From Sounds to Language

CS 4706

Julia Hirschberg

Studying Linguistic Sounds

- Who studies speech sounds?
  - Linguists (phoneticians, phonologists, forensic), speech engineers (ASR, speaker id, dialect and language ID), speech pathologists, lexicographers, language teachers, singers, marketing experts,
- What questions do they ask?
  - What is the sound inventory of a language X?
  - How are they produced?
  - What sounds are *shared* by languages X and Y? Which are not?
  - How do particular sounds vary in context?

DRILLING DOWN

$2.22? Gosh, That Sounds Expensive

28.1 Average percentage discount perceived on a $3 product cut to $2.33.

24.13 Percentage discount perceived when the product is cut to $2.22.

Researchers have known for 80 years about a symbolic connection between speech and size: back-of-the-mouth vowels like the "o" in "two" make people think of large sizes, whereas people associate front-of-the-mouth vowels like "ee" with diminutiveness. Marketers can use this effect to make consumers think a discount is bigger or smaller than it truly is, according to a study soon to be published in The Journal of Consumer Research by Keith Coulter of Clark University and Robin Coulter of the University of Connecticut.

In one experiment, researchers told consumers the regular and sale prices of a product, asked them to repeat the sale price to themselves, and then, a few minutes later, told them to estimate the size of the discount in percentage terms. Products with "small-sounding" sale prices (like $2.33) seemed like better deals than products with "big-sounding" sales prices (like $2.22).

In another experiment, the researchers used a pair of sale prices — $7.88, which sounds "big" in English, and $7.01, which sounds "small" — but are the other way around in Chinese. Chinese and English speakers had opposite perceptions of the products' relative value.

ALEX MINDLIN

LOOKING AHEAD

...s include: corp, Goldman Sachs, Google, ...lth

Rena

By DAV

PARIS — C... chief executi... sought to defus... conflict with th... ment, telling ... Sarkozy over ... the automaker ... duction of its n... tween France ... would produce ... vehicle entirel... Paris.

The French ... largest share... of a little m... Mr. Sarkoz... minister, C... been pres... maintain ... French f... summon... meeting ... to a st... Palace ... dustri...

"I ... pres... the ... Mr ... te... "... p... s...

How do we represent speech sounds?

- Why do we need to have representations?
  - Translating between sounds and words (ASR, TTS), learning pronunciation, talking about language similarities and differences,…
- How should we represent sounds?
  - Regular orthography
  - Special-purpose symbol sets
  - Abstract sound classes based upon sound similarities

Trying Orthographic Representation

- **A single letter may have many different acoustic realizations, e.g., in English**
  - o comb, tomb, bomb
  - oo blood, food, good
  - c court, center, cheese
  - s reason, surreal, shy
- **A single sound may have different orthographic correspondences**
  - [i] sea, see, scene, receive, thief
  - [s] cereal, same, miss
  - [u] true, few, choose, lieu, do
  - [ay] prime, buy, rhyme, lie
- **Is orthography a good choice for English?**

Solution: Phonetic Symbol Sets

- International Phonetic Alphabet (IPA)
  - Single character for each sound
  - Represents all sounds of the world's languages but is quite large and requires special fonts
- ARPAbet, TIMIT, …
  - Multiple characters for sounds but ASCII
  - English specific, so new symbol sets required for each new language to be represented

IPA Symbol	ARPAbet Symbol	Word	IPA Transcription	ARPAbet Transcription
[p]	[p]	parsley	[ˈparsl̩i]	[p aa r s l iy]
[t]	[t]	tarragon	[ˈtærəgɑn]	[t ae r ax g aa n]
[k]	[k]	catnip	[ˈkætn̩ip]	[k ae t n ix p]
[b]	[b]	bay	[beɪ]	[b ey]
[d]	[d]	dill	[dɪl]	[d ih l]
[g]	[g]	garlic	[ˈgɑrlɨk]	[g aa r l ix k]
[m]	[m]	mint	[mɪnt]	[m ih n t]
[n]	[n]	nutmeg	[ˈnʌtmɛg]	[n ah t m eh g
[ŋ]	[ng]	ginseng	[ˈdʒɪnsɨŋ]	[jh ih n s ix ng]
[f]	[f]	fennel	[ˈfɛnl̩]	[f eh n el]
[v]	[v]	clove	[kloʊv]	[k l ow v]
[θ]	[th]	thistle	[ˈθɪsl̩]	[th ih s el]
[ð]	[dh]	heather	[ˈhɛðɚ]	[h eh dh axr]
[s]	[s]	sage	[seɪdʒ]	[s ey jh]
[z]	[z]	hazelnut	[ˈheɪzl̩nʌt]	[h ey z el n ah t]
[ʃ]	[sh]	squash	[skwɑʃ]	[s k w a sh]
[ʒ]	[zh]	ambrosia	[æmˈbroʊʒə]	[ae m b r ow zh ax]
[tʃ]	[ch]	chicory	[ˈtʃɪkɚi]	[ch ih k axr iy]
[dʒ]	[jh]	sage	[seɪdʒ]	[s ey jh]
[l]	[l]	licorice	[ˈlɪkɚɨʃ]	[l ih k axr ix sh]
[w]	[w]	kiwi	[ˈkiwi]	[k iy w iy]
[ɹ]	[r]	parsley	[ˈpɑrsli]	[p aa r s l iy]
[j]	[y]	yew	[yu]	[y uw]
[h]	[h]	horseradish	[ˈhɔrsrædɪʃ]	[h ao r s r ae d ih sh]
[ʔ]	[q]	uh-oh	[ʔʌʔoʊ]	[q ah q ow]
[ɾ]	[dx]	butter	[ˈbʌɾɚ]	[b ah dx axr]
[ɾ̃]	[nx]	wintergreen	[wɪɾ̃ɚgrin]	[w ih nx axr g r i n]
[l̩]	[el]	thistle	[ˈθɪsl̩]	[th ih s el]

