Content

- Phonetic Transcription
- Phonemes
- Speech production
- Speech Analysis

Phonetic Transcription

Phonetic transcription

- Graphemes cannot represent speech sounds
  - cat vs. cite vs. cello
  - cat vs. kite
  - thin, graph, attend

Phonetic transcription

- IPA (International Phonetic Alphabet)
- cat vs. cite vs. cello
  /kæt/ vs. /saɪt/ vs. /tʃeloʊ/
- cat vs. kite
  /kæt/ vs. /kaɪt/
- thin, graph, attend
  /θɪn/, /græf/, /ətend/

Phonemes

Phoneme

- The smallest distinctive unit of sound
- Can change the meaning of a word
- Abstract representations of sounds

Phoneme

- bite vs. light ---> /baɪt/ vs. /laɪt/
- body vs. buddy ---> /bɑdi/ vs. /bʌdi/
- budge vs. buzz ---> /bʌdʒ/ vs. /bʌz/

English consonants

p	pie	pea		lowercase *p*
t	tie	tea		lowercase *t*
k	kye	key		lowercase *k*
b	by	bee		lowercase *b*
d	dye	D		lowercase *d*
g	guy			lowercase *g*
m	my	me	ra*m*	lowercase *m*
n	nigh	knee	ra*n*	lowercase *n*
ŋ			ra*ng*	eng (or angma)
f	fie	fee		lowercase *f*
v	vie	V		lowercase *v*
θ	thigh			theta
ð	thy	thee		eth
s	sigh	sea	listen	lowercase *s*
z		Z	mizzen	lowercase *z*
ʃ (š)	shy	she	mission	esh (or long *s*)
ʒ (ž)			vision	long *z* (or yogh)
l	lie	lee		lowercase *l*
w	why	we		lowercase *w*
r	rye			lowercase *r*
j (y)		ye		lowercase *j*
h	high	he		lowercase *h*

Note also the following:

tʃ (tš)	chi(me)	chea(p)
dʒ (dž)	ji(ve)	G

English vowels

monophthongs / diphthongs

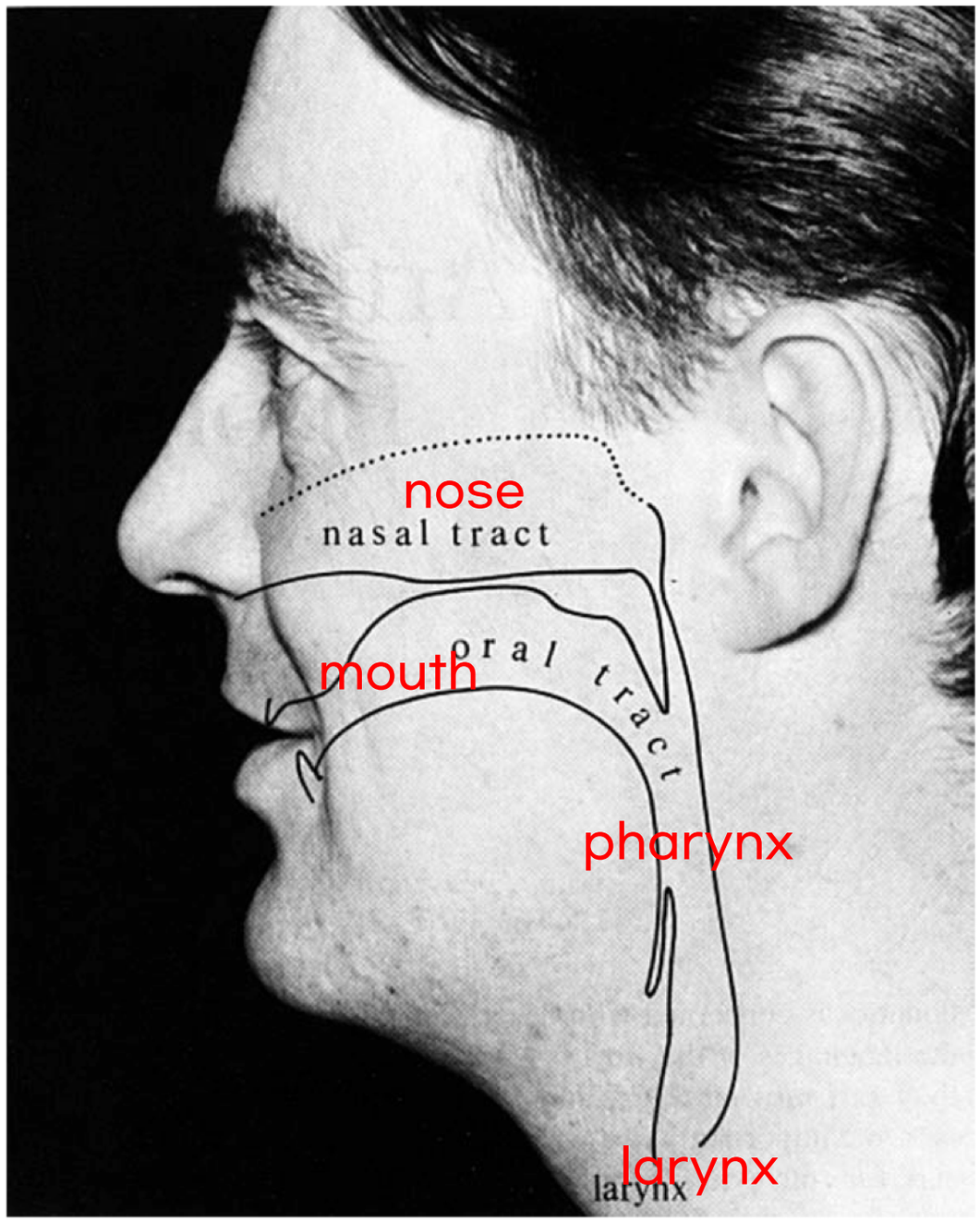
1	2						
i	i	heed	he	bead	heat	keyed	lowercase *i*
ɪ	ɪ	hid		bid	hit	kid	small capital *I*
eɪ	eɪ	hayed	hay	bayed	hate	Cade	lowercase *e*
ɛ	ɛ	head		bed			epsilon
æ	æ	had		bad	hat	cad	ash
ɑ	ɑ	hard		bard	heart	card	script *a*
ɑ	ɒ	hod		bod	hot	cod	turned script *a*
ɔ	ɔ	hawed	haw	bawd		cawed	open *o*
ʊ	ʊ	hood				could	upsilon
oʊ	əʊ	hoed	hoe	bode		code	lowercase *o*
u	u	who'd	who	booed	hoot	cooed	lowercase *u*
ʌ	ʌ	Hudd		bud	hut	cud	turned *v*
ɝ	ɜ	herd	her	bird	hurt	curd	reversed epsilon
aɪ	aɪ	hide	high	bide	height		lowercase a (+I)
aʊ	aʊ		how	bowed		cowed	(as noted above)
ɔɪ	ɔɪ		(a)hoy	Boyd			(as noted above)
ɪr	ɪə		here	beard			(as noted above)
ɛr	ɛə		hair	bared		cared	(as noted above)
aɪr	aə	hired	hire				(as noted above)

Note also:

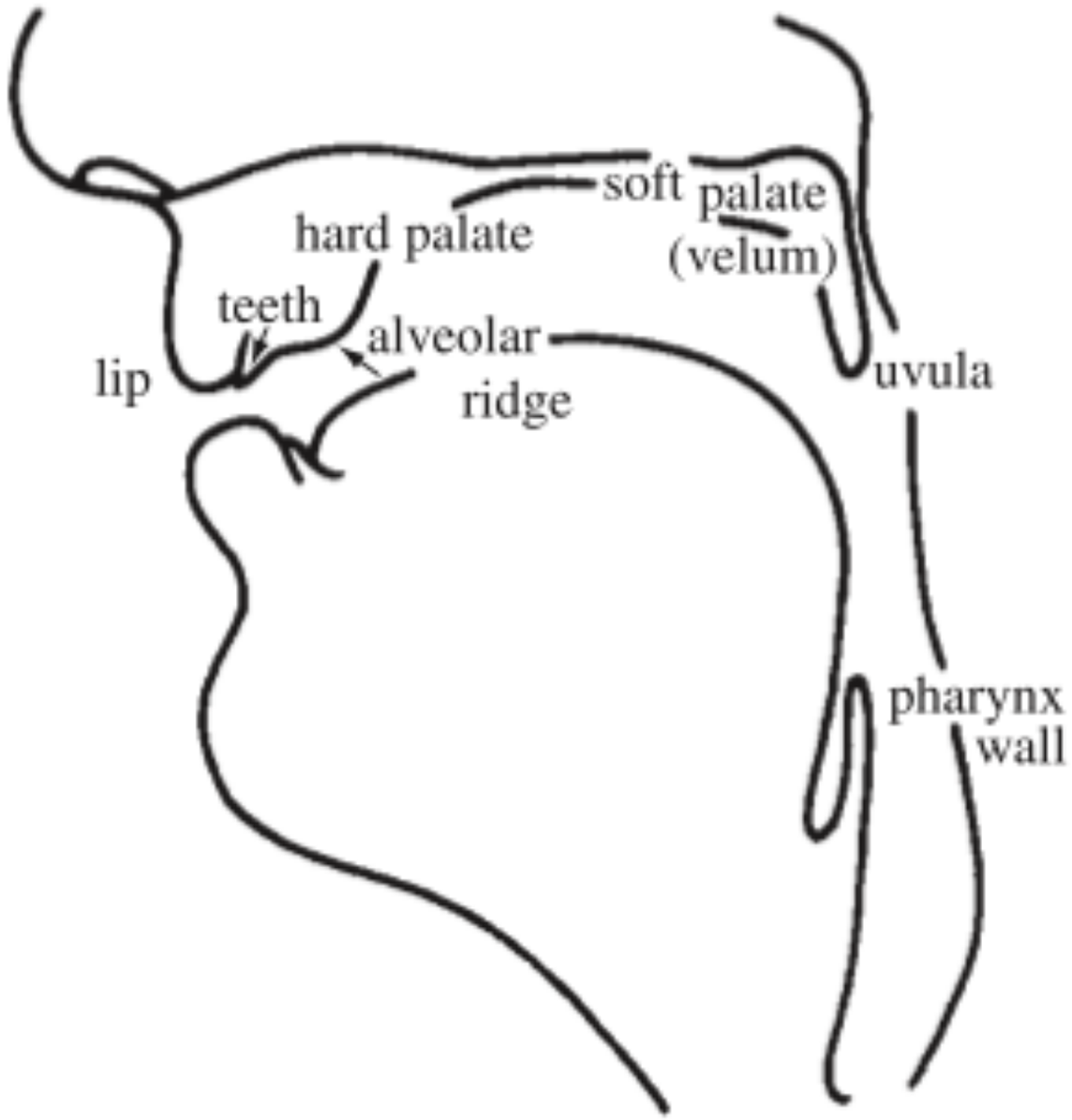
ju	ju	hued	hue	Bude		cued	(as noted above)

