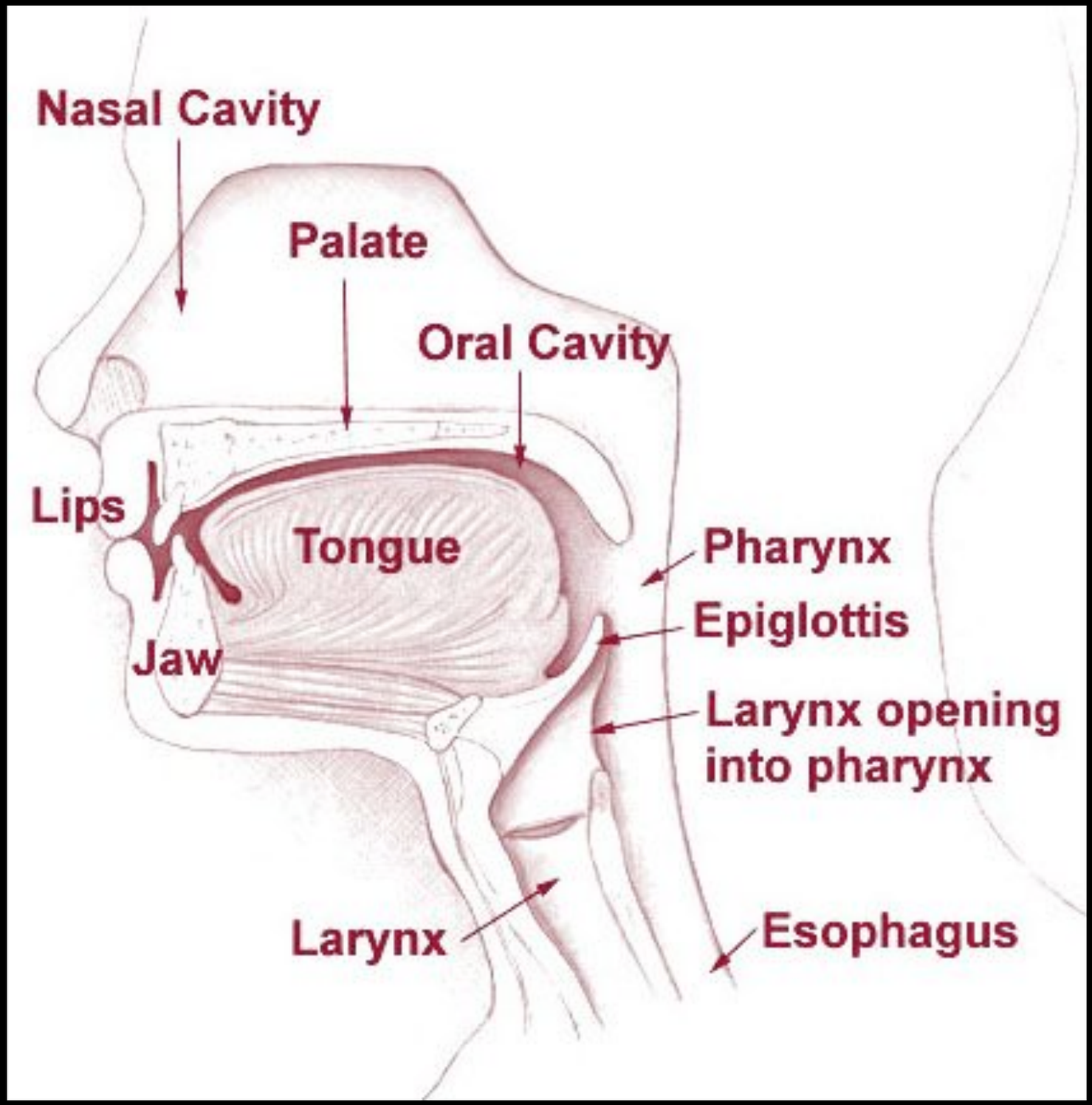
The Human Voice

Alexander Refsum Jensenius
MUS2800 - Sound Theory 1 - H2010
University of Oslo