IPA Symbol	ARPAbet Symbol	Word	IPA Transcription	ARPAbet Transcription
[i]	[iy]	lily	[ˈlɪli]	[l ih l iy]
[ɪ]	[ih]	lily	[ˈlɪli]	[l ih l iy]
[eɪ]	[ey]	daisy	[ˈdeɪzi]	[d ey z i]
[ɛ]	[eh]	poinsettia	[pɔɪnˈsɛɹiə]	[p oy n s eh dx iy ax]
[æ]	[ae]	aster	[ˈæstɚ]	[ae s t axr]
[ɑ]	[aa]	poppy	[ˈpɑpi]	[p aa p i]
[ɔ]	[ao]	orchid	[ˈɔrkɨd]	[ao r k ix d]
[ʊ]	[uh]	woodruff	[ˈwʊdrʌf]	[w uh d r ah f]
[oʊ]	[ow]	lotus	[ˈloʊɾəs]	[l ow dx ax s]
[u]	[uw]	tulip	[ˈtulɨp]	[t uw l ix p]
[ʌ]	[uh]	buttercup	[ˈbʌɾɚˌkʌp]	[b uh dx axr k uh p]
[ɝ]	[er]	bird	[ˈbɝd]	[b er d]
[aɪ]	[ay]	iris	[ˈaɪɹɨs]	[ay r ix s]
[aʊ]	[aw]	sunflower	[ˈsʌnflaʊɚ]	[s ah n f l aw axr]
[ɔɪ]	[oy]	poinsettia	[pɔɪnˈsɛɹiə]	[p oy n s eh dx iy ax]
[ju]	[y uw]	feverfew	[fivɚfju]	[f iy v axr f y u]
[ə]	[ax]	woodruff	[ˈwʊdrəf]	[w uh d r ax f]
[ɨ]	[ix]	tulip	[ˈtulɨp]	[t uw l ix p]
[ɚ]	[axr]	heather	[ˈhɛðɚ]	[h eh dh axr]
[ʉ]	[ux]	dude[1]	[dʉd]	[d ux d]

Exercise:
Write your full name in English orthography and in ARPAbet.

Sound Categories

- Phone: Basic speech sound of a language
  - A minimal sound difference between two words (e.g. *too, zoo*)
  - Not every human sound is phonetic, e.g.
    - Sniffs, laughs, coughs,…
- Phoneme: Class of speech sounds
  - Phoneme may include several phones (e.g. the /t/ in *top, stop, little, butter, winter*)
- Allophone: the set of phonetic variants that comprise a phoneme, e.g. {[t], [ɾ],…}

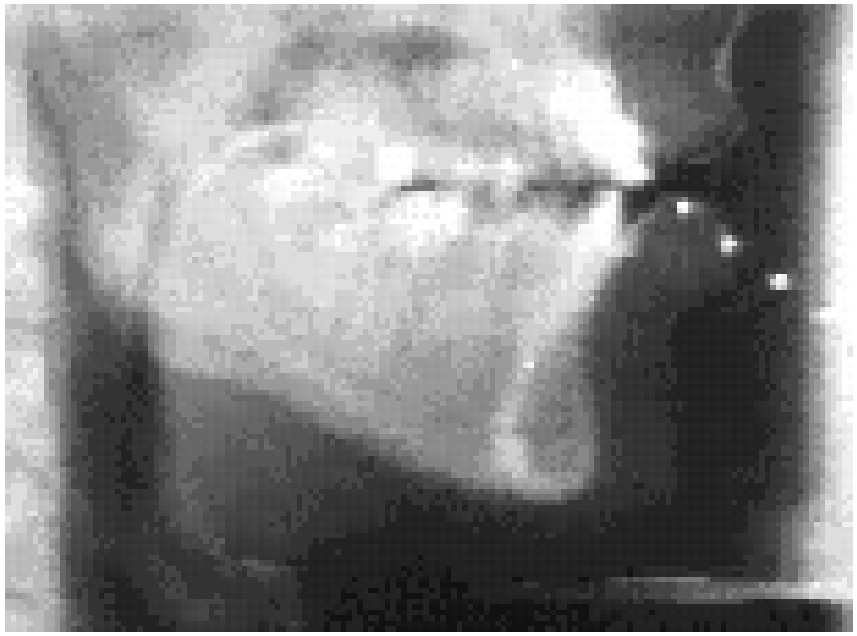
Articulatory Phonetics: How do people produce speech?