Speech Production

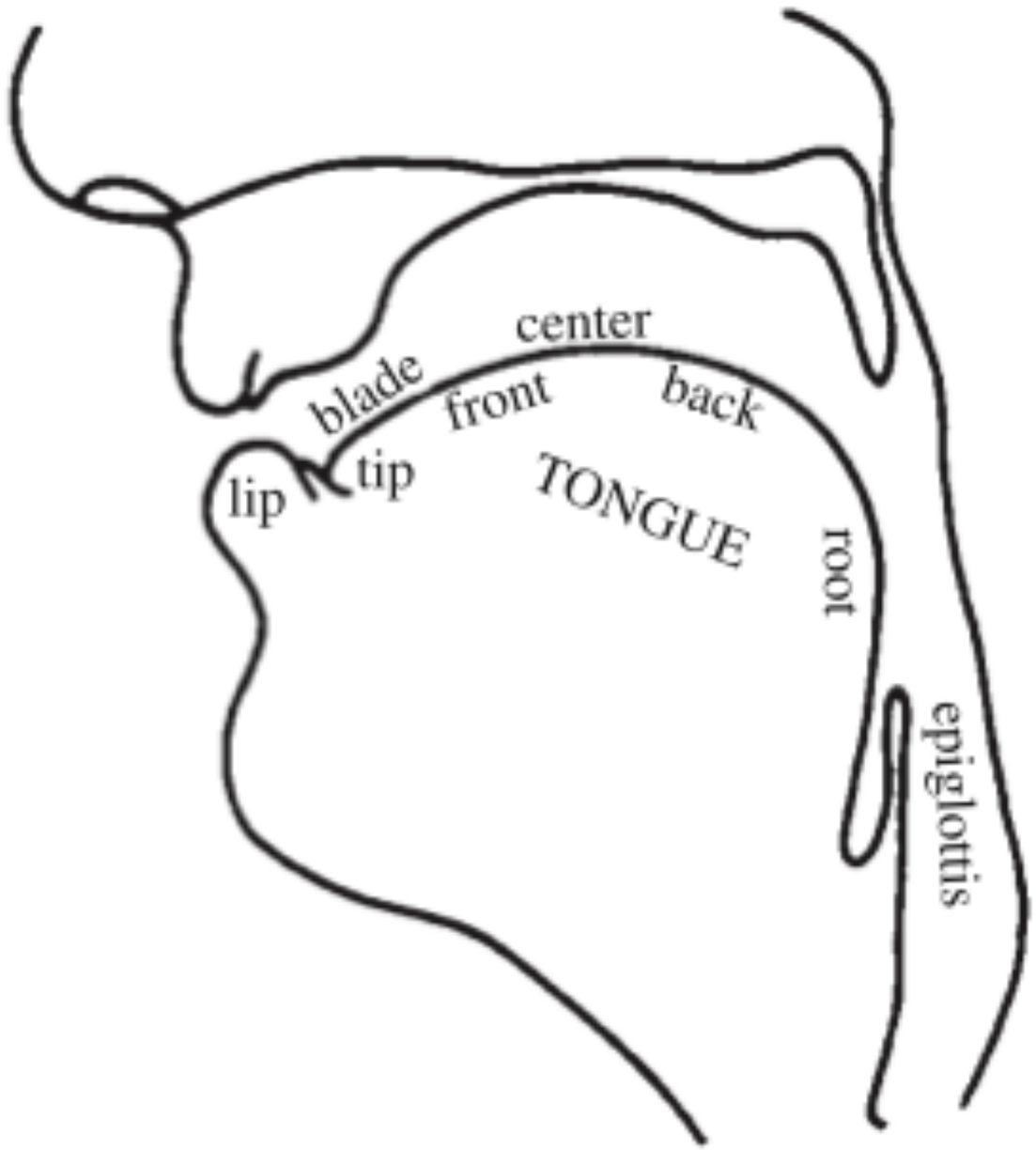
Vocal tract



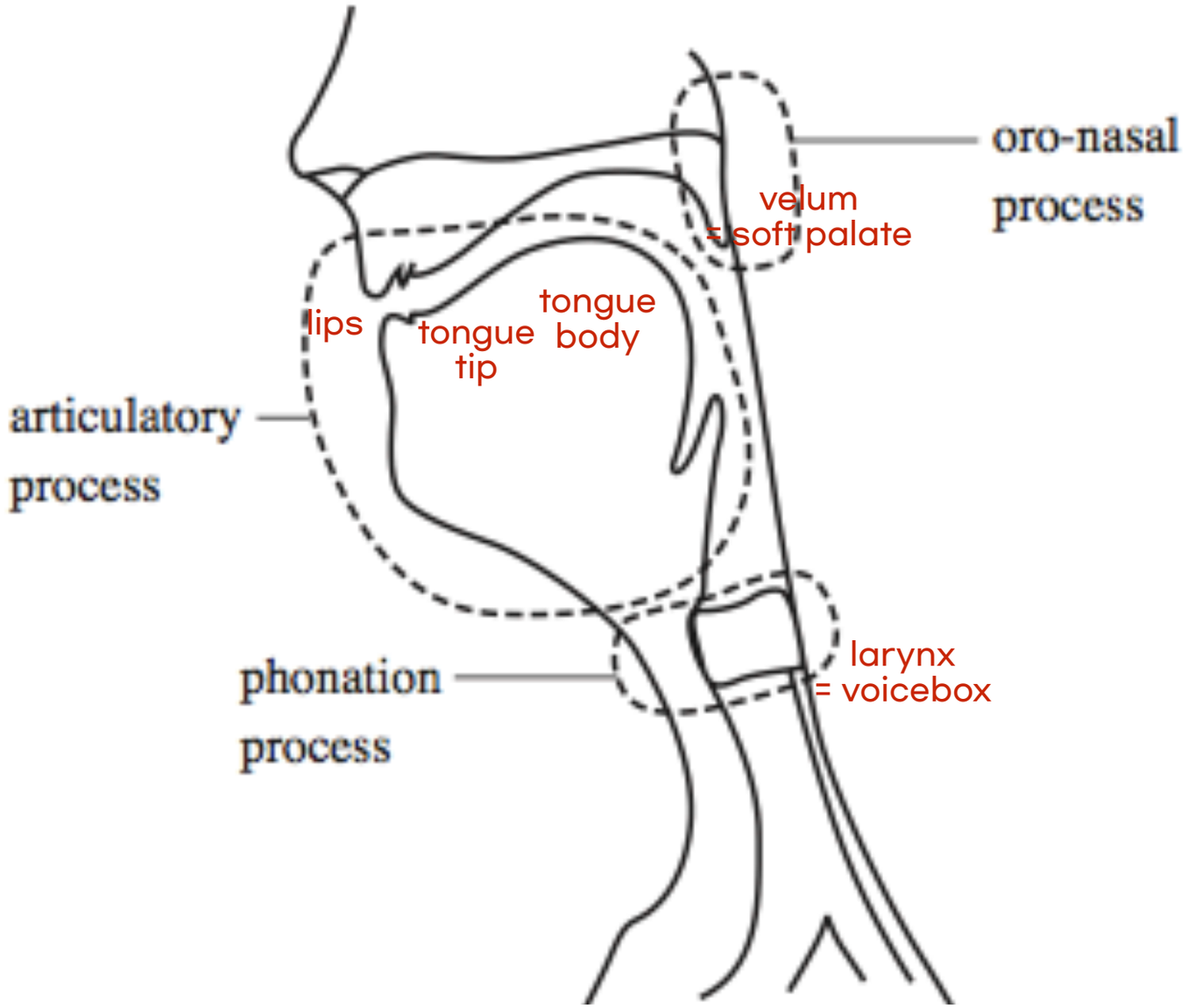
Vocal tract (upper)



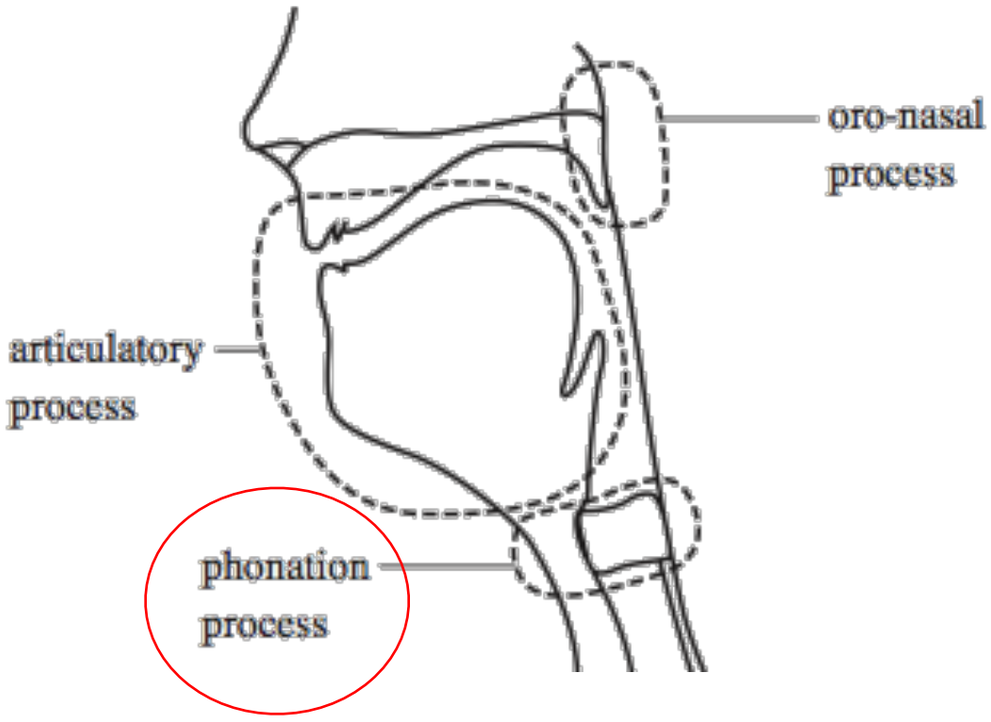
Vocal tract (lower)