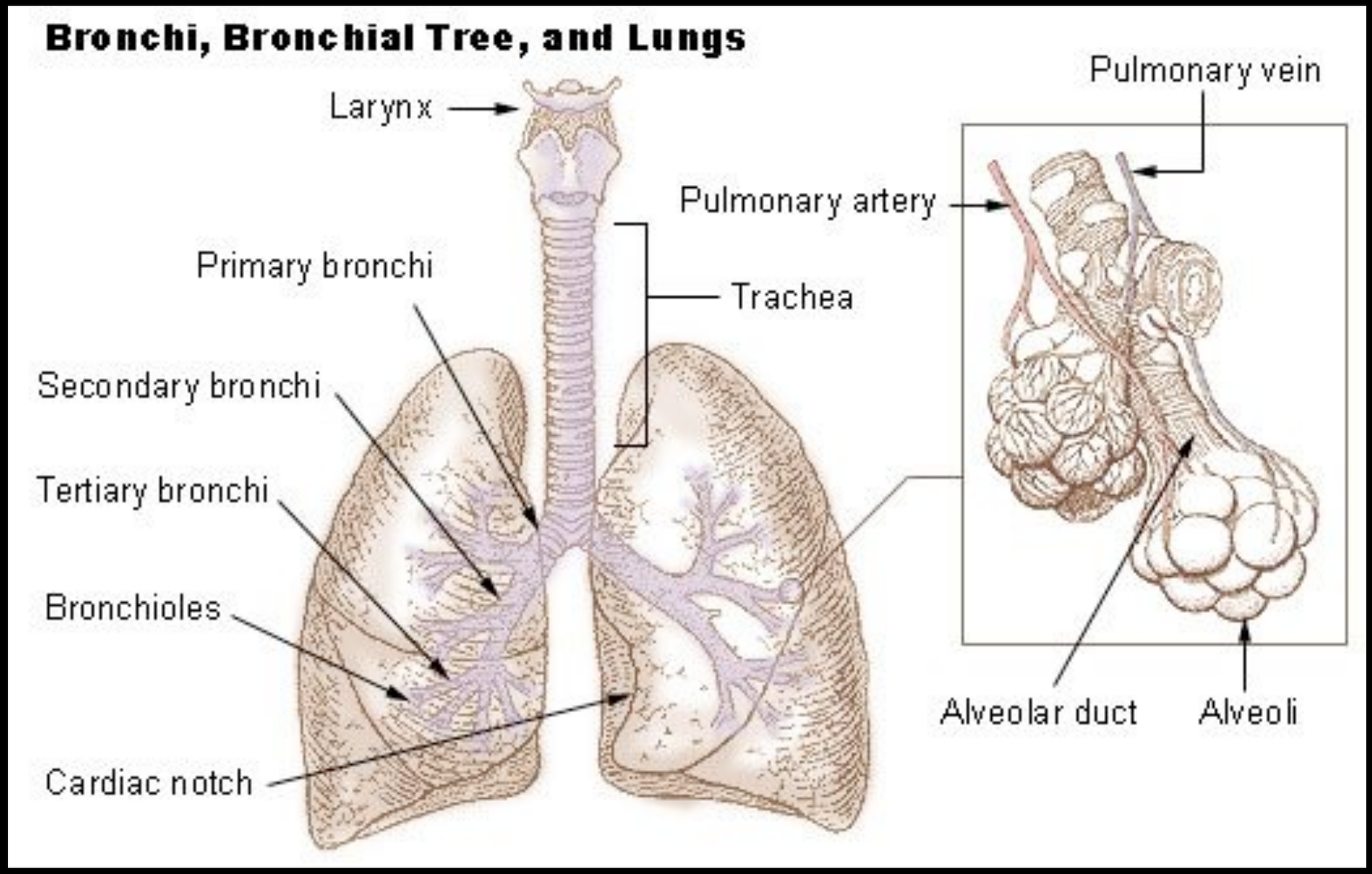
Speech production


glottal flow
vocal tract
mouth

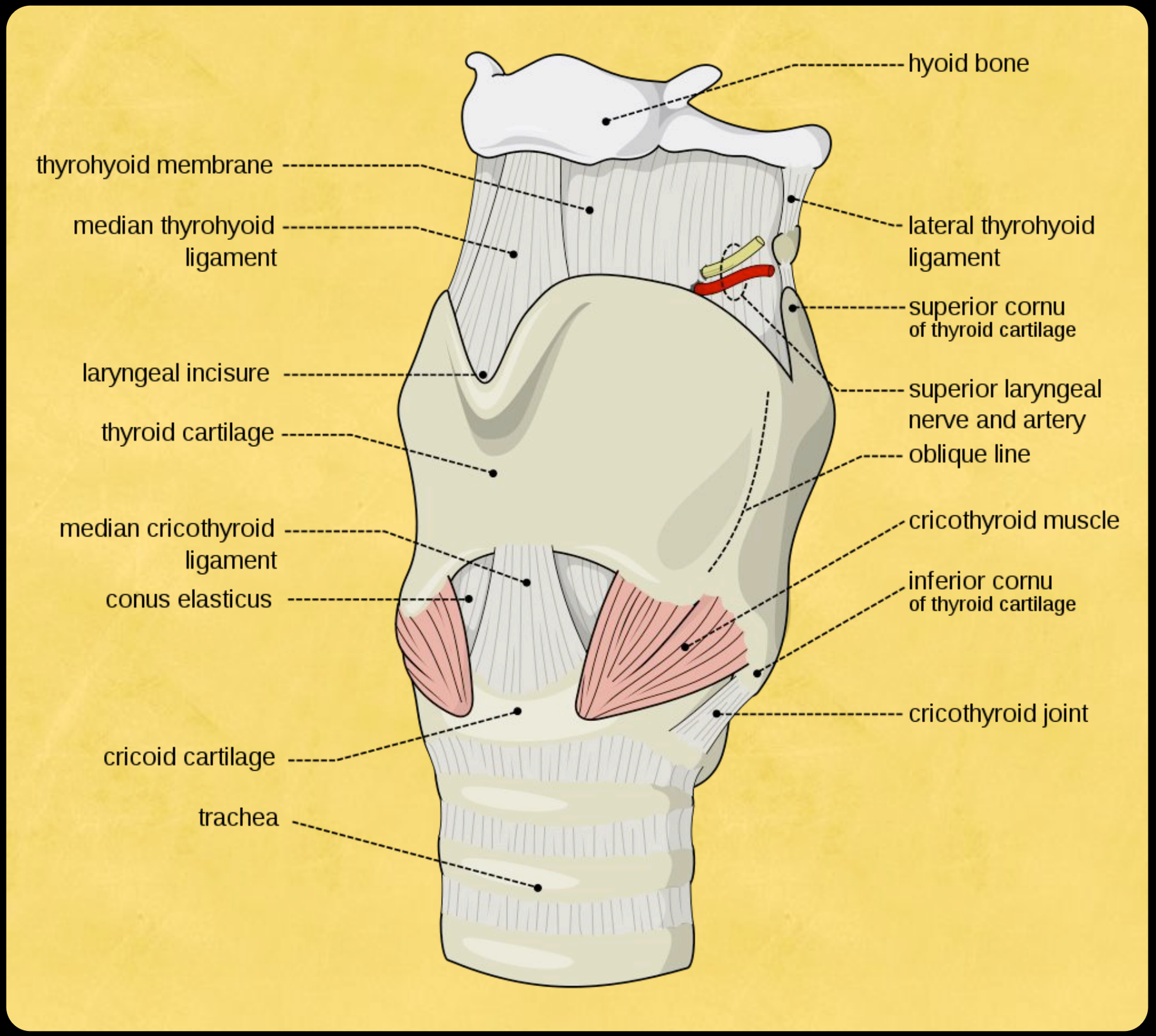

source
filter
radiator


Nasal Cavity
Palate
Oral Cavity
Lips
Tongue
Pharynx
Epiglottis
Jaw
Larynx opening into pharynx
Larynx
Esophagus