- The articulatory organs
- General process:
  - Air expelled from lungs through windpipe (trachea) leaving via mouth (mostly) and nose (nasals) (e.g. [m], [n])
  - Air passing thru trachea goes thru larynx, which contains vocal folds – space between them is glottis
  - When vocal folds vibrate, we get voiced sounds (e.g. [v]); o.w. voiceless (e.g. [f])

Vocal fold vibration

[UCLA Phonetics Lab demo]

Articulators in action



(Sample from the Queen's University / ATR Labs X-ray Film Database)

"Why did Ken set the soggy net on top of his deck?" Other examples

How do we capture articulatory data?

- X-ray/pellet film archive
- X-Ray Microbeam Database
  - Sample output (English: *light*)
- Electroglottography
- Electromagnetic articulography (EMMA)
  - 3 transmitters on helmet produce alternating magnetic fields at different frequencies, forming equilateral triangle
  - Creates alternating current in 5-15 sensors to calculate sensor positions via XY coordinates
  - Sample output

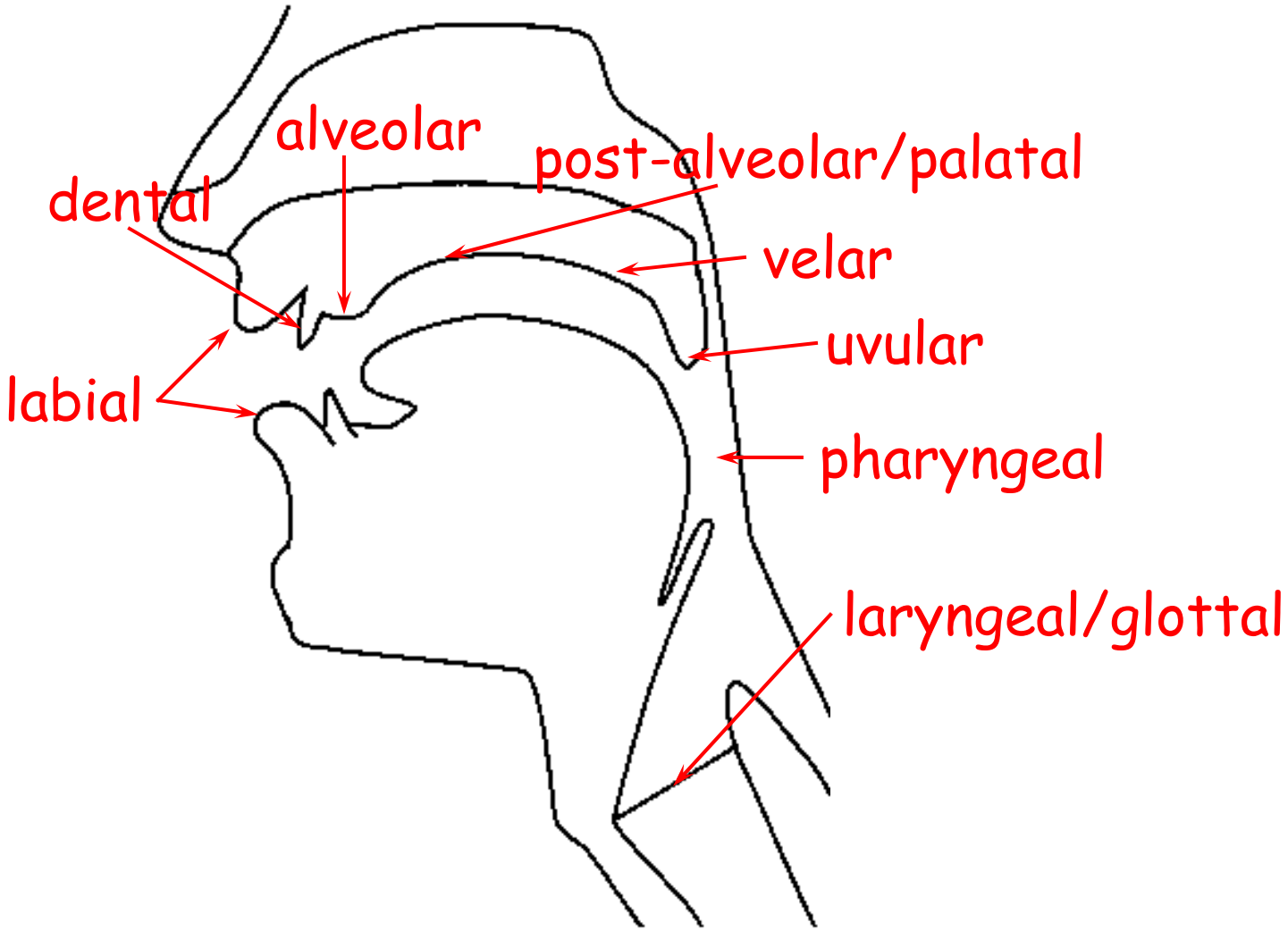
Classes of Sounds

- Consonants and vowels:
  - Consonants:
    - Restriction/blockage of air flow (e.g. [s])
    - Voiced or voiceless [s] vs. [z]
  - Vowels:
    - Generally voiced, less restriction (e.g. [u]
  - Semivowels (glides): [w], [y]

