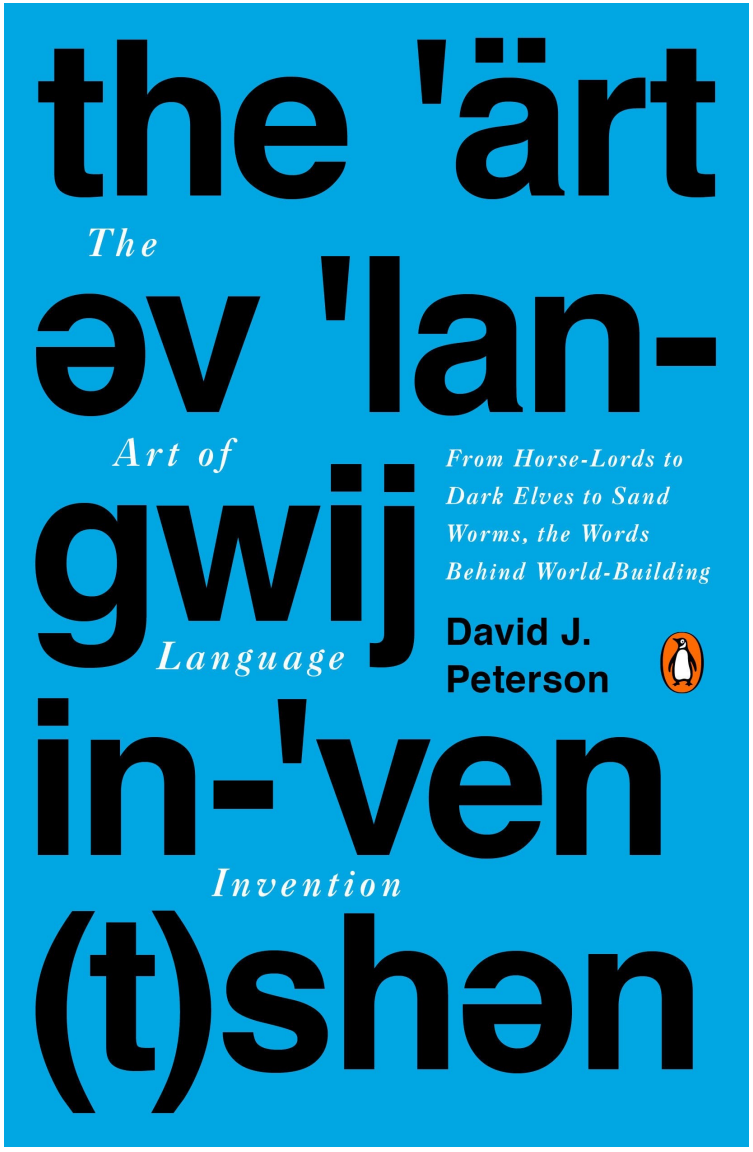
Table from Jennifer Venditti

The art of language invention

- Fun, informative book on phonetics and phonotactics across languages.

  Great audio book!

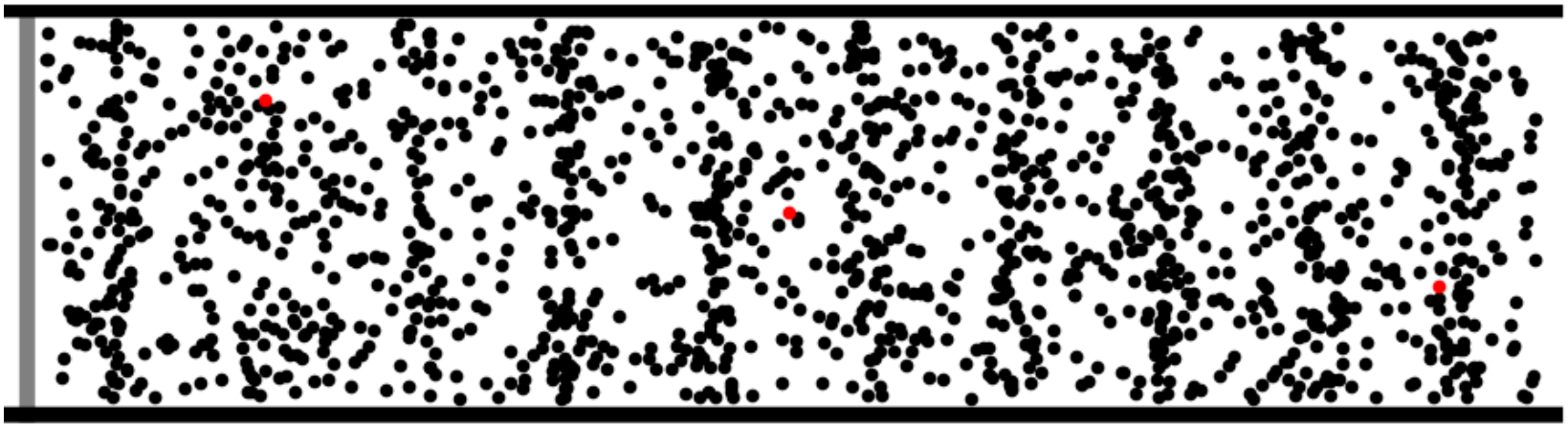
  Talk video



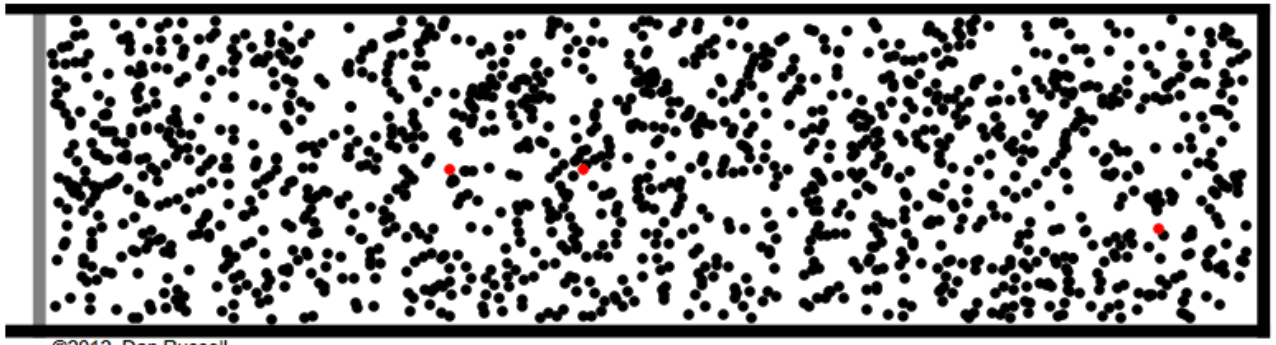
Phonetics

- ARPAbet
  - An alphabet for transcribing American English phonetic sounds.
- Articulatory Phonetics
  - How speech sounds are made by articulators (moving organs) in mouth.
- **Acoustic Phonetics**
  - **Acoustic properties of speech sounds**

