VOWELS

Mr. R. DHARMALINGAM

Assistant Professor of English

Jamal Mohamed College (Autonomous)

Tiruchirappalli – 20.

VOWELS

Next major topic: **Vowel articulation and (a little bit) the acoustics of vowel production.**

What's a vowel? A speech sound produced with a (relatively) unimpeded air stream.

What's a consonant? A speech sound produced with air stream impeded, constricted, or obstructed.

Bottom line: **Vowels are *open-ish*, consonants are *closed-ish.***

bababababababa (CVCVCV… closed, open, closed, open …)

shashashashashasha (CVCVCV… closed, open, closed, open …)

hahahahahahahaha (CVCVCV… closed, open, closed, open …)

Constriction can be anywhere – in the case of [h], it's way down at the larynx. Still counts as a constriction, which makes [h] a consonant.

Not all vowels are equally open. Compare the vowels in *"keep"* and *"cop"*: The */ɑ/* of *"cop"* is much more open that the [i] of *"keep"* (note how much further your jaw drops for [ɑ]).

Not all consonants are equally closed. **Compare the consonants in *"beep"* and *"weep":* The [b] of *"beep"* is much more closed that the [w] of *"weep"*; i.e., for [b] the closure at the lips is complete, while for [w] the air steam is merely impeded.**

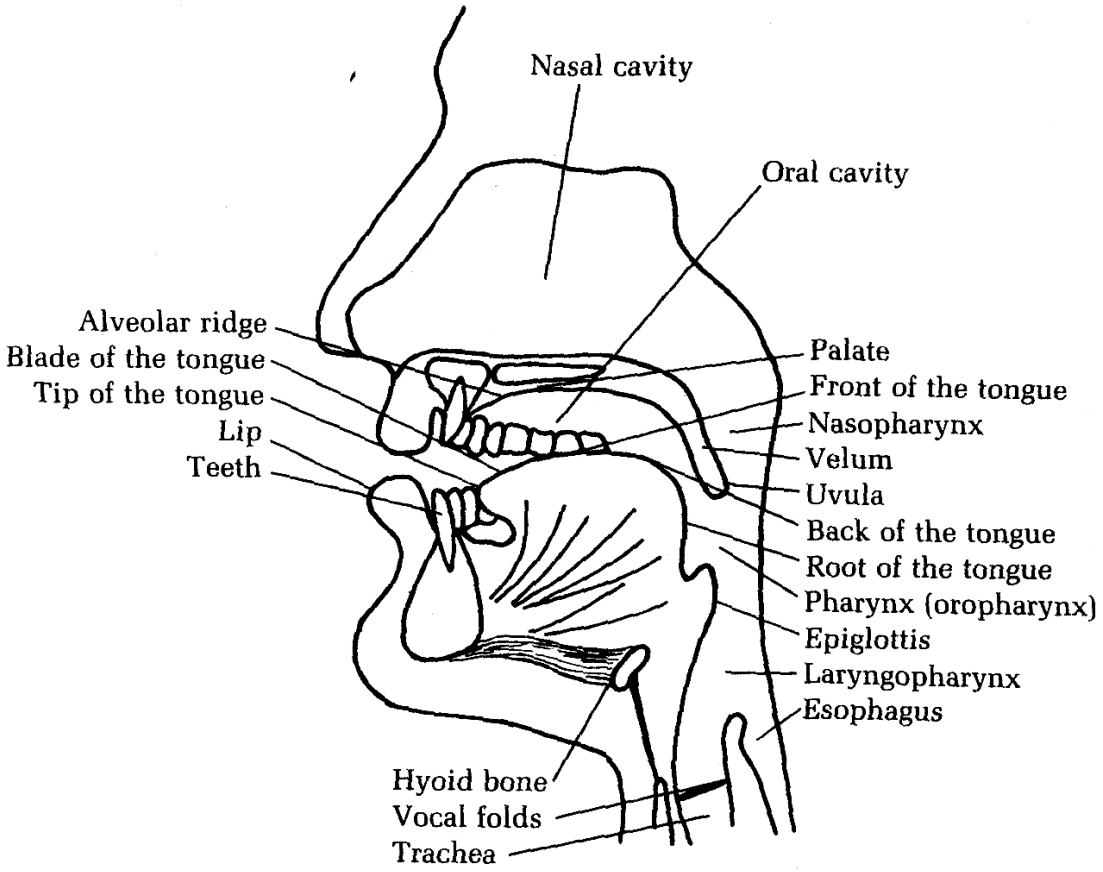
So, what's the formal definition of a vowel; i.e., how open does the vocal tract have to be for a sound to qualify as a vowel? Or, how closed does the vocal tract have to be for a sound to qualify as a consonant?

Don’t have a formal definition or dividing line; don’t really need one. Your intuitions from grade school about vowels and consonants will *almost* always be right. (Vocalic /r/ is the one tricky case – we’ll talk about it.)

One grade school idea to get out of your head – *the vowels of English are not A, E, I, O, U.* These are the letters that are used to represent vowels in English orthography. *English has many more than 5 vowels.*

A QUICK TOUR OF THE ARTICULATORY STRUCTURES

teeth/dental
alveolar ridge/alveolar
(hard) palate/palatal
velum/soft palate (velar)
pharynx/naso-/oroph-/laryng
uvula/uvular
epiglottis
larynx/glottis/vocal folds
trachea
lips/labia/labial/bilabial
tongue/lingual
apex/tongue tip/apical
blade
front
back/dorsum/dorsal
root

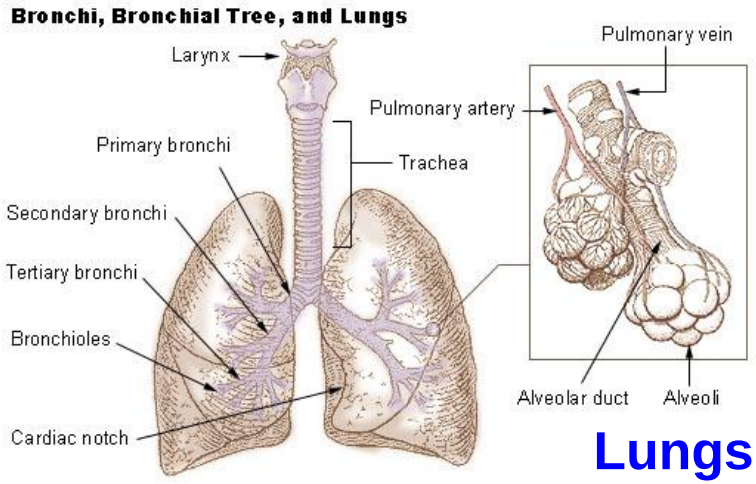


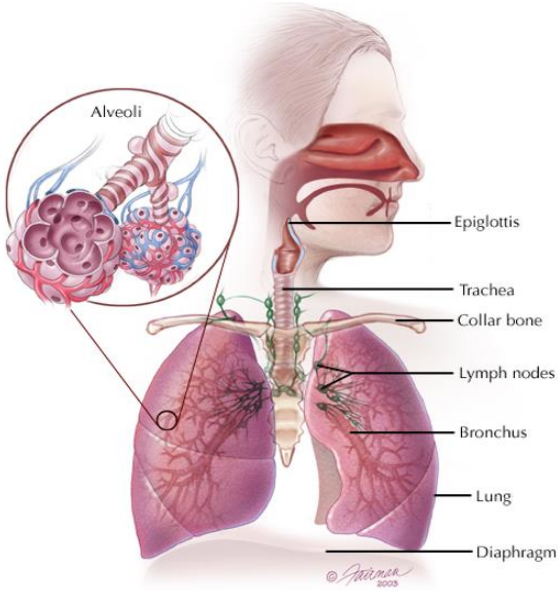
Next step: A *Functional Model of Vowel Production* What's a functional model?

Analogy: Pretend our goal is to understand the modern automobile on the left. It's a lot to take on all at once, so we'll start by understanding the simpler go-kart on the right. The go-kart is a simplified model of the car – *it has all the essential elements of the car but is easier to understand.*

A Function Model of Vowel Production

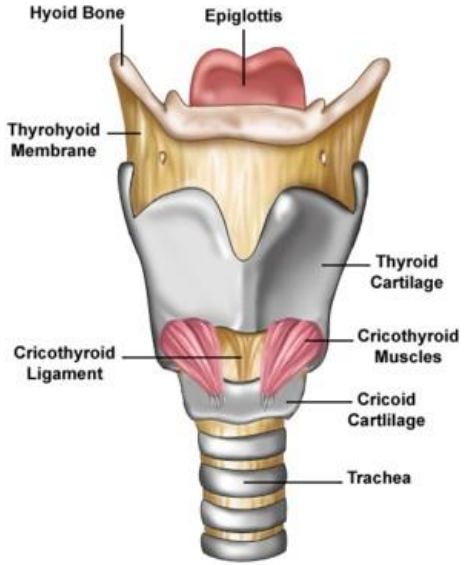
A functional model purposely strips away as much anatomical detail as possible. Purpose: To get the big picture of how the thing works.

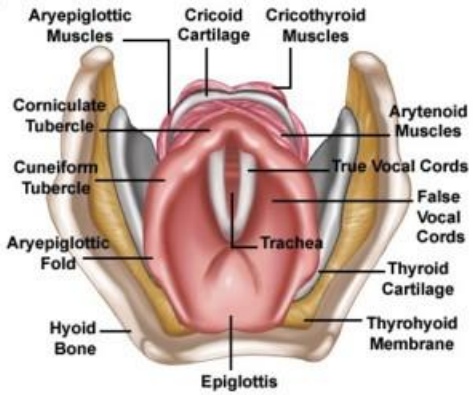




Lungs have a very intricate internal structure. Understanding this structure is ultimately important, but for the purpose of understanding how speech production works *we can ignore most of that detail.* From a *functional* point of view the *lungs behave like a bellows.*

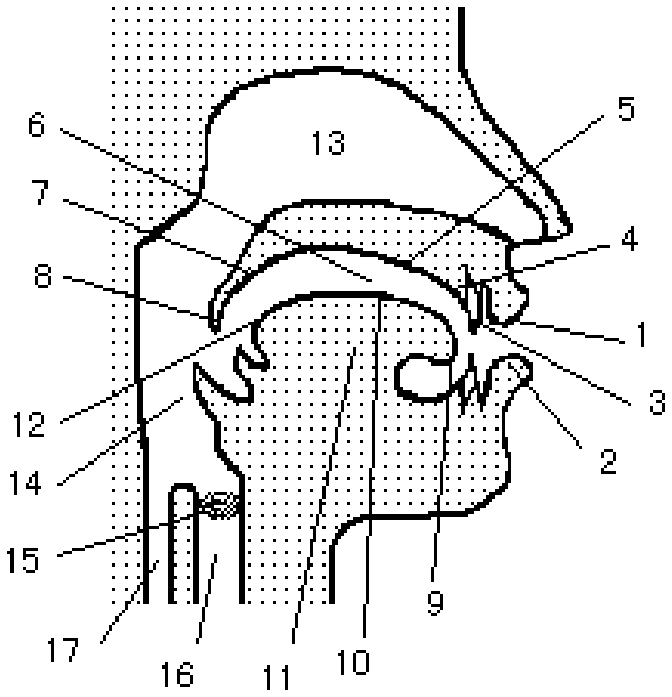
Same deal with the larynx – lots of detail, all of it important. But let's purposely strip away most of the detail to get the big picture of how the system works.



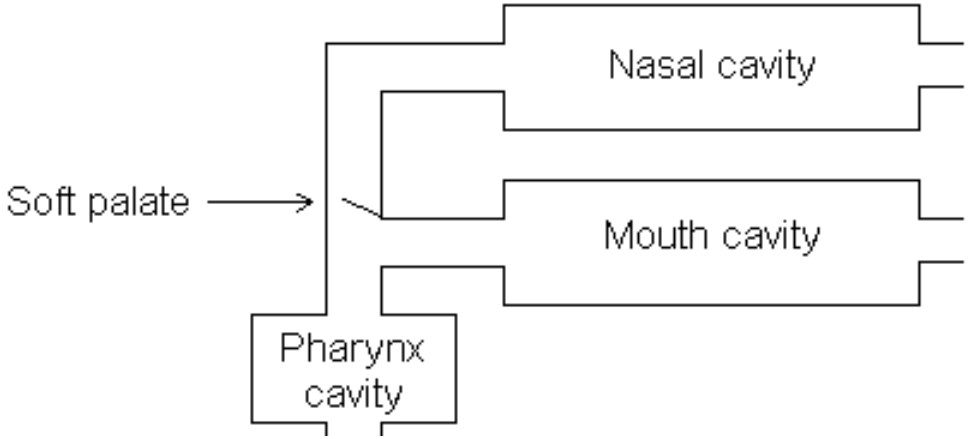


We'll replace the complicated structure to the left with a simple tube (representing the trachea) and 2 flaps of flexible tissue jutting into the tube (representing the *vocal folds*).

Finally, the *vocal tract.* It has a pretty complicated shape. To get an idea of how the system works, we can replace the complex shape on the left with the simpler shape on the right.



The real thing

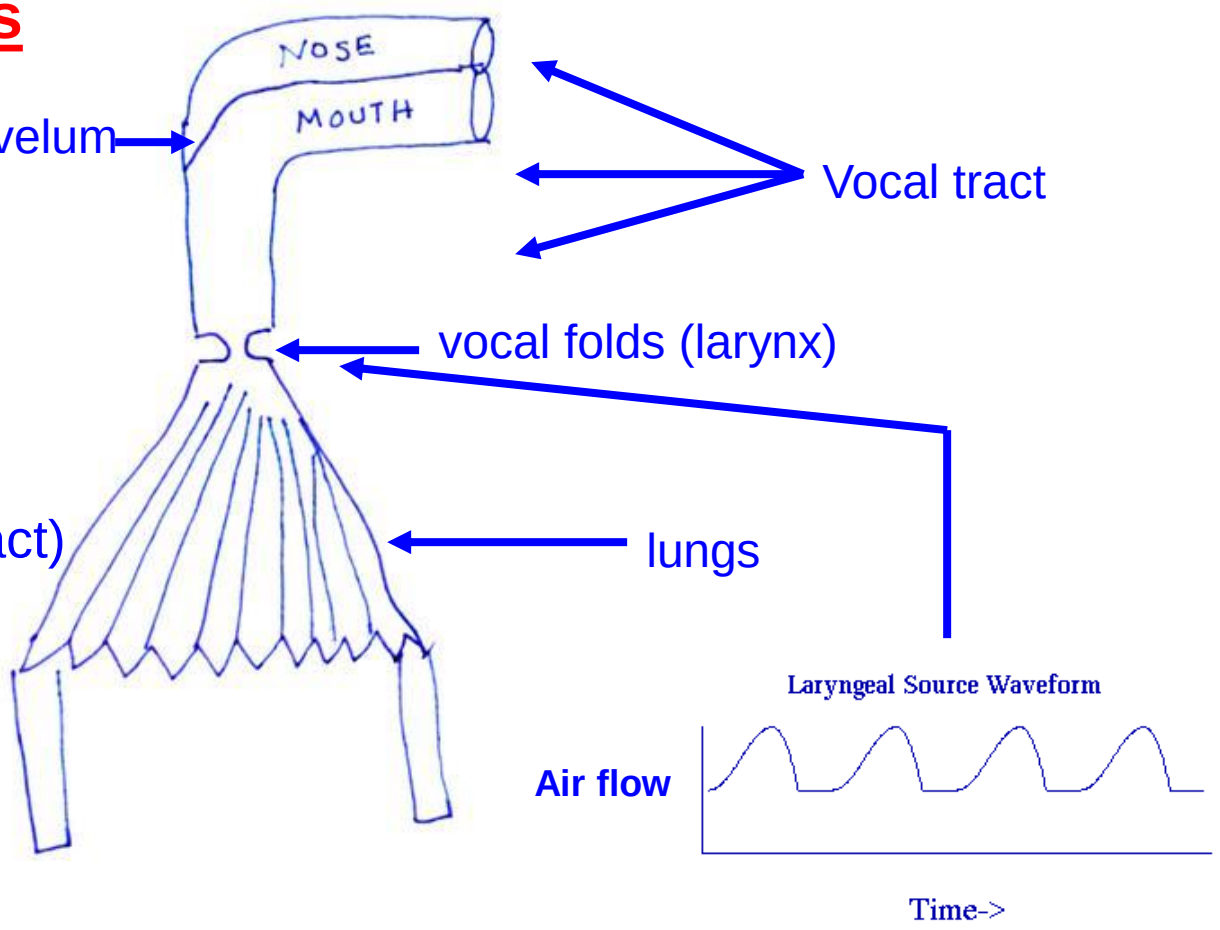


Simplified functional model of the real thing

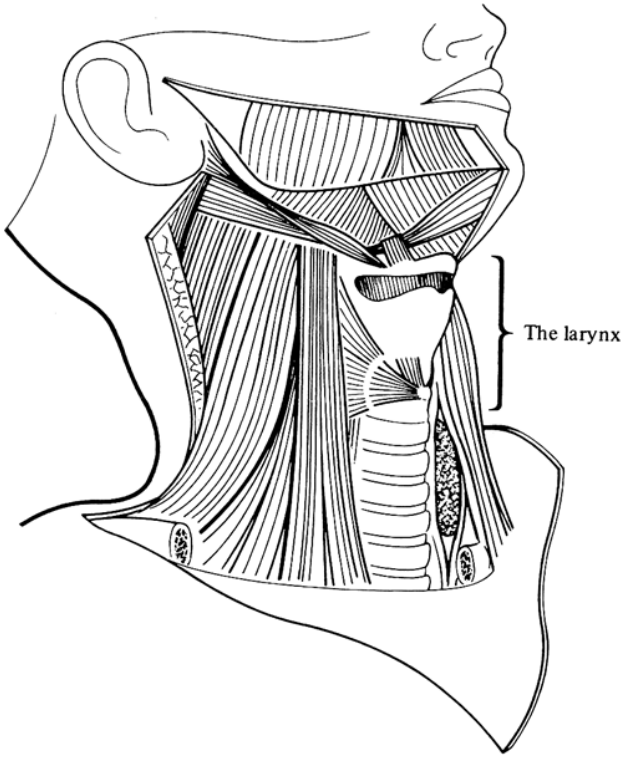
Functional Model of Vowel Production

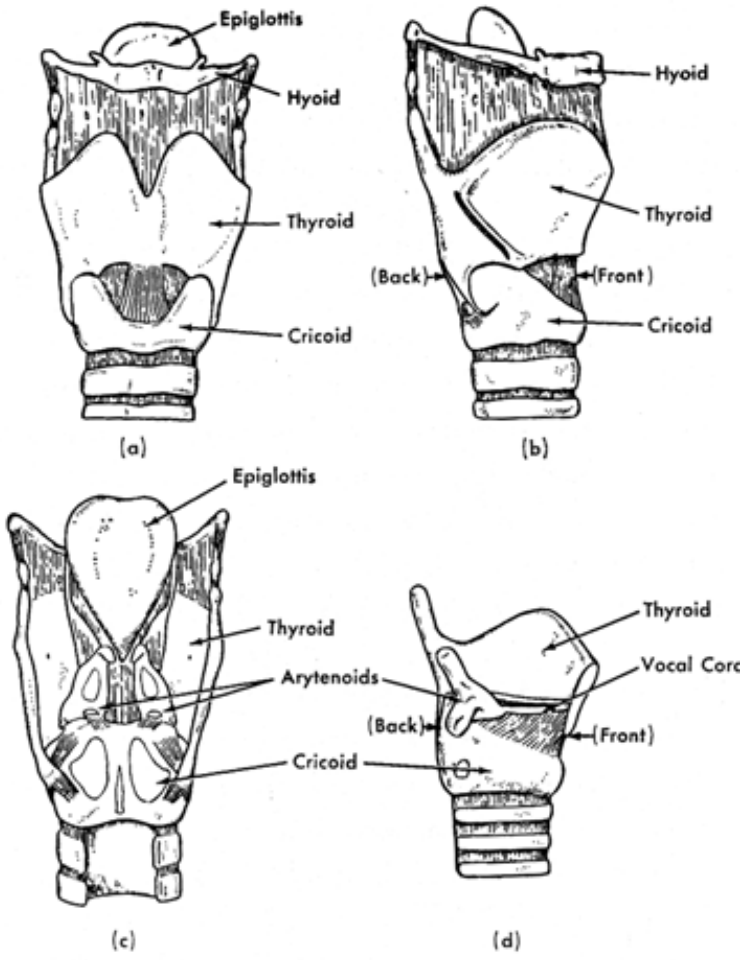
Three Subsystems

1. ***Power supply*** (respiratory system)
2. ***Buzz generator*** (larynx)
3. ***Variable filter/Variable resonator*** (vocal tract)