Consonants: *Place of Articulation*

- What is the point of maximum (air) restriction?
  - Labial: bilabial [b], [p]; labiodental [v], [f]
  - Dental: [θ], [ð] thief vs. them
  - Alveolar: [t], [d], [s], [z]
  - Palatal: [ʃ], [tʃ] shrimp vs. chimp
  - Velar: [k], [g]
  - Glottal: [?] glottal stop

Places of articulation



http://www.chass.utoronto.ca/~danhall/phonetics/sammy.html

Consonants: *Manner of Articulation*

- ***How*** is the airflow restricted?
  - Stop: [p],[t],[g],… aka plosive
    - Airflow completely blocked (closure), then released (release)
    - Glottal stop, e.g. before word-initial vowels in English after pause (*extra*)
  - Nasal: air released thru nose [m],[ng],…
  - Fricative: [s], [z], [f] air forced thru narrow channel
  - Affricates [tʃ] begin as stops and end as fricatives

- Approximant: [w],[y]
  - 2 articulators come close but don't restrict much
  - Between vowels and consonants
  - Lateral: [l]
- Tap or flap: [ɾ] e.g. *butter*

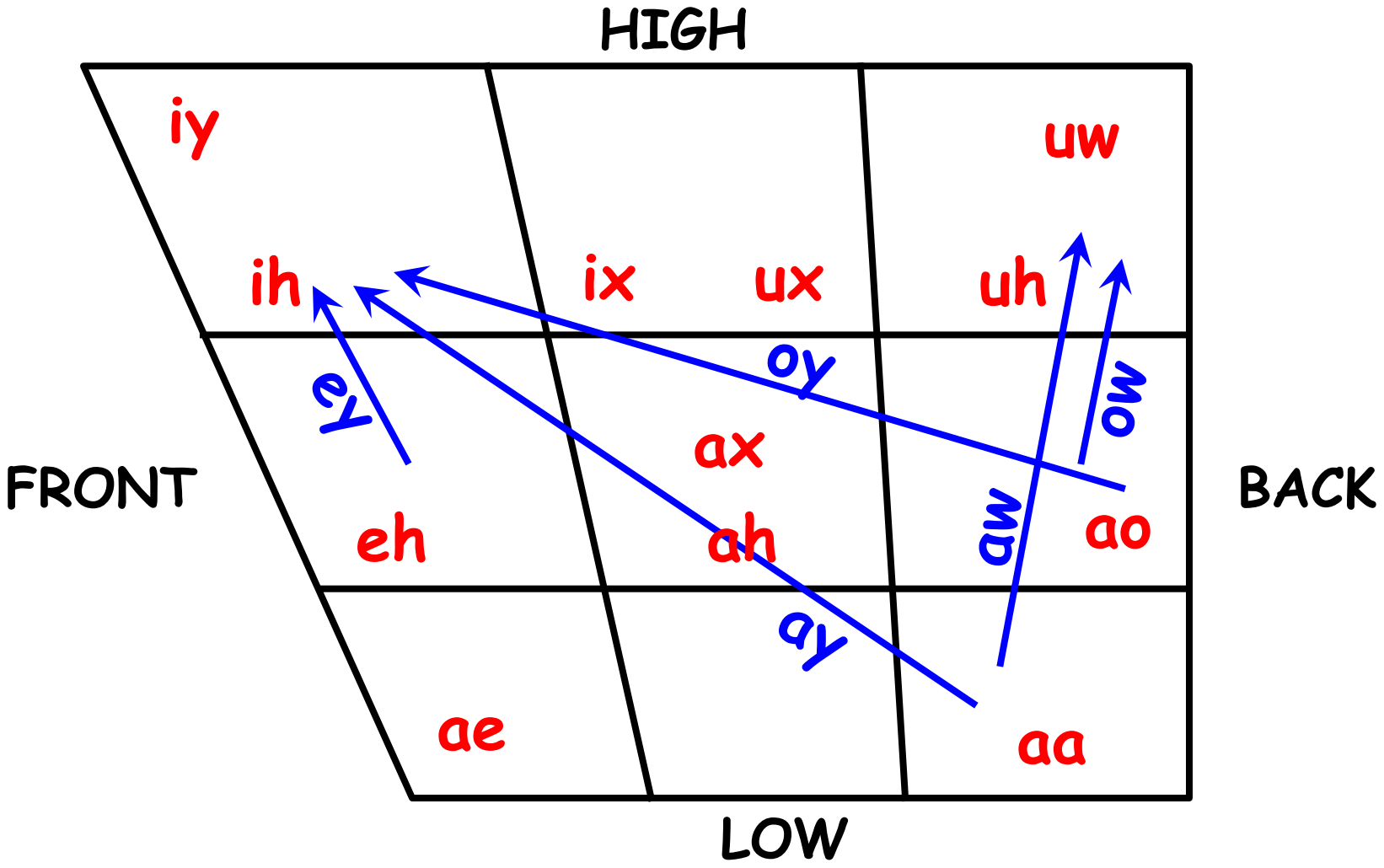
		PLACE OF ARTICULATION													
		bilabial		labio-dental		inter-dental		alveolar		palatal		velar		glottal	
MANNER OF ARTICULATION	stop	p	b					t	d			k	g	q	
	fric.			f	v	th	dh	s	z	sh	zh			h	
	affric.									ch	jh				
	nasal		m						n				ng		
	approx		w						l/r		y				
	flap							dx							