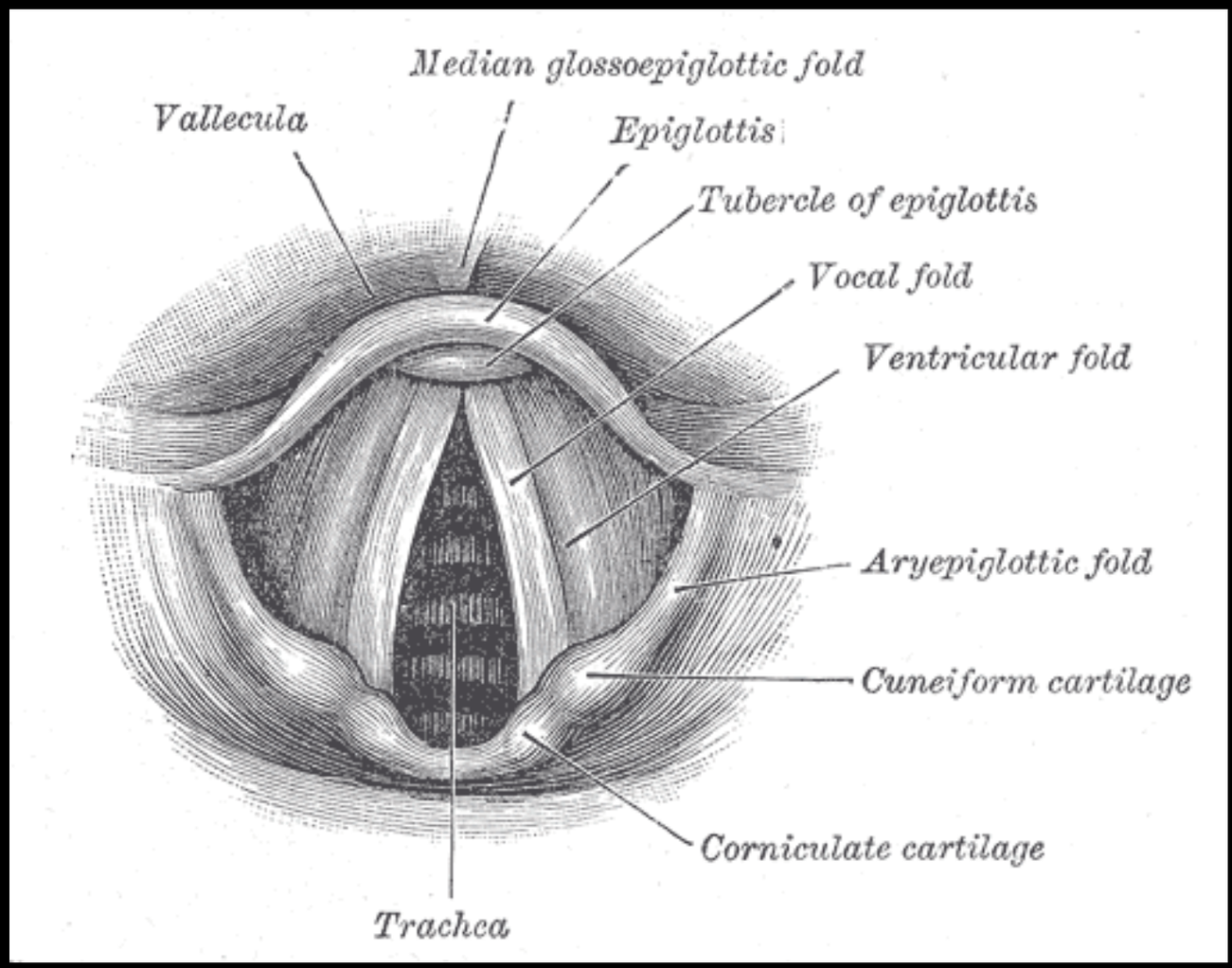
Lungs


Bronchi, Bronchial Tree, and Lungs
Larynx
Primary bronchi
Secondary bronchi
Tertiary bronchi
Bronchioles
Cardiac notch
Trachea
Pulmonary vein
Pulmonary artery
Alveolar duct
Alveoli

Larynx (strupehode)