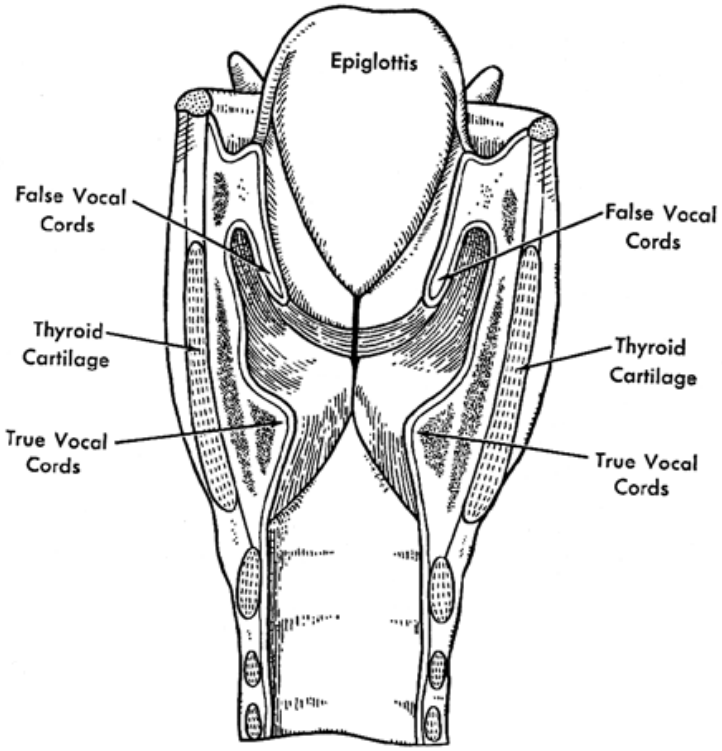


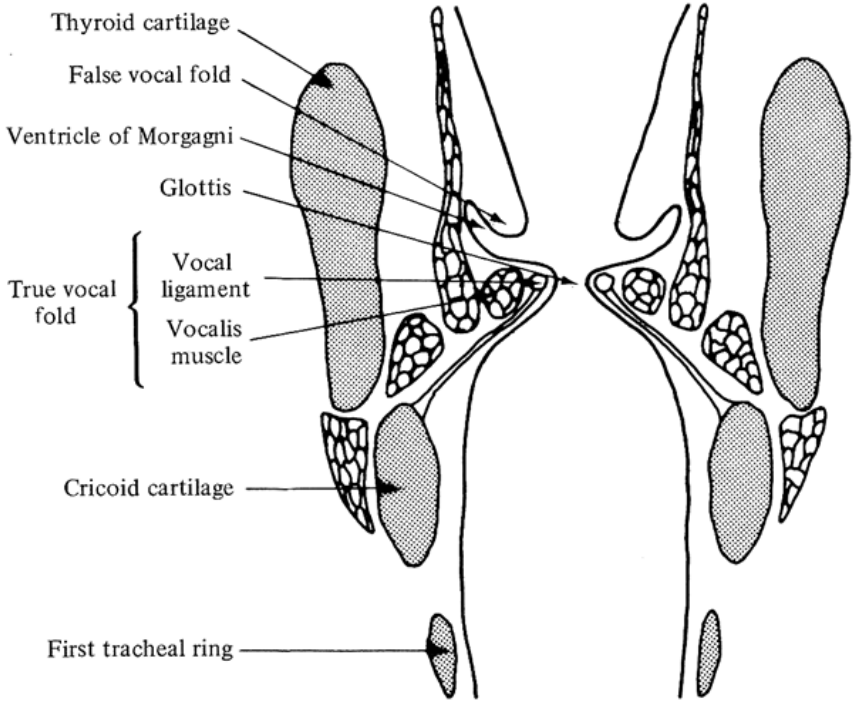
The Buzz Generator (Larynx)



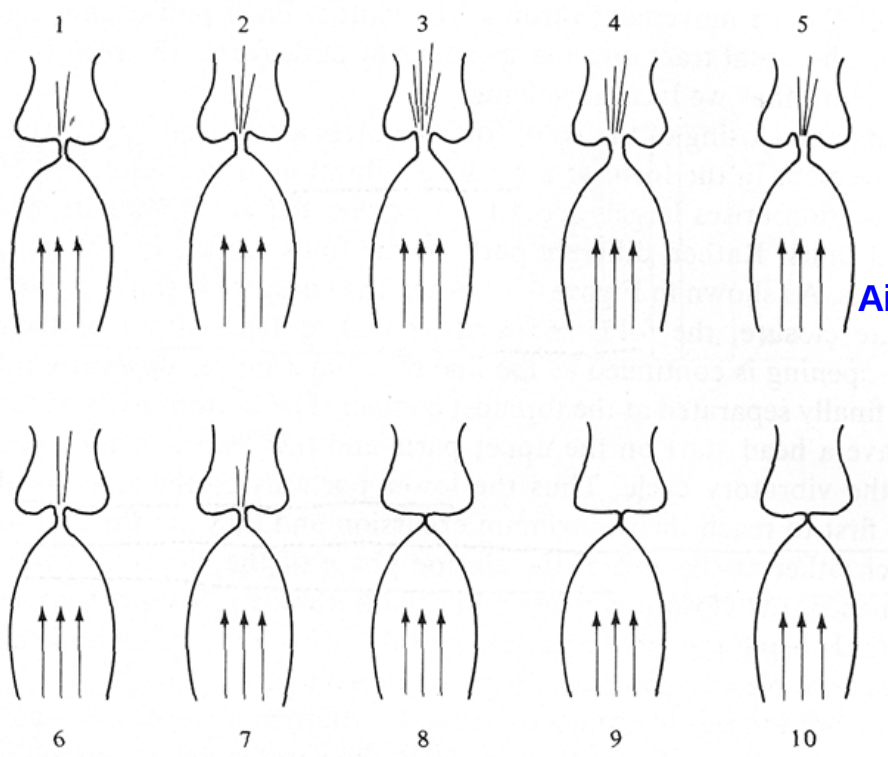


Larynx (cont'd)

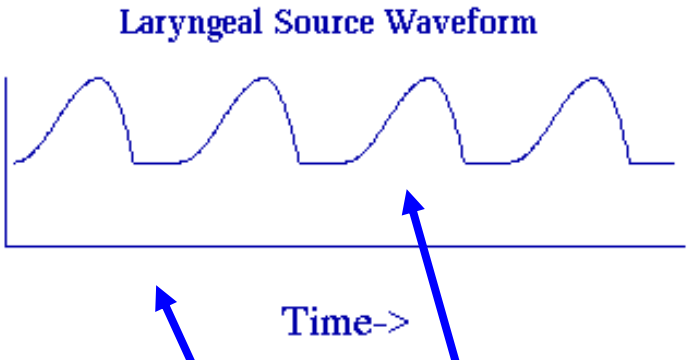




Larynx (cont'd)



Air flow



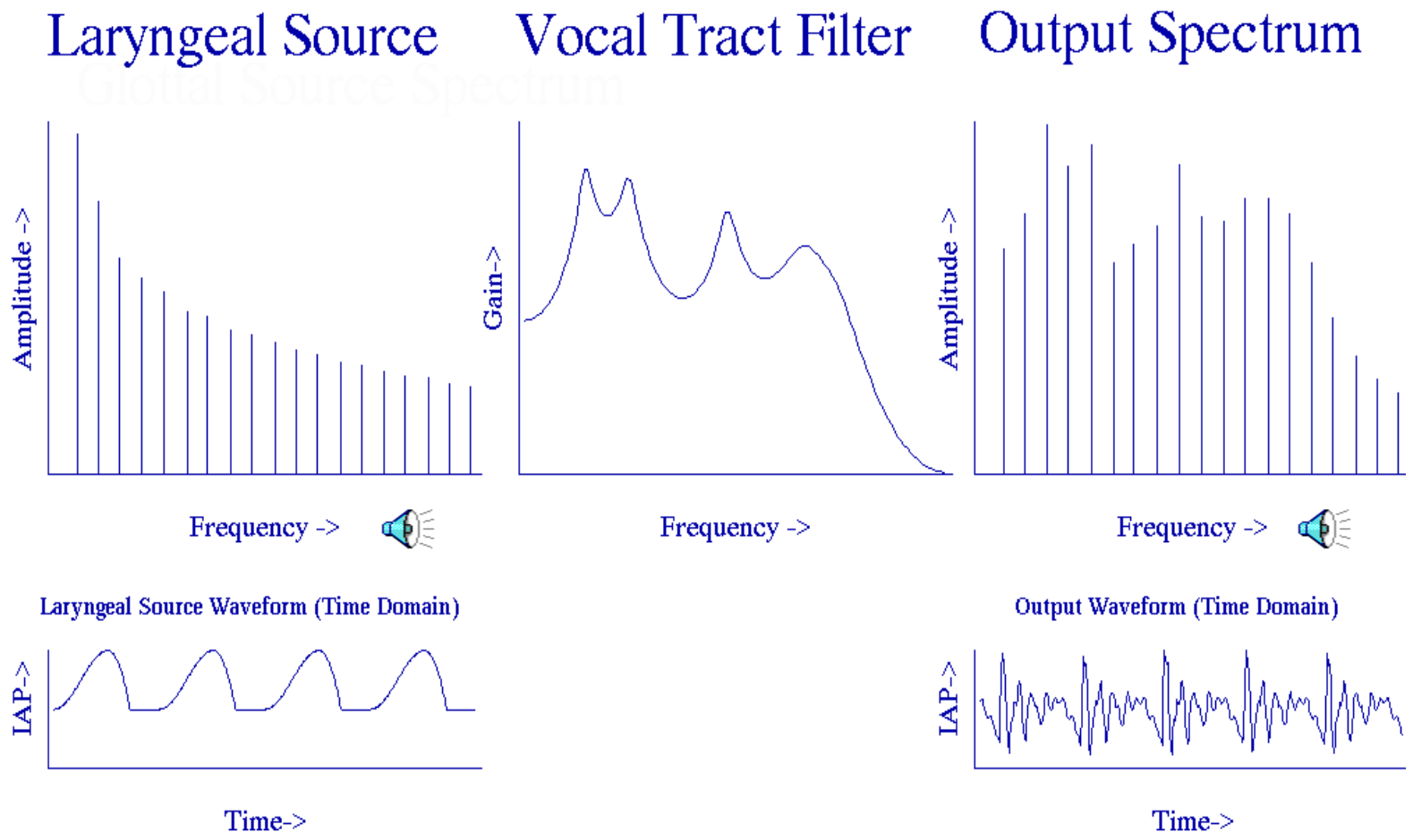
***Sounds like a buzz*, not like speech. Doesn't sound like speech until the v.t. gets hold of it.**

Opening & closing of vocal folds modulates (controls/ regulates) air flow through the glottis, like turning a faucet on and off. Air flow thru the glottis looks like *this*.

Larynx (cont'd)

Vocal fold vibration/voicing/phonation

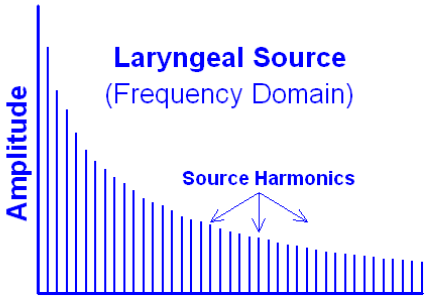
SOURCE-FILTER MODEL OF SPEECH PRODUCTION (simplified)

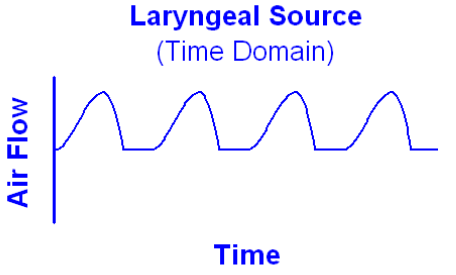


SOURCE

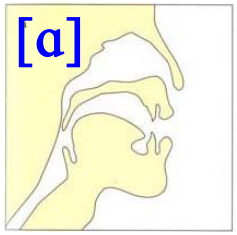
FILTER

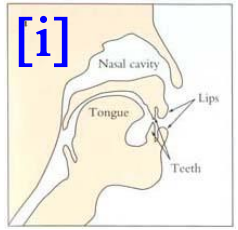
OUTPUT

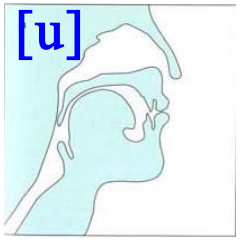


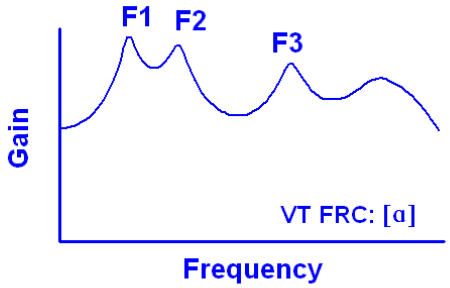


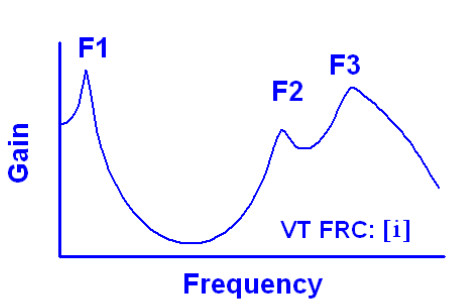
Identical Buzz Energizes Three Different Vocal Tract Shapes *with Three Different FR Curves*

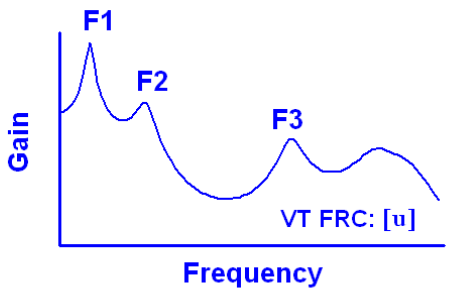


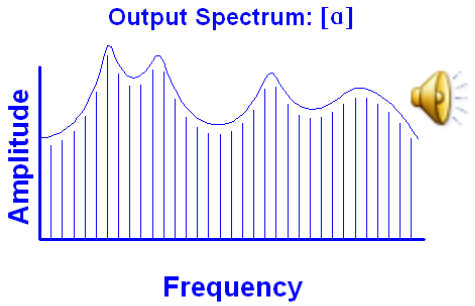


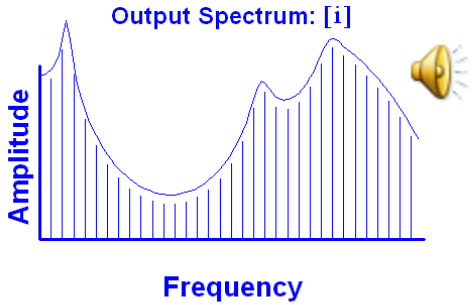


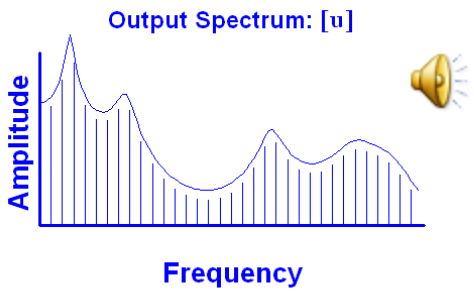


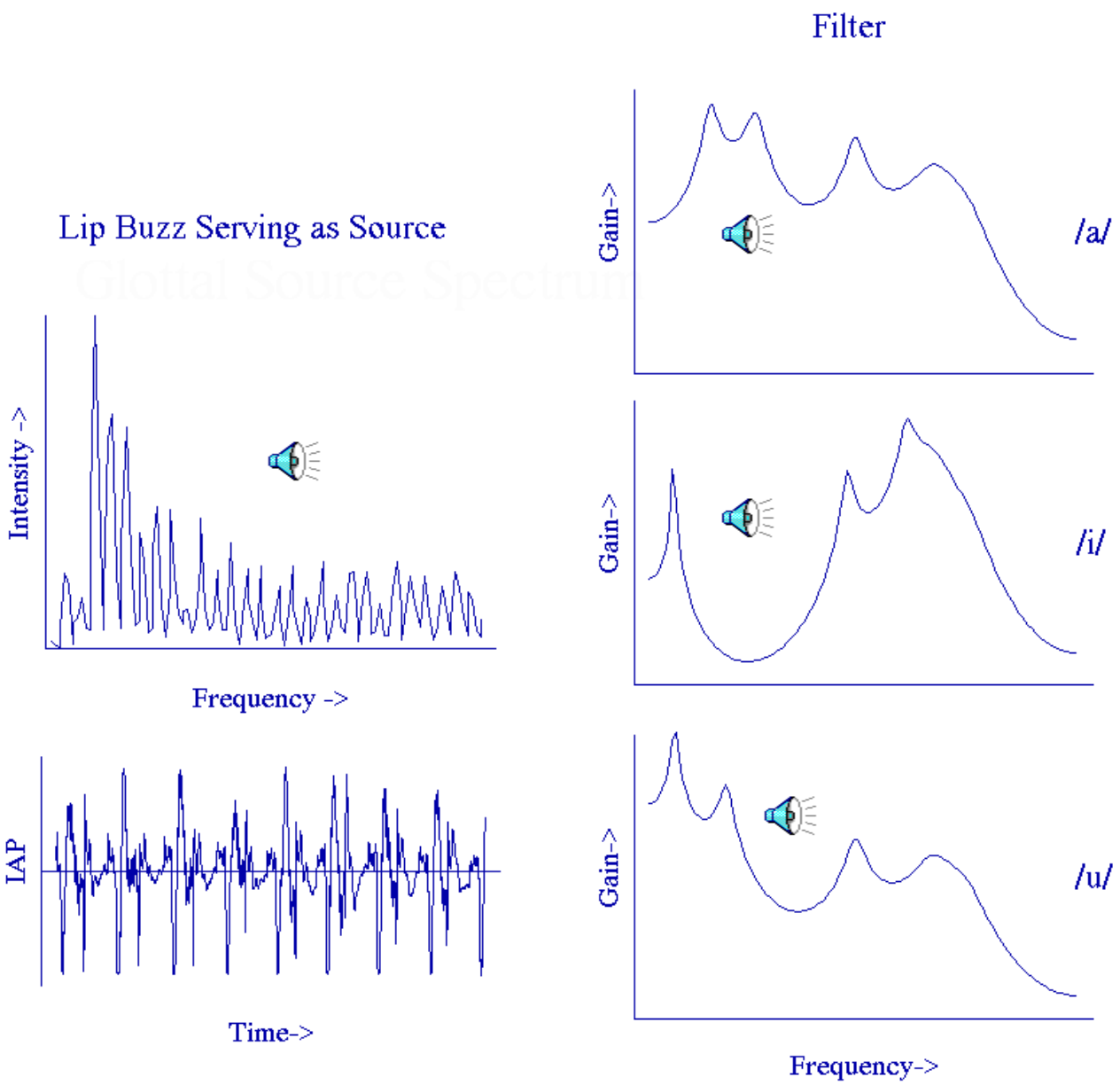











Filter
Lip Buzz Serving as Source
Glottal Source Spectrum
Intensity ->
Frequency ->
IAP
Time->
Gain->
/a/
Gain->
/i/
Gain->
/u/
Frequency->

Source-Filter Theory's BIG Idea

Quiz:

1. **Fundamental frequency (f_0, *OR* the rate of vibration of the vocal folds, *OR* *the pitch of the voice*) is controlled by:**

   a. the source

   b. the filter

   c. a combination of the source and the filter

2. **Vowel quality ([ɑ] vs. [i] vs. [u], etc.) is controlled mainly by the frequencies of the formants (especially F_1 & F_2). The formants, in turn, are controlled by:**

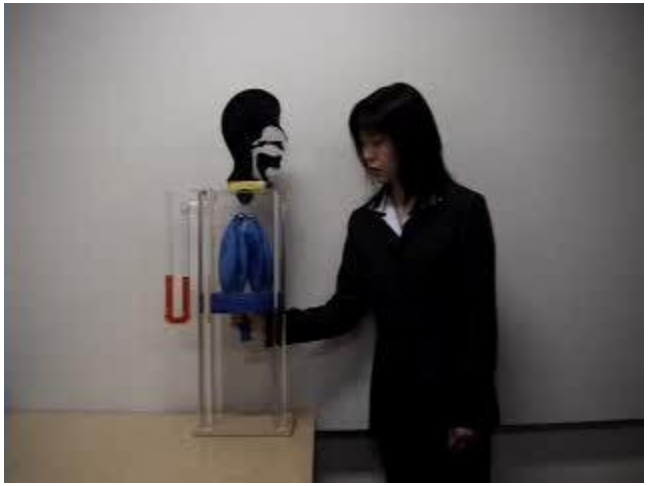
   a. the source

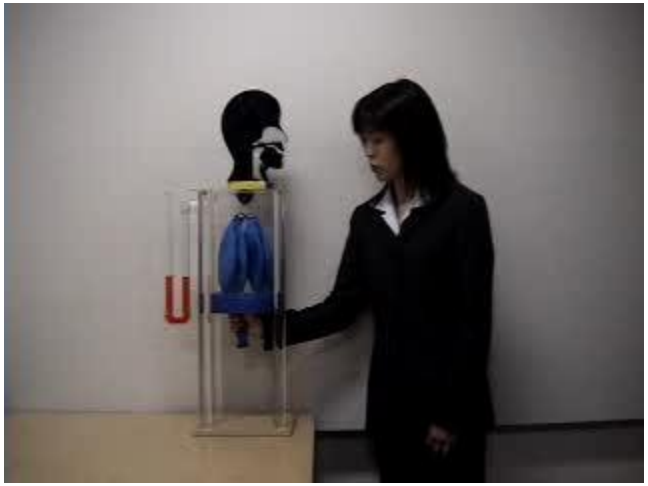
   b. the filter

   c. a combination of the source and the filter

(1: A, 2: B)


U


U


U

Source-Filter Theory Demonstration: Peter Frampton's Talk Box

http://www.youtube.com/watch?v=g9D-kUEp03c

Source-Filter Theory's BIG Idea

A Familiar Quiz:

1. **Fundamental frequency (f_0, OR the rate of vibration of the vocal folds, OR *the pitch of the voice*) is controlled by:**

   a. the source

   b. the filter

   c. a combination of the source and the filter

2. **Vowel quality ([ɑ] vs. [i] vs. [u], etc.) is controlled mainly by the frequencies of the formants (especially F_1 & F_2). The formants, in turn, are controlled by:**

   a. the source

   b. the filter

   c. a combination of the source and the filter

(Answers: See previous identical quiz)

A Slightly Harder Quiz:

1. In the production of a sustained vowel, when the tongue changes from a position high in the mouth to a position lower in the mouth, what about the speech sound changes?

   a. the formant frequencies
   b. the vocal tract frequency-response curve
   c. the quality of the vowel ([ɑ] vs. [i] vs. [u], etc.)
   d. the fundamental frequency
   e. the pitch of the voice
   f. some of the above – which ones?

(A, B, & C)

A Slightly Harder Quiz (cont'd):

2. In the production of a sustained vowel, when the lips change from a retracted (smiley) posture to a rounded posture, what about the speech sound changes?

   a. the formant frequencies
   b. the vocal tract frequency-response curve
   c. the quality of the vowel ([ɑ] vs. [i] vs. [u], etc.)
   d. the fundamental frequency
   e. the pitch of the voice
   f. some of the above – which ones?