VOICING: voiceless | voiced

Vowels

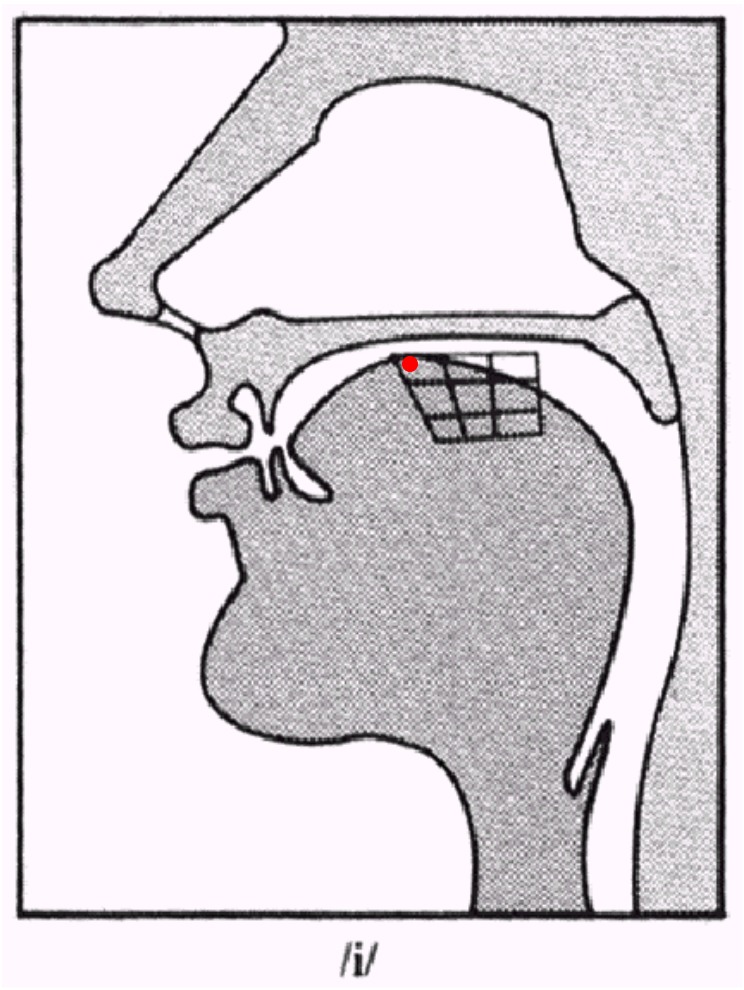
- All voiced
- Vowel height
  - How ***high*** is the ***tongue***? high or low vowel
  - ***Where*** is its highest point? front or back vowel
- How ***rounded*** are the ***lips***?
- Mono- [eh] vs. diphthong, e.g. [ey]
  - 1 vowel sound or 2?