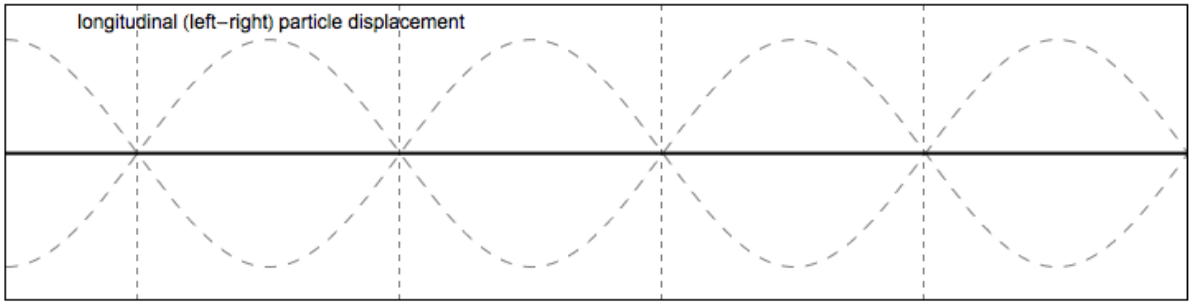
Sound waves are longitudinal waves



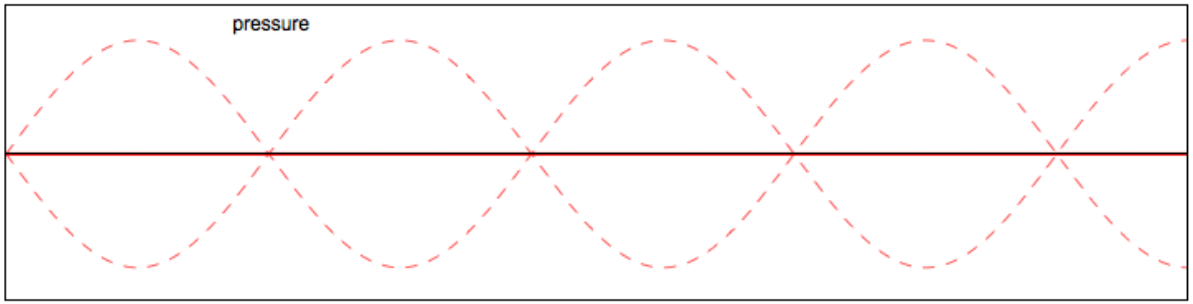
Dan Rusell Figure



particle dispacment



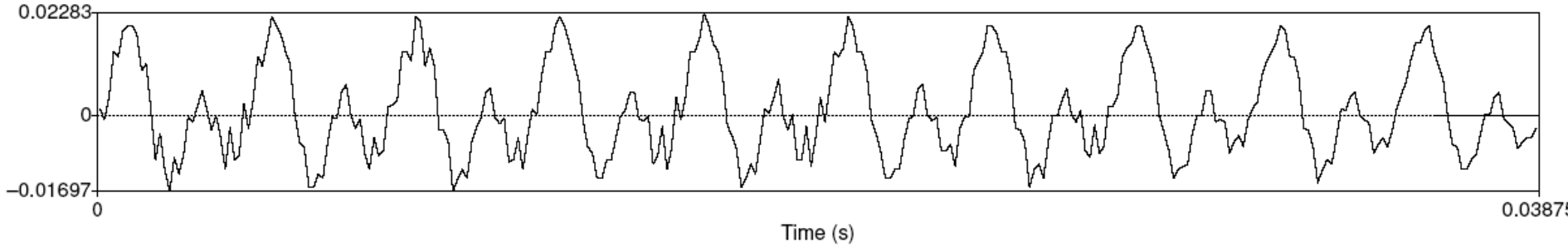
pressure



Dan Rusell Figure

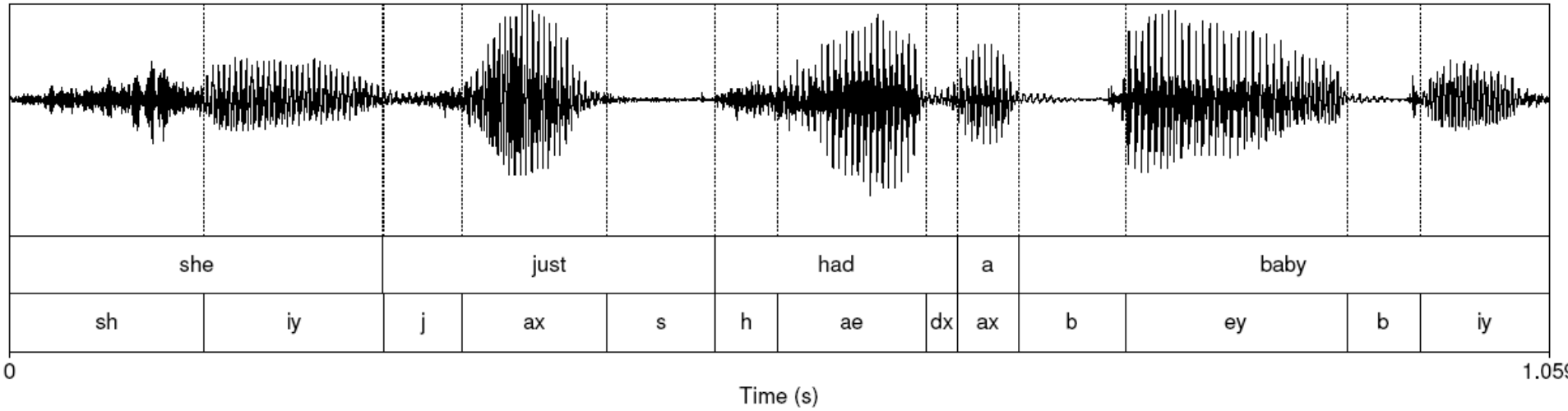
Back to waves: Fundamental frequency

- Waveform of the vowel [iy]