5 speech organs
=constrictors
=articulators

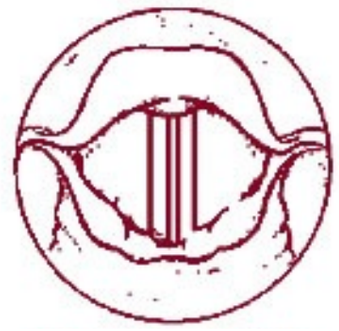


Larynx = voicebox



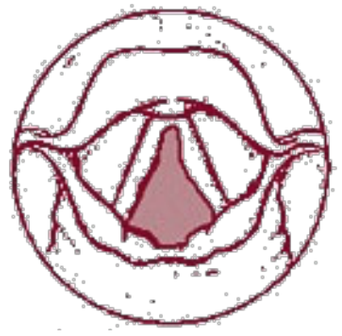
- voiced: can feel vibration

  e.g. v, z, l, m, a, i, …

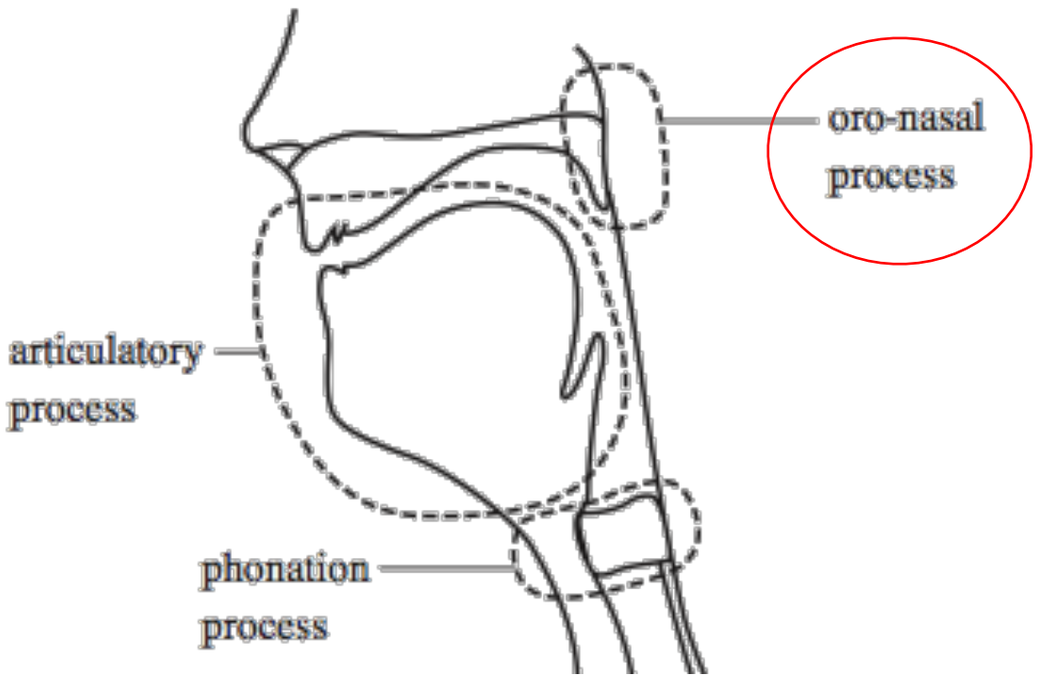


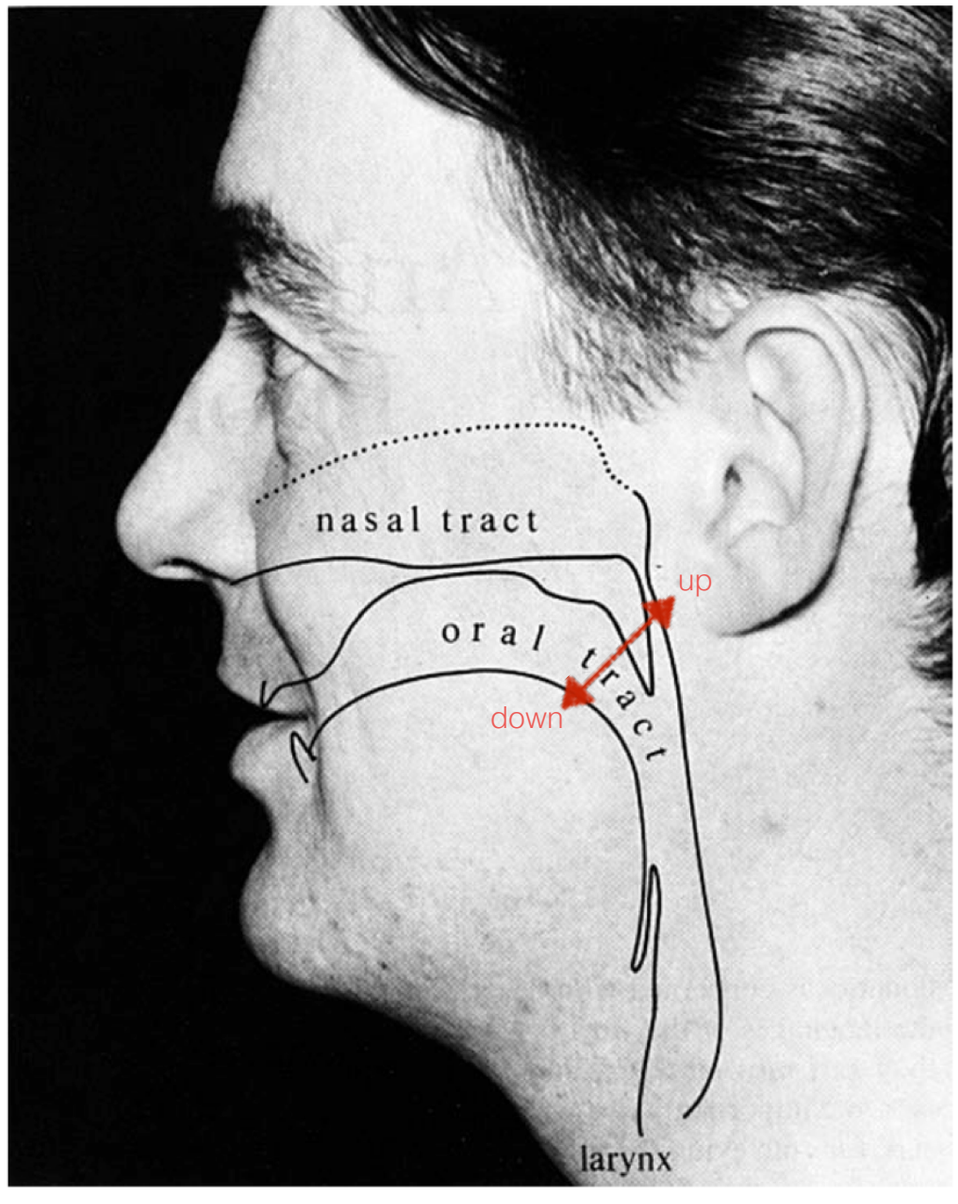
- voiceless: can't feel vibration

  e.g. f, s, k, p, h, …



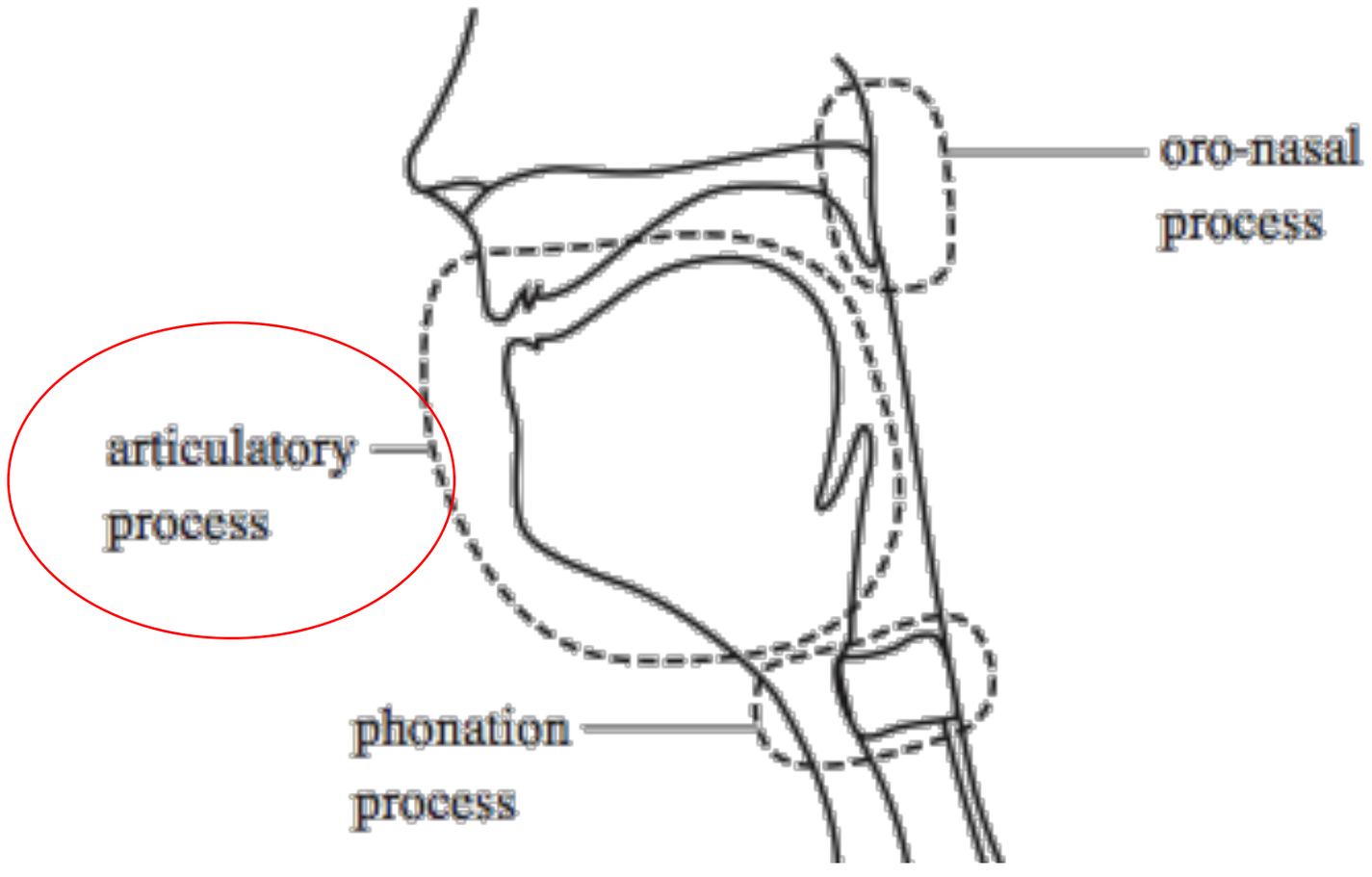
Whether velum is lowered





- nasals: m n ŋ …

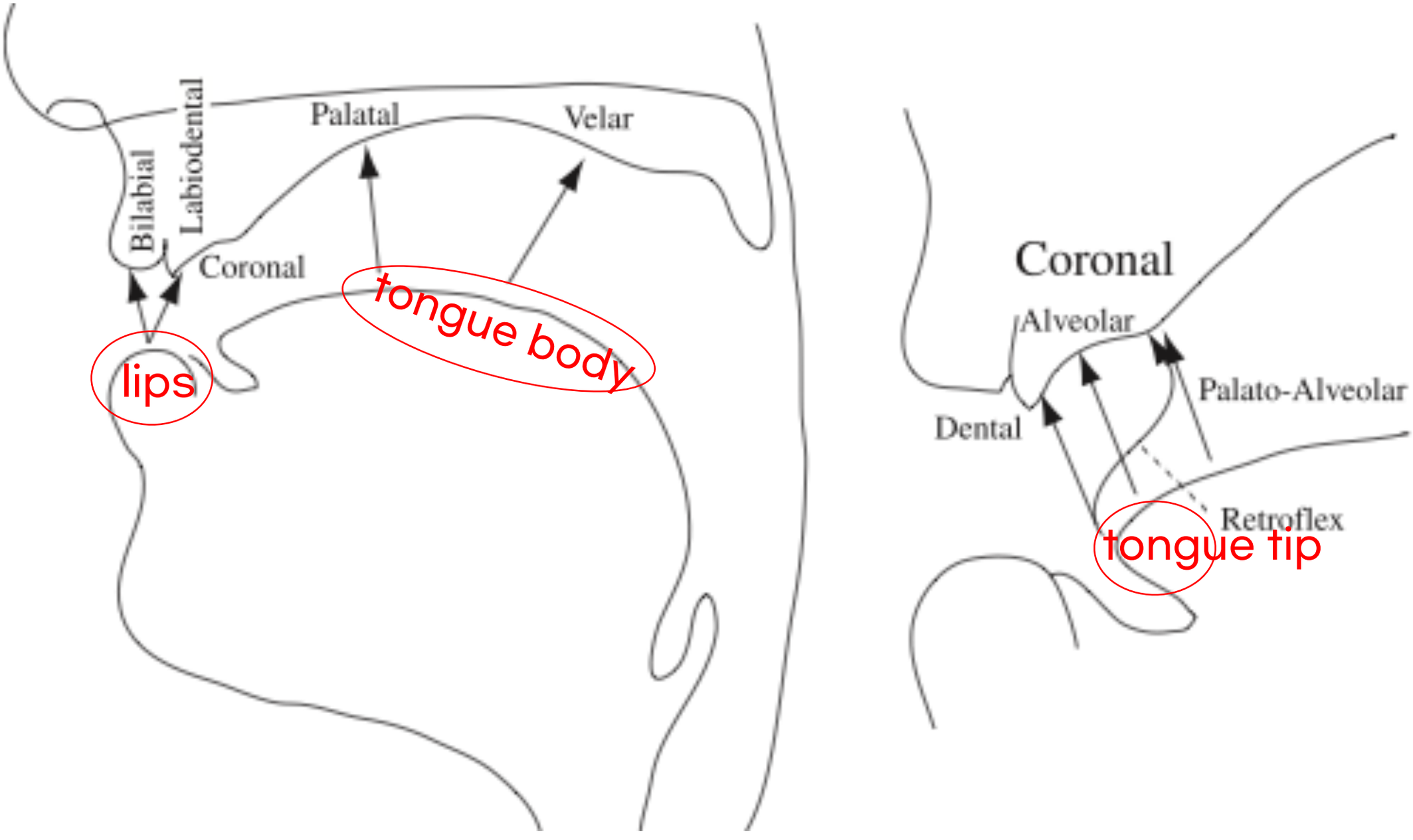
Articulation



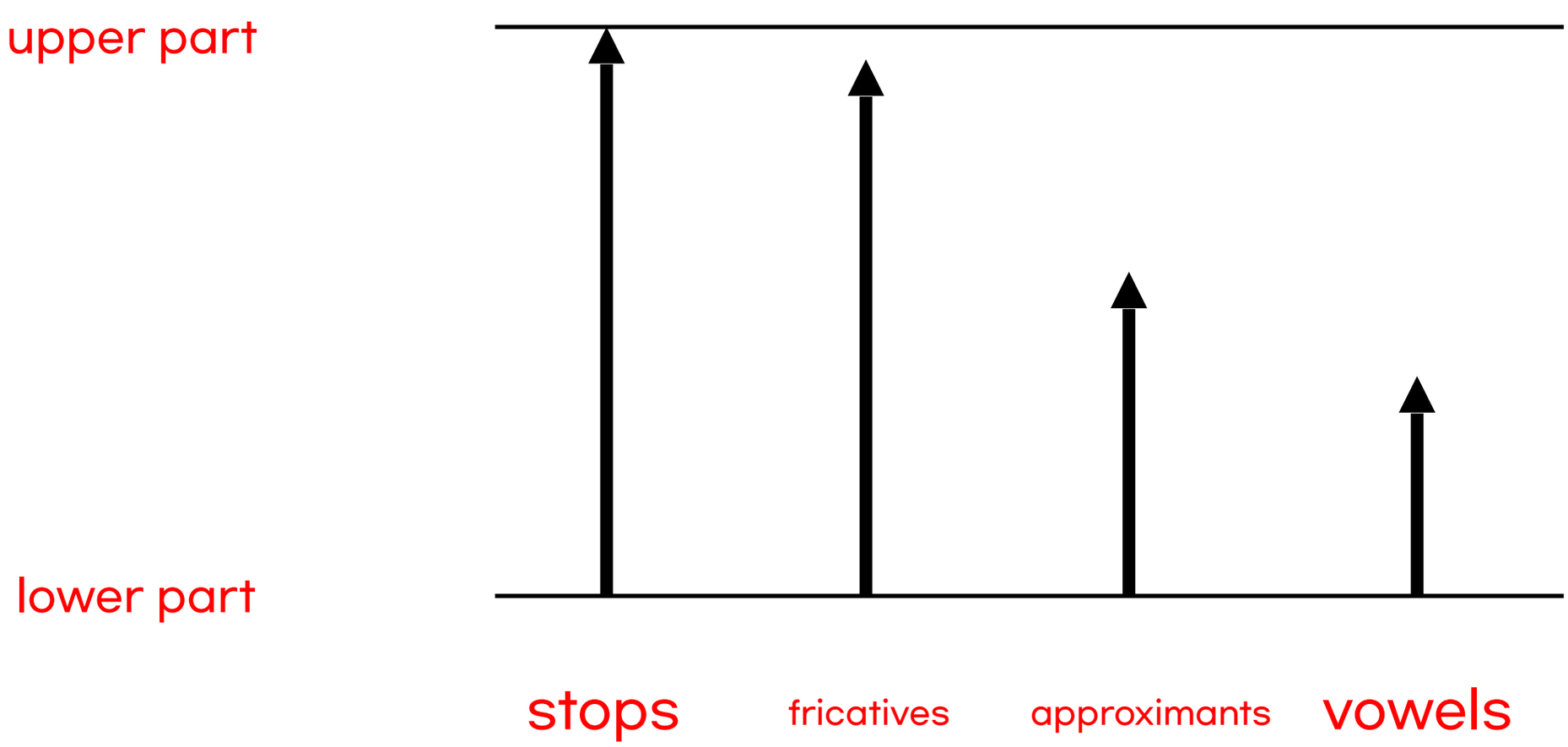
Constrictor details

- Constriction location (CL): where exactly?
- Constriction degree (CD): how much exactly?

Constriction location (CL)



Constriction degree (CD)

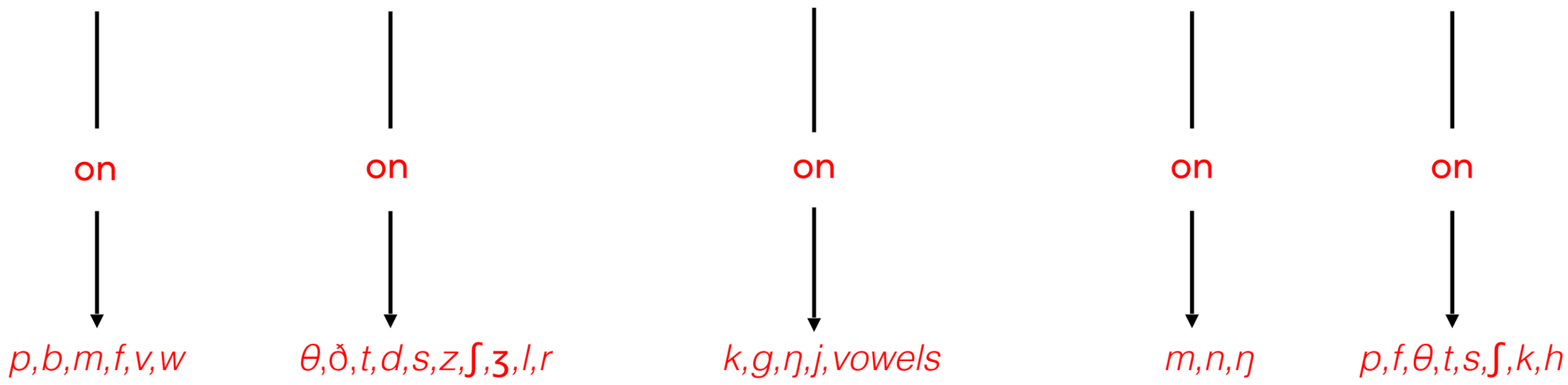


How to produce English consonants & vowels?

By specifying constrictors, CD, and CL

Phonemes

- Individual sounds that form words
- a combination of speech organs' actions
- lips / tongue tip / tongue body / velum / larynx



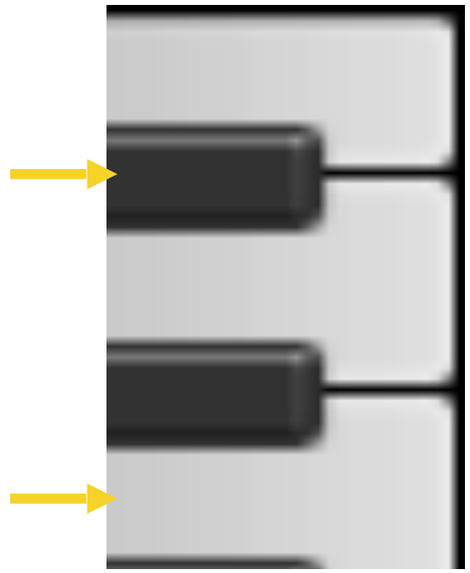
on = active = effortful

off = inactive = effortless =default

Piano playing analogy

- e.g., /t/

speech organ	on/off
lips	off
tongue tip	on
tongue body	off
velum	off
larynx	on



- Five speech organs analogous to 5 keyboards on piano
- Producing a phoneme analogous to hitting a certain set of keys

Phonemes (practice)

- e.g., /t/

speech organ	on/off
lips	off
tongue tip	on
tongue body	off
velum	off
larynx	on

- e.g., /d/ ?

Phonemes (practice)

- e.g., /m/

speech organ	on/off
lips	on
tongue tip	off
tongue body	off
velum	on
larynx	off

- e.g., /n/ ?

Phonemes (practice)

- e.g., /g/

speech organ	on/off
lips	off
tongue tip	off
tongue body	on
velum	off
larynx	off

- e.g., /k/ ?

/t/ vs. /s/?

speech organ	on/off
lips	off
tongue tip	on
tongue body	off
velum	off
larynx	on

speech organ	on/off
lips	off
tongue tip	on
tongue body	off
velum	off
larynx	on

- not different! More specification needed.

Constriction Degree (CD)

- How much constriction?
  - stops: full constriction *(e.g. p, t, k, b, d, g, m, n, ŋ, …)*
  - fricatives: critical constriction - *(e.g. s, z, f, z, ʃ, ʒ, θ, ð, …)*
  - approximants: little constriction - *(e.g. r, l, w, y, …)*
  - vowels: no constriction - *(e.g. vowels)*
- For active (=on) speech organs only among lips, tongue tip, tongue body.

/t/

speech	on/off	CD
lips	off	-
tongue tip	on	stop
tongue body	off	-
velum	off	-
larynx	on	-

/t/ vs. /s/

speech	on/off	CD
lips	off	-
tongue tip	on	stop
tongue body	off	-
velum	off	-
larynx	on	-

speech	on/off	CD
lips	off	-
tongue tip	on	fric.
tongue body	off	-
velum	off	-
larynx	on	-

/t/ vs. /r/?

speech	on/off	CD
lips	off	-
tongue tip	on	stop
tongue body	off	-
velum	off	-
larynx	on	-

speech	on/off	CD
lips	off	-
tongue tip	on	approx.
tongue body	off	-
velum	off	-
larynx	off	-

/a/ vs. /g/?

speech	on/off	CD
lips	off	-
tongue tip	off	-
tongue body	on	vowel
velum	off	-
larynx	off	-

speech organ	on/ off	CD
lips	off	-
tongue tip	off	-
tongue body	on	stop
velum	off	-
larynx	off	-

/s/ vs. /ʃ/?

speech	on/off	CD
lips	off	-
tongue tip	on	fric.
tongue body	off	-
velum	off	-
larynx	on	-

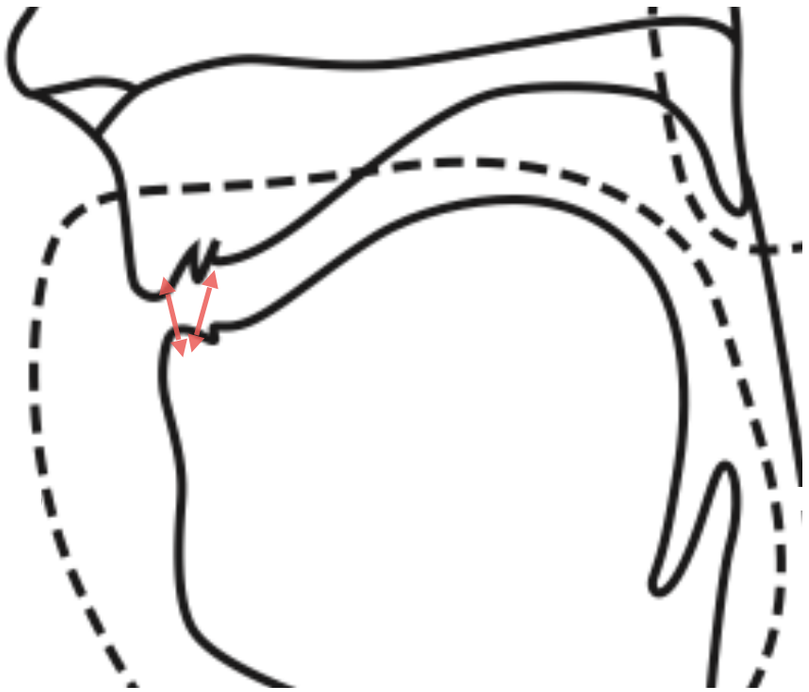
speech organ	on/ off	CD
lips	off	-
tongue tip	on	fric.
tongue body	off	-
velum	off	-
larynx	on	-

- not different! More specification needed.

Constriction Location (CL)

- Where does constriction occur at a speech organ?