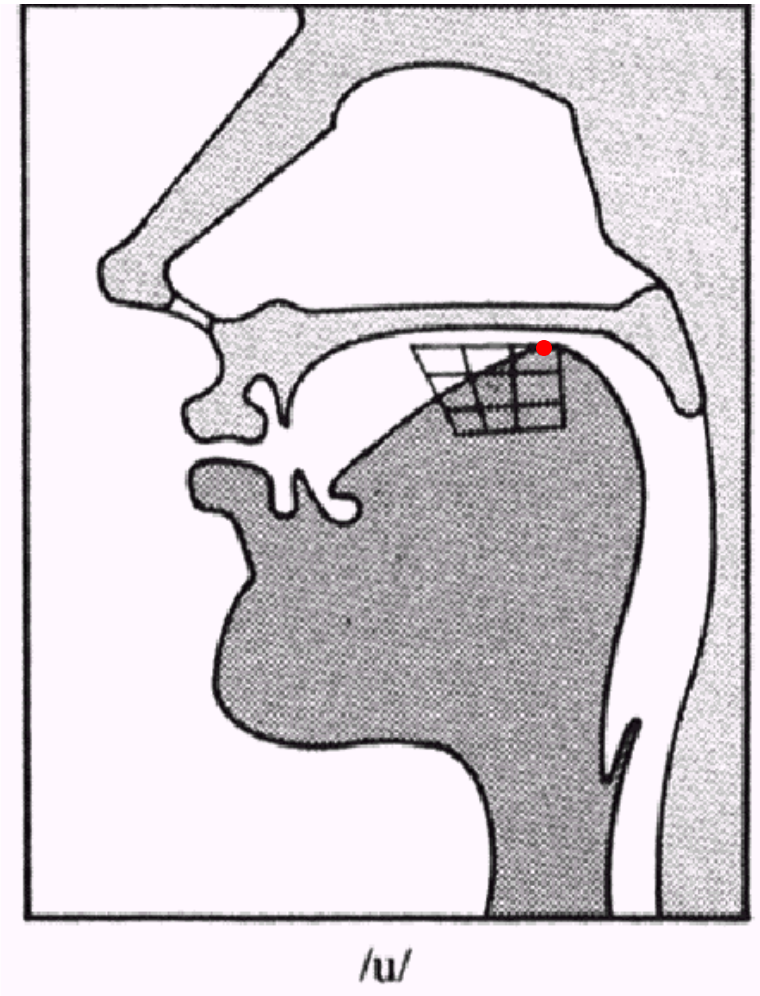
American English vowel space



- Compare to British English, Indian English, Swedish, Spanish, Japanese, Mandarin?

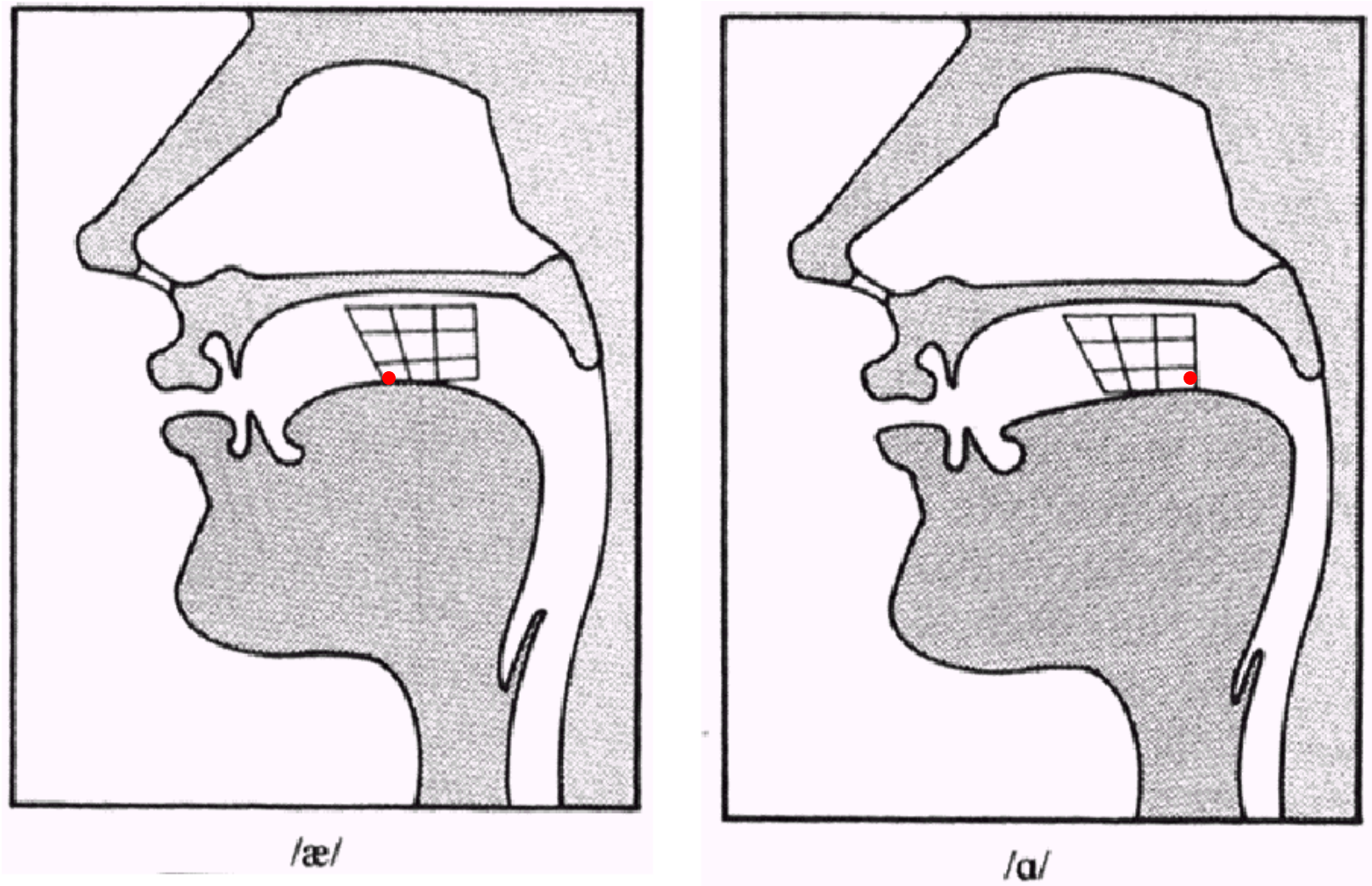
[iy] vs. [uw]