(A, B, & C)

Again With the Slightly Harder Quiz:

3. In the Peter Frampton Talk Box demo, the *source* of source-filter theory was:

   a. Frampton's vocal tract
   b. Frampton's guitar
   c. both his guitar *and* his vocal tract – you can't get intelligible speech without both source *&* filter

4. In the Peter Frampton Talk Box demo, the *filter* of source-filter theory was:

   a. Frampton's vocal tract
   b. Frampton's guitar
   c. both his guitar *and* his vocal tract – you can't get intelligible speech without both source *&* filter

(3: B, 4: A)

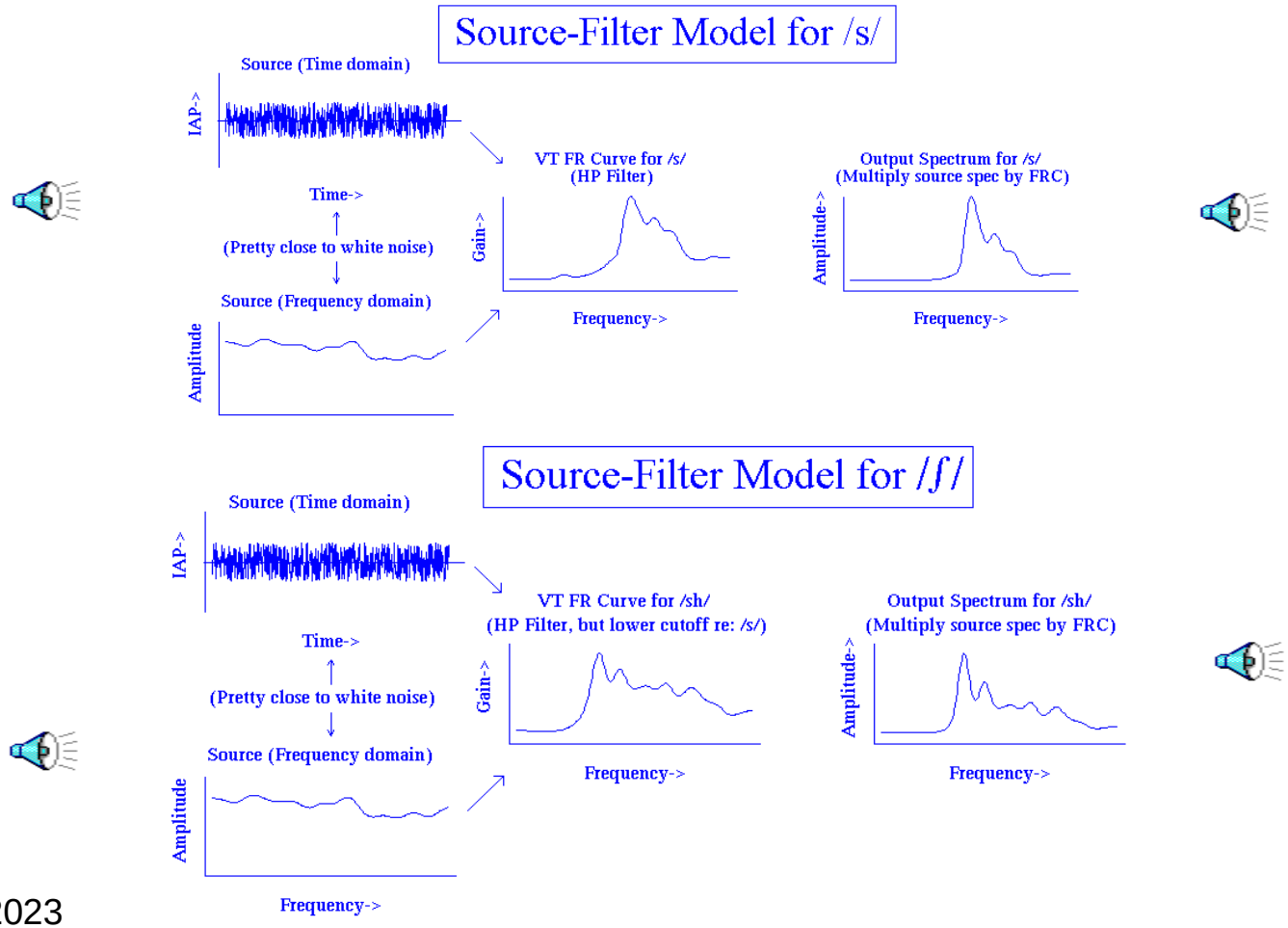
One More Quiz:

5. When a speaker produces the same vowel at two different pitches (fundamental frequencies, or f0s), the pitch difference that you hear is due to:

   a. the filter
   b. the source
   c. both the source and the filter
   d. the larynx
   e. the position of some combination of the tongue, the lips, and the jaw
   f. some of the above – which ones?

(B and D)

The source *does not have to be the larynx* and it *does not have to be periodic*. For example, the source for unvoiced fricatives is *produced in the vocal tract* rather than the larynx and is (roughly) *white noise*. The pictures below show SF models for /s/ and /ʃ/.



1. Vocal Tract Synthesis Demo (ArticSyn)

2. Real-time frequency analysis demo

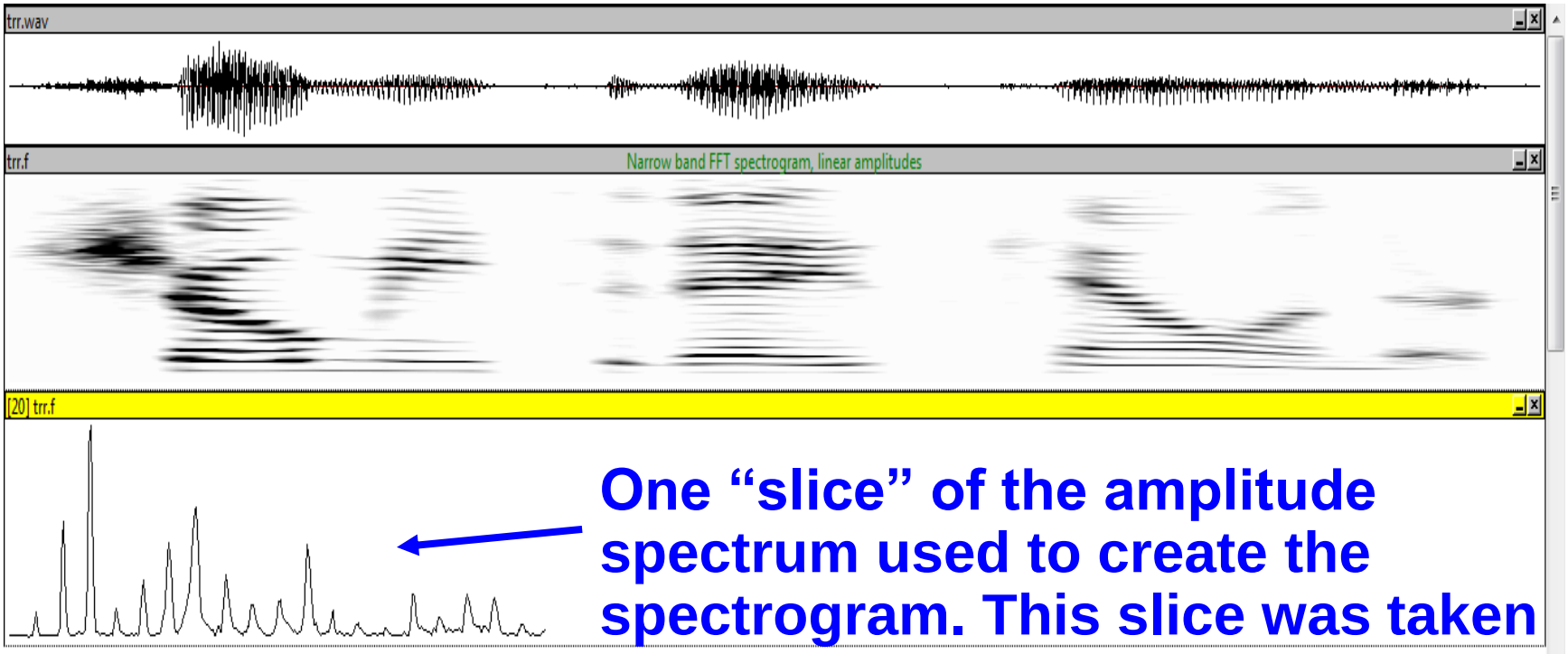
3. Narrow and broadband spectrograms
   (ztool – ah.wav, trr.wav)

   (This slide is nothing; it's just a reminder to me.)

Narrowband Spectrogram

Produced with a *high-resolution, fine-grained frequency analysis* – fine enough to resolve voice harmonics.

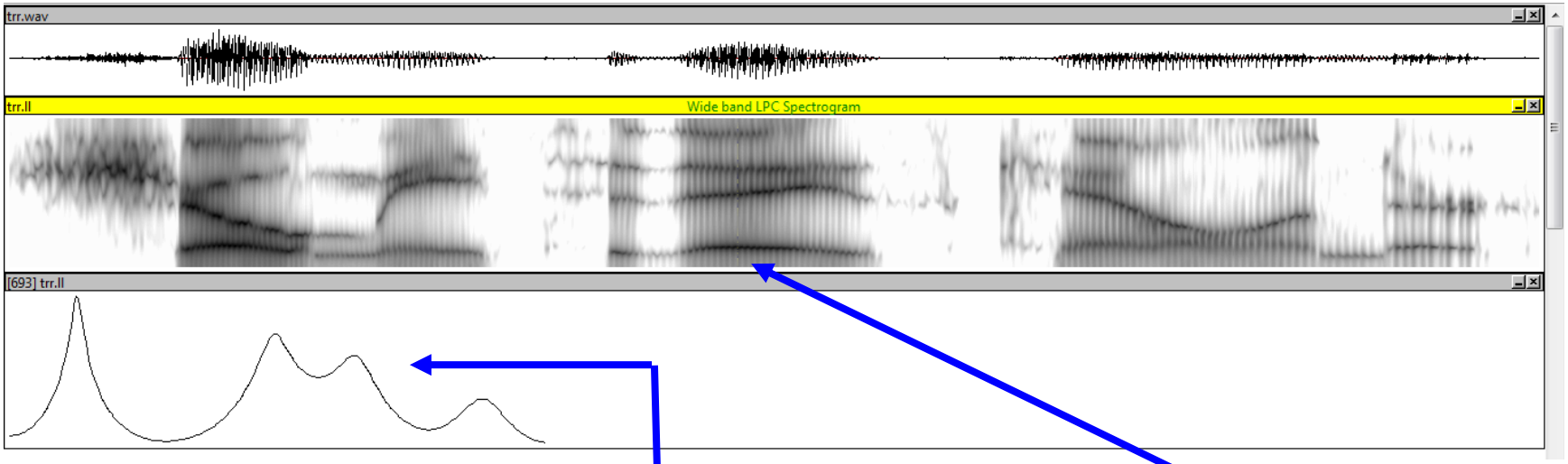
"Show me today's calendar."



One "slice" of the amplitude spectrum used to create the spectrogram. This slice was taken during the 1st vowel. *Note the harmonics on both the amplitude spectrum and the NB spectrogram.*

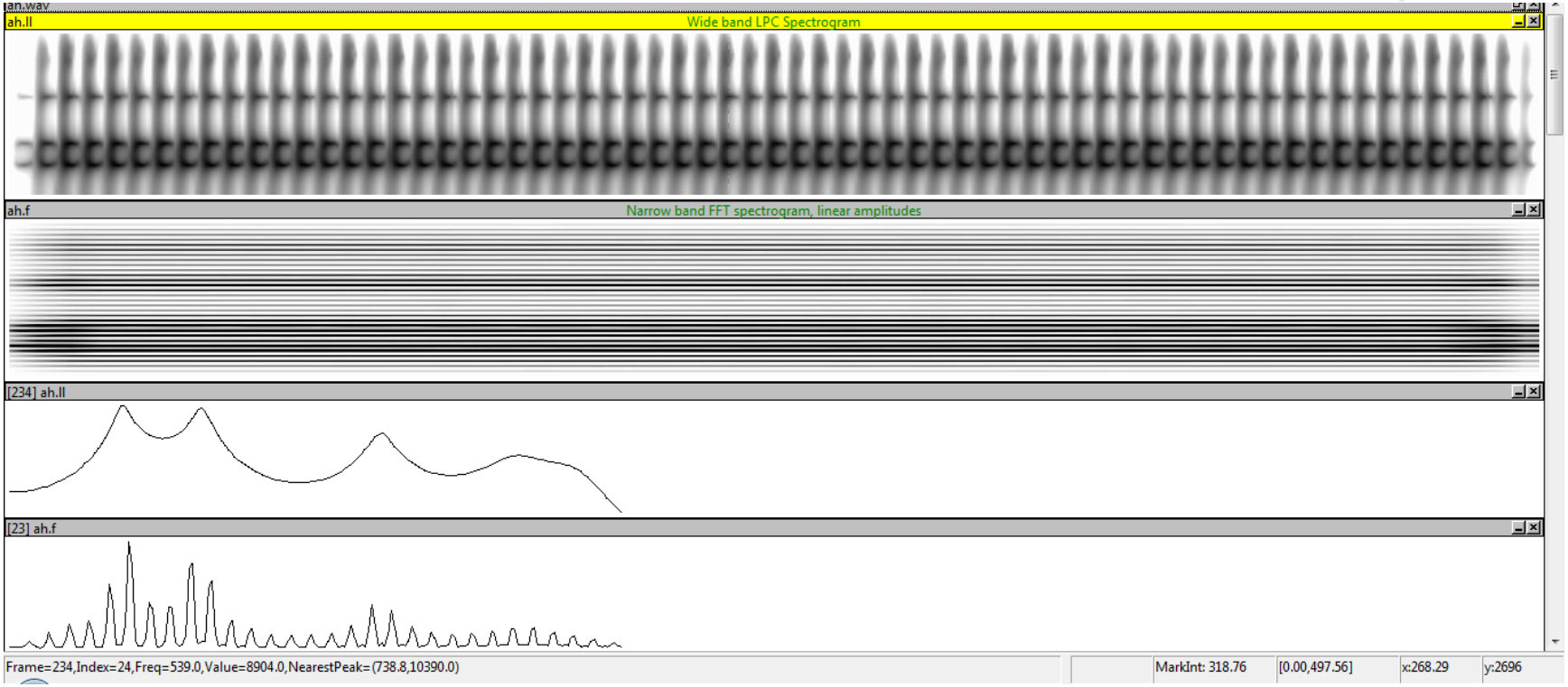
Broadband (wideband) Spectrogram

Produced with a purposely *low-resolution, coarse frequency analysis – too coarse to resolve voice harmonics, but very good at showing formants.*



One "slice" of the amplitude spectrum used to create the spectrogram. This slice was taken during the 3rd vowel. *Note that the peaks in the amplitude spectrum correspond to the formants that are seen in the BB spectrogram;* i.e., the amp spec shows low freq formant peak (F_1), a valley, then a 2nd peak (F_2). These peaks correspond to the formants that are seen in the BB spectrogram.

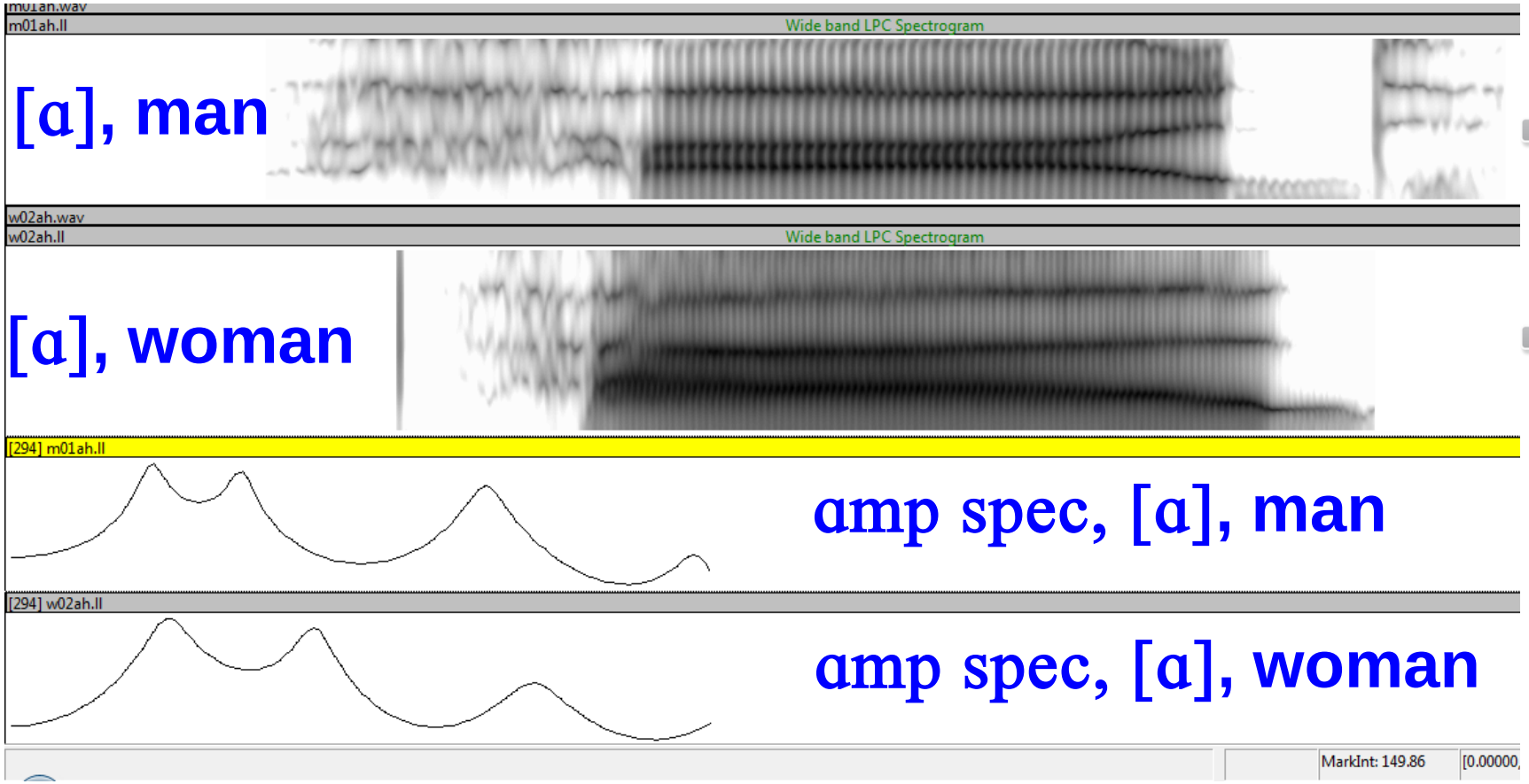
One More Spectrogram Fact



Broad and narrowband spectrograms of the same signal – sustained [ɑ]. (1) ***Which is BB and which is NB?*** **(2)** ***Which of the spectral slices (amplitude spectra) go with the BB and which with the NB? How do you know?*** **(3)** ***Notice the regularly spaced vertical lines, seen only on the BB gram.***

Any idea what these are? **(Answer: These vertical lines correspond to individual glottal pulses. Every time the vocal folds slap together, a pulse of energy is generated – like running a stick along a picket fence. Each one of these vertical lines corresponds to the slapping together of the vocal folds.)**

One More SF Theory Quiz:



When a man and a woman produce the same vowel, the *formant frequencies of the woman will usually be shifted up* relative to the man.

1. **What explains this upward shift in formant frequencies for women relative to men?**

   a. **on average, women have *shorter, lighter vocal folds* than men**

   b. **on average, women have *shorter vocal tracts than men***

   c. **a *combination* of *a* and *b***

2. **On average, women have higher fundamental frequencies than men. What explains this fact?**

   a. **on average, women have *shorter, lighter vocal folds* than men**

   b. **on average, women have *shorter vocal tracts than men***

   c. **a *combination* of *a* and *b***

(B, A)

Vowel Symbols

/i/	heed	small i
/ɪ/	hid	cap i, or small cap i
/e/	hayed, bait	small e
/ɛ/	head	epsilon
/æ/	had	ash
/ɑ/	hod, pod	script a (note the difference between /ɑ/ and /a/)
/ɔ/	hawed, caught	open o
/o/	hoed, boat	small o
/ʊ/	hood	upsilon
/u/	who'd, boot	small u
/ʌ/	hud, but	caret or wedge or turned v
/ɚ/	heard	schwar
/ə/	about, mantra	schwa

Vowel Articulation

Dimensions of vowel production (Slightly different list from MacKay; Major dimensions 1-3 in red; 4-8: secondary dimensions)

(*Major dimensions in red*)

1. **Tongue height [e.g., [i] ("*beet*") vs. [æ] ("*bat*")]**
2. **Frontness or advancement [e.g., [æ] ("pat") vs. [ɑ] ("pot")]**
3. **Lip rounding (e.g., [u] vs. [ɑ])**
4. **Tense vs. lax (e.g., [i] vs. [ɪ], [u] vs. [ʊ], [e] vs. [ɛ])**
5. **The special case of vocalic R ([ɚ] or [ɝ] as in "*bird" or "sir"*)**

Vowel Articulation (cont'd)

6. **Length/Duration/Quantity: Any vowel can be spoken at any duration, but different vowels have different *typical* or *inherent* durations (e.g., *beet-bit, bait-bet, suit-soot, bat-bet*, etc).**

7. **Phonation/Breathy/Whisper: Vowels can be phonated, whispered, or anywhere in between (part buzz *and* part hiss = *breathy*)**

1. Tongue height

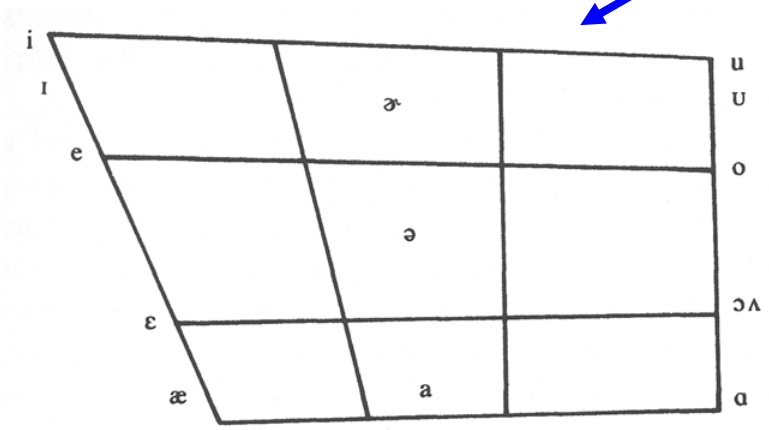
a. Compare tongue/jaw position for [i] (“*beet*”) and [æ] (“*bat*”)

b. Attend to tongue/jaw position for this sequence: [i ɪ e ɛ æ]