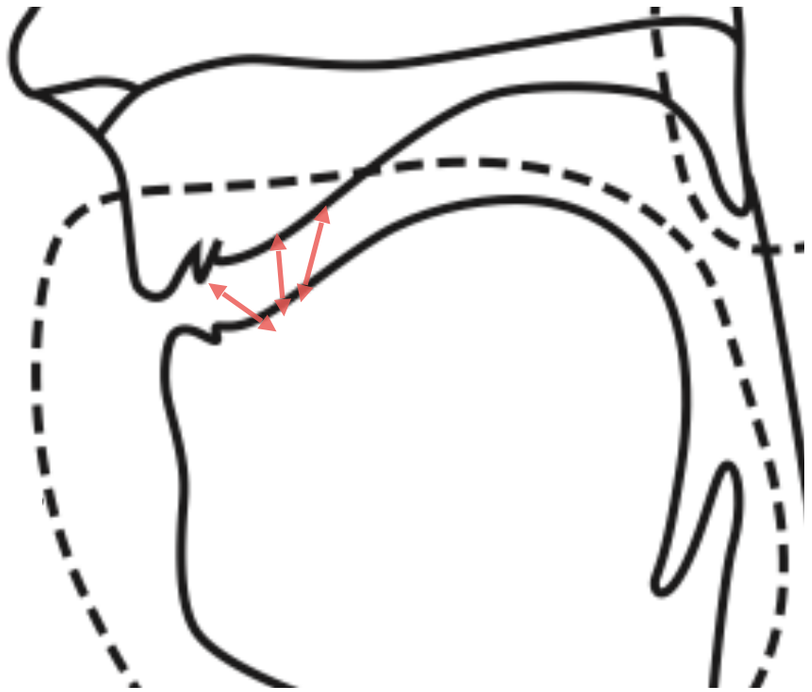
lips



labial / labiodental

p,b,m,w... f,v...

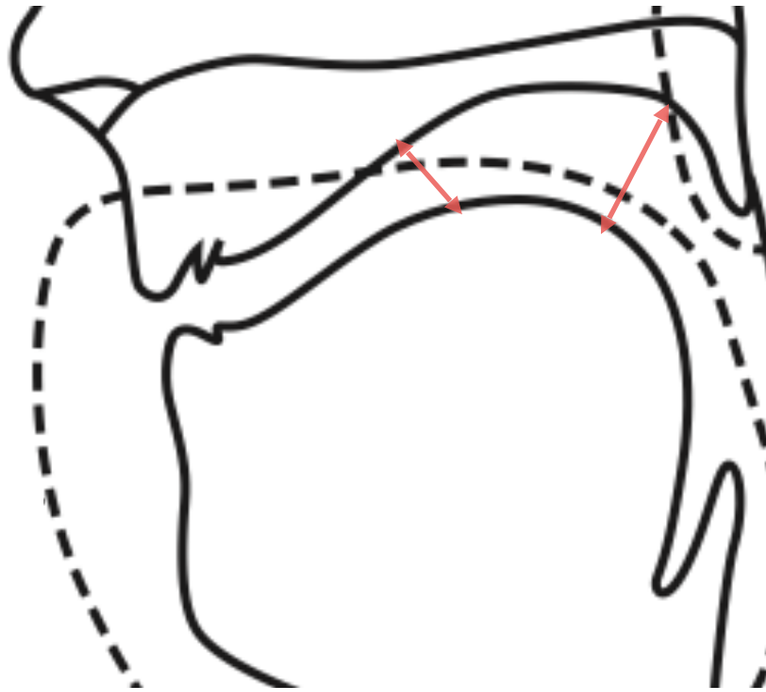
tongue tip



dental / alveolar / palato-alveolar

θ,ð... t,d,s,z,n, l,r... ʃ,ʒ...

tongue body



palatal / velar

j... k,g,ŋ...

/s/ vs. /ʃ/

speech organ	on/ off	CD	CL
lips	off	-	-
tongue tip	on	fric.	alv.
tongue body	off	-	-
velum	off	-	-
larynx	on	-	-

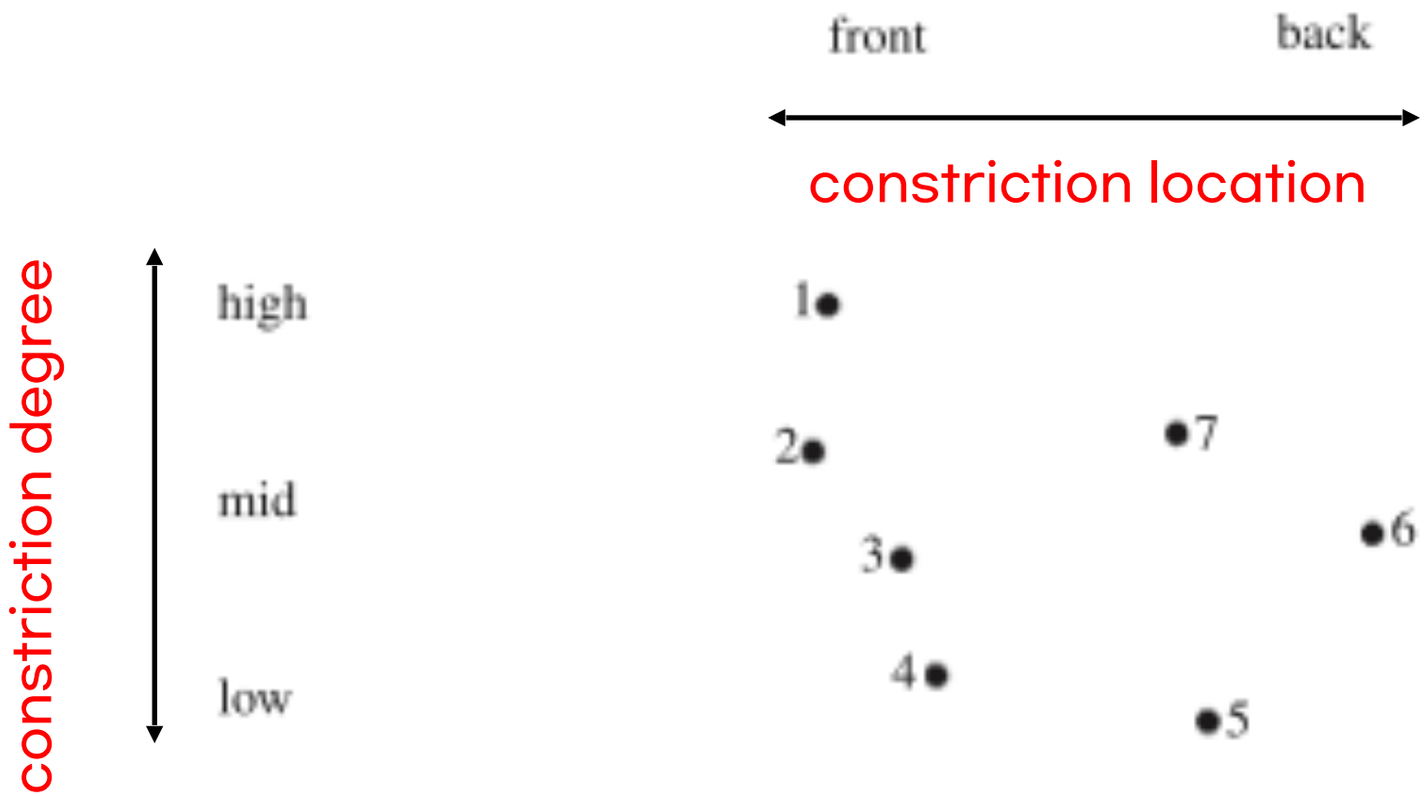
speech	on/off	CD	CL
lips	off	-	-
tongue tip	on	fric.	pal-alv
tongue body	off	-	-
velum	off	-	-
larynx	on	-	-

Vowels
by tongue body only

Figure 1.12 The positions of the vocal organs for the vowels in the words 1 *heed*, 2 *hid*, 3 *head*, 4 *had*, 5 *father*, 6 *good*, 7 *food*. The lip positions for vowels 2, 3, and 4 are between those shown for 1 and 5. The lip position for vowel 6 is between those shown for 1 and 7.

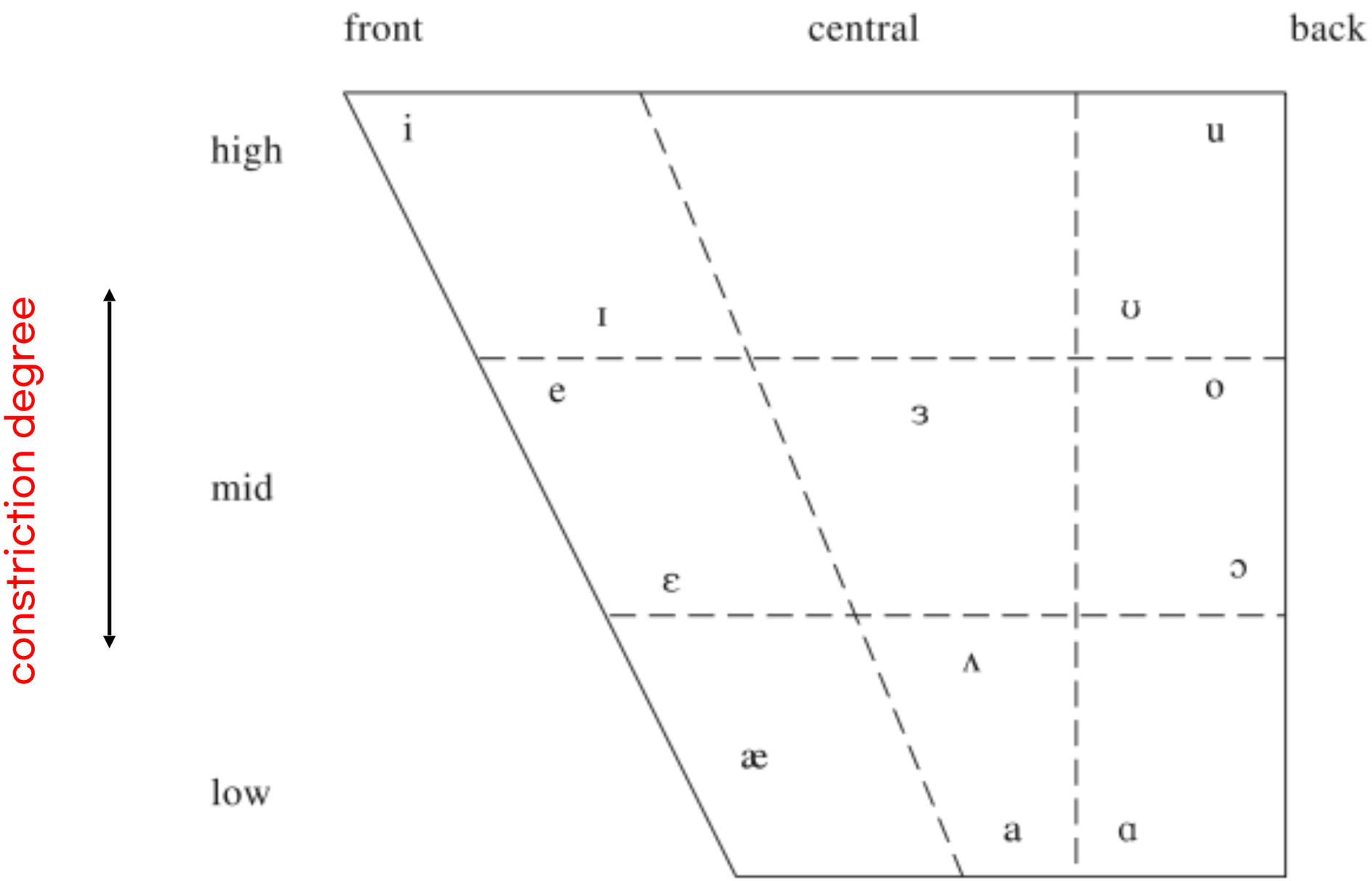
CL & CD specifications

Figure 1.13 The relative positions of the highest points of the tongue in the vowels in 1 *heed*, 2 *hid*, 3 *head*, 4 *had*, 5 *father*, 6 *good*, 7 *food*.



Vowel chart

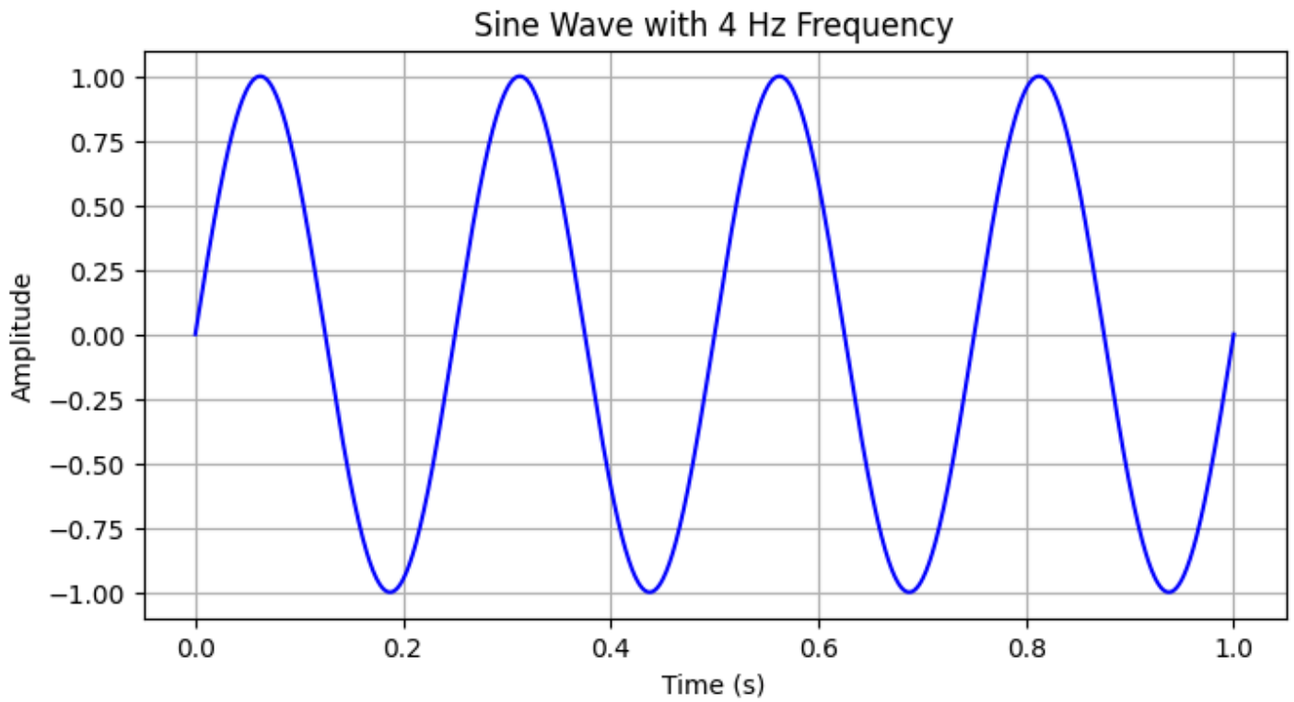
constriction location



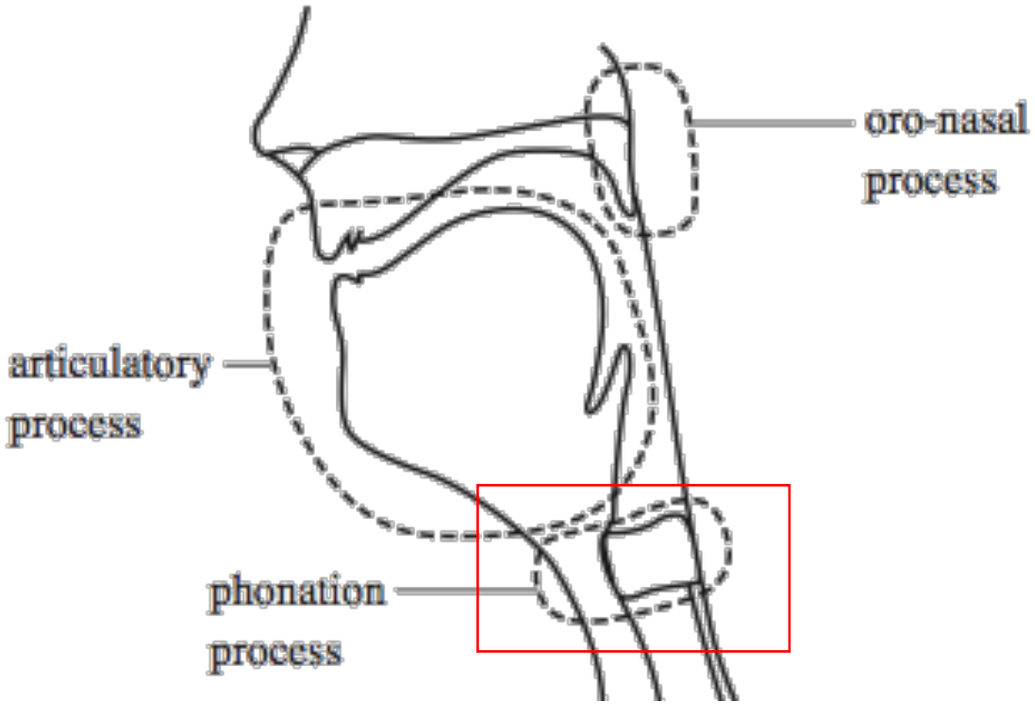
Geometry of tongue body!!