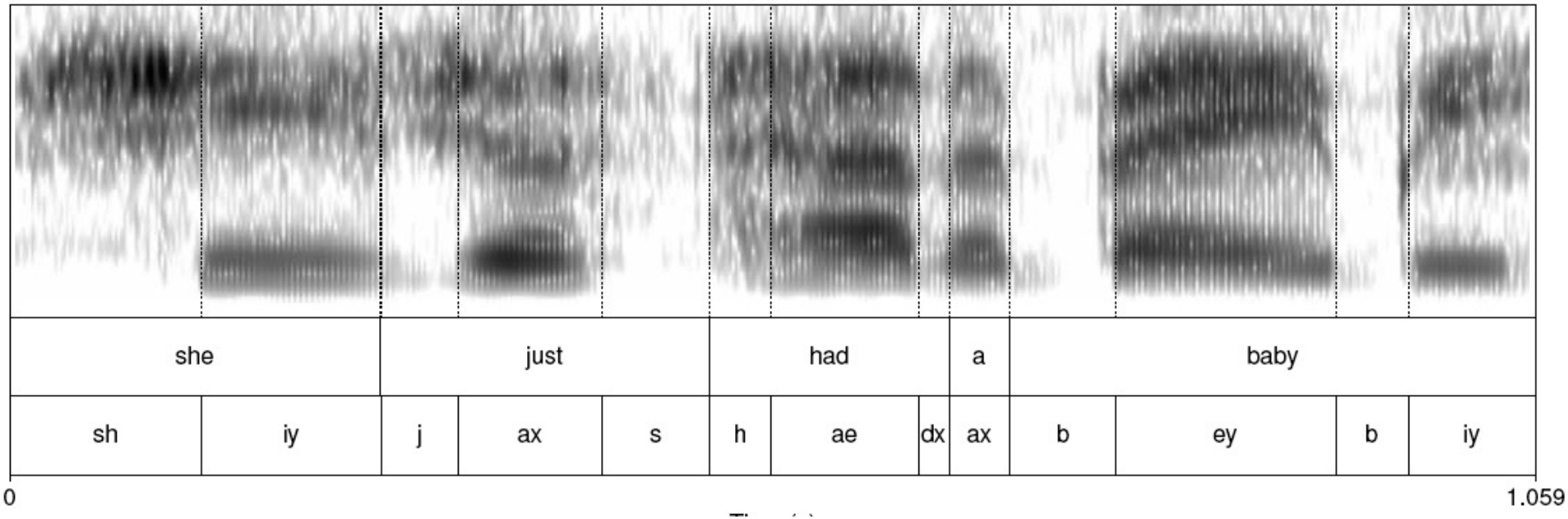
- Frequency: 10 repetitions / .03875 seconds = 258 Hz
- This is speed that vocal folds move, hence voicing
- Each peak corresponds to an opening of the vocal folds
- The low frequency of the complex wave is called the fundamental frequency of the wave or F0

She just had a baby



- Note that vowels all have regular amplitude peaks
- Stop consonant
  - Closure followed by release
  - Notice the silence followed by slight bursts of emphasis: very clear for [b] of “baby”
- Fricative: noisy. [sh] of “she” at beginning

Spectrogram: spectrum + time dimension

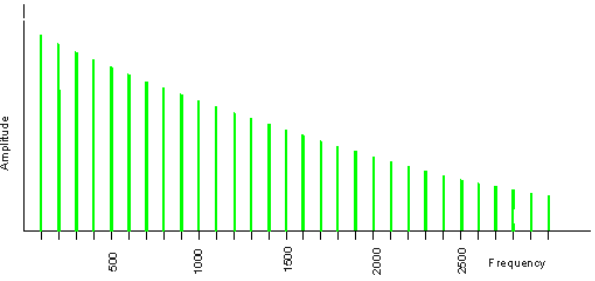


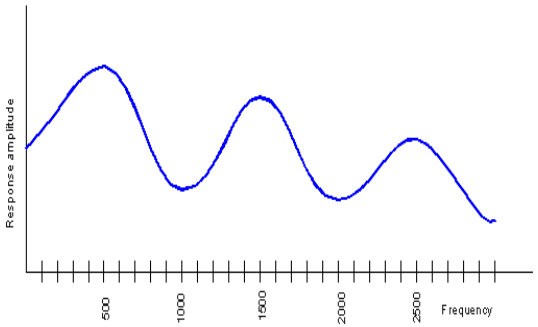
Source filter model of vowels

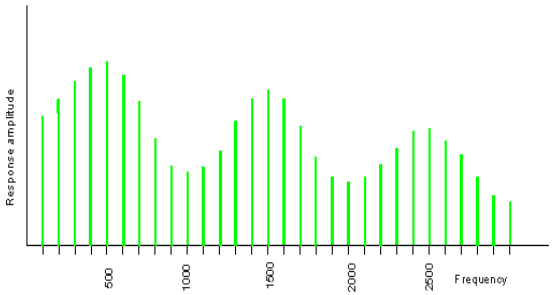
- Any body of air will vibrate in a way that depends on its size and shape.
- Vocal tract as "amplifier"; amplifies certain harmonics
- Formants are result of different shapes of vocal tract.

Source-filter model of speech production

Input → Filter → Output







Glottal spectrum

Vocal tract frequency response function

Source and filter are independent, so:

Different vowels can have same pitch

The same vowel can have different pitch

Figures and text from Ratree Wayland slide from his website

Resonances of the vocal tract

- The human vocal tract as an open tube

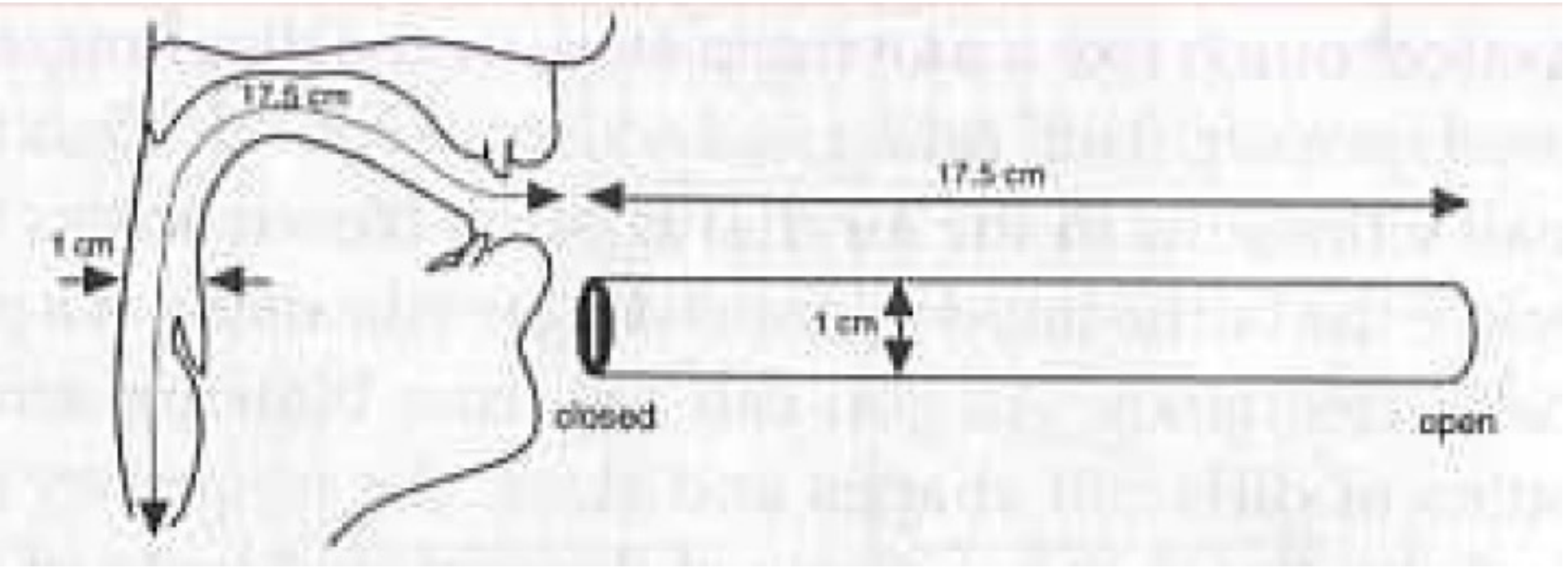
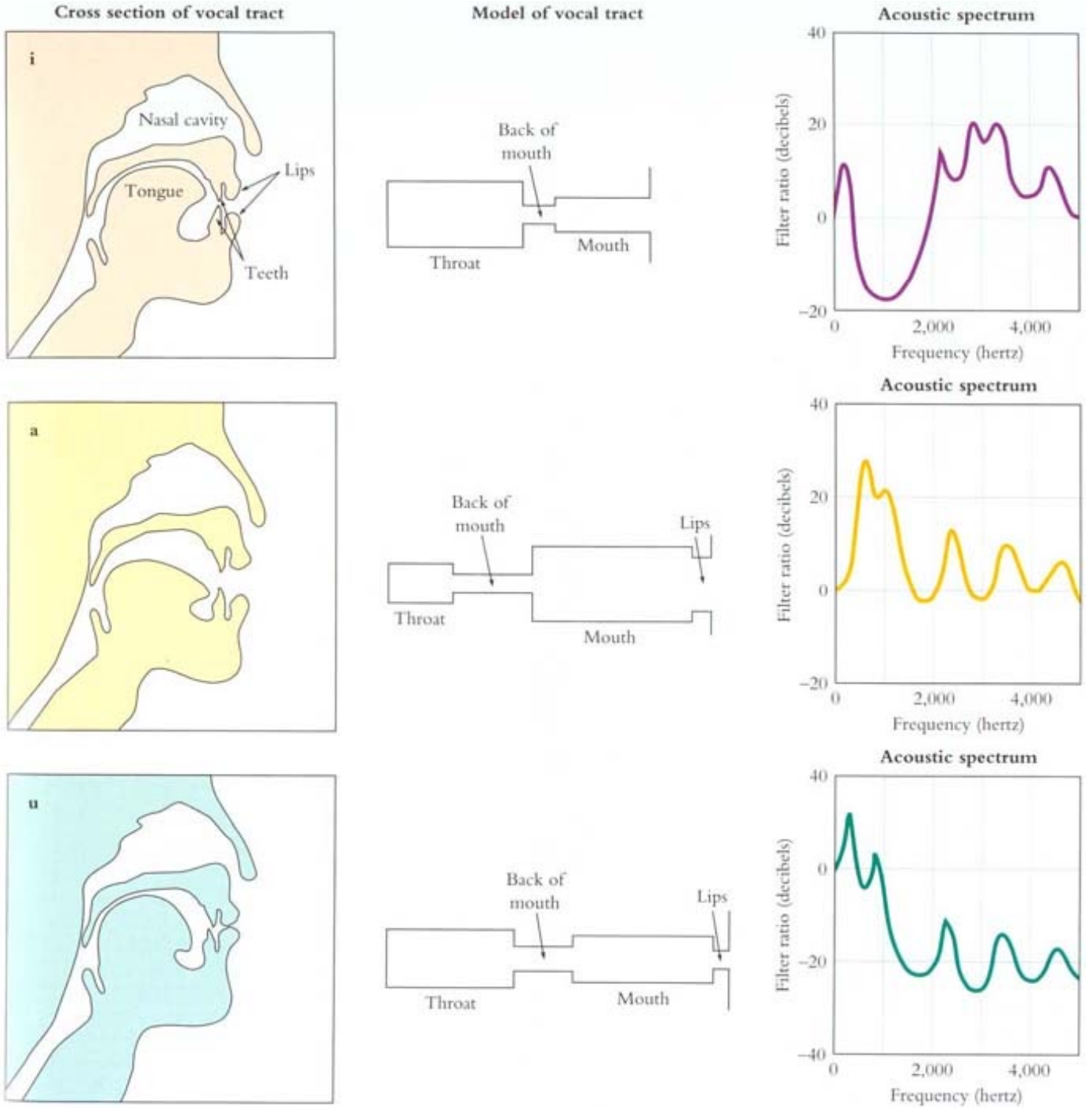