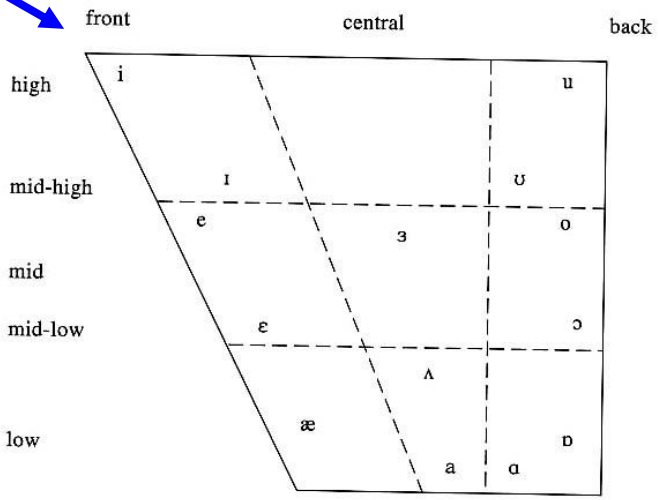
c. Attend to tongue/jaw position for this sequence: [ɑ ɔ o ʊ u]

Vowel Quadrilateral

MacKay



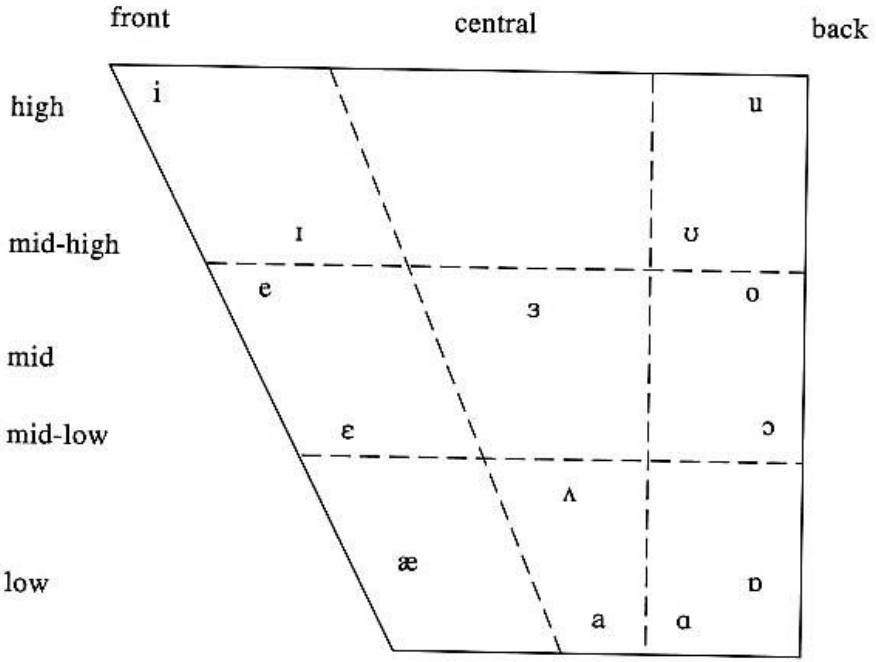
Ladefoged (many others)



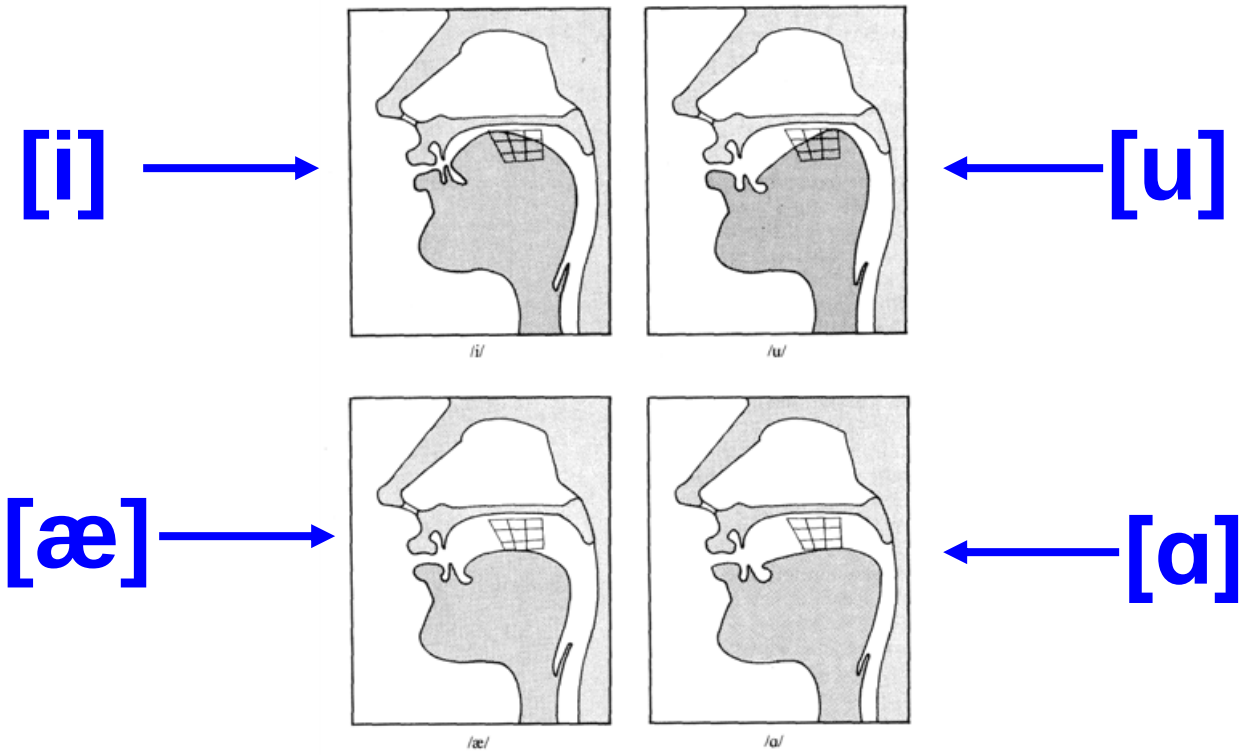
For now, attend only to [i ɪ e ɛ æ] and [ɑ ɔ o ʊ u]. Slight difference of opinion on [ʊ] and [ɪ]; otherwise these two vowel quadrilaterals are quite similar (for these 10 vowels).

Terminology: Vowels differing in tongue height are classified as *high, mid, or low*. Note: When we're talking about tongue height, vowels like [e ɛ o ɔ] are *mid*, not *central.* The term *central* distinguishes vowels based on *advancement* or *frontness*, not tongue height.

Mid* = midway between high & low; *Central* = midway between front and back. *You just have to memorize it.



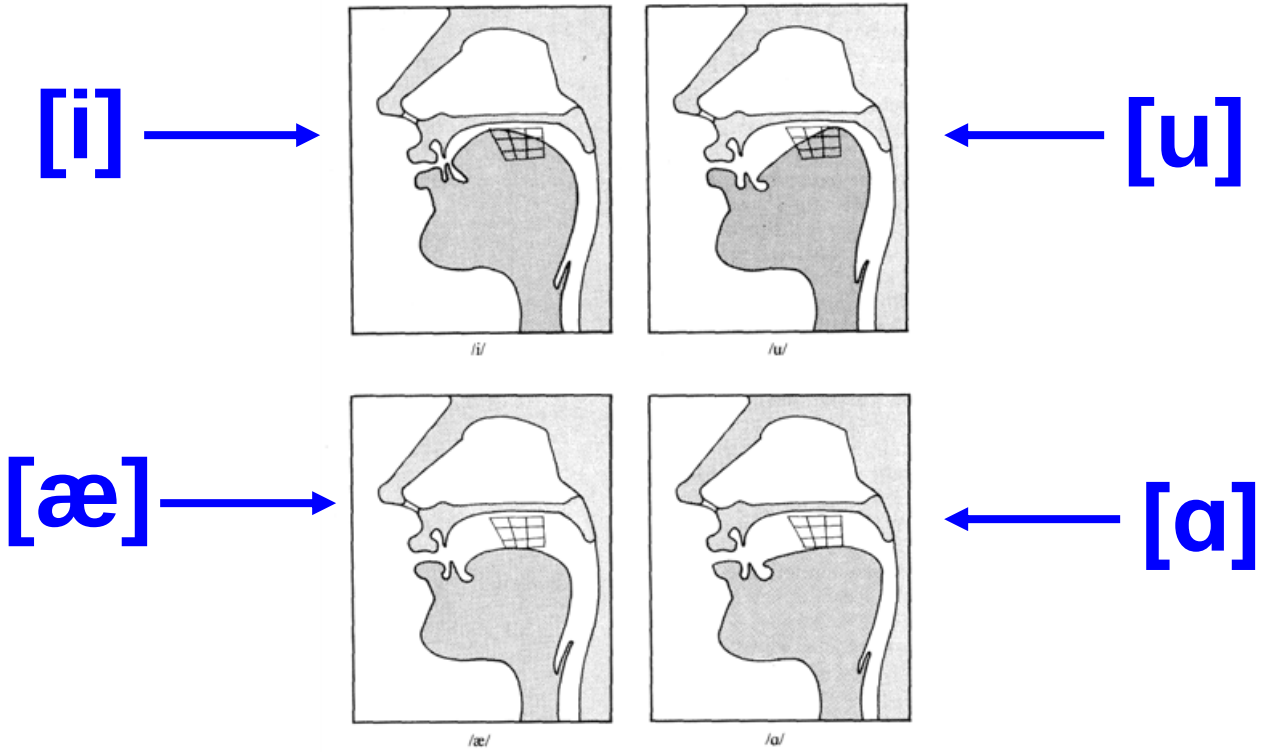
Note the *tongue height* differences between: (a) [i] and [æ] and, (b) [ɑ] and [ʊ]. There are other differences that distinguish /i/-/æ/ and /ɑ/-/u/. For now, focus on tongue-height differences only.



2. Tongue advancement or frontness

a. Compare front-to-back tongue position for [ɑ] ("*pot*") and [æ] ("*bat*"). You should notice that the tongue is further forward for [æ] than [ɑ].

b. Compare front-to-back tongue position for [i] ("*beet*") and [u] ("*boot*") [ignore lip shape – for now]. You should notice that the tongue is further forward for [i] than [u].



3. Lip rounding

Produce the sequence: [ɑ ɔ o ʊ u]. What is happening to the lip shape from [ɑ] to [u]?

Lip rounding is important In English, but *it is not an independent parameter* of vowel articulation:

1. **Running through the sequence [ɑ ɔ o ʊ u], lip rounding is *not the only parameter that is changing.* What else is changing?**
2. **There is no pair of English vowels that differs only in lip rounding (except maybe */ɔ/* vs. */ʌ/* -- though there isn't good agreement on this).**

***Lip rounding is an independent articulatory parameter in some languages*: e.g.,**

- **French has a high, front retracted (i.e., unrounded) vowel ([i]) and a high, front rounded vowel ([y] or [ü]).**
- **German also has *rounded and unrounded versions of vowels with the same height and advancement.***

The vowel in the French word "*tu*" ("*You*.") is not [u], but [y] (also transcribed [ü]); i.e., a high, front, *rounded* vowel.

4. The Tense-Lax Distinction

English has several pairs of acoustically and articulatorily similar vowels that differ according to a distinction called *tense-lax*.

[i] vs. [ɪ]	[e] vs. [ɛ]	[u] vs. [ʊ]
beet-bit	*bait-bet*	*boot-book*

Tense-Lax (cont'd)

[i] vs. [ɪ]	[e] vs. [ɛ]	[u] vs. [ʊ]
beet-bit	*bait-bet*	*boot-book*

The *tense vowels* of the pair are:

1. **Produced with a *higher tongue position***
2. **Produced with a more *advanced tongue root* (ATR) (expands the pharyngeal cavity; *thought by some to be the key tense-lax feature*)**
3. ***Longer***
4. **Spoken with *increased muscular tension* – once thought to be the key feature; it's *irrelevant to sound quality*, but it's how the dimension got its name.**

The tense-lax feature is important because it <u>figures into the phonological rule system</u> of English.

What that means is that there are *sound-pattern rules that apply broadly to all tense vowels as a group, or all lax vowels as a group.*

One simple example: In *most dialects of English, words are not allowed to end in a lax vowel.**

/bi/ ("*be*") is allowed but not something like /bɪ/
/du/ ("*do*") is allowed, but not something like not /dʊ/
/be/ ("*bay*") is allowed, but not something like /bɛ/

***Southern dialects allow lax some vowels to end some words: e.g., Former President *Jimmih Cahtuh*" [dʒɪmɪ kɑtə].**

This is one example. There are others.

5. The Special Case of /ɚ/ (schwar; also transcribed as /ɝ/)

This is the vowel in words like *"bird," "learn," "nerd," "sir"*

Symbol: /ɚ/ (schwar) or /ɝ/

MacKay prefers /ɚ/; I'll stick with it, but you may see both symbols in use.

Some phoneticians distinguish stressed /ɚ/ or /ɝ/ of /bɝd/ ("*bird*") from the unstressed vowel of /bʌɾɚ/ ("*butter*"). MacKay doesn't; we won't either.

There's a very important term that applies to the *sound quality* of vocalic /r/ in words like *"bird," "learn,"* etc. – *as they are spoken in General American*. This type of vowel is called *rhotic*. (Note: the squiggle (tilde: ~) on /ɚ/ and /ɝ/ means *rhotic*.)

***Rhotic** (General American):* [bɚd] or [bɝd]

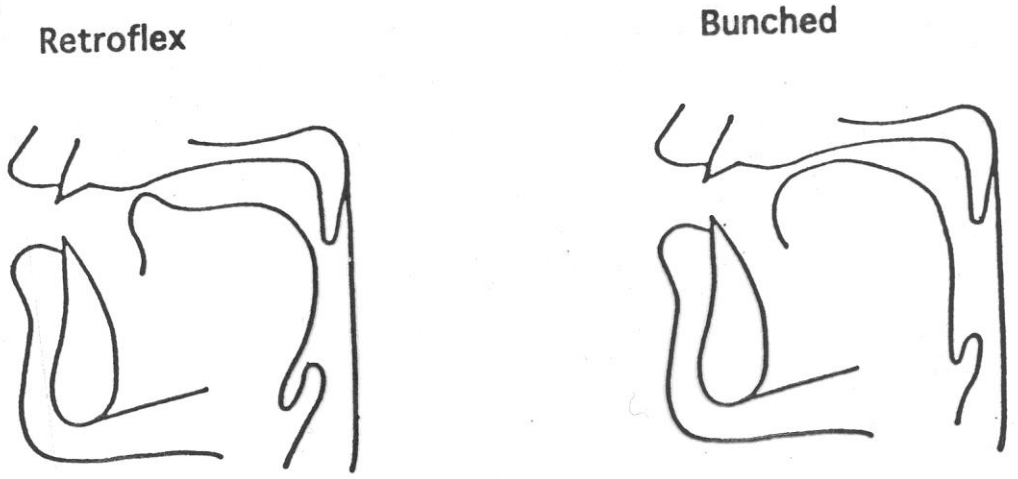
***Non-rhotic** (British RP):* [bɜd] (symbol looks like a '3', but without the tilde; tilde=rhotic)

The term rhotic also applies to consonantal /r/ ("rabbit" "red," "rain," etc.).

***The term rhotic is a reference to sound quality, not articulation* – if it *sounds* like the vowel of (General American) "*bird*", then it's rhotic; if it sounds like the /r/ of "*robot*," then it's rhotic.**

Articulation of /ɚ/

Classified as a mid-central. Not exactly inaccurate, but the articulatory facts are more complicated than this. There are two very different articulatory postures that perfectly good, *perceptually indistinguishable* rhotic vowels: *retroflex* and *bunched*.



Most speakers appear to use *both retroflex and bunched:* compare tongue postures for "*dirt*" vs. "*girl*". Tendency is to use *retroflex for /dɚ/* and *bunched for /gɚ/. (Any idea why?)*

Articulation of /ɚ/ (cont'd)

Which posture (retroflex or bunched) would you expect to be used for words like "sir," "learn," "nerd," "turtle," "zircon," "sure" "German"?

Which posture (retroflex or bunched) would you expect to be used for words like "curb," "girder," "kernel," "gurney," "Kirsten"?

Retroflex

Bunched

Are retroflex and bunched variants of /ɚ/ members of the same phoneme class or are they distinct phonemes?

One more very important fact: *Rhotic sounds are wicked hard to learn.* (This almost certainly explains why rhotics are so rare in the world's languages.)

Rhotic sounds are *auditorily* (i.e., perceptually) *very distinctive*; there's a specific acoustic feature that is seen *always and only* for rhotic sounds, so /ɚ/ is rarely confused with other sounds.

But, *the articulatory trick that produces the rhotic quality is hard to find*. This explains why kids say "*wabbit*," for "*rabbit*", "*wed*" for "*red*", and /lɜn/ for /lɚn/ ("*learn*").

1.Rhotic sounds (vowels and consonants) are usually among the last sounds acquired by kids learning English* (and other languages with rhotic sounds) – typically around age 7 or so *(with a lot of variability). SLPs spend a lot of time with /ɹ,ɚ/.

2.Rhotics are on a very short list of sounds that can remain problematic into adulthood.

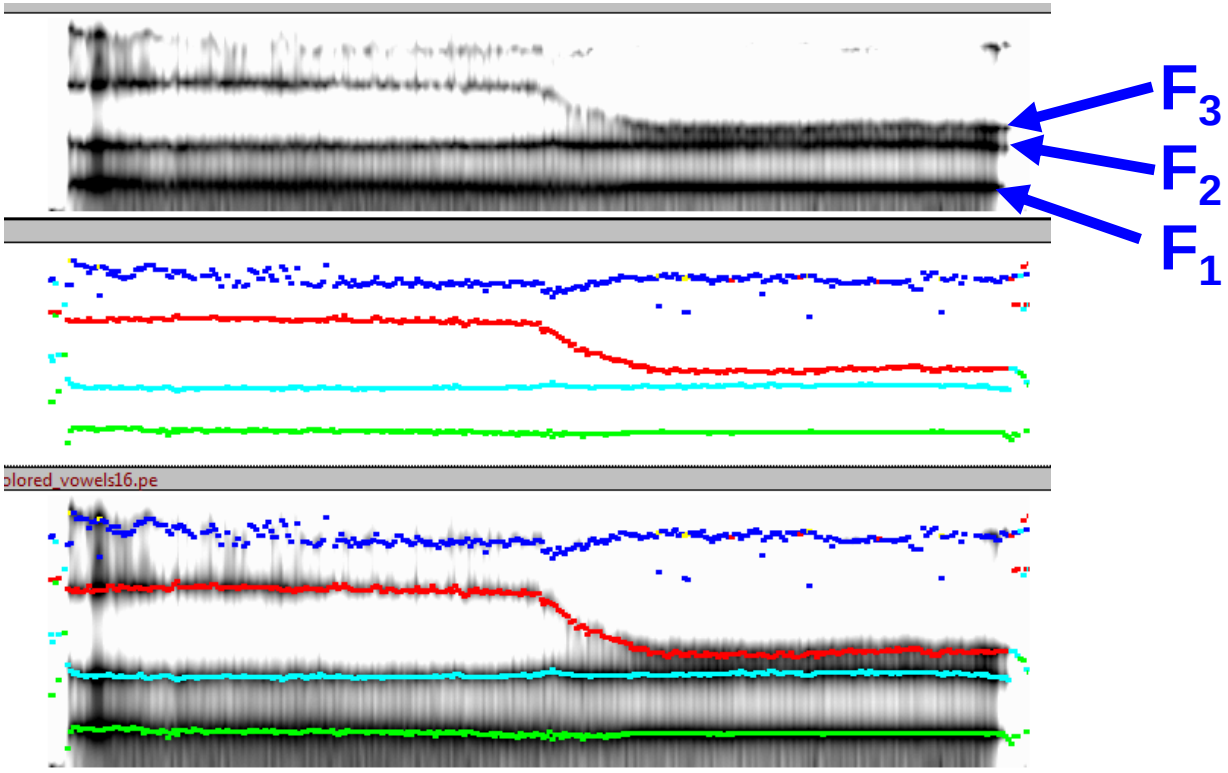
3. **Rhotic sounds also tend to be very *difficult for adults trying to learn English as a 2nd language* if the native tongue does not include rhotic sounds – and most languages do not have them. *Rhotic sounds are found in just 1% of the world's languages*.**

   So, if you're learning English as a 2nd language, what are the odds that your native language *does not* have rhotic sounds? Really high – ~99%.

   Is it likely that these sounds will be hard to learn? [Yes]

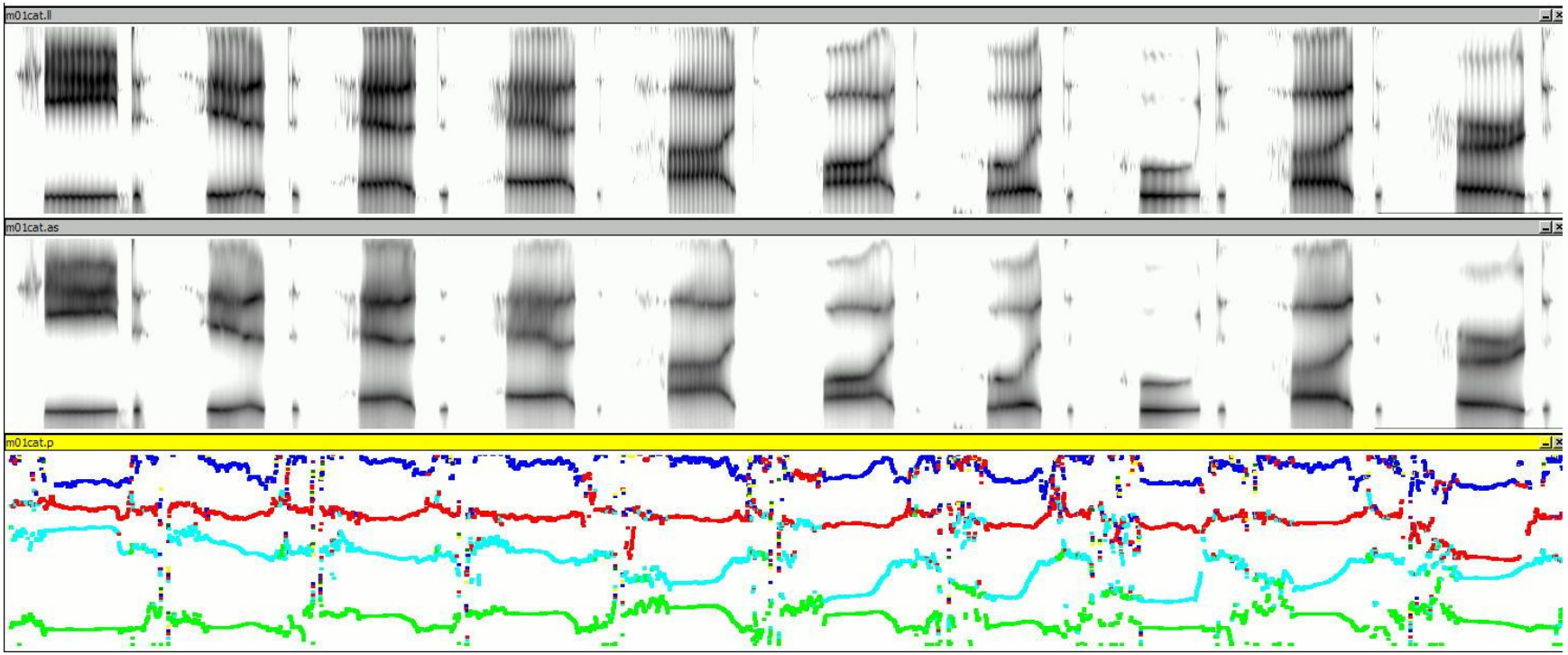
What is the specific acoustic feature that is seen *always and only* for rhotic sounds, giving /ɚ/ such a distinctive sound quality?