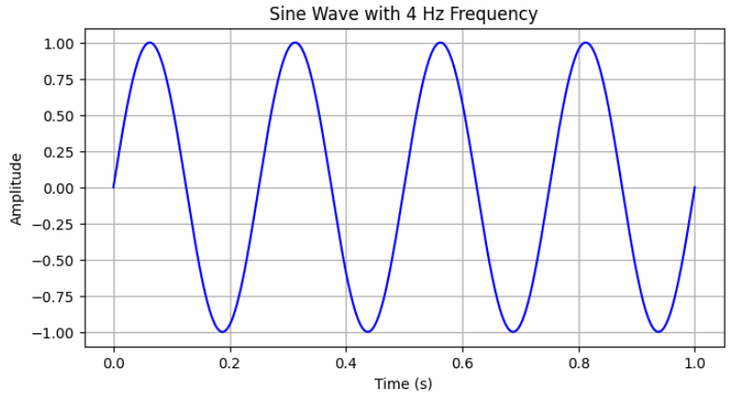
Speech Analysis

Waveform

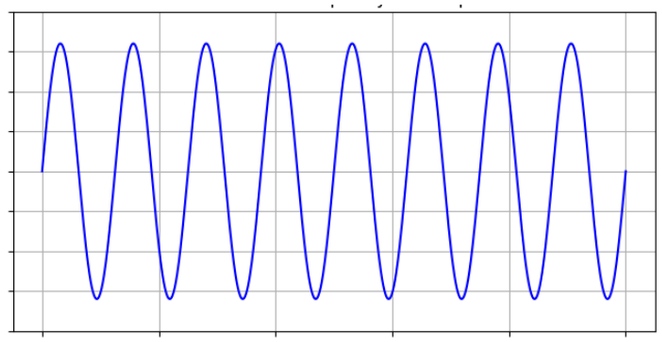

Sine Wave with 4 Hz Frequency
Amplitude
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)

Source

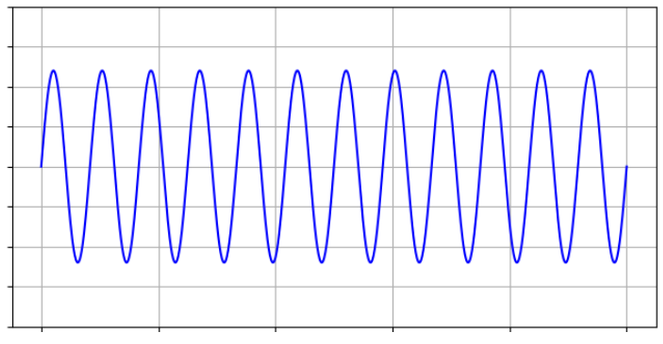

oro-nasal
process
articulatory
process
phonation
process


Sine Wave with 4 Hz Frequency
Amplitude
Time (s)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.2
0.4
0.6
0.8
1.0