Figure from Ladefoged(1996) p 117

From Mark Liberman's Web site

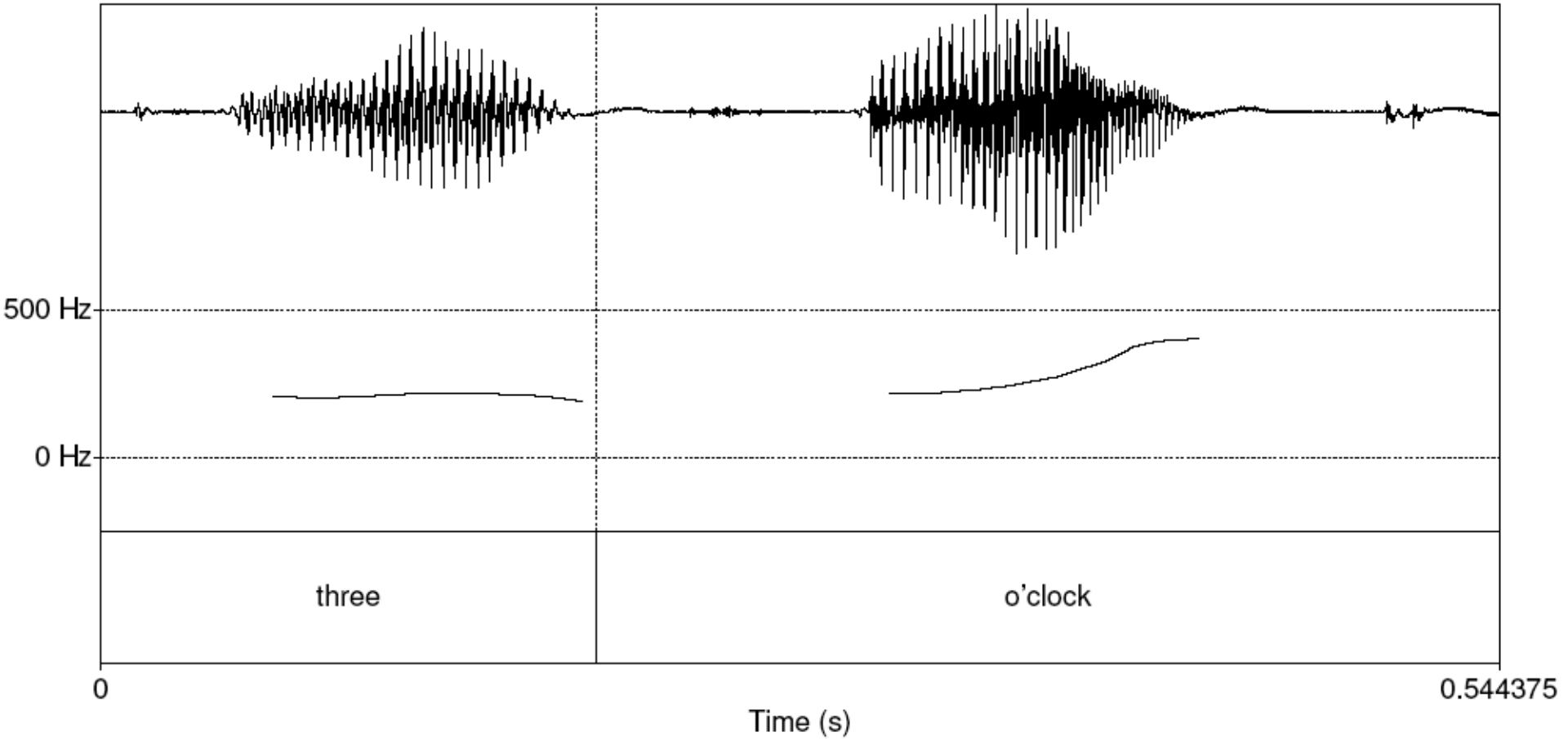


Defining Intonation

- Ladd (1996) "Intonational phonology"
- "The use of suprasegmental phonetic features

  Suprasegmental = above & beyond the segment/phone
  - F0
  - Intensity (energy)
  - Duration
- to convey sentence-level pragmatic meanings"
  - I.e. meanings that apply to phrases or utterances as a whole, not lexical stress, not lexical tone.