[ʌɚ]



What's different about the /ɚ/? (Hint: *Pay attention to the 3rd formant.*)

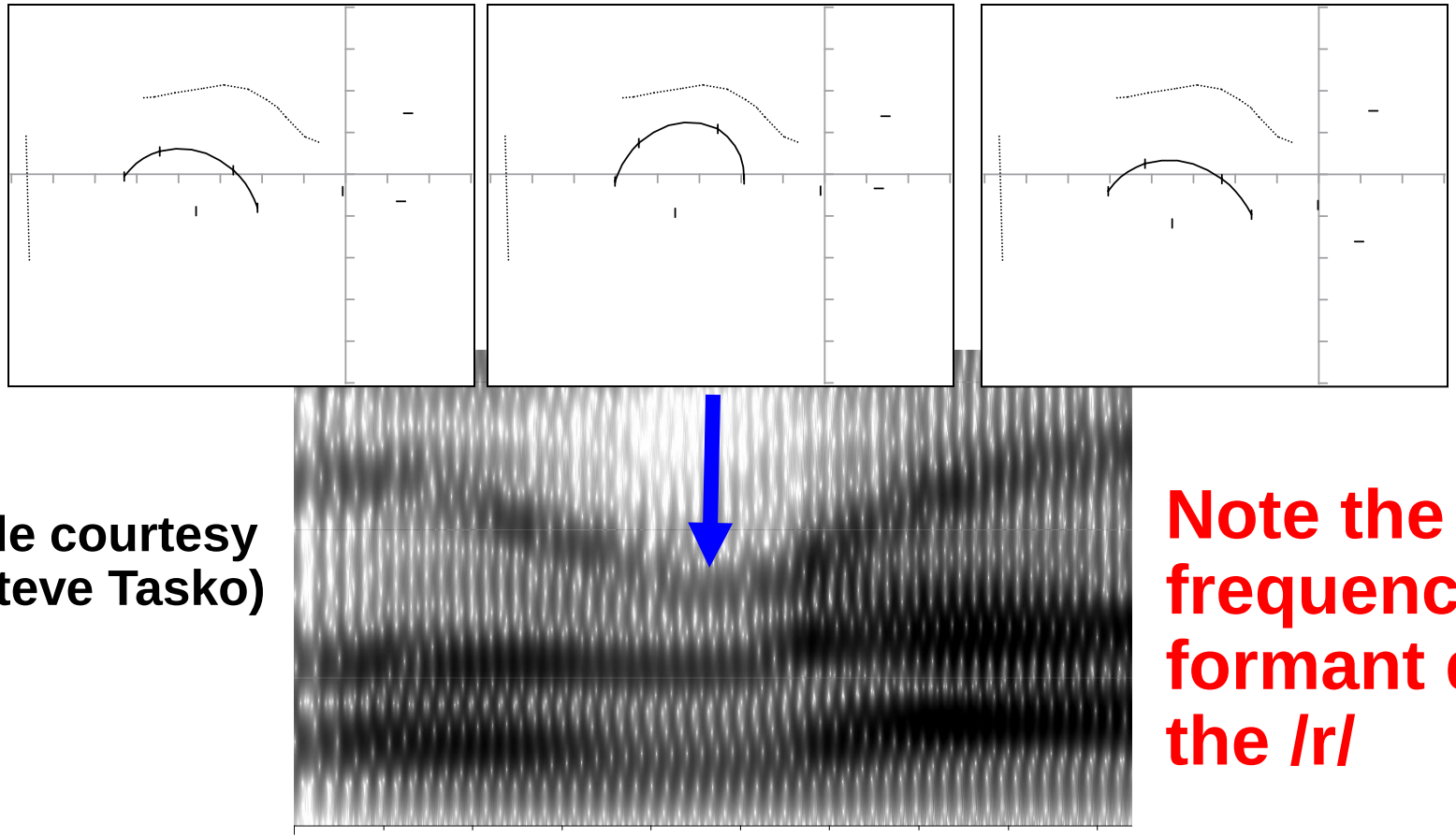
What is the specific acoustic feature that is seen *always and only* for rhotic sounds, giving /ɚ/ such a distinctive sound quality?



What's different about the /ɚ/? (Hint: *Pay attention to the 3rd formant.*)

“Bunched” /r/

(Note that only a portion of the tongue is show here.)

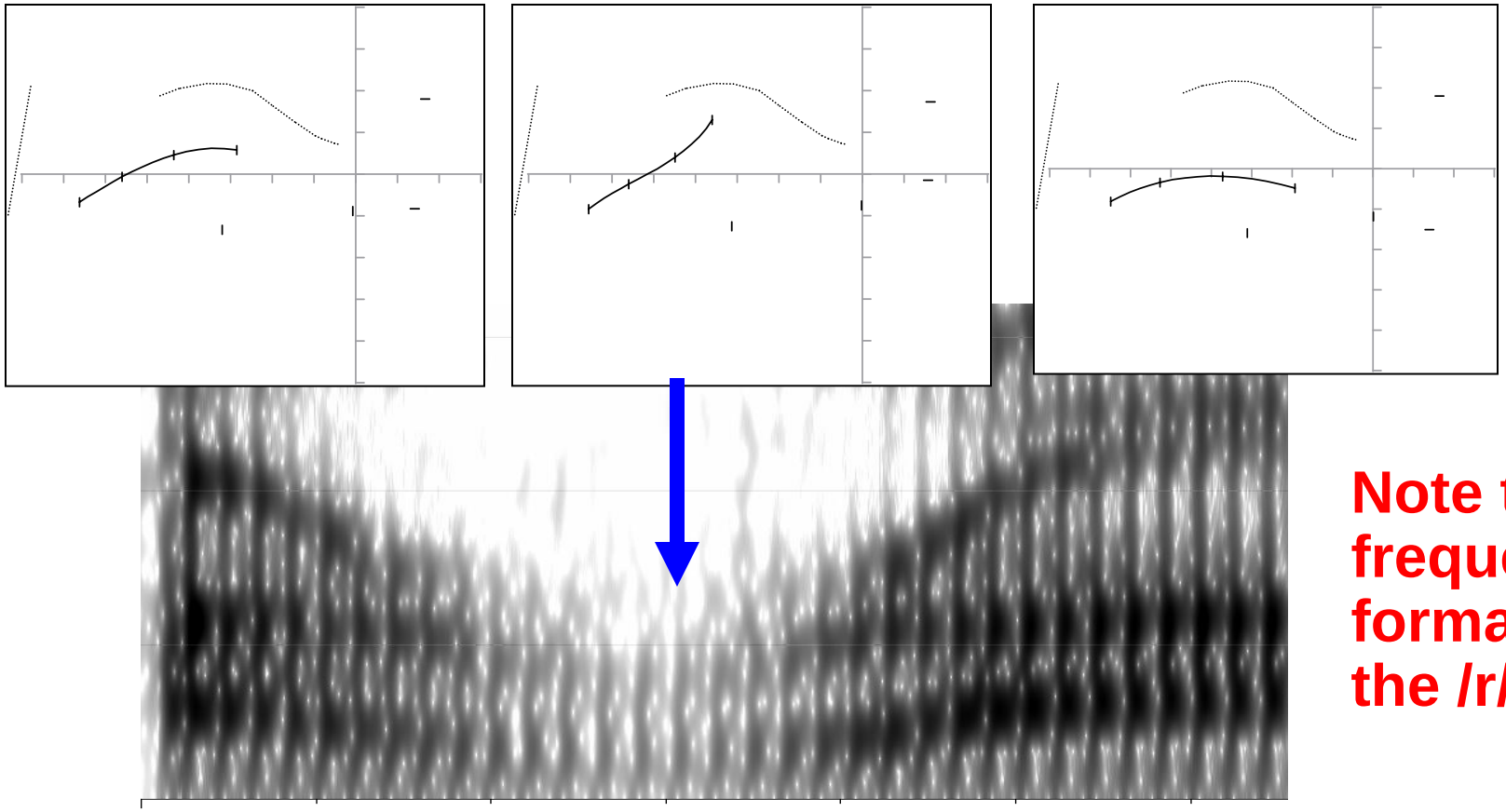


(Slide courtesy of Steve Tasko)

Note the low frequency 3rd formant during the /r/

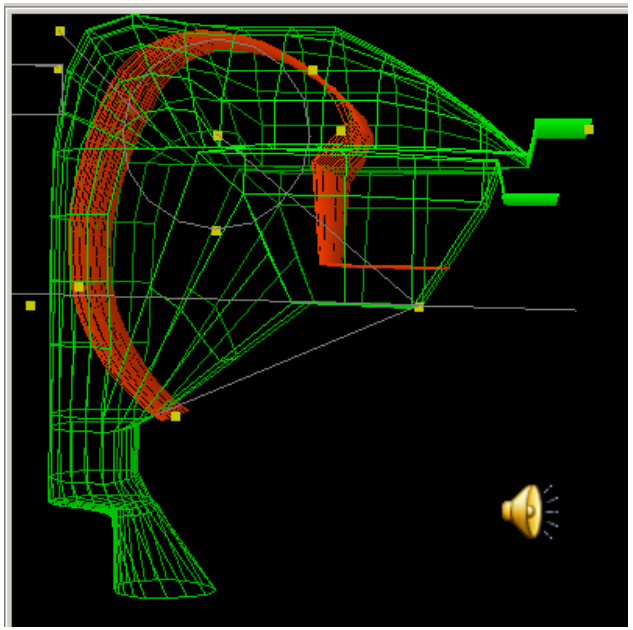
“Retroflex” /r/

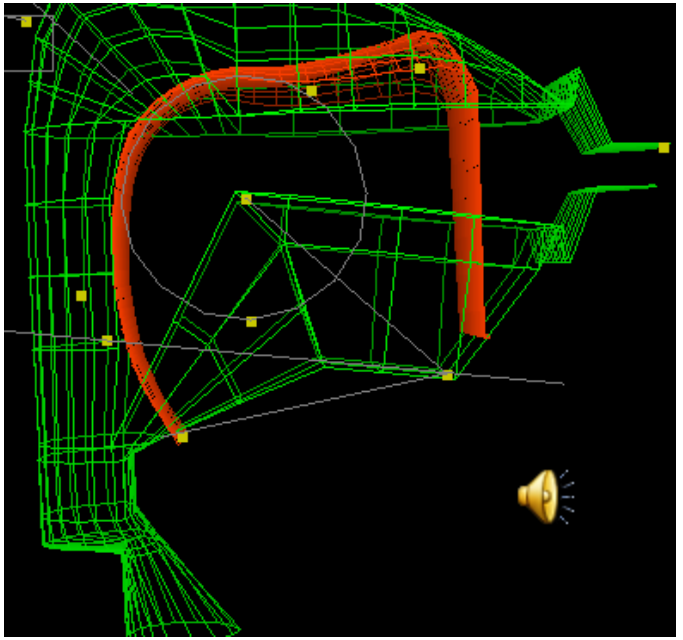
(Note that only a portion of the tongue is show here – tongue root is not shown.)

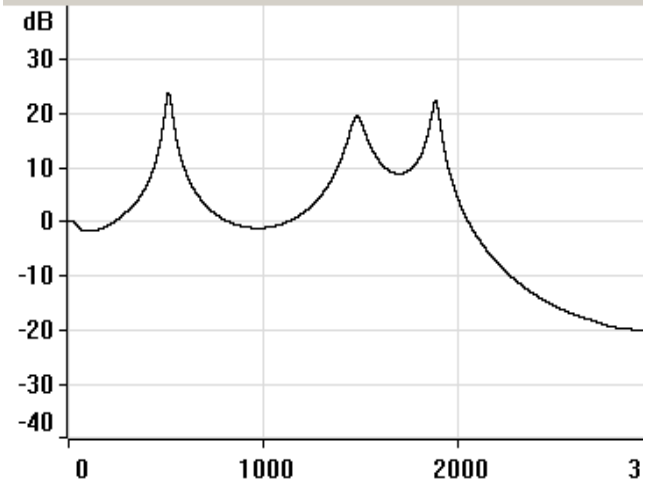


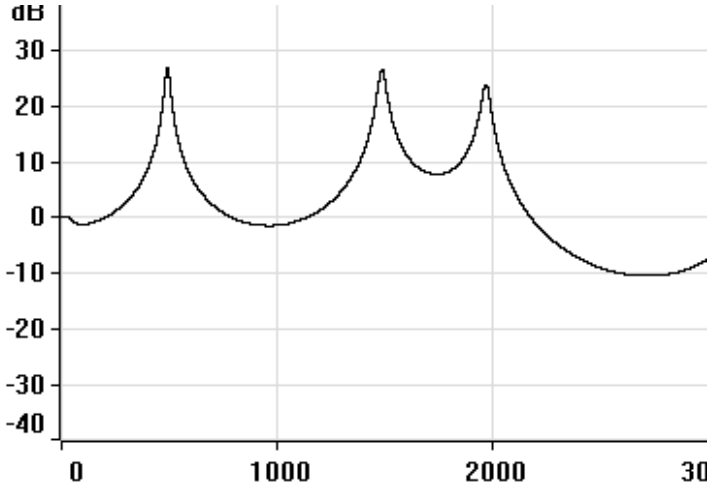
Note the low frequency 3rd formant during the /r/

(Slide courtesy of Steve Tasko)

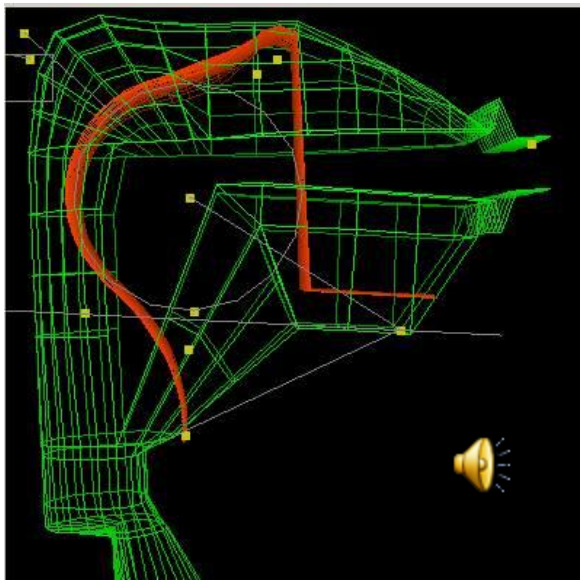
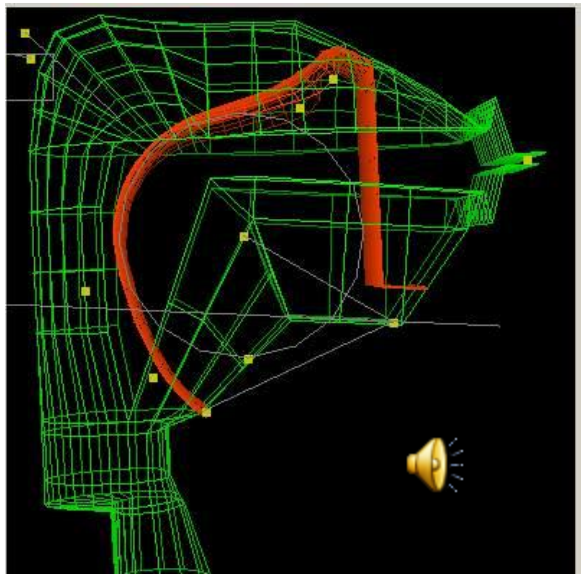
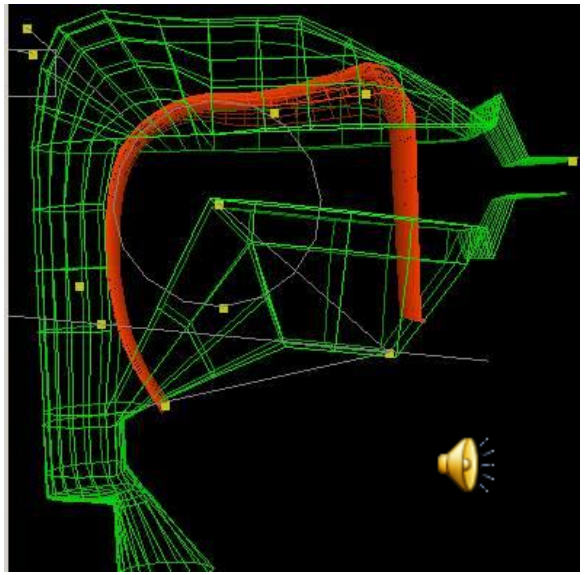








Two very different tongue (and lip) postures, but: (a) similar formants, (b) similar rhotic sound quality.

All three tongue postures are *retroflex*, but also quite different. Note similar rhotic sound quality in all three.

This is all done with a model, but *you see the same kind of variability in real speakers.*

What should we make of this?

(1) Is there *a* way to produce [ɚ] (or [ɹ])? The case of [ɚ,ɹ] is the most obvious, but *this is true of all vowels and most (maybe all) consonants.* The descriptions you see in phonetics texts (including MacKay) are just a guide.

(2) What kind of luck might you run into trying to teach [ɚ,ɹ] by giving a child (or adult) advice about tongue or lip placement?

Rhotic and non-rhotic dialects of English

General American is called a *rhotic dialect*. Others, like British English, some Southern dialects, some New England dialects (e.g., some areas of Boston), some NYC dialects, Australian & New Zealand English, are called *non-rhotic*.

The term non-rhotic is misleading – some underlying /r/'s in this dialect are rhotic; others are not. In rhotic dialects, though, *all /r/ sounds (/r/ and /ɚ/) are rhotic*.

Here's the question*: In "non-rhotic" dialects, *is there a rule* that determines which /r/ sounds are rhotic and which ones are not, *or is it random?

Sample of British RP (Received Pronunciation):

What's the rule here?

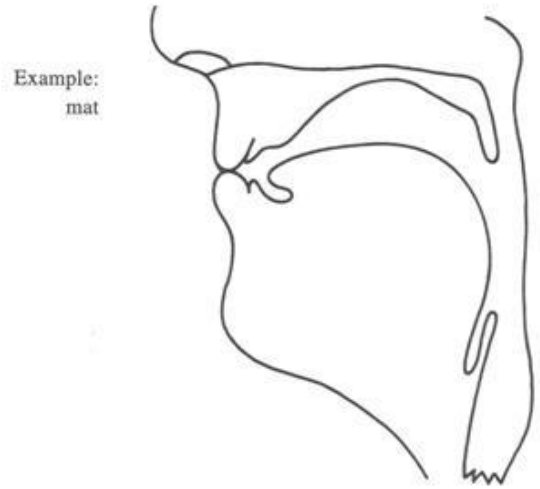
In "non-rhotic" dialects, /ɚ,r/ is not pronounced:
Unless a vowel follows.

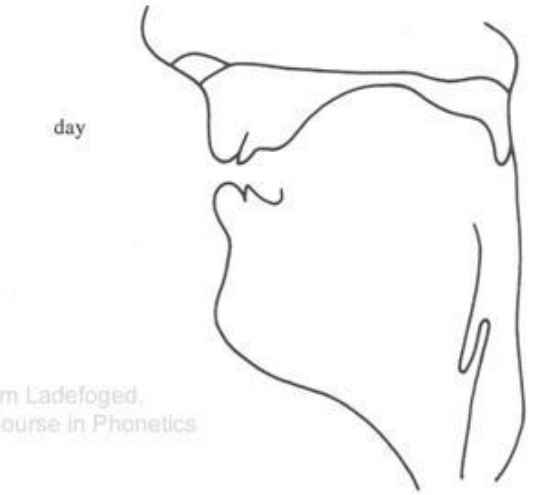
What do you think of this dialect? Rhotic/Non-rhotic? Does the speaker follow the rule we derived for British RP?

Guesses about where this speaker might be from?

6. Nasalized vowels

Any vowel can be nasalized – just drop the velum.





From Ladefoged,
A Course in Phonetics

Vowel nasalization is *not phonemic in English*; i.e., it serves no contrastive function. Vowel nasalization is predictable: If a vowel is adjacent to a nasal consonant, it will be nasalized. Otherwise, it won't.

Vowel nasalization *is* phonemic in some languages – e.g., Portuguese, French. See your notes for a discussion of this.

7. Length/Duration/Quantity

American English Vowels Have Different Typical Durations

/i/	>	/ɪ/
/u/	>	/ʊ/
/æ/	>	/ɛ/
/ɑ/	>	/ʌ/
/ɔ/	>	/ɑ/

Basic idea is simple: **Any vowel can be spoken at any duration, but some vowels are *typically* longer than others. This is called the vowel's *inherent duration* or *typical duration*.**

Vowel length can be indicated in phonetic transcriptions with a colon:

[kɪt] vs. [kɪːd] (“***kit***” **vs.** “***kid***”)

Note that the longer vowel is indicated with a colon. **(A half-colon can be used for “half-long” vowels.)**

Despite these systematic differences in inherent duration, *vowel duration/vowel length is not phonemic in General American English*; i.e., there aren’t pairs of words that different only in vowel length.