+



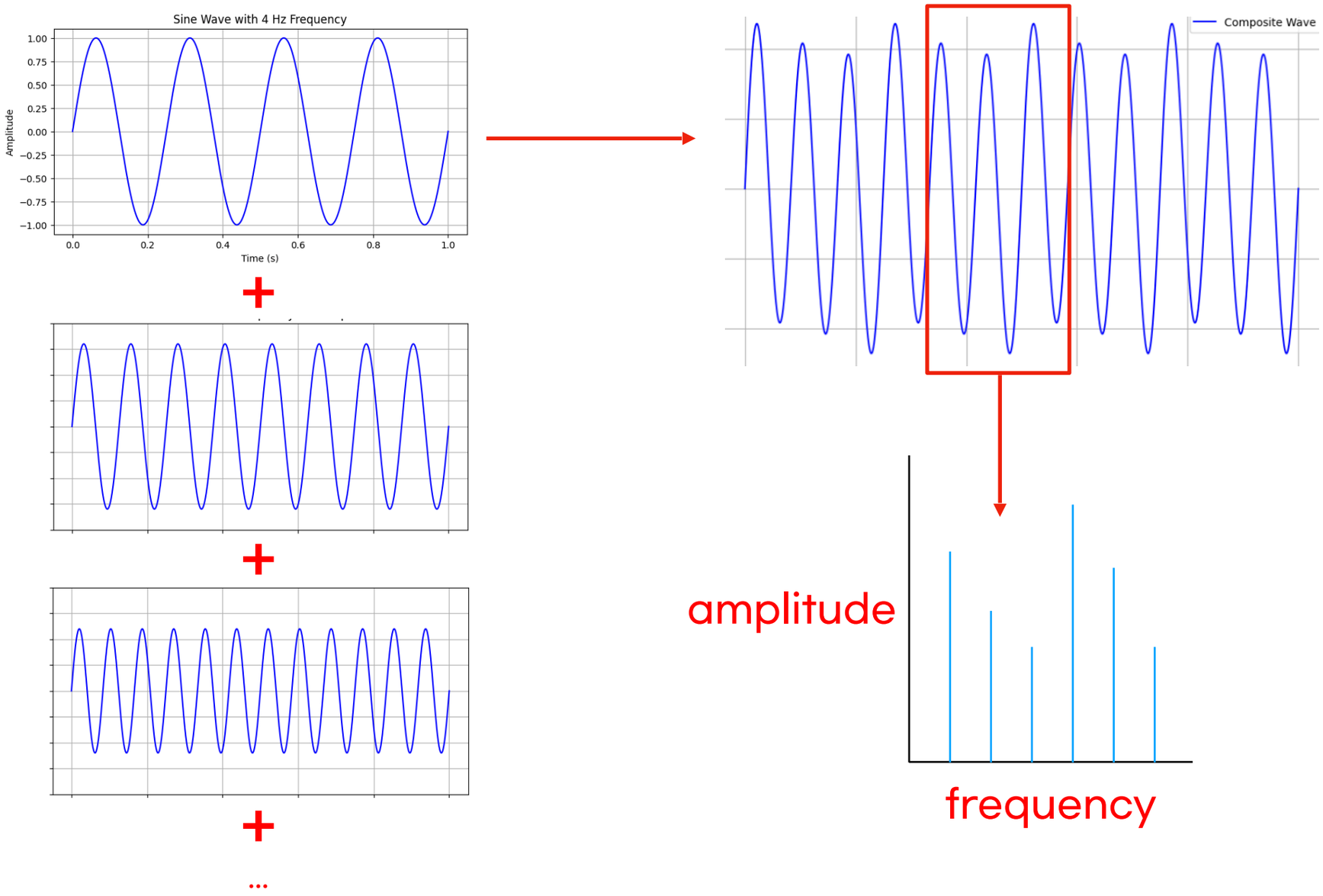
+



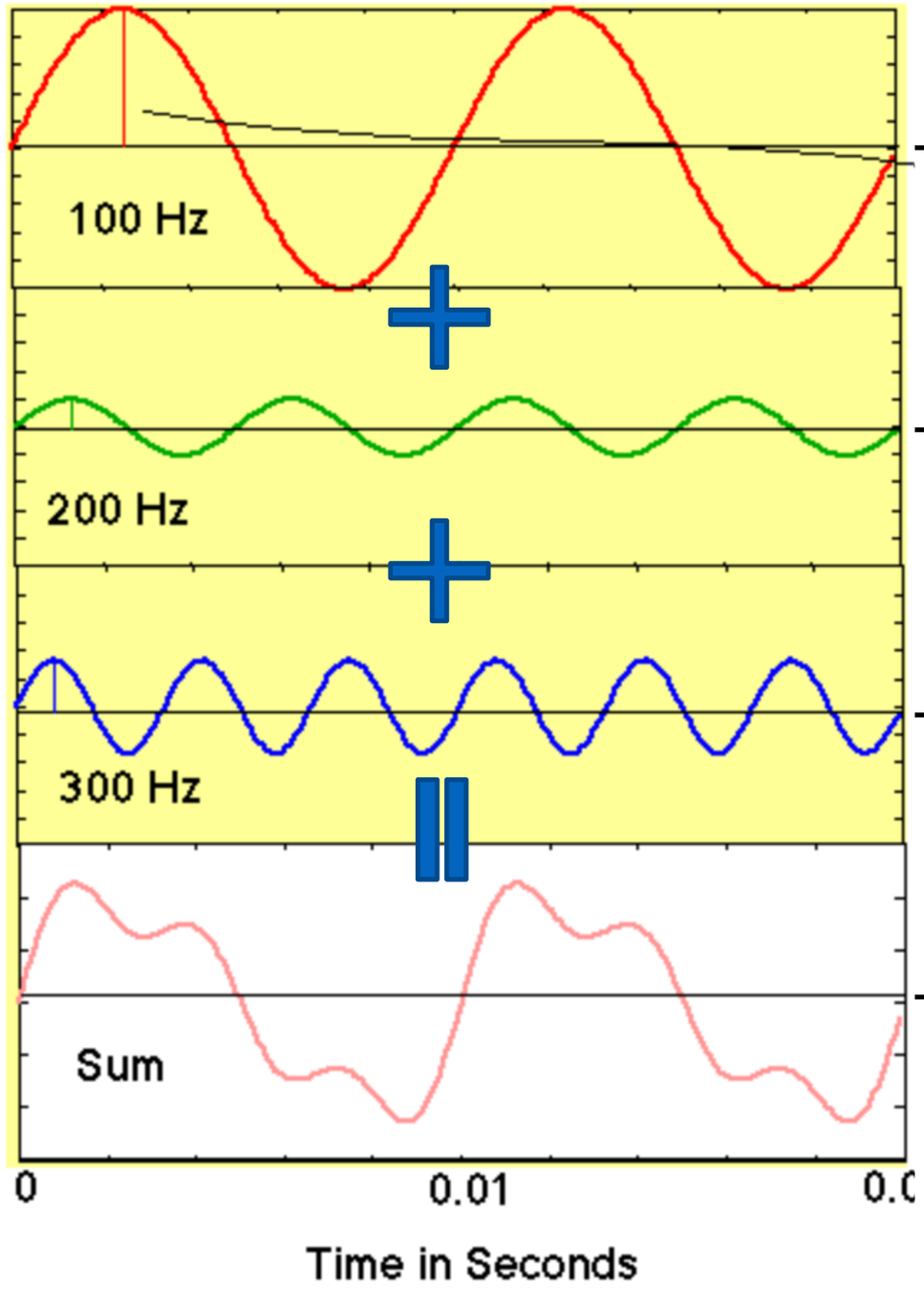
+

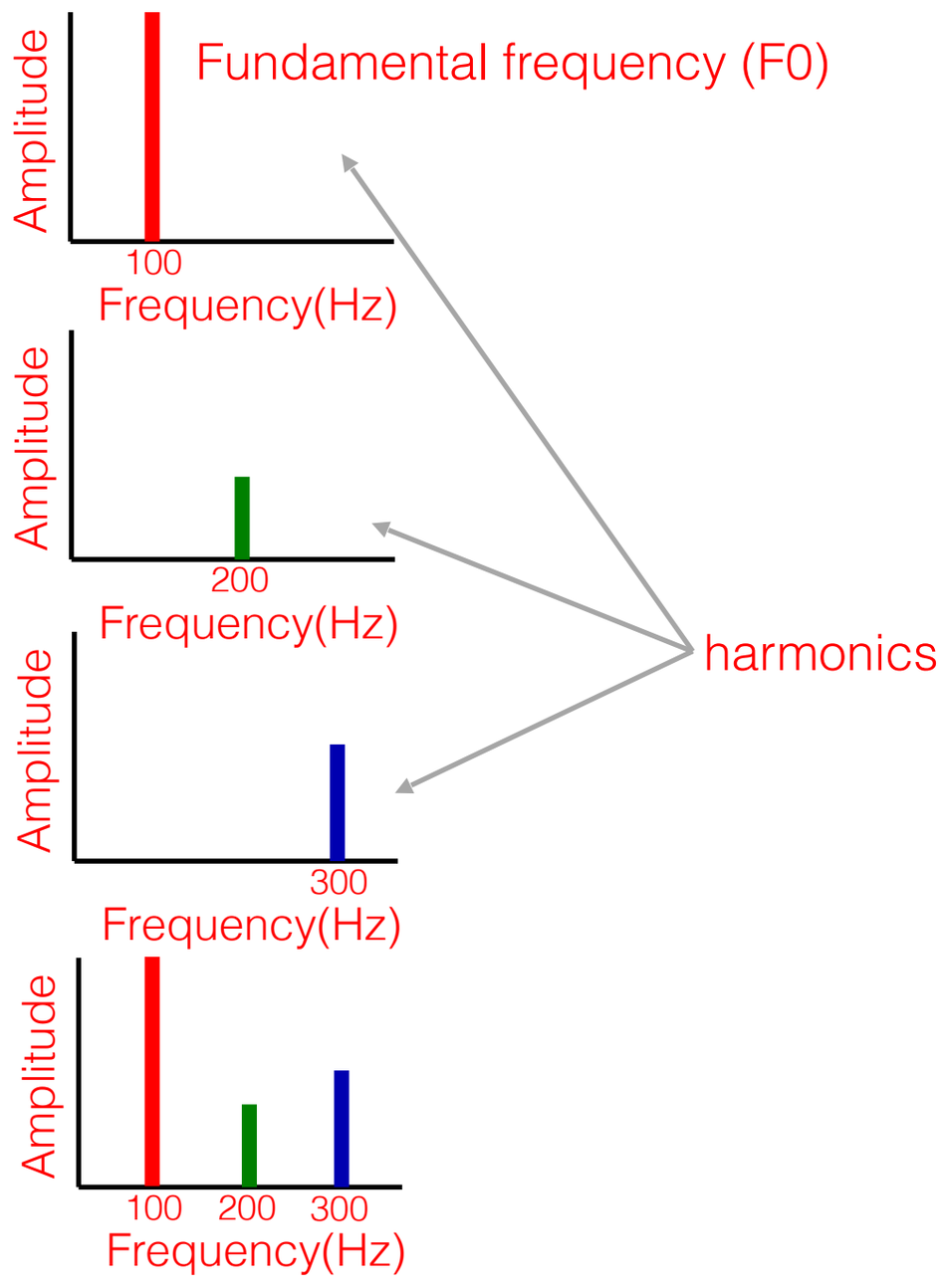
…

Spectrum

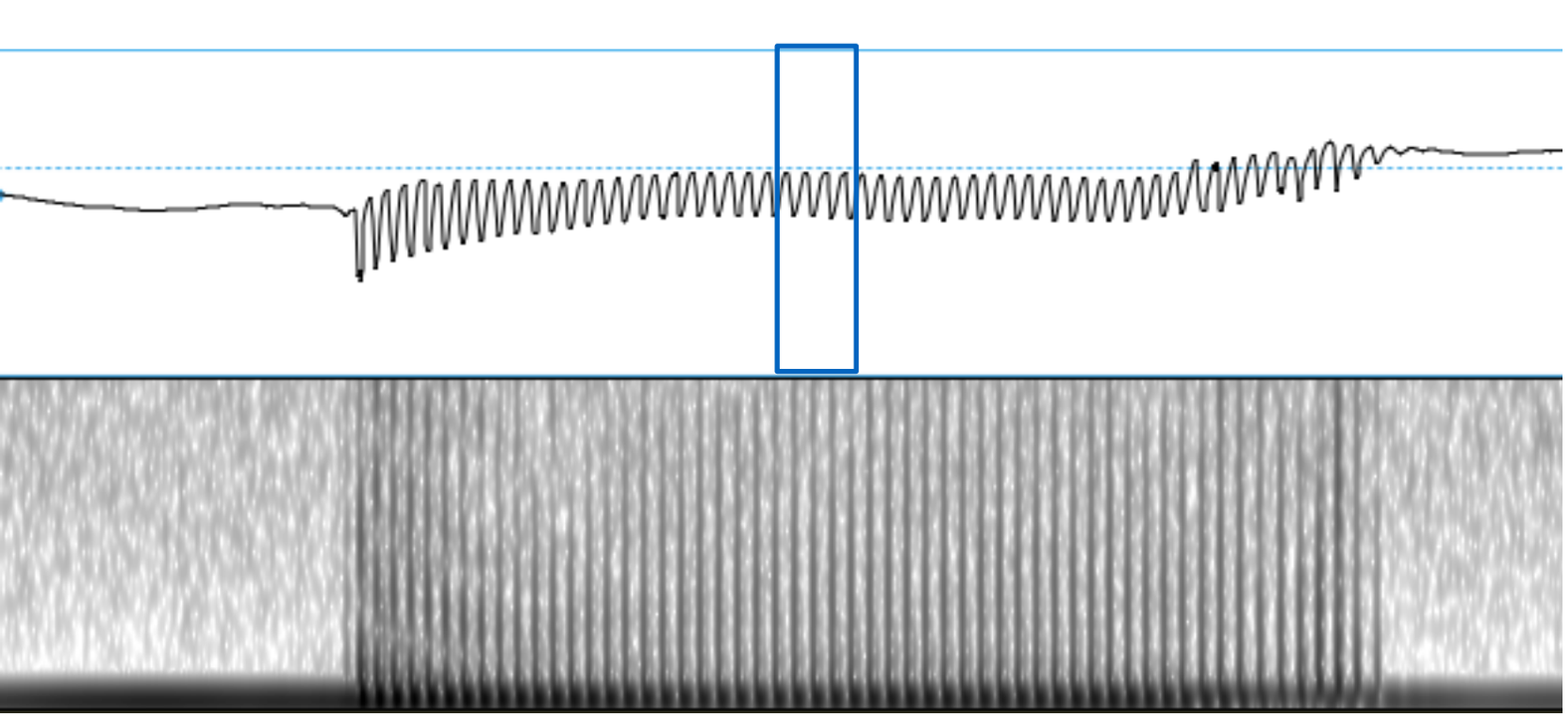

Sine Wave with 4 Hz Frequency
Amplitude
Time (s)
Composite Wave
amplitude
frequency