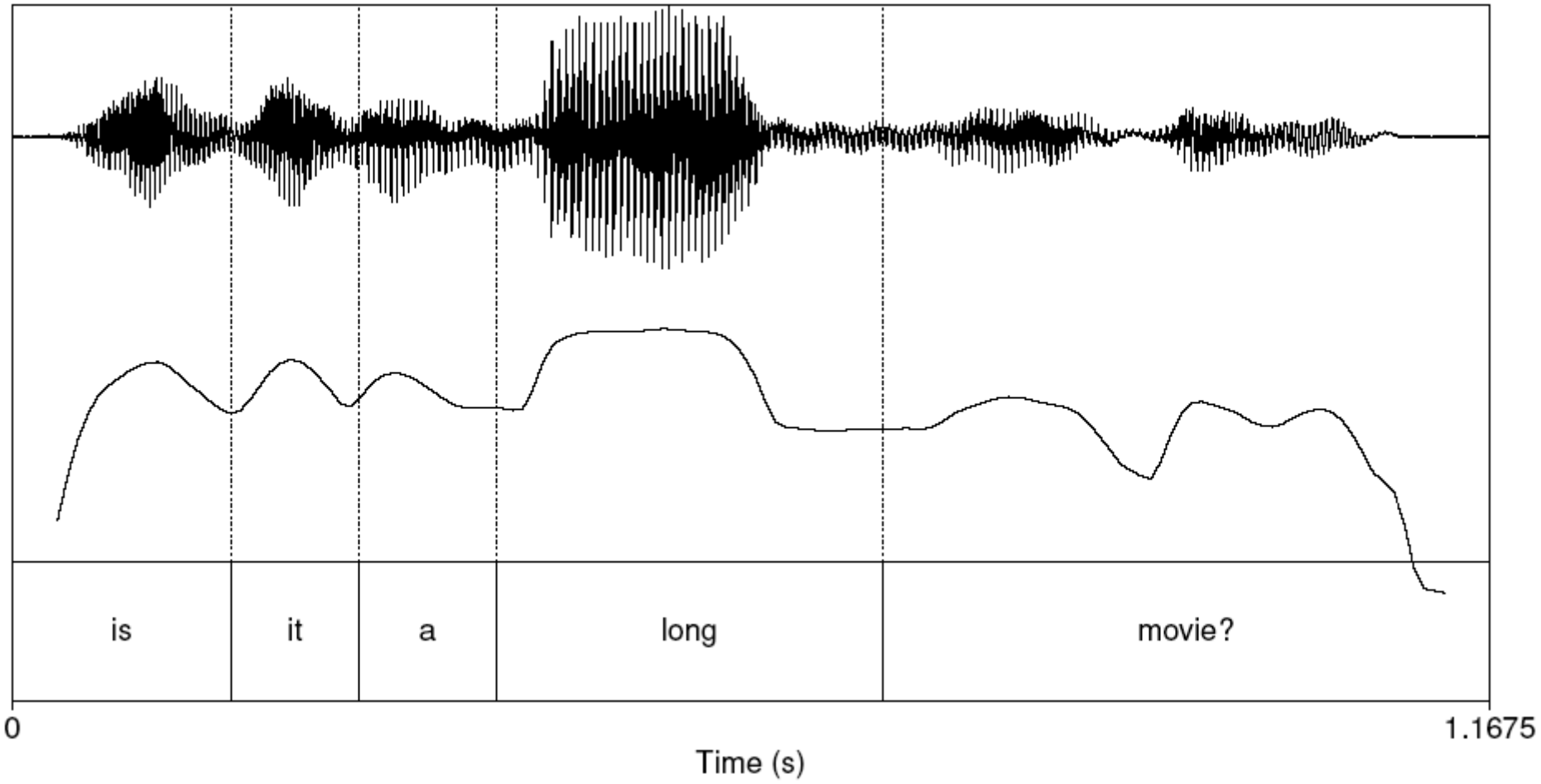
Pitch track



Pitch is not Frequency

- Pitch is the mental sensation or perceptual correlated of F0
- Relationship between pitch and F0 is not linear;
  - human pitch perception is most accurate between 100Hz and 1000Hz.
    - Linear in this range
    - Logarithmic above 1000Hz
- Mel scale is one model of this F0-pitch mapping
  - A mel is a unit of pitch defined so that pairs of sounds which are perceptually equidistant in pitch are separated by an equal number of mels
  - Frequency in mels = 1127 ln (1 + f/700)

Plot of Intensity



Three aspects of prosody

- Prominence: some syllables/words are more prominent than others
- Structure/boundaries: sentences have prosodic structure
  - Some words group naturally together
  - Others have a noticeable break or disjuncture between them
- Tune: the intonational melody of an utterance.

From Ladd (1996)

Prosodic Boundaries

I met Mary and Elena's mother at the mall yesterday.
I met Mary and Elena's mother at the mall yesterday.

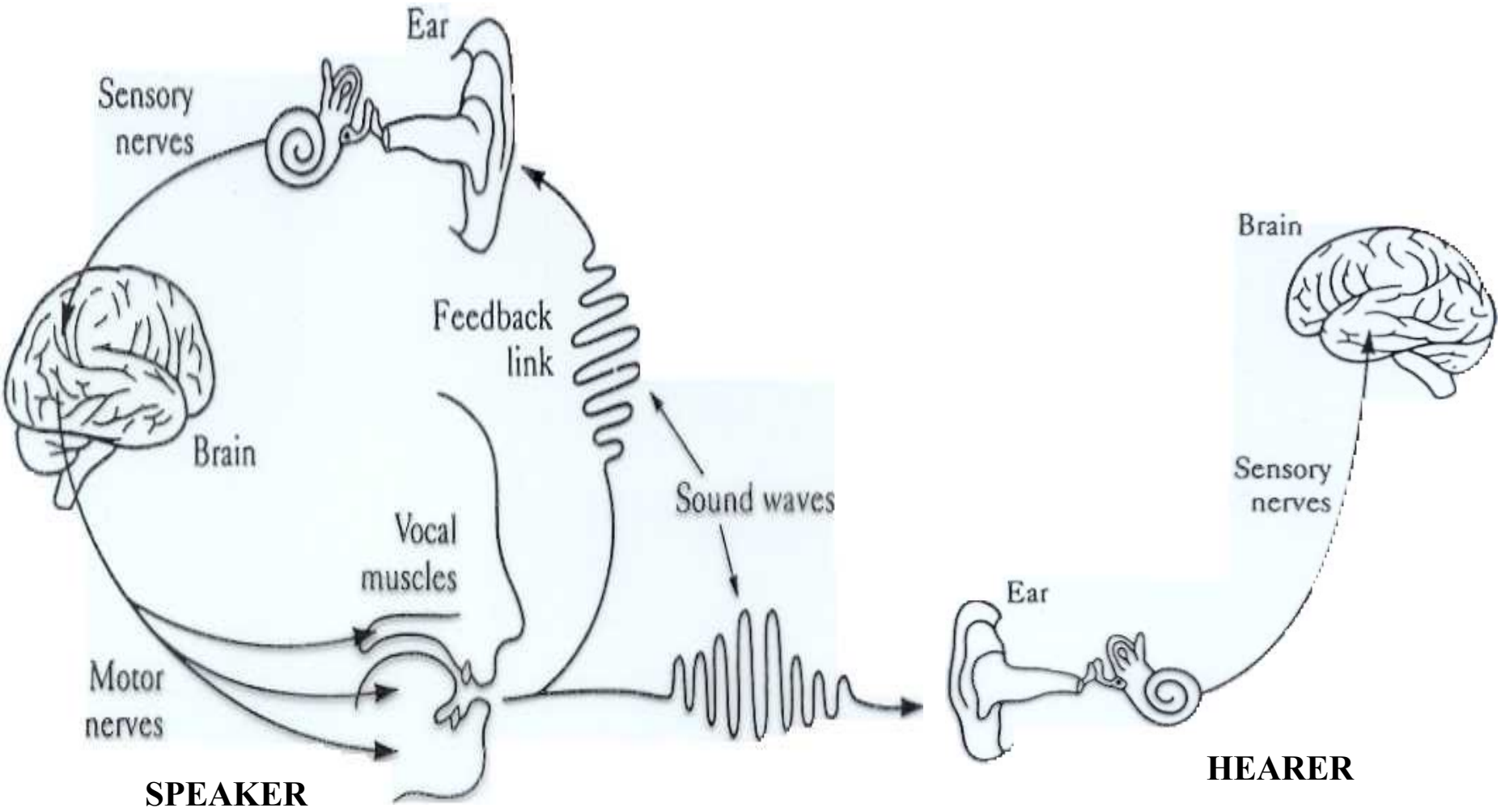
French [bread and cheese]