The vowel length difference between [kɪt] vs. [kɪːd] **is an example of *allophonic variation*: (a) these two words *do not differ solely on the basis of vowel length*, (b) *vowel length is predictable* (based on the voicing of the final consonant) and therefore *not contrastive.***

Vowel length is phonemic in some languages **– e.g., Japanese, Czech, Hawaiian, Finnish, Thai, Old English, Estonian, many others.**

Examples from Finnish:

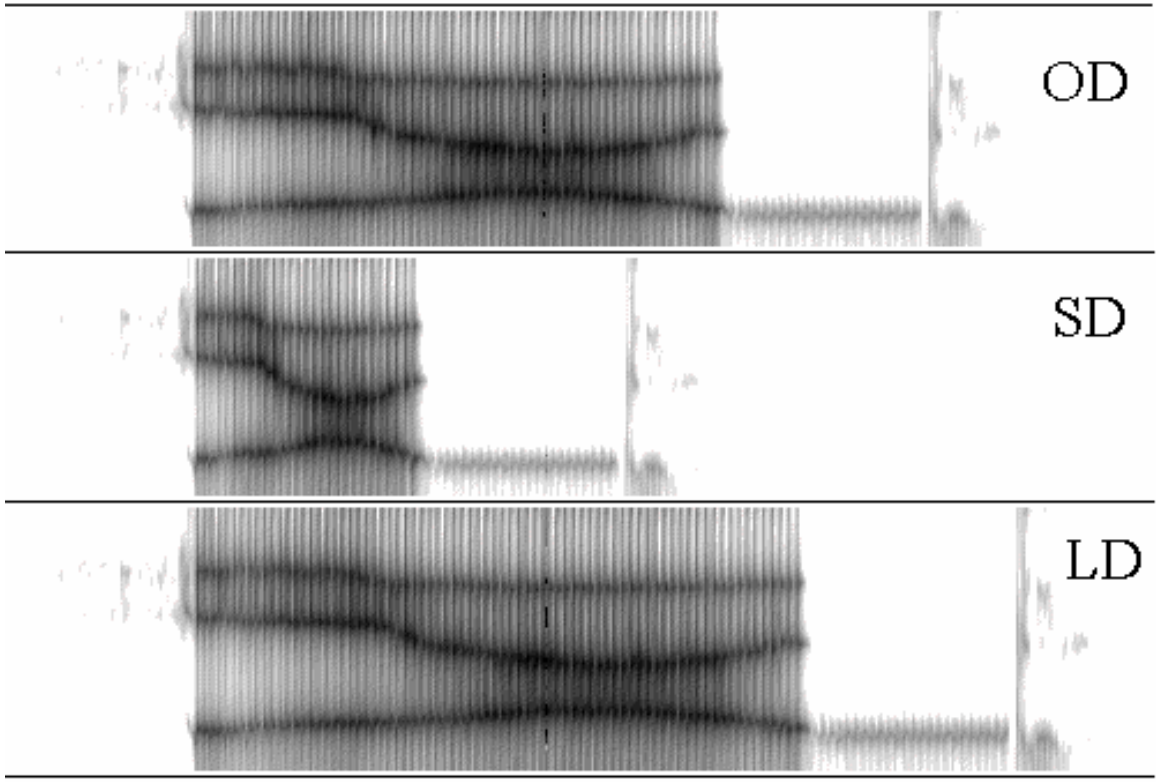
- ***taka (back)***
- ***takaa (from behind)***
- ***takka (fireplace)***
- ***taakka (burden)***

Note also the *long vs. short consonants* – e.g., compare "*taka*" and "*takka*."

(Aside: Are the stops here aspirated or unaspirated?)

OK, so vowel duration is not phonemic in Gen. American. Does that mean it plays no role in vowel perception? (Hint: No)

OK, vowel duration is not phonemic, but does that mean it doesn't matter? Do listeners pay any attention to it in deciding what vowel was spoken or do they ignore it because it is not phonemic?



Original Duration

Short Duration

Long Duration

Logic: If duration plays no role in vowel recognition, the 4 signal types ought to be equally intelligible; i.e., artificially modifying duration will not affect what vowel is heard. On the other hand, *if duration plays a role in vowel perception, the OD signals ought to be more intelligible than any of the duration-modified signals*.

Also, there are *specific kinds of changes in vowel identity that we would expect*. For example:

Shortened /i/ ought to be heard as /ɪ/
Lengthened /ɪ/ ought to be heard as /i/

Shortened /æ/ ought to be heard as /ɛ/
Lengthened /ɛ/ ought to be heard as /æ/

Shortened /u/ ought to be heard as /ʊ/
Lengthened /ʊ/ ought to be heard as /u/

Shortened /ɑ/ ought to be heard as /ʌ/
Lengthened /ʌ/ ought to be heard as /ɑ/

RESULTS

Original Duration:	**96.0%**
Short Duration:	**91.4%**
Long Duration:	**90.9%**

Overall vowel intelligibility drops by ~5%. So, vowel duration matters, but not much – *on average*.

Turns out that average is a little misleading. You get that ~5% figure by averaging cases in which duration does not matter at all with other cases in which duration matters a whole lot.

Effects of Duration on Vowel Perception

original duration: 

long duration: 

short duration: 

original duration: 

long duration: 

short duration: 

original duration: 

long duration: 

short duration: 

original duration: 

long duration: 

short duration: 

original duration: 

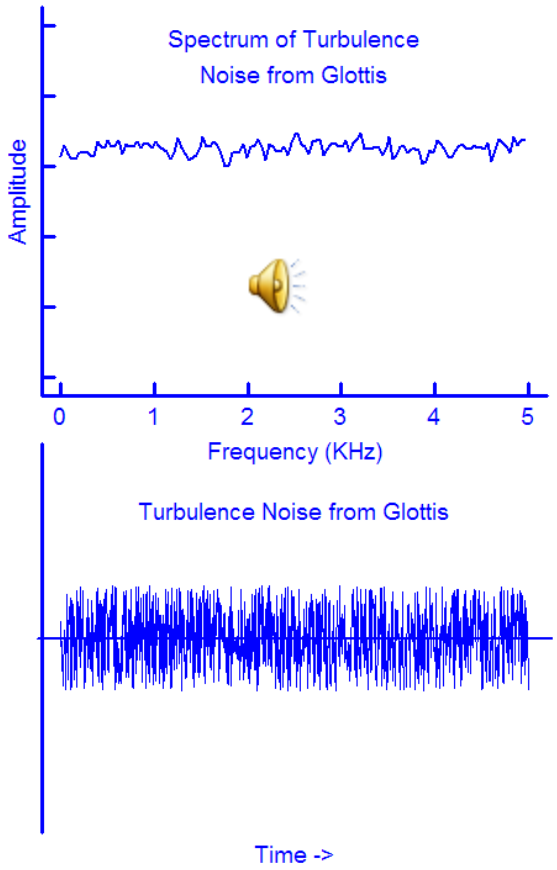
long duration: 

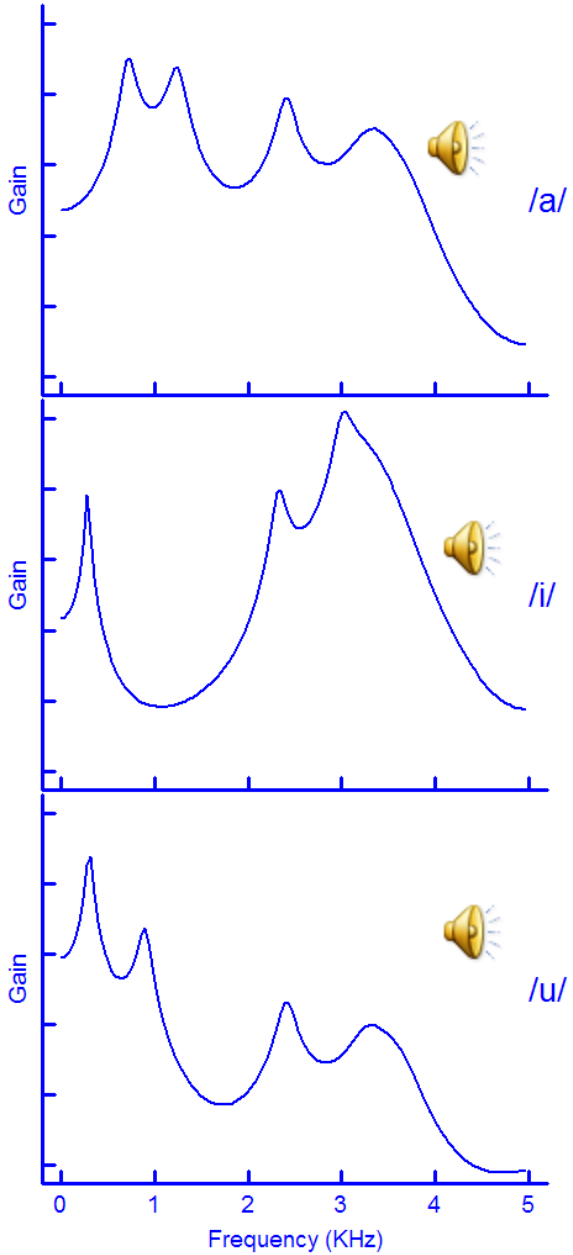
short duration: 

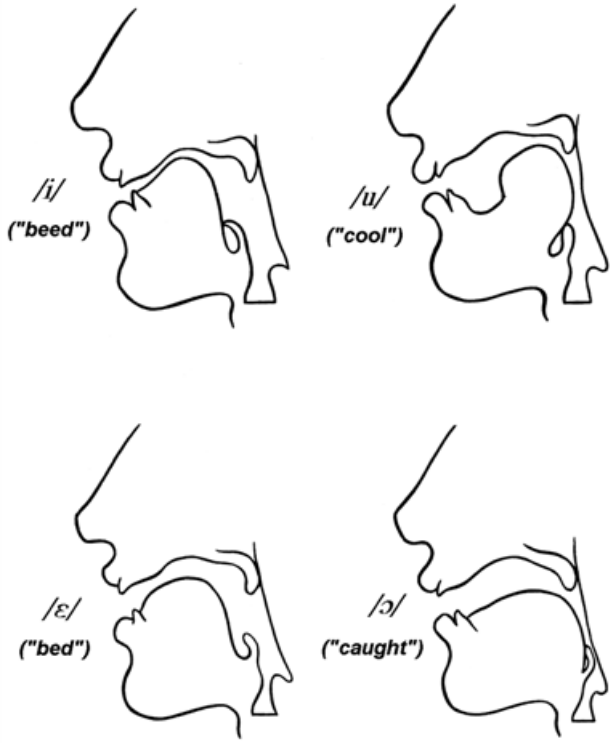
CONCLUSIONS

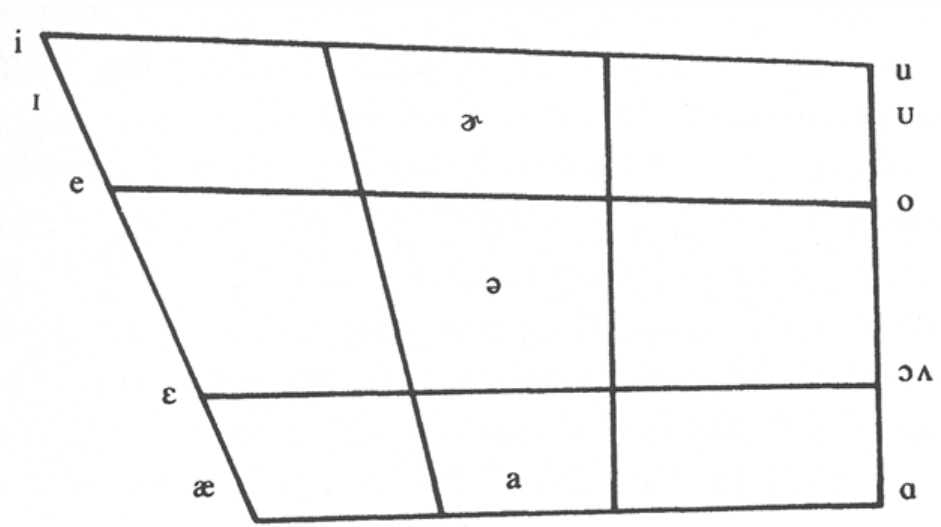
1. Duration has a measurable but fairly small *average* effect on vowel perception.
2. Vowel Shortening: ~5% drop in intelligibility, *averaged across all vowels*
3. Vowel Lengthening: ~5% drop in intelligibility, *averaged across all vowels*
4. Vowels *Most* Affected: /ɑ/-/ɔ/-/ʌ/, /æ/-/ɛ/
5. Vowels *Not* Affected: /i/-/ɪ/, /u/-/ʊ/

8. Phonation/Breathy/Whisper



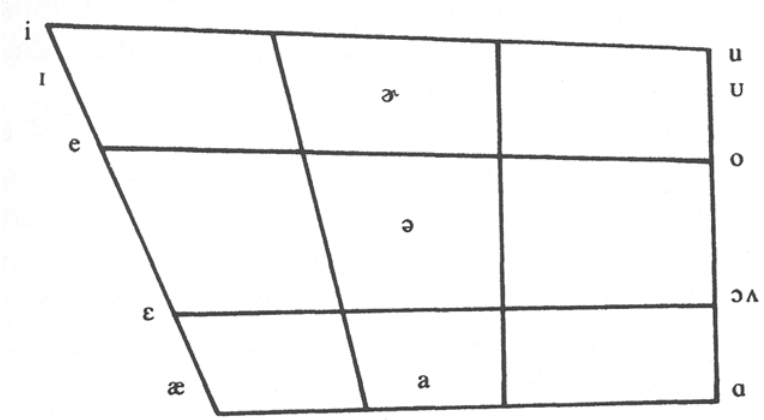




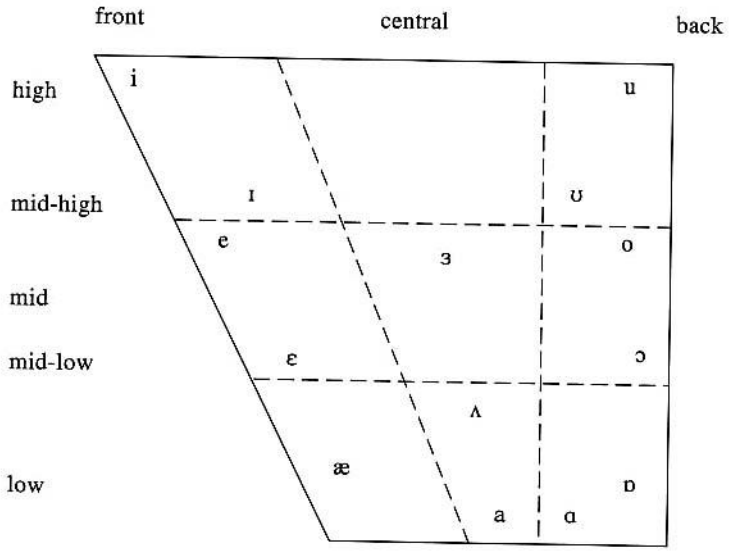


(I don't remember what this slide was supposed to show. I'm leaving it in here in case I remember.)

MacKay



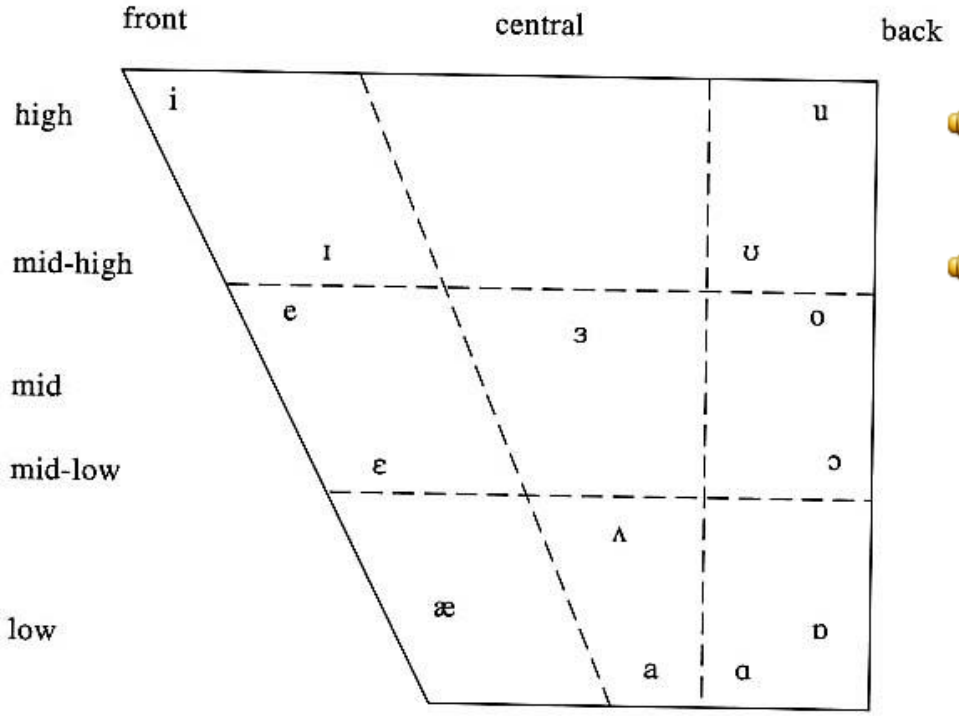
Ladefoged (many others)



Note especially difference in location of [ʌ], but see also [ɪ], [ʊ], [ɑ]. Who's right? Not sure. Probably Ladefoged. Some may reflect dialect differences.

The Strange Case of /æ/ in Lower Michigan (and Syracuse and Buffalo and Toledo and Chicago …)

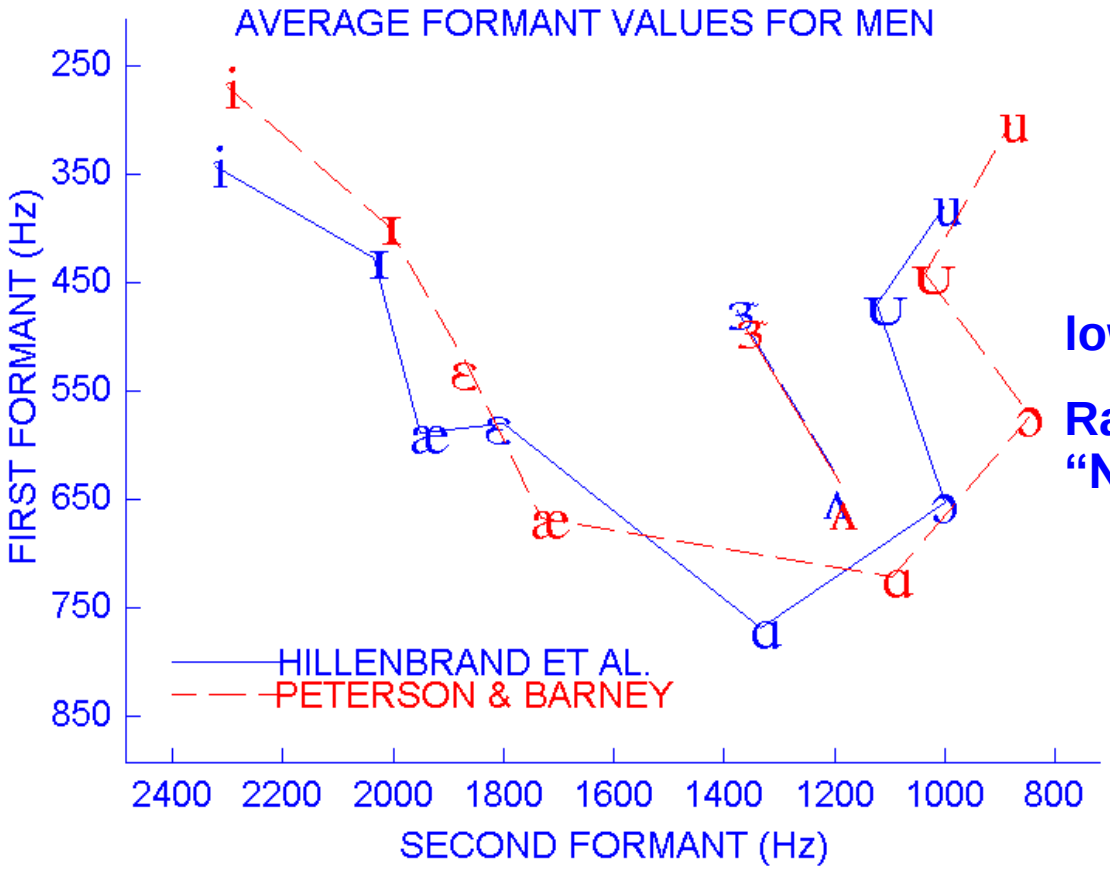
According to the diagram below, what kind of a vowel is /æ/? (Answer: *low* front)



low /æ/

Raised, fronted "Northern Cities" /æ/

Peterson & Barney (Mostly Mid-Atlantic) vs. Hillenbrand et al. (Upper Midwest/Northern Cities)

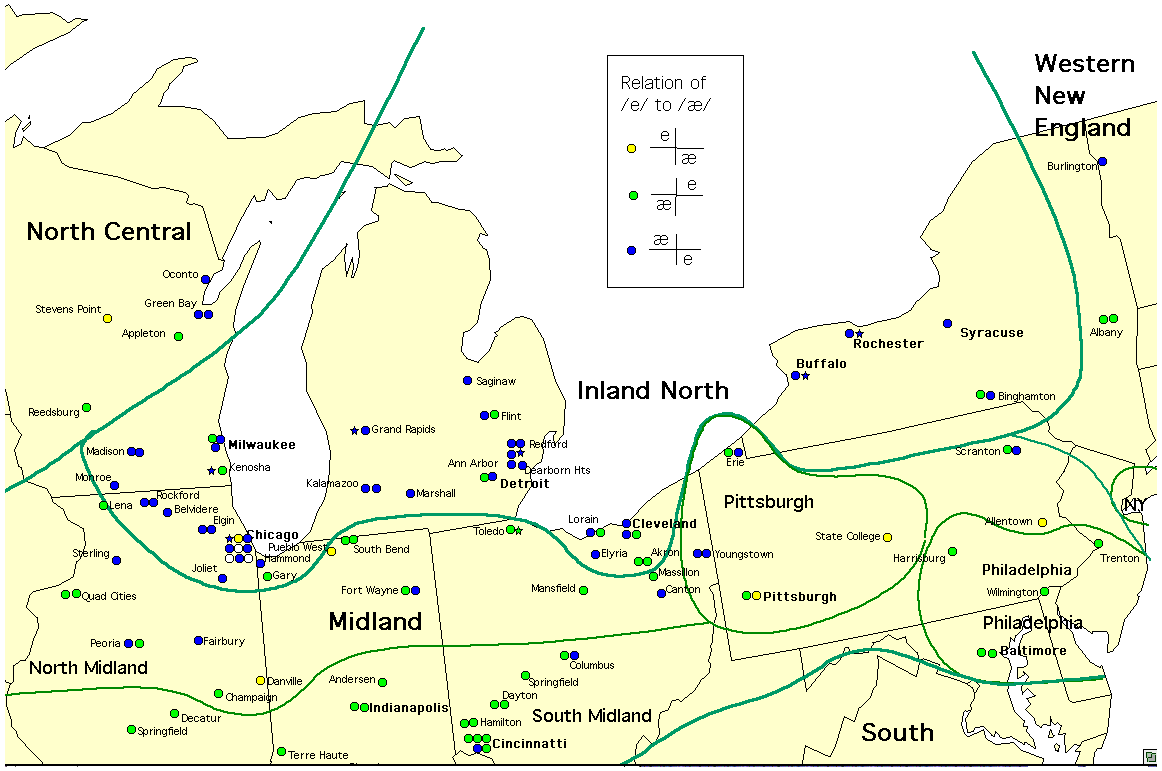


low */æ/*

Raised, fronted "Northern Cities" */æ/*

1. **/æ/ is *raised* and *fronted* in Northern Cities data**
2. **Low back vowels (/ɑ,ɔ/) fronted and much lower in Northern Cities data**
3. **High vowels (/i/ /ɪ/ /u/ /ʊ/) not quite as high in Northern Cities data**

The Northern Cities dialect pattern is observed in a broad region that extends roughly from *upstate NY* (Syracuse, Rochester, Buffalo), *Northern Pennsylvania* (Erie), *Northern Ohio* (Cleveland, Toledo), *Southern Michigan* (Detroit, Flint, Jackson, Kalamazoo, Grand Rapids), *Northern Indiana* (South Bend, Gary), *Northern Illinois* (e.g., Chicago, Rockford) through *south-central Wisconsin* (e.g., Madison).