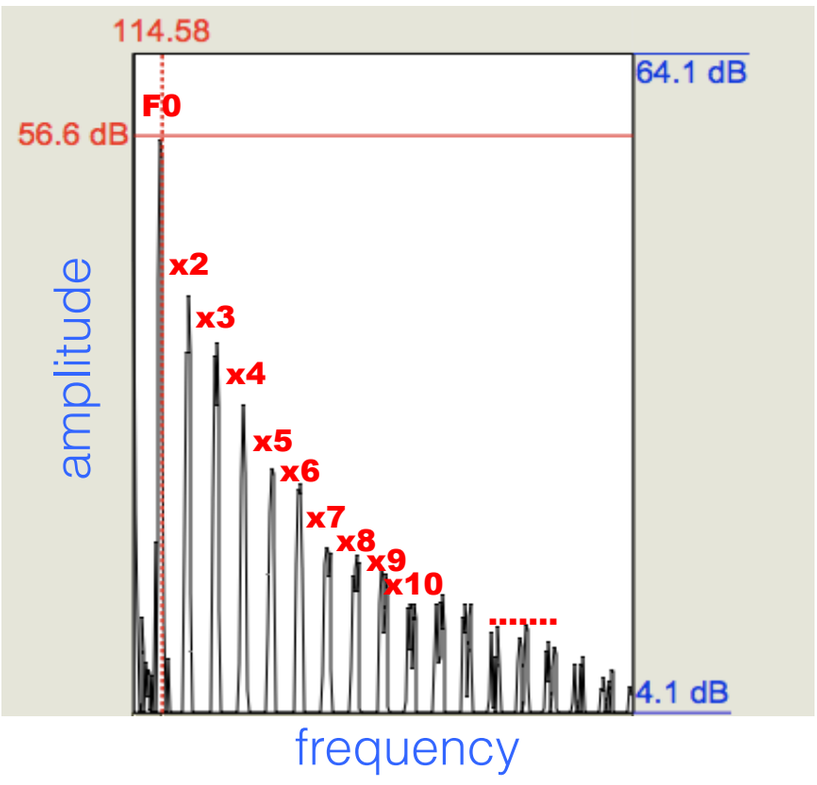
Complex tone in spectrum





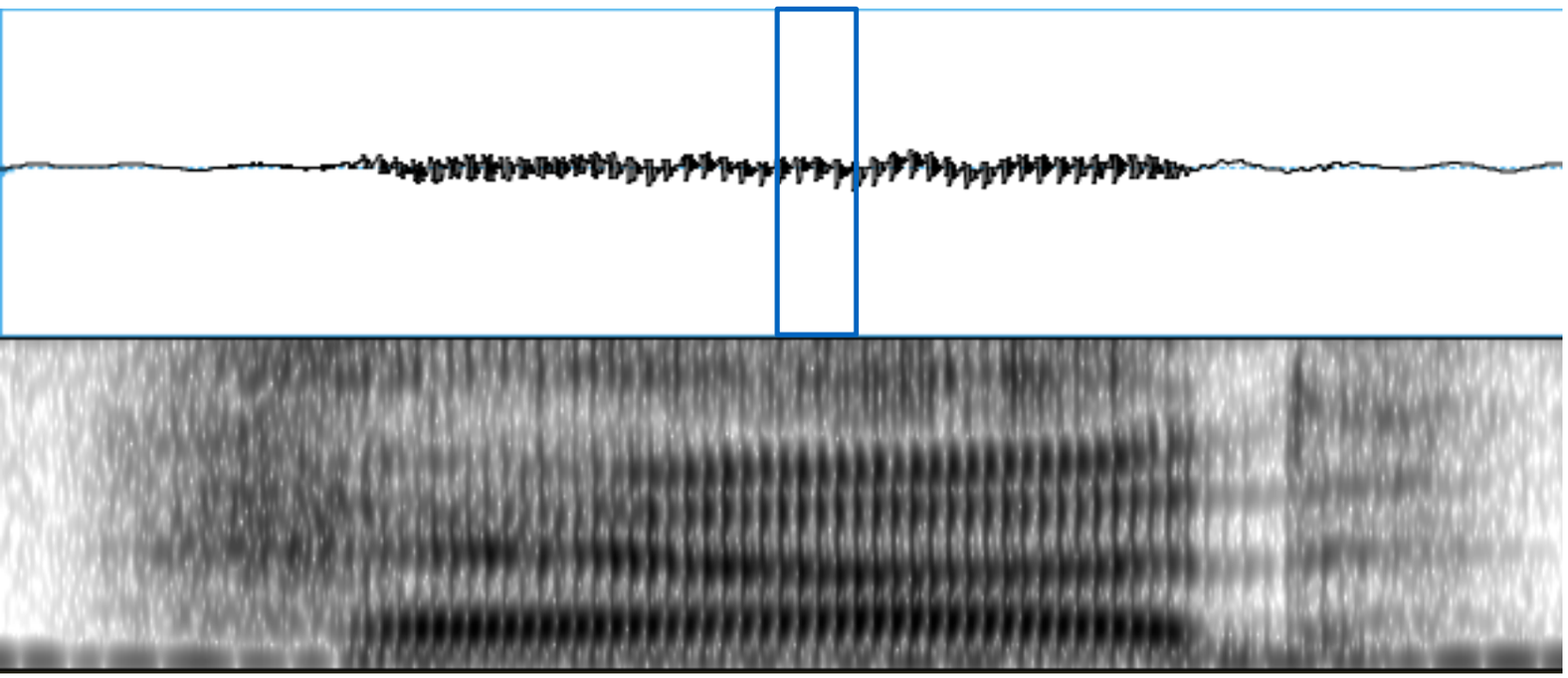
Human voice source

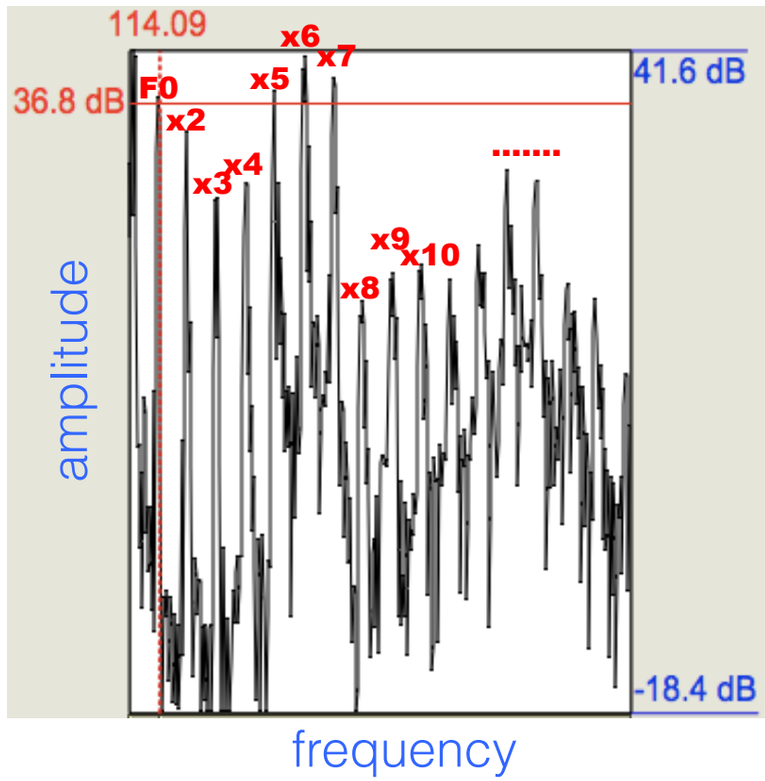




gradually decreasing

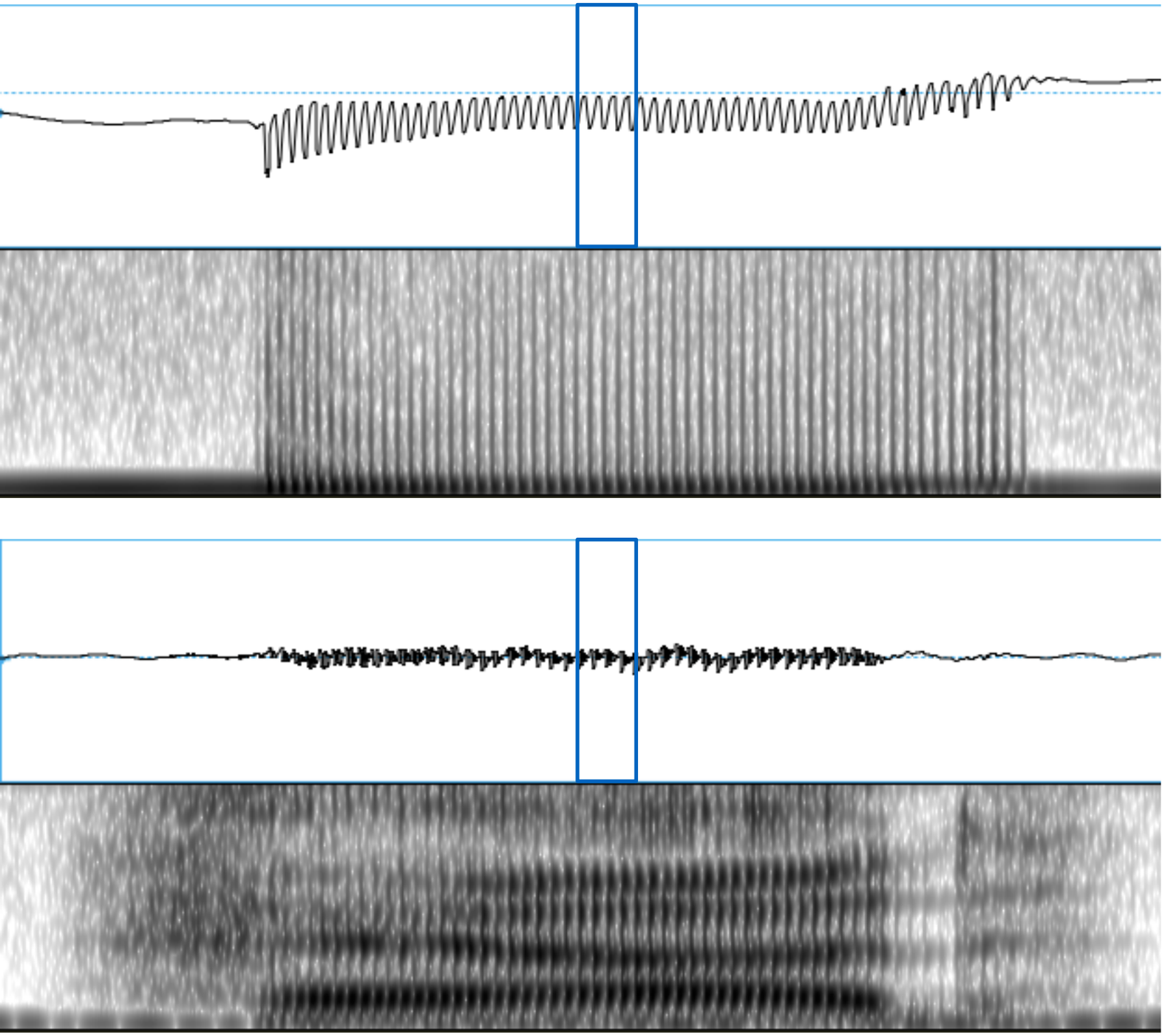
Filtered by vocal tract

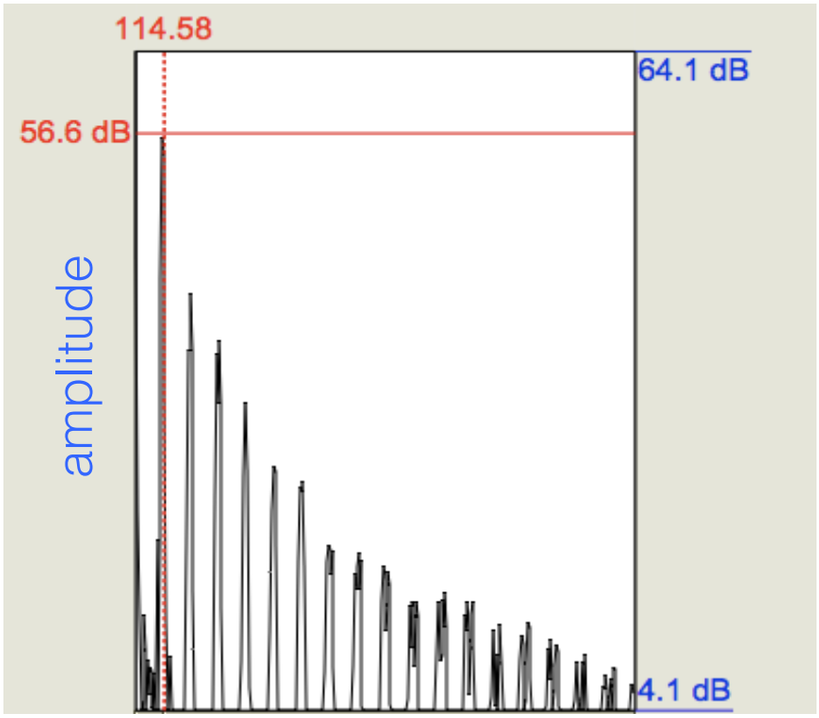


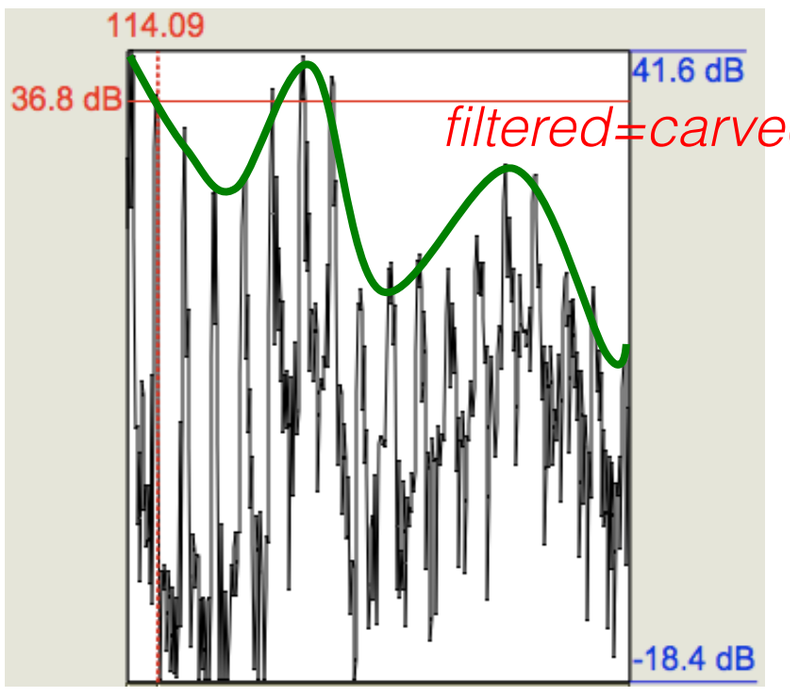


Jigjagging
with peaks and valleys

source & filter

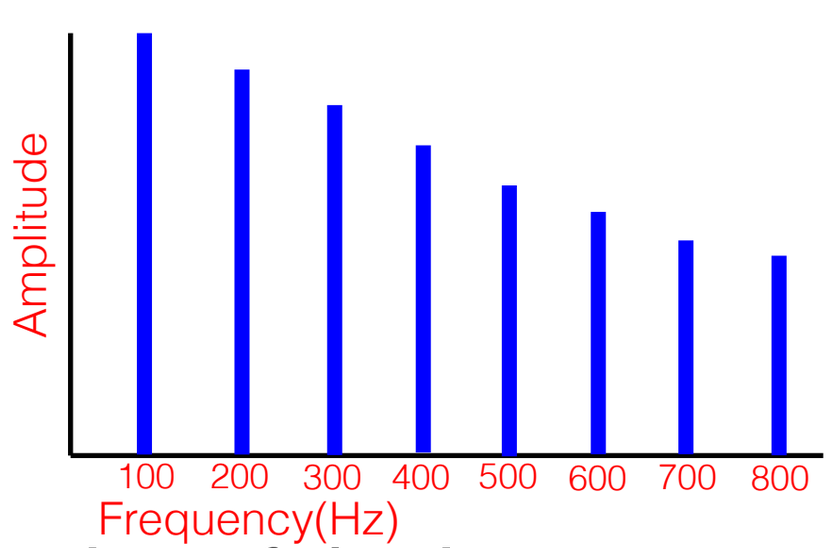






source

- Human voice source consists of *harmonics*
- A complex tone = sum of pure tones at integer multiples of the lowest pure tone
  - the lowest pure tone
    - *Fundamental frequency (F0)*
    - rate of vibration of the larynx
    - the number of opening-closing cycles of the larynx per second
- Amplitude of pure tones gradually decreases



filter

- audio: peaks/mountains and valleys
- Because it is filtered by the vocal tract (VT)
  - peaks/mountains: frequencies VT likes = formants
  - valleys: frequencies VT does NOT like

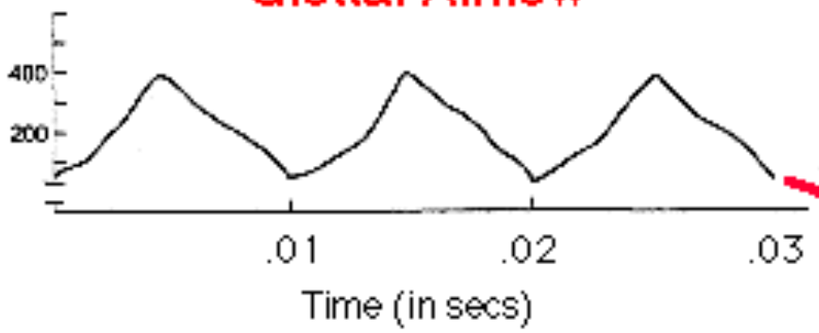
Source-filter theory

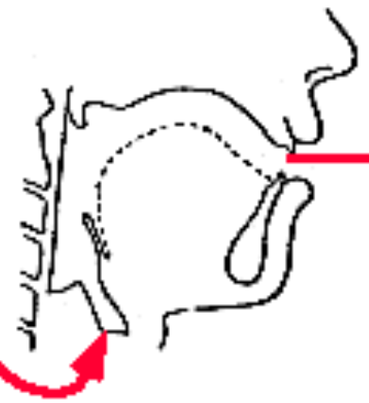
(from larynx) (by vocal tract)