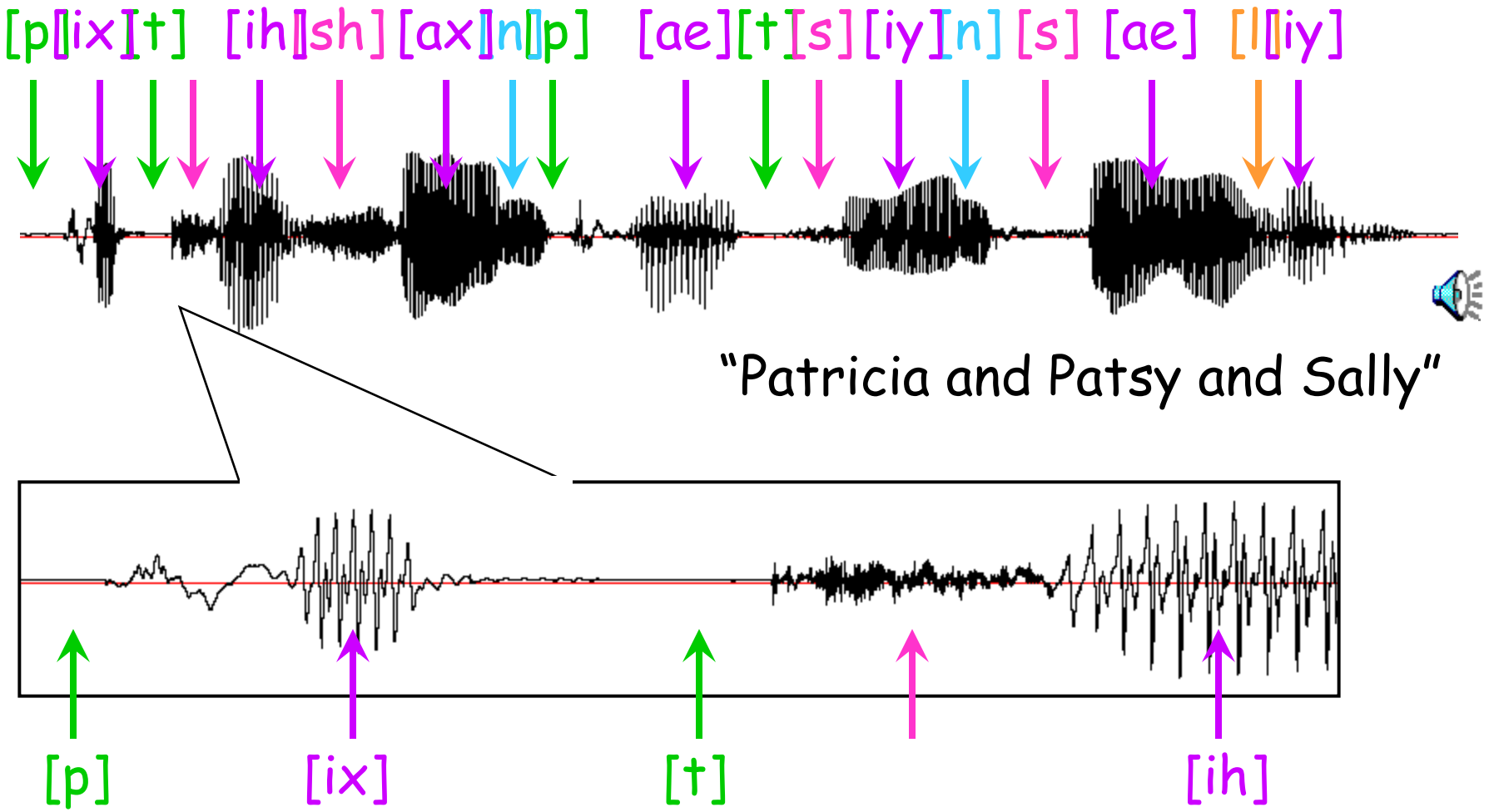
(From a lecture given by Rochelle Newman)