The NC dialect pattern *divides many states north-to-south;* e.g., the NC pattern is seen in Cleveland but not Cincinnati; Chicago but not Bloomington; S. Bend but not Evansville.

U Penn Linguist William Labov was the 1st to identify this pattern.

He calls it a dialect, but *dialect group* may be a better term – speech patterns vary quite a bit within the NC region; e.g., the speech patterns in Syracuse and Chicago are quite different, though there are features in common.

The features that characterize the NC dialect group are *seen in their most extreme form in the largest cities* in this region – e.g., the raising of /æ/ and the lowering and fronting of /ɑ/ and /ɔ/ are seen more in Detroit, Flint & Chicago than in some of the smaller cities and towns.

Schwa

Schwa is called the *neutral vowel* – because it's as centralized as a vowel can get: *tongue is not front and not back, not high and not low; lips are not rounded and not spread/retracted* (as in /i/). *The vocal tract is said to be in a neutral position.*

Symbol: [ə] (upside down 'e' – *practice it*)

Examples: about, Cuba, mantra, abut, exclamation (more later)

The schwa in English is unstressed.

Quality seems similar to /ʌ/ or sometimes a very short /ɪ/, but *always very brief and unstressed*.

Phonetician Bill Tiffany: *Schwa is one step away from a burp.*

Which of these words contains schwa, and where is it?

amazing

[əmezɪŋ] **(velar nasal [ŋ] is new – later)**

altogether

[ɔltəgɛðɚ]

another

[ənʌðɚ]

habit

[hæbət] or [hæbɪt]

mantra

[mɑntrə]

feeder

[fidɚ] **(no schwa here, right?)**

Which of these words contains schwa, and where is it?

fascination

[fæsəneʃən]

buzz

[bʌz] **(no schwa)**

bunker

[bʌŋkɚ] **(no schwa; note velar nasal [ŋ] again)**

laboratory

[læbrətɔri]

elephant

[ɛləfənt]

cantor

[kæntɚ] **(no schwa)**

Which of these words contains schwa, and where is it?

sponsor

[spɑnsɚ] **(no schwa)**

plus

[plʌs] **(no schwa)**

blanket

[bleŋkət] **or** [bleŋkɪt] **(note velar nasal again)**

synonym

[sɪnənɪm]

liquify

[lɪkwəfai] **(*/*ai*/* is a diphthong we haven't seen yet)**

laminate

[læmənet]

Which of these words contains schwa, and where is it?

butter

[bʌɾɚ] **(no schwa – note the /ɾ/ symbol (flap); later)**

musket

[mʌskət] or [mʌskɪt]

Mackinaw

[mækənɔ]

pasted

[pestəd]

Kansas

[kænzəs] **(note the** [z] **– this is not** [kænsəs]

candies

[kændiz] **(no schwa)**

Which of these words contains schwa, and where is it?

practice

[præktəs]

pansies

[pænziz] **(no schwa)**

lasses

[læsəz]

busted

[bʌstəd]

nationality

[næʃənælərɹi]

mustache

[mʌstæʃ] **(no schwa)**

Diphthongs

di **= two;** ***phthong*** **= sound; so, two sounds**

Different from monophthong = 1 sound

/*i*/, ***/*u*/***, ***/*ɪ*/*** **, etc. =** ***monophthongs***

/*aɪ*/ **(*buy*),** ***/*aʊ*/*** **(*bow*),** ***/*ɔɪ*/*** **(*boy*) =** ***diphthongs*****; i.e., two sounds, or two vowels; one vowel quality slides into another one:**

"*buy*": /a/->/i/; "*now*": /a/->/u/; "*boy*": /ɔ/->/i/

Note: The word is *diph*thong, *NOT dip*thong; similar to *diph*theria, NOT *dip*theria.

Similarly, mono*phthong*, NOT mono*thong*

That 'fth' combination feels and sounds odd, but that's what it is.

Diphthongs are also called *vowels of changing color* or *vowels of changing timbre*. The terms *vowel color* and *vowel timbre* here mean that same as vowel quality – i.e., whether the vowel sounds like */i/* or */ɪ/* or */u/* or */æ/*.

These terms all mean the same thing: *The perceptual quality of the vowel changes throughout the course of the vocoid* (remember vocoid? – vowelish thing including both monophthongs and diphthongs).

Transcription of Diphthongs

What two vowel qualities are in "*buy*," "*high*," "*lie*"?

[ɑi]? [ɑɪ]? [ai]? [aɪ]?

At an abstract linguistic (i.e., phonemic) level, it's probably /ɑi/. At a narrow phonetic level it may be closer to [ai] or [aɪ]. You'll see several variations on this. MacKay uses [ai]; we'll stick with it.

What two vowel qualities are in "*boy*," "*toy*," "*noise*"?

[oi]? [oɪ]? [ɔi]? [ɔɪ]?

At an abstract linguistic (i.e., phonemic) level, it's probably /oi/. At a narrow phonetic level it may be closer to [ɔi] or [ɔɪ]. You'll see several variations on this too. MacKay uses [ɔi]; we'll stick with it.

What two vowel qualities are in "*now*," "*house*," "*mouse*"?

[ɑu]? [au]? [ɑʊ]? [aʊ]?

At an abstract linguistic (i.e., phonemic) level, it's probably /ɑu/. At a narrow phonetic level it may be closer to [au] or [aʊ]. You'll see several variations on this too. MacKay uses [au]; we'll stick with it.

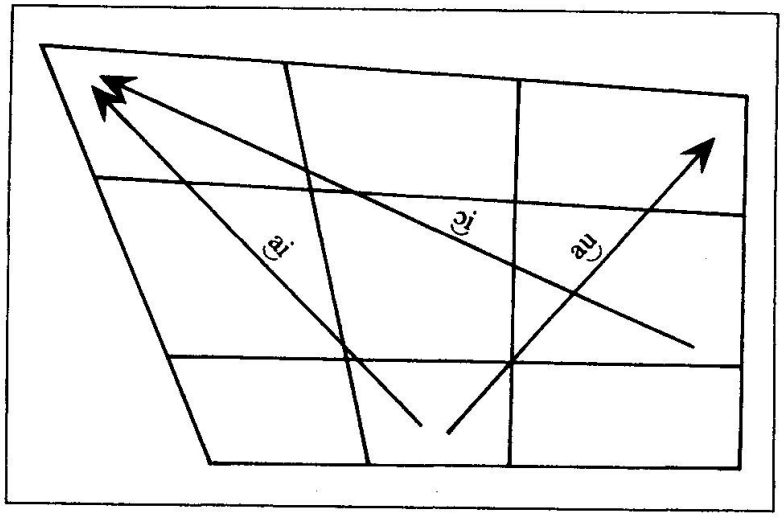


Figure 4-25. Diphthong vectors for the three English diphthong phonemes.

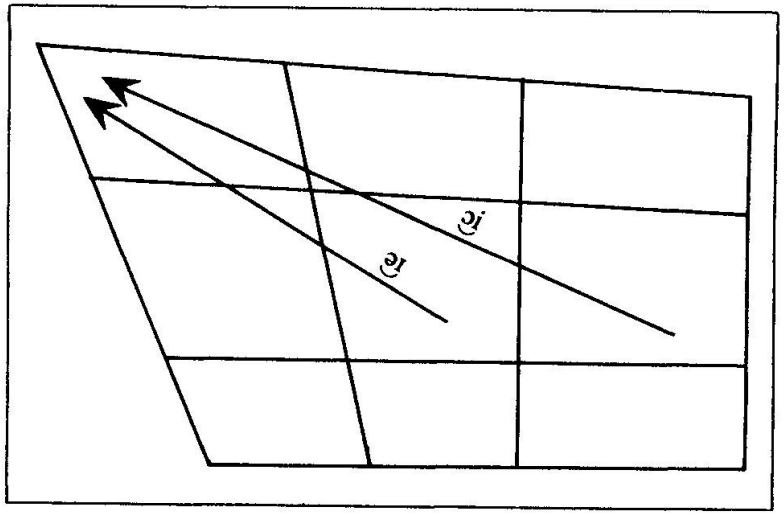


Figure 4-26. Diphthong vectors for Brooklynese [ɔɪ] (for standard /ɚ/) as compared to standard English /ɔɪ/.

***Diphthongs can be represented as arrows (vectors) on the standard vowel quadrilateral*, to show what kinds of vowel movements are occurring. The start and end points should be thought of as approximate. They're usually based on the intuitions of the phonetician who drew the vectors.**

What's going on with these words:

bide, silo, lied, fiber, imbibe, libel, Nile, scribe, eyes

Versus these words:

bite, viper, right, light, pipe, like, ripe, Nike, nice

Are the diphthongs in the two groups pronounced the same way? In many but not all dialects:

[ʌi] or [əi] before syllable-final voiceless consonants

[ai] in other environments

Listen to the diphthongs in these pairs:

lied-light, eyes-ice, hide-height, lies-lice, bide-bite

Rhotic Diphthongs

English has quite a few diphthongs consisting of some initial vowel followed by /ɚ/

Some examples:

beer near fear rear

bear dare care chair

floor door war sore

*poor boor (**some dialects only**) tour contour*

tar bar far car

Three quick points: (1) the 2nd sound is /ɚ/ in all cases; the key question is the phonetic quality of the 1st vowel; (2) pronunciation of these diphthongs varies quite a bit across dialect; (3) transcription practices vary across phoneticians, even when dialect is not an issue.

That said, what do you think the 1st vowel is in:

beer near fear rear

/iɚ/ **or** ***/ɪɚ/*** **-- both are used; the 1st vowel is not** ***/i/*** **and not** ***/ɪ/*****; it's** ***intermediate*** **(but closer to** ***/i/*****)**

bear dare care chair

/eɚ/ **or** ***/ɛɚ/*** **-- both are used; the 1st vowel is not** ***/e/*** **and not** ***/ɛ/*****; it's** ***intermediate*** **(but closer to** ***/e/*****)**

floor door war sore

/oɚ/ **or** ***/ɔɚ/*** **-- both are used; the 1st vowel is not** ***/o/*** **and not** ***/ɔ/*****; it's** ***intermediate*** **(but closer to** ***/o/*****)**

poor boor (**some dialects only**) tour contour

/uɚ/ or /ʊɚ/ -- both are used; the 1st vowel is not /u/ and not /ʊ/; it's *intermediate* (but closer to /u/)

In many dialects (including mine), *poor* and *boor* are pronounced with the same diphthong as *door* and *floor*:

/doɚ/ /floɚ/ /poɚ/ /boɚ/

In others:

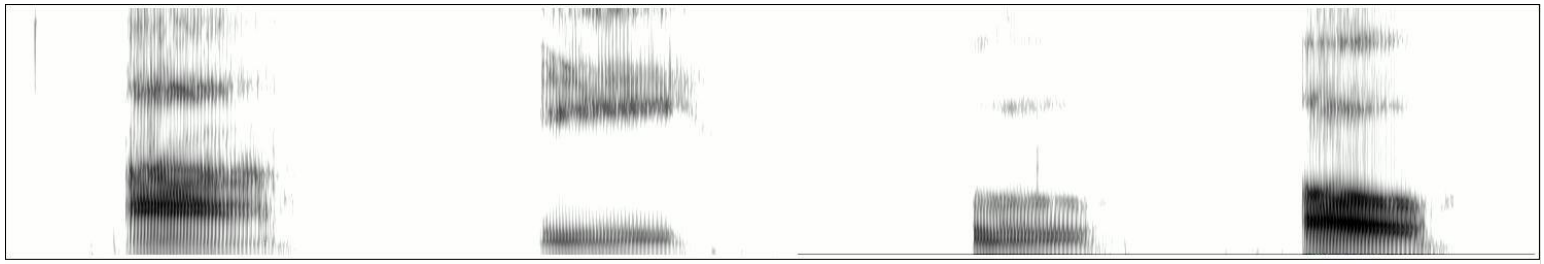
/doɚ/ /floɚ/ **versus** /puɚ/ /buɚ/ /tuɚ/ /kɑntuɚ/

tar bar far car

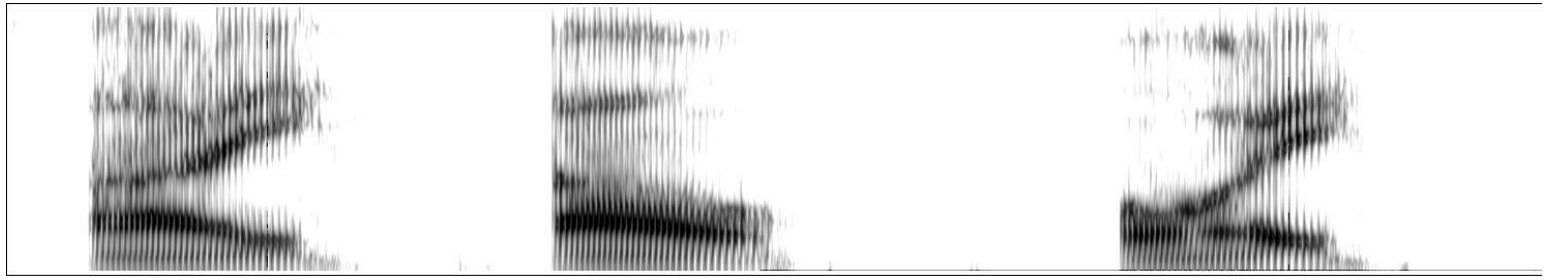
/ɑɚ/ (or maybe /aɚ/; this one sounds like a normal /ɑ/ to me, at least in my speech)

Summary

- **English has lots of /ɚ/ diphthongs.**
- **Pronunciation varies quite a bit across dialects.**
- **Transcription of that 1st vowel is difficult because the vowel quality *does not correspond exactly to any monophthongal vowel*; that is, they are usually *intermediate between two vowels*.**
- **Phoneticians do not agree on how these should be transcribed. *You may well see transcriptions like /ɑɹ/, /iɹ/, /eɹ/,* etc.**

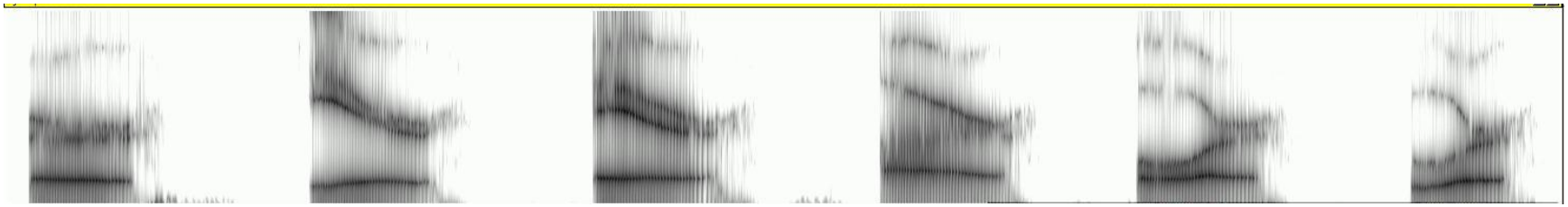


/ɑ/ /i/ /u/ /o/



/ai/ /au/ /oi/

Notice that the formants for the monophthongs are fairly stationary (*this won't always be the case – more later*). The diphthongs, on the other had, show quite a bit of formant movement. ***The formants change frequency because the articulators are moving*** throughout the course of the diphthong.



/ɚ/ /iɚ/ /eɚ/ /aɚ/ /oɚ/ /uɚ/

Notice that, as with the other diphthongs, ***the formants are moving rather than stationary***. Note also the ***drop in the frequency of F_3 during the /ɚ/.***

Last point on diphthongs: ***How are diphthongs different from monophthongs?***

Easy, right?

For monophthongs the articulators remain more-or-less stationary throughout the course of the vocoid, but for diphthongs they move. So:

Diphthongs:

(1) the articulators move

(2) therefore, the formants change frequency over time

(3) therefore, vowel quality (timbre) changes (e.g., a>i; a>u; o>i)

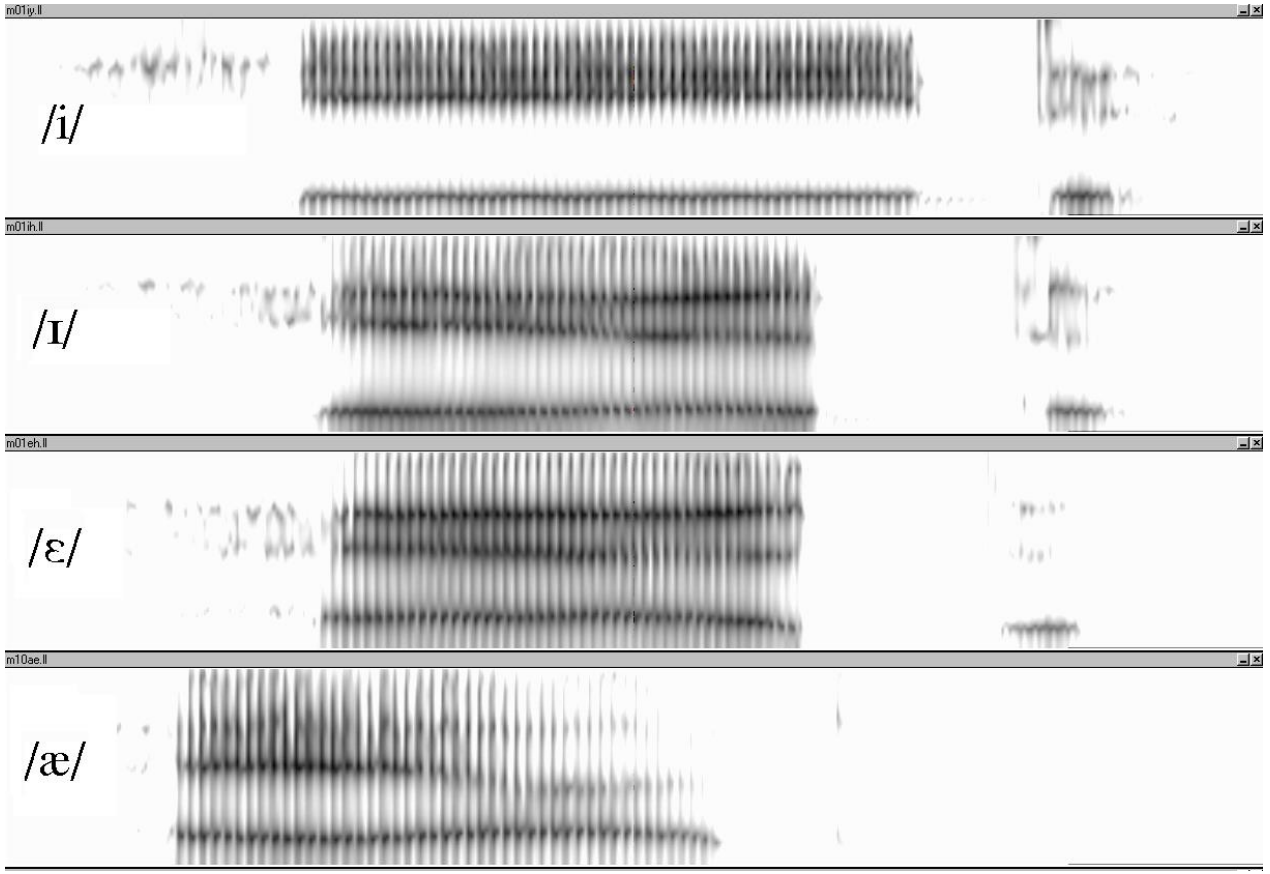
Monophthongs:

(1) the articulators remain more-or-less stationary

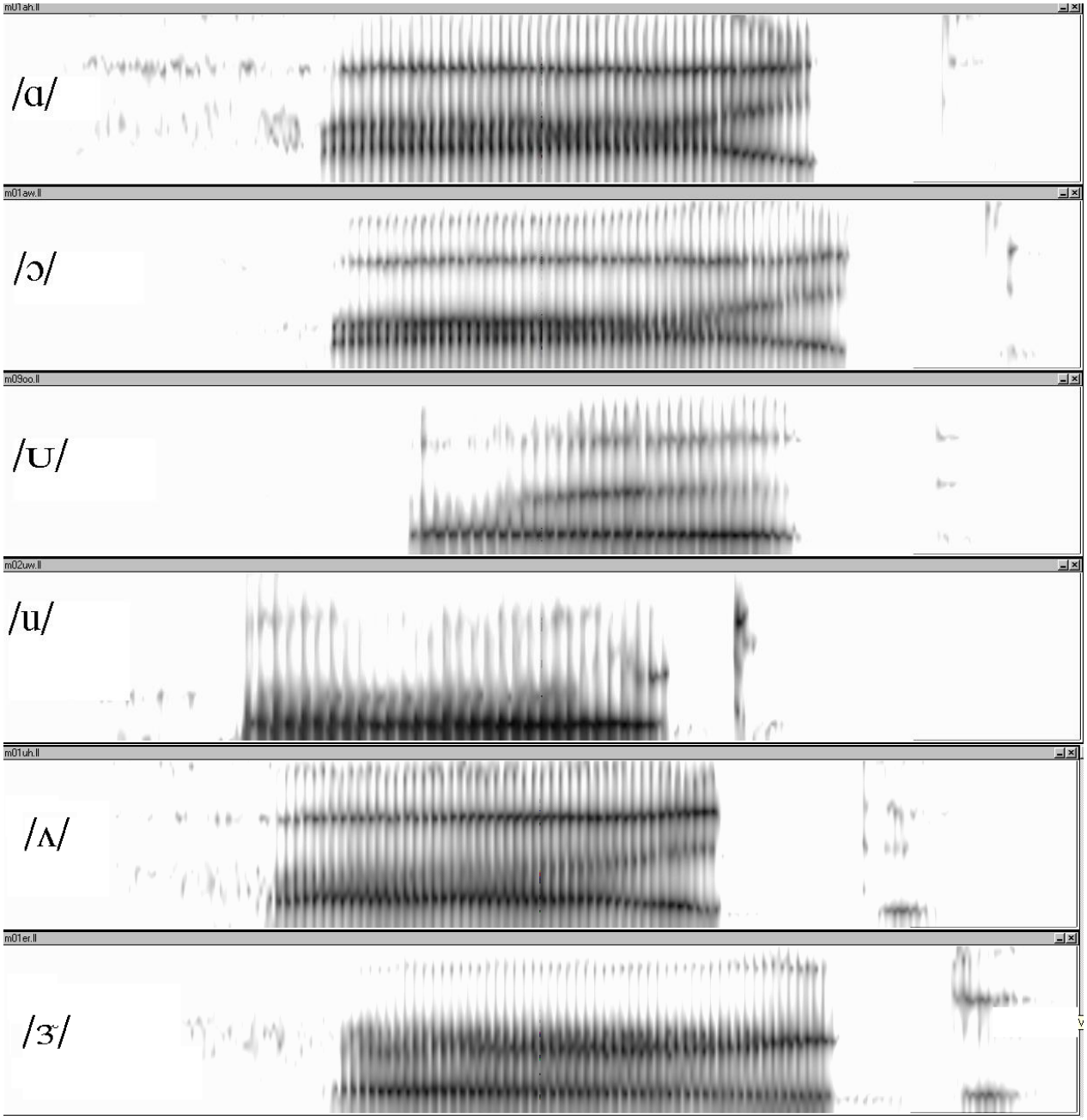
(2) therefore, the formants change don't frequency over time

(3) therefore, vowel quality (timbre) remains fairly constant

Is the story really that simple? Vowels stay put, diphthongs move?