[French bread] and [cheese]

Slide from Jennifer Venditti

Appendix

The Speech Chain (Denes and Pinson)



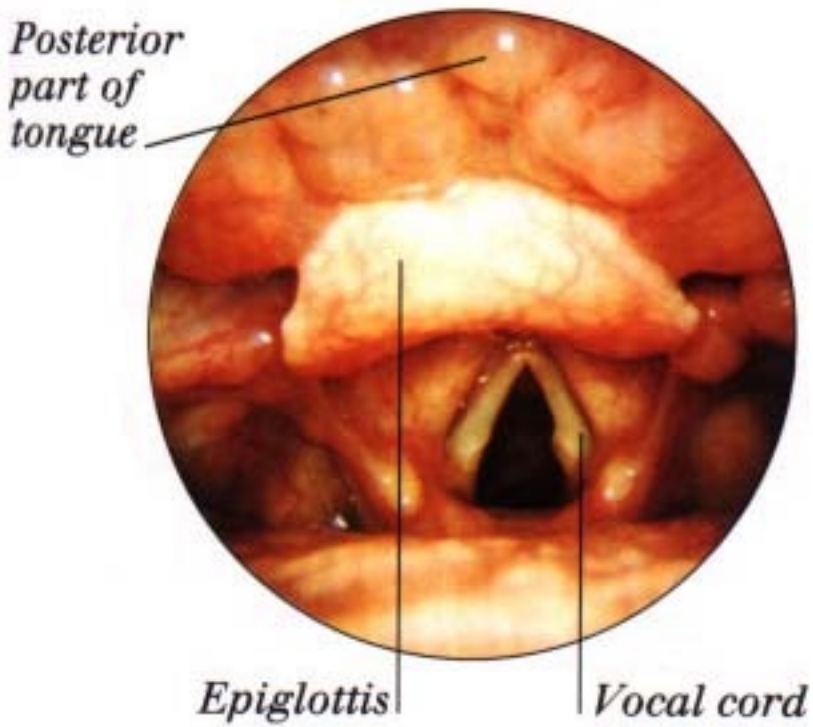
ARPAbet

- http://www.stanford.edu/class/cs224s/arpabet.html

- The CMU Pronouncing Dictionary
- http://www.speech.cs.cmu.edu/cgi-bin/cmudict

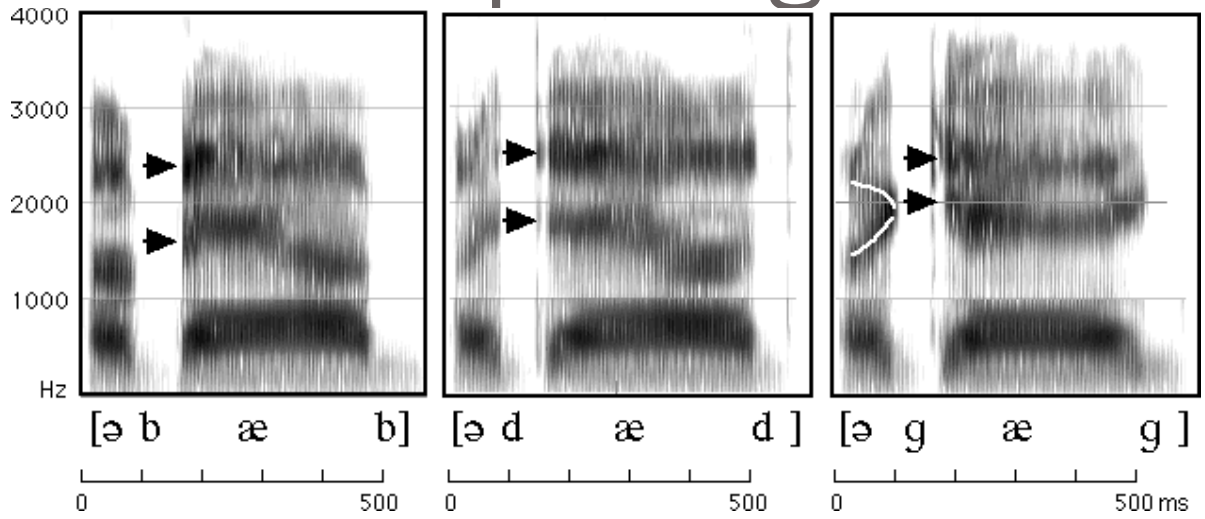
- International Phonetic Alphabet:
- http://en.wikipedia.org/wiki/International_Phonetic_Alphabet