[ae] vs. [aa]



(From a lecture given by Rochelle Newman)

Acoustic landmarks



A Problem: Coarticulation

- Same phone produced differently depending on phonetic context
- Occurs when articulations overlap as articulators are moving in different timing patterns to produce different adjacent sounds
  - Eight vs. Eighth
    - Place of articulation moves forward as /t/ is dentalized
  - Met vs. Men
    - Vowel is nasalized

IPA consonants

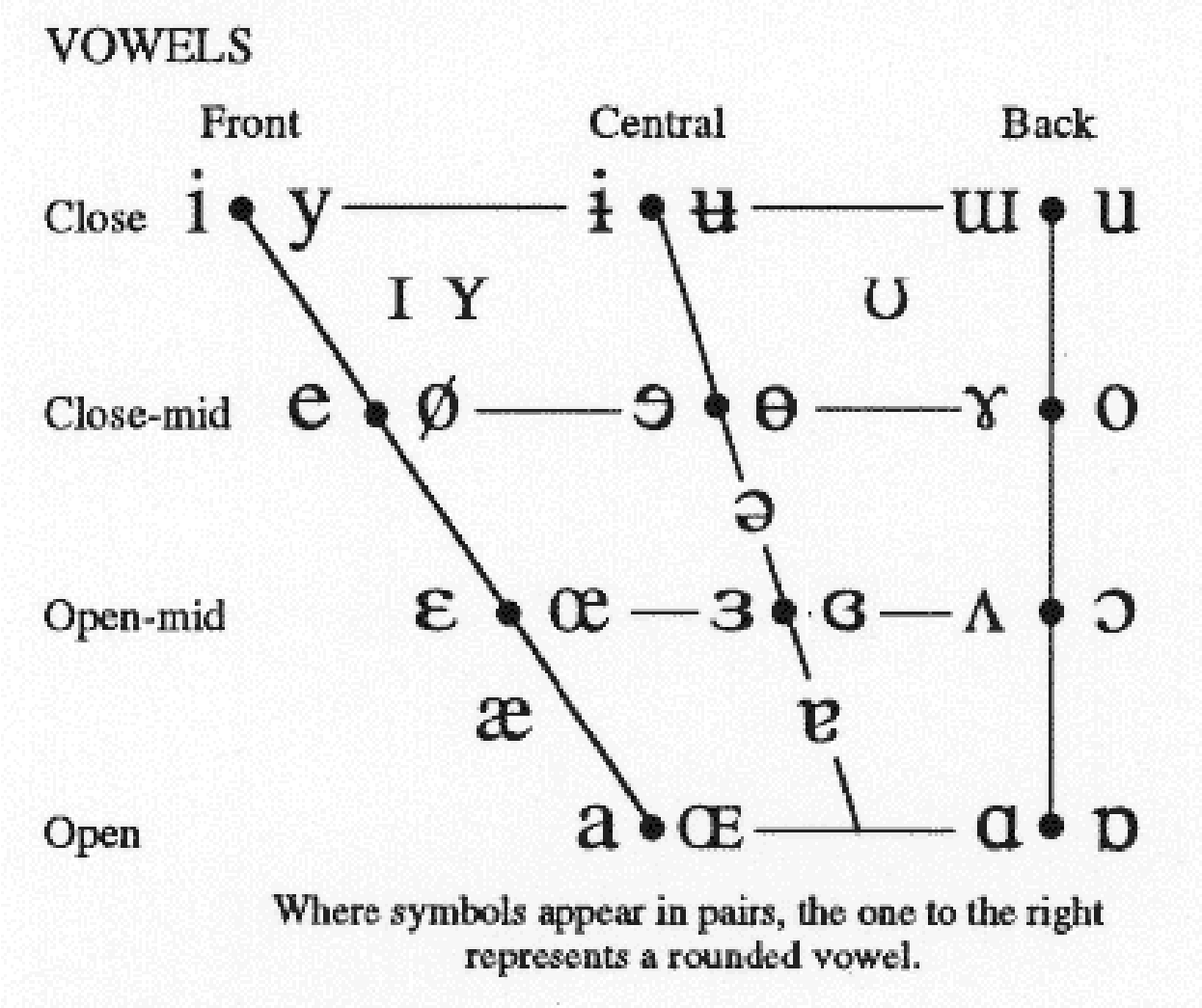
CONSONANTS (PULMONIC)

	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				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			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

(Distributed by the International Phonetics Association.)

IPA vowels



(Distributed by the International Phonetics Association.)

Representations for Sounds

- Now we have ways to represent the sounds of a language (IPA, Arpabet…) and to classify similar sounds
  - Automatic speech recognition
  - Speech synthesis
  - Speech pathology, language id, speaker id
- But…how can we recognize different sounds automatically?
  - Acoustic analysis and tools

Next Class

- Readings: Acoustics of Speech Production (J&M 7.4, *Johnson Ch 1-2)