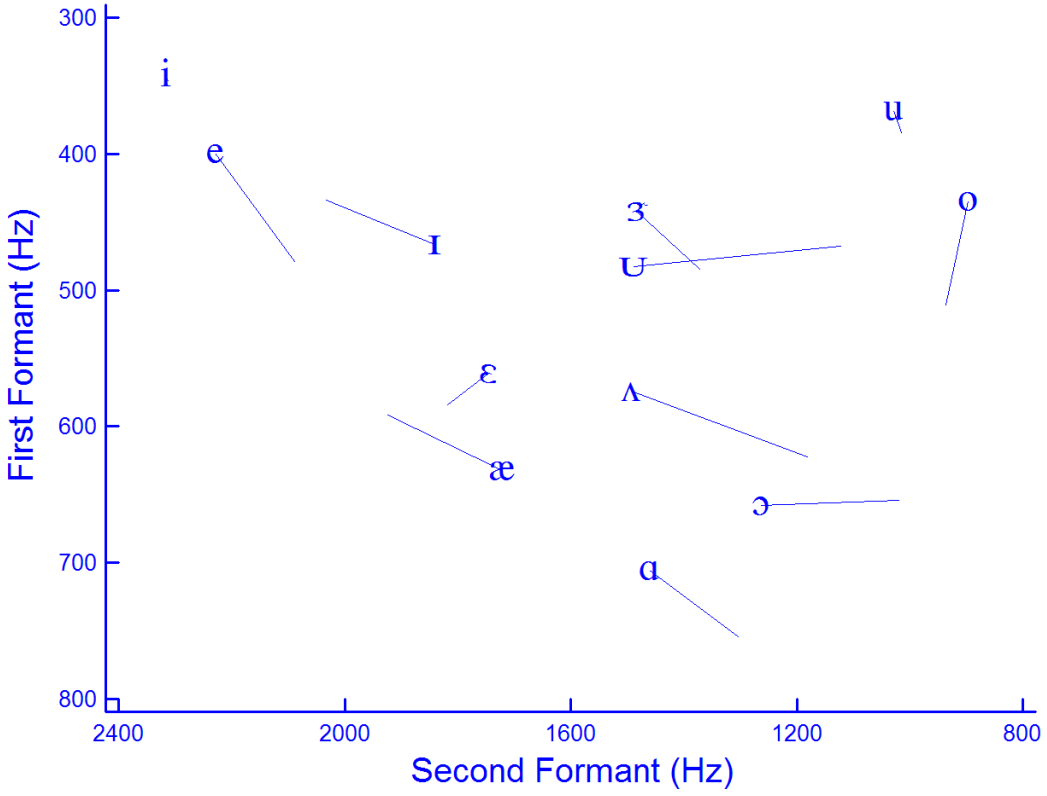


Notice that ***some vowels – especially /æ/ and /ɪ/– show a fair amount of change*** in formant freq's throughout the course of the vowel. Is it possible that these formant movements are perceptually significant?



More examples. Note especially the rise in *F*2 for /ʊ/ and /ʌ/.

Here's another way of looking at diphthong-like movements for "monophthongs".

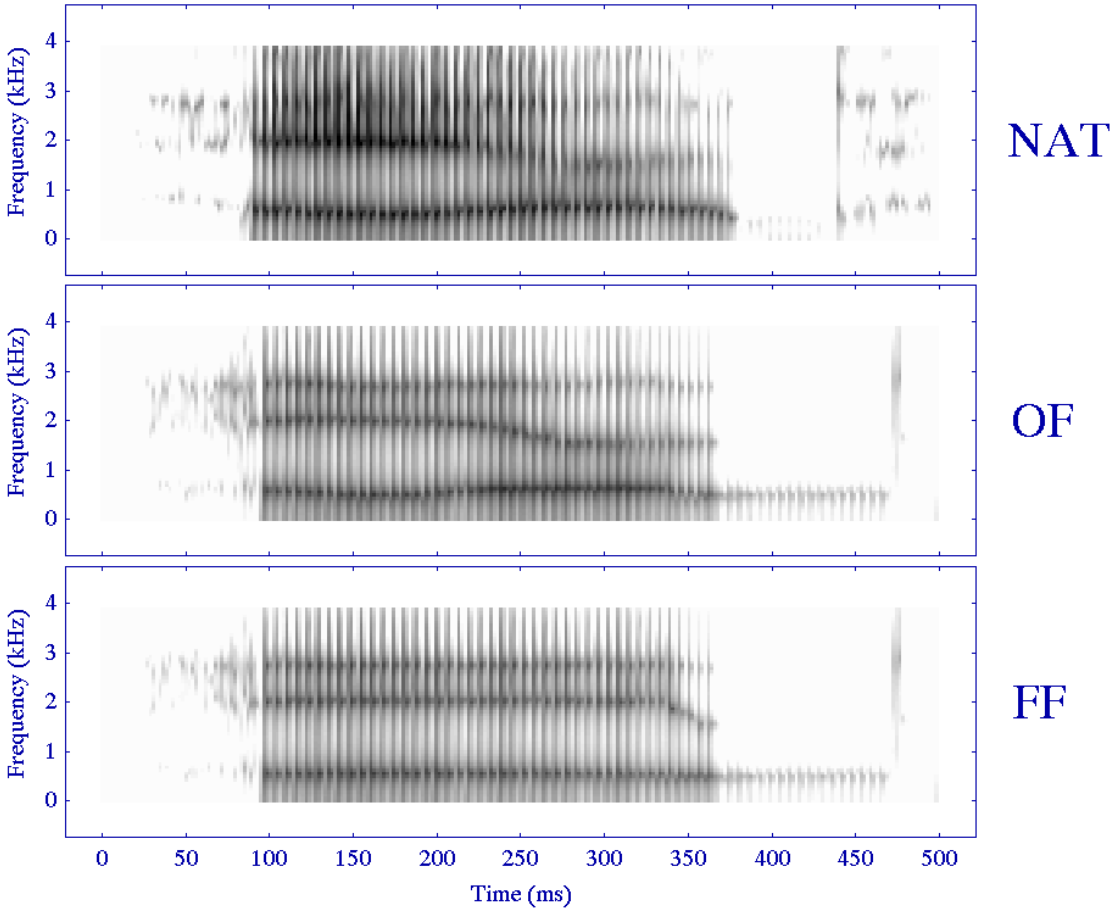


This figure shows how the articulators change through time for 12 American English vowels. (Note: These are formant frequency changes, but the formants can't change unless the articulators move. ***You can interpret this figure as a vowel quadrilateral.***) Each of the lines is drawn from 20% of vowel duration to 80% of vowel duration – the phonetic symbols are drawn at the ***end*** of the vowel.

Do these "monophthongs" remain stationary?

/i/ and /u/ don't show much movement. What about the others? What about MacKay's statement that long but not short vowels show diphthong-like movement?

Do these dynamic diphthong-like movements matter to listeners? Maybe they're minor little changes that are ignored by listeners, or maybe not even noticed. An experiment was run to test this.



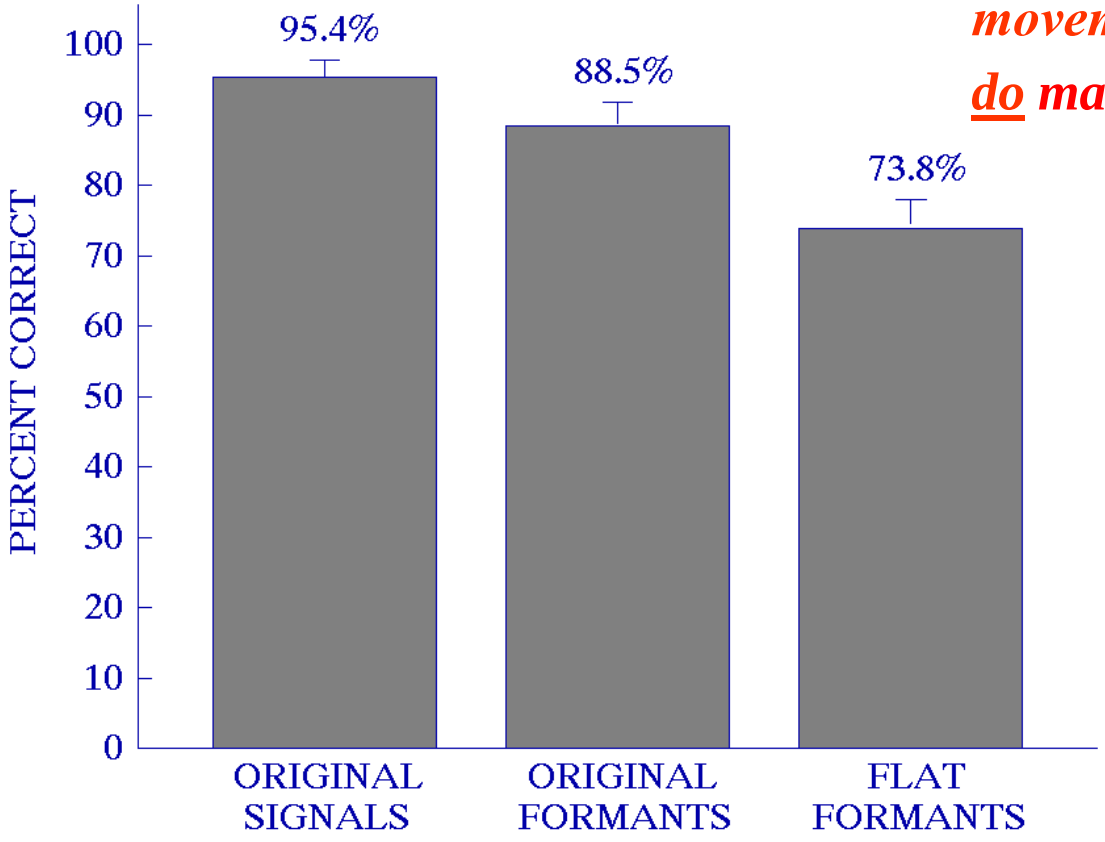
NAT: ***Naturally spoken*** /hæd/

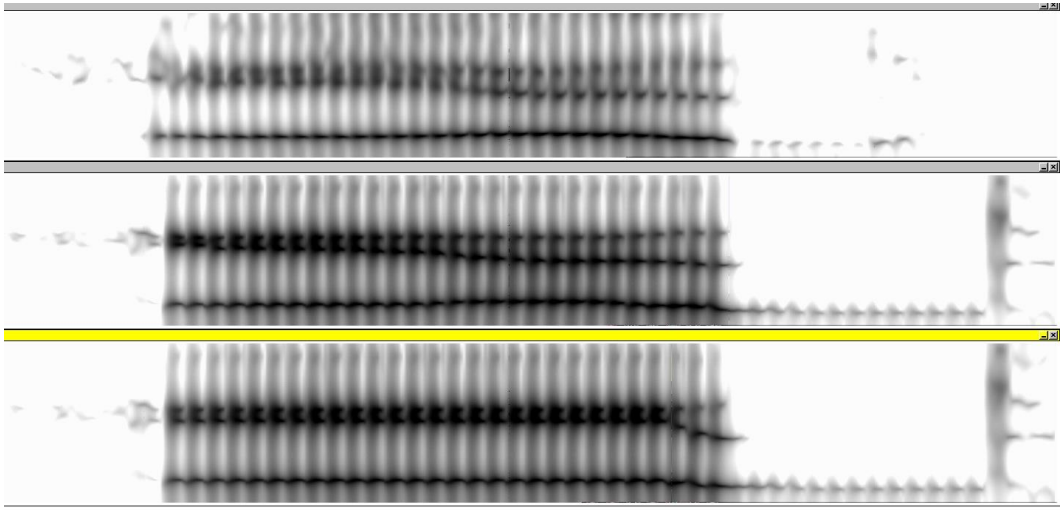
OF: Synthesized, preserving ***original formant contours***

FF: Synthesized with ***flattened formants***

Key comparison is OF vs. FF: If the formant movements don't matter, flattening the formant contour will not affect the vowel percept, and the recognition rates for OF and FF should be very similar. On the other hand, ***if the diphthong-like formant movements are important, the FF signals will be less intelligible than the OF signals.***

Conclusion: *Diphthong-like movements for "monophthongs" do matter.*





NAT

OF

FF

Vowel intelligibility drops ~15% (~89% vs. ~74%) ***on average*** when dynamic changes are artificially removed, but you get this 15% figure by averaging some vowels for which these dynamic features are unimportant with other vowels for which these same features are very important.

NAT

OF

FF

Bottom line:

1. Nearly all "monophthongal" vowels show a good deal of movement – the only real exceptions are /i/ and /u/.

2. MacKay's suggestion that these diphthong-like movements characterize only long vowels is not true.

3. ***These dynamic changes are quite important to vowel intelligibility:*** When dynamic changes are artificially removed using synthesis tricks, vowel intelligibility suffers.

So, if it's not movement vs. no movement, ***how are diphthongs different from monophthongs?*** Answer: ***Only monophthongs can be spoken without articulatory movement.***

Try making prolonged, stationary versions of vowels like /ɪ/, /ʊ/, /o/, /e/, and /æ/. Not perfect, but it can be done.

How about /ai/, /au/, and /oi/? Can these be spoken without moving your tongue and/or jaw and/or lips?

<u>Conclusion</u>: Dynamic changes are important <u>secondary</u> cues for monophthongs. They are essential for diphthongs.

The Cardinal Vowels

Cardinal vowel concept was developed by phonetician ***Daniel Jones*** to address a very particular problem: There is a nearly infinite range of vowel qualities, but the IPA has a finite set of symbols to describe these qualities.

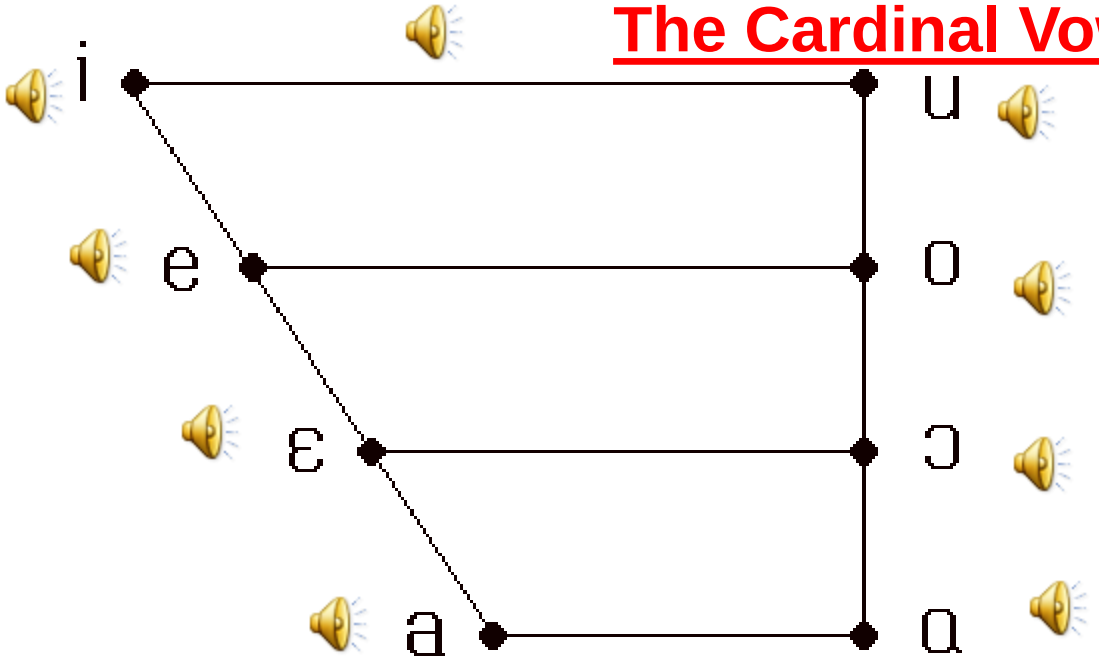
It's common for there to be vowels with different auditory (perceptual) qualities that are assigned the same symbol.



Jones didn't think it could be solved just by adding more symbols -- the ear will always be able to hear more distinctions than you'll have symbols for.

Jones' idea: ***Define a set of standard vowels that can be used as common reference points, allowing a phonetician to define vowel qualities in relation to these reference points*** – e.g., *this vowel is lower and fronter than cardinal vowel 6*.

The Cardinal Vowels



Phonetician Daniel Jones

These recordings of the 8 “primary” cardinal vowels were made by DJ his actual self in 1956. There are also secondary cardinal vowels defining other reference points in vowel space.

Cardinal vowels are ***highly specific vowel qualities***; i.e., it is inaccurate to say the /u/, for example, is cardinal vowel 5. It is a very specific instance of /u/ -- the one spoken by DJ – that is the cardinal vowel.

Opinions differ about the usefulness of the cardinal vowels as reference points for defining vowel quality.

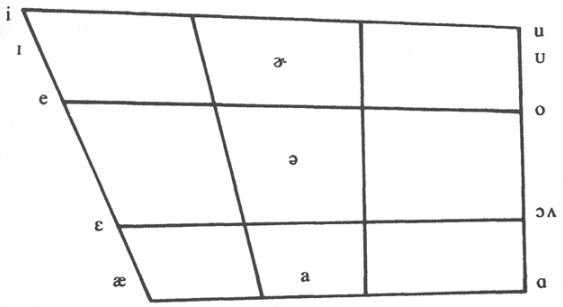
Acoustic measurements (formants) may provide a better solution. More later.

Last thing: How much can we trust these articulatory facts about vowels; i.e., is it really true that /ɑ/ is low-back, /o/ is mid-high back, etc.?

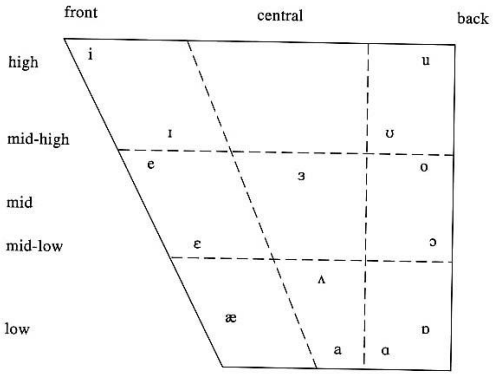
Below – difference of opinion. Who's right? Is it possible they're all wrong? All right, but for different dialects?

Vowel Quadrilateral

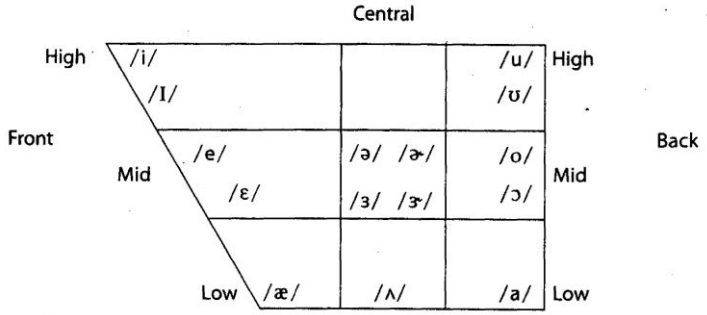
MacKay



Ladefoged (many others)



Ferrand



THANK YOU