Vocal folds open during breathing



From Mark Liberman's web site, from Ultimate Visual Dictionary

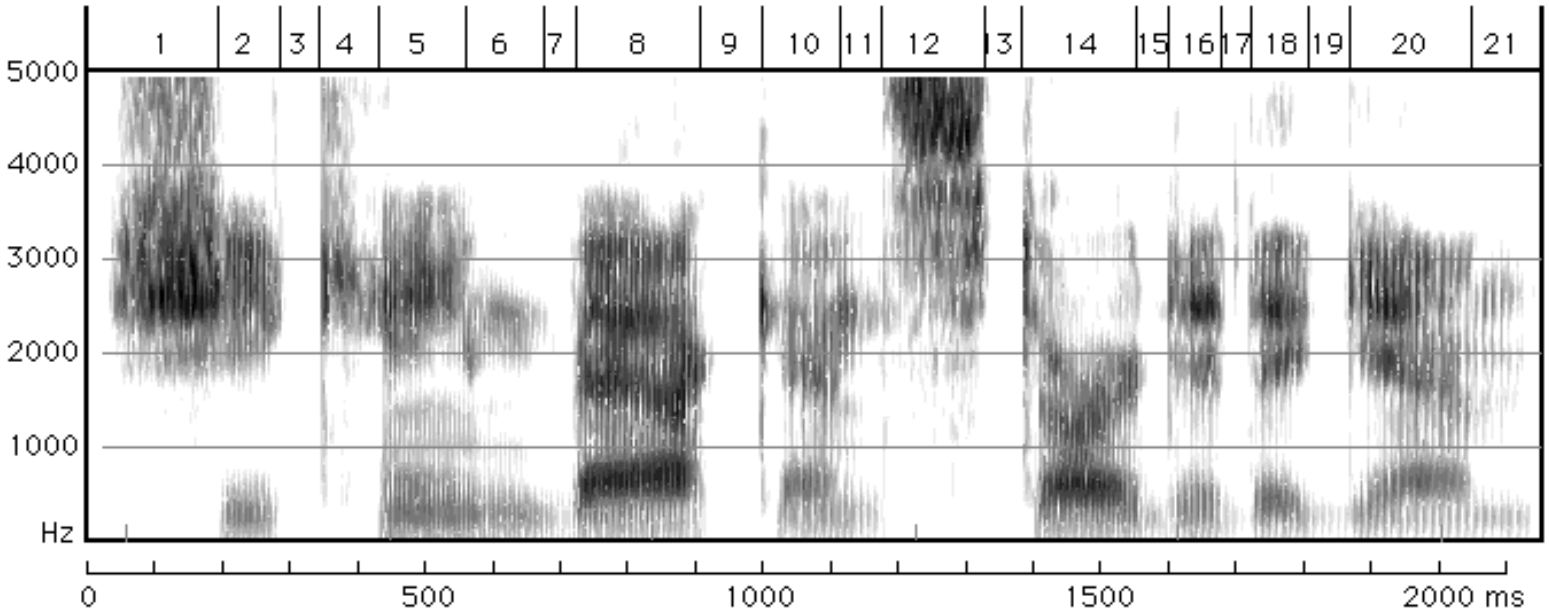
How to read spectrograms



- bab: closure of lips lowers all formants: so rapid increase in all formants at beginning of "bab"
- dad: first formant increases, but F2 and F3 slight fall
- gag: F2 and F3 come together: this is a characteristic of velars. Formant transitions take longer in velars than in alveolars or labials

From Ladefoged "A Course in Phonetics"

She came back and started again



- 1. lots of high-freq energy
- 3. closure for k
- 4. burst of aspiration for k
- 5. ey vowel;faint 1100 Hz formant is nasalization
- 6. bilabial nasal
- 7. short b closure, voicing barely visible.
- 8. ae; note upward transitions after bilabial stop at beginning
- 9. note F2 and F3 coming together for "k”

From Ladefoged “A Course in Phonetics”

More on manner of articulation of consonants

- Tap or flap
  - Tongue makes a single tap against the alveolar ridge
    - dx in “butter”
- Affricate
  - Stop immediately followed by a fricative
    - ch, jh

Vowels

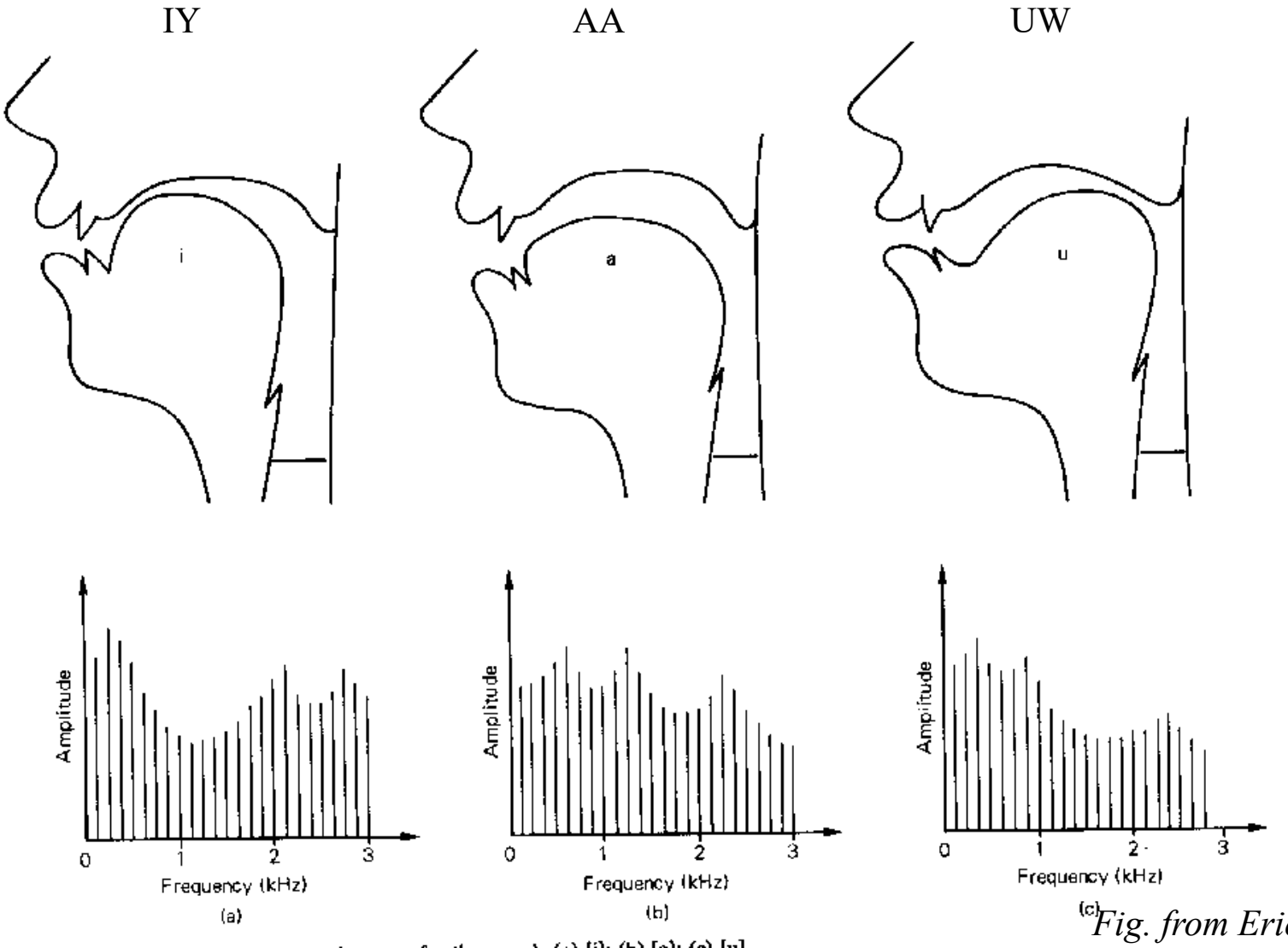


Fig. from Eric Keller

The oral cavity amplifies some harmonics