Voice Source

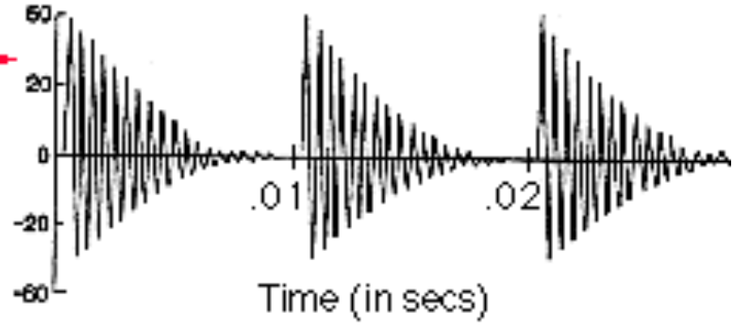
waveform

Glottal Airflow

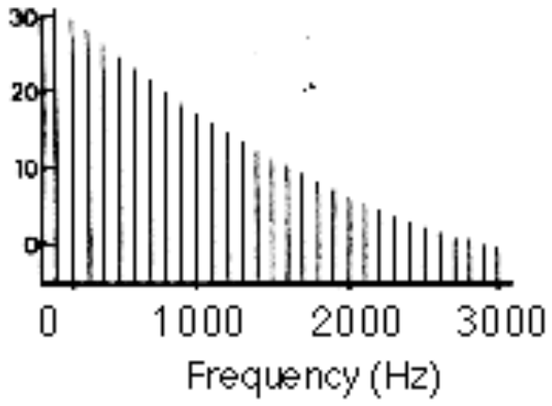




Output from lips



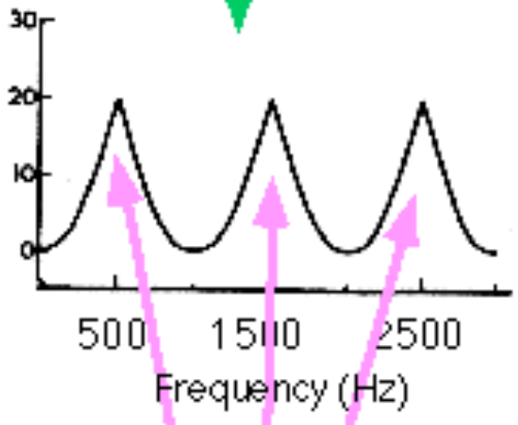
spectrum



Source Spectrum

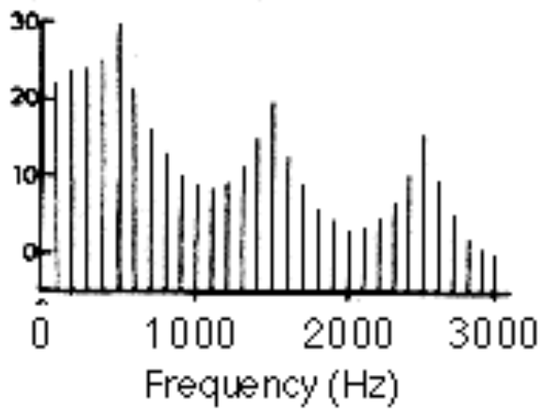
X

Vocal Tract
Filter Function



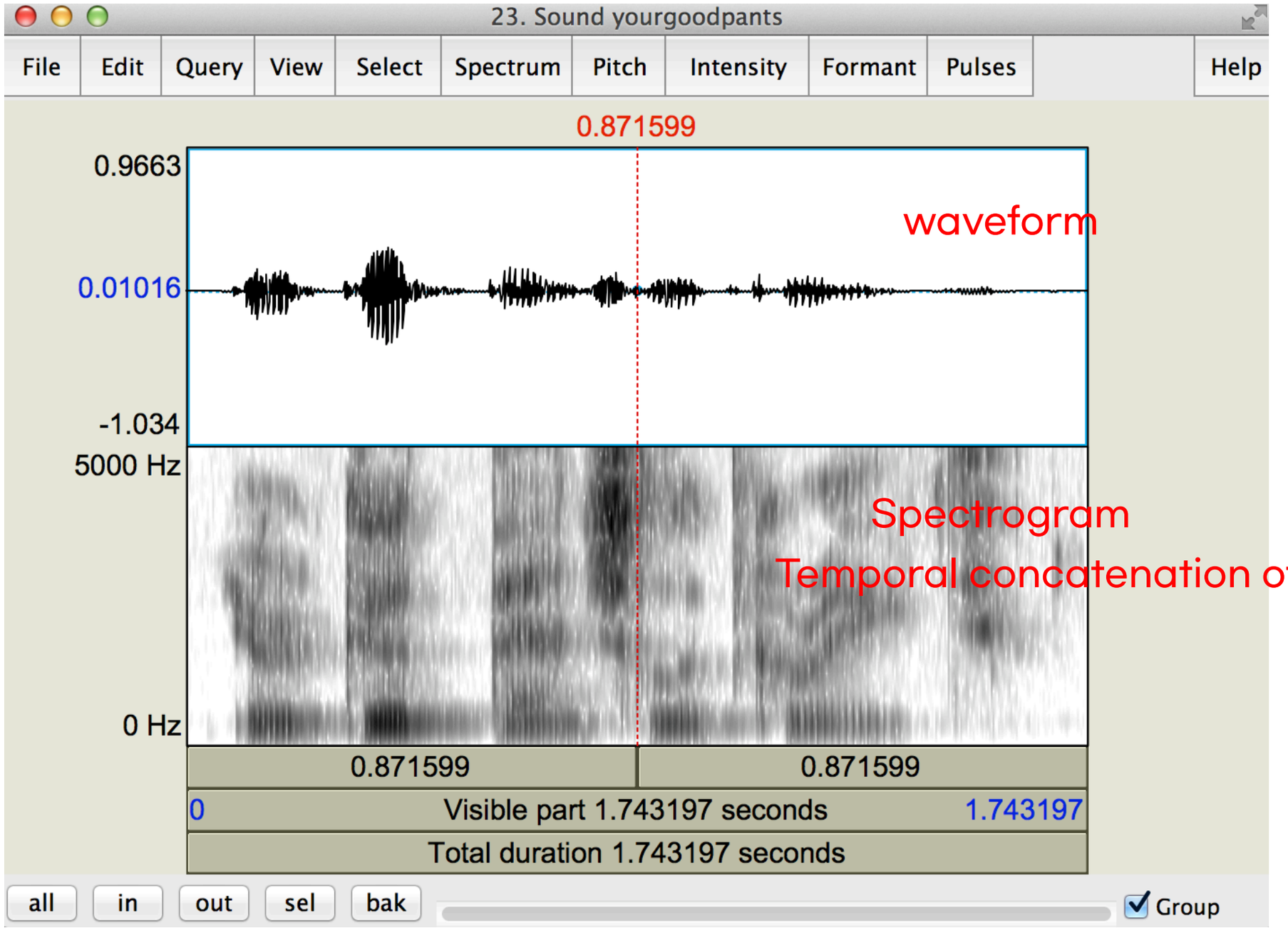
Resonances =
Formant Frequencies

=

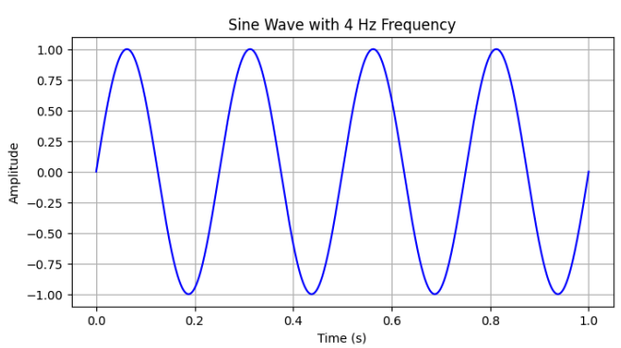


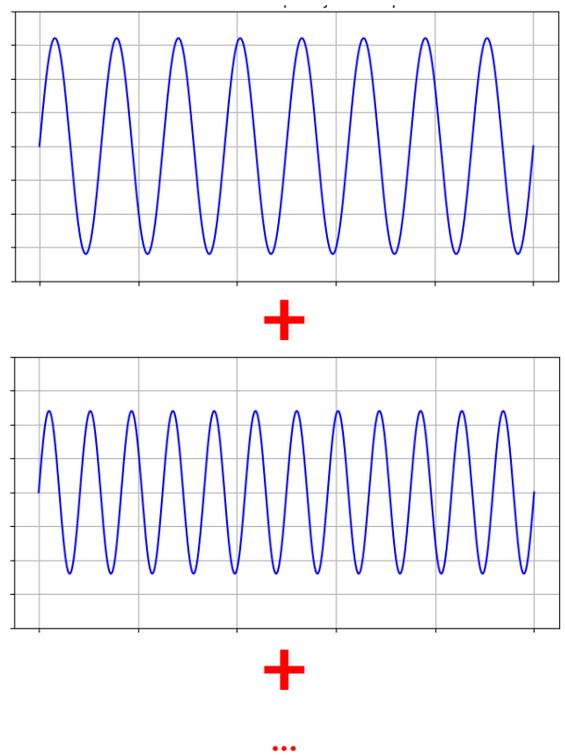
Output Spectrum

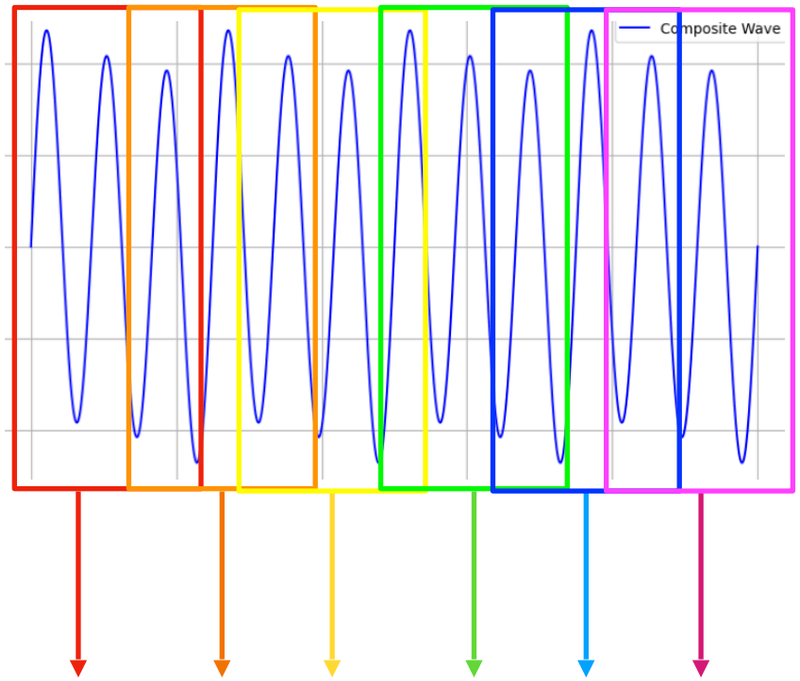
Spectrogram



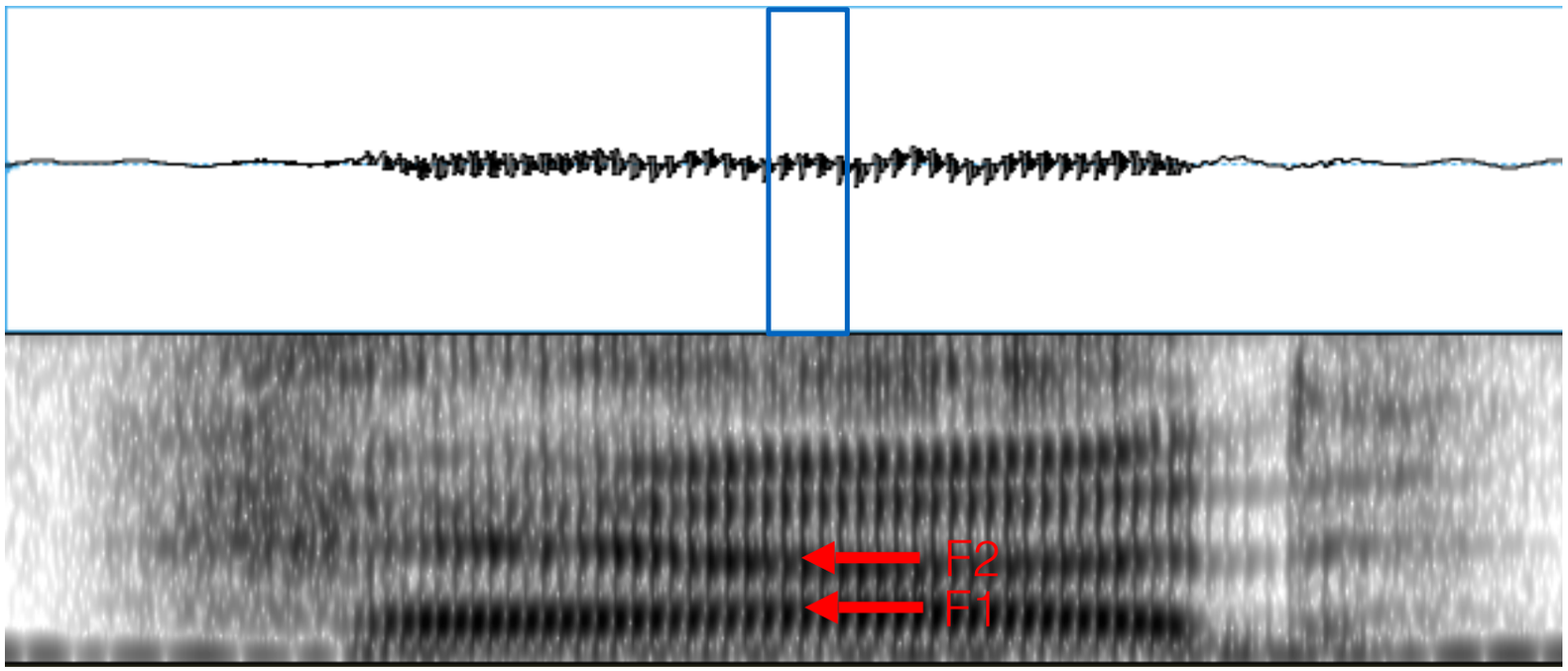
Spectrogram

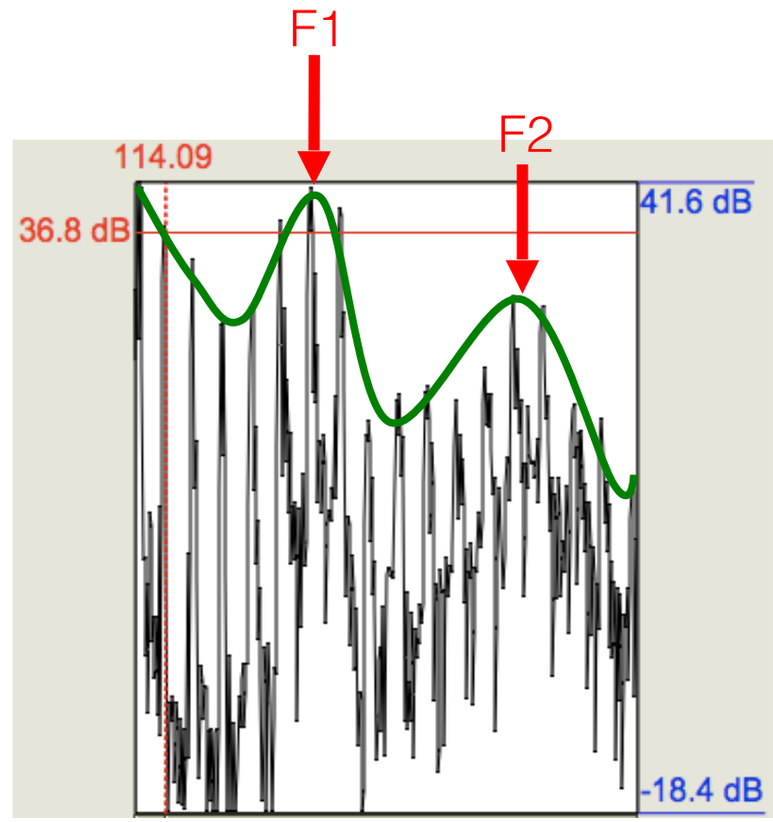






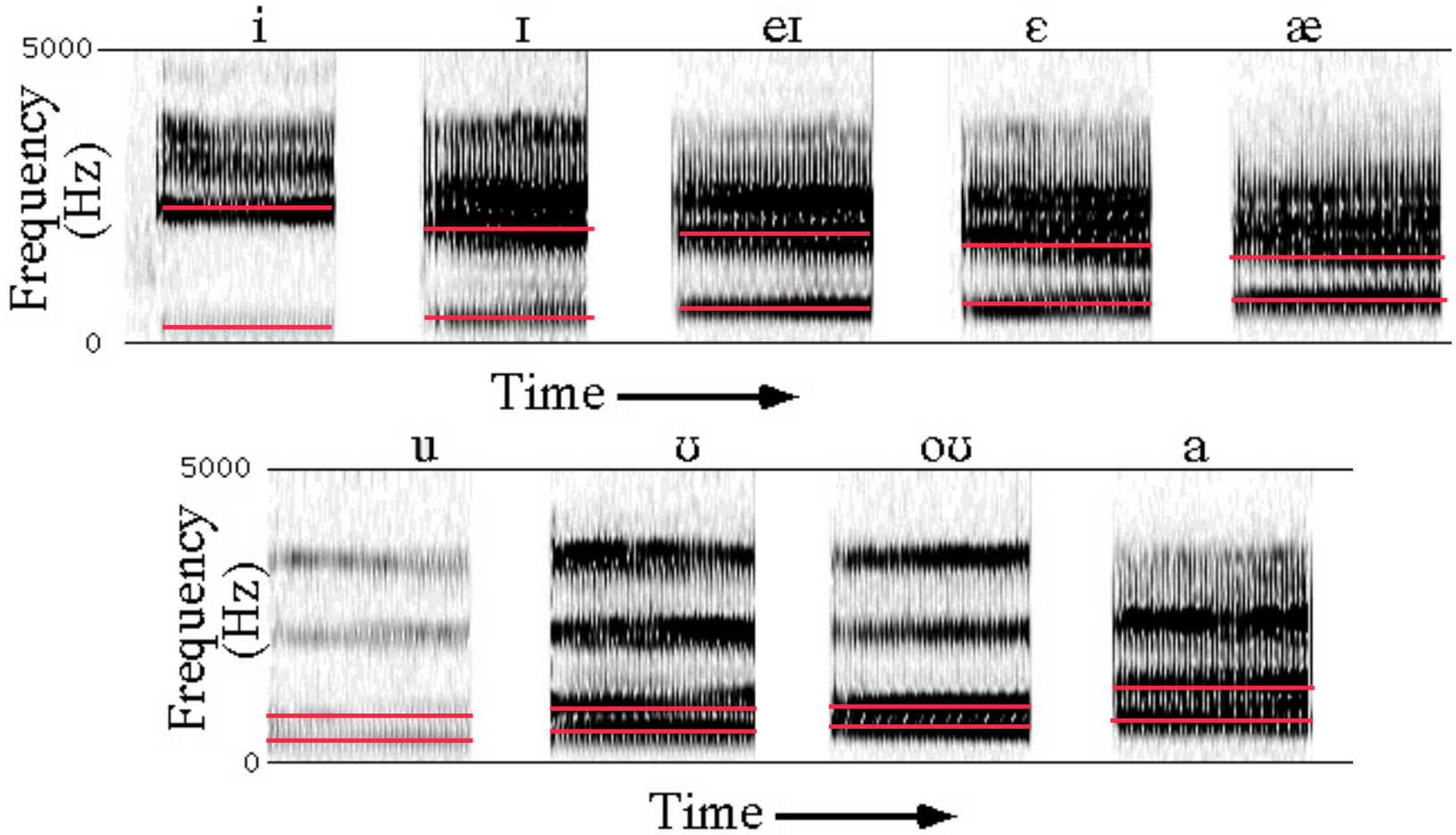
Spectrogram





- Airplane view of temporal concatenation of spectrum!
- Dark band: mountains = Formants

Formants



F1 and F2 are enough to disambiguate vowels.

Vowel space & Formants

front-back

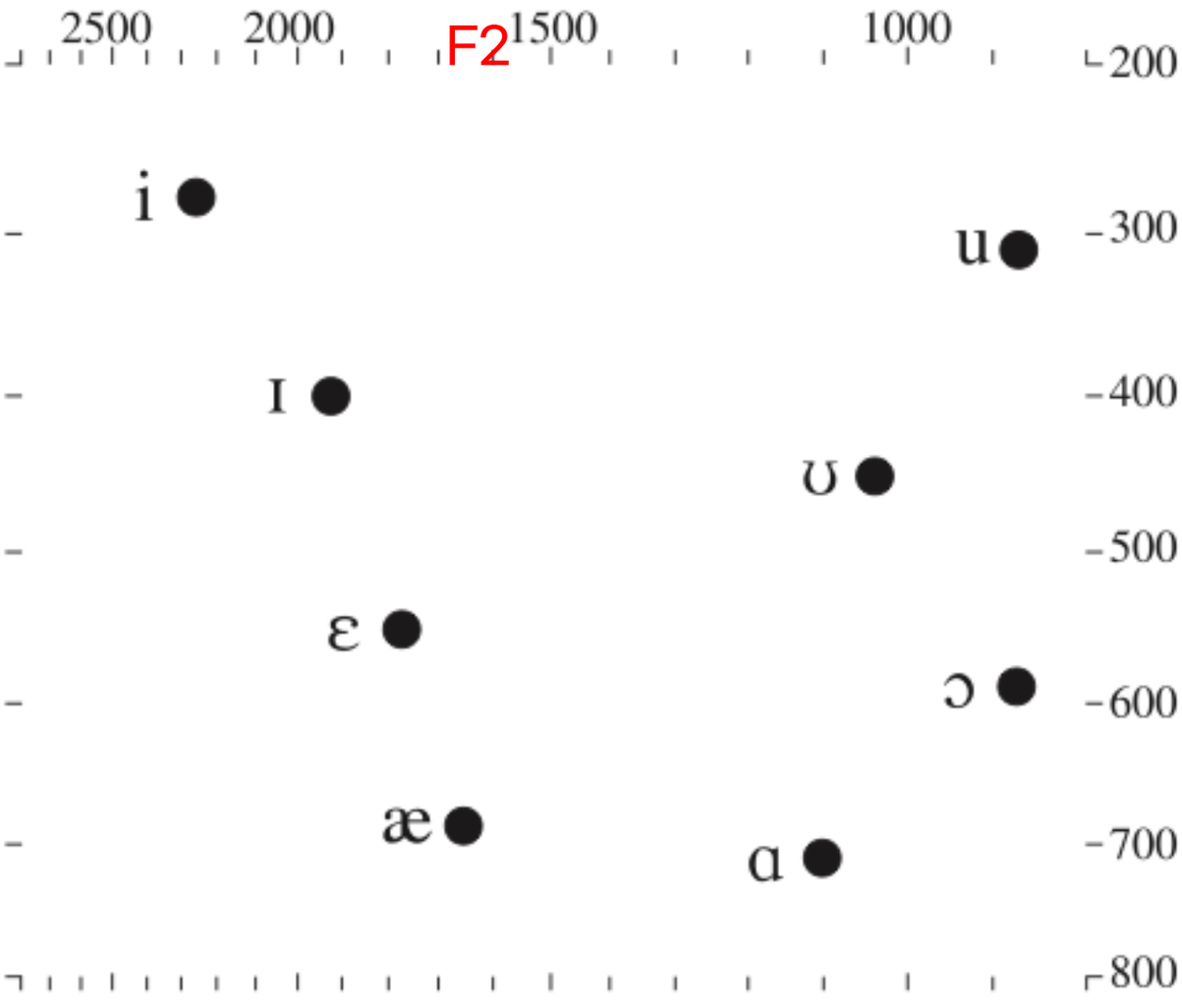


Figure 1.12 The positions of the vocal organs for the vowels in the words 1 *heed*, 2 *hid*, 3 *head*, 4 *had*, 5 *father*, 6 *good*, 7 *food*. The lip positions for vowels 2, 3, and 4 are between those shown for 1 and 5. The lip position for vowel 6 is between those shown for 1 and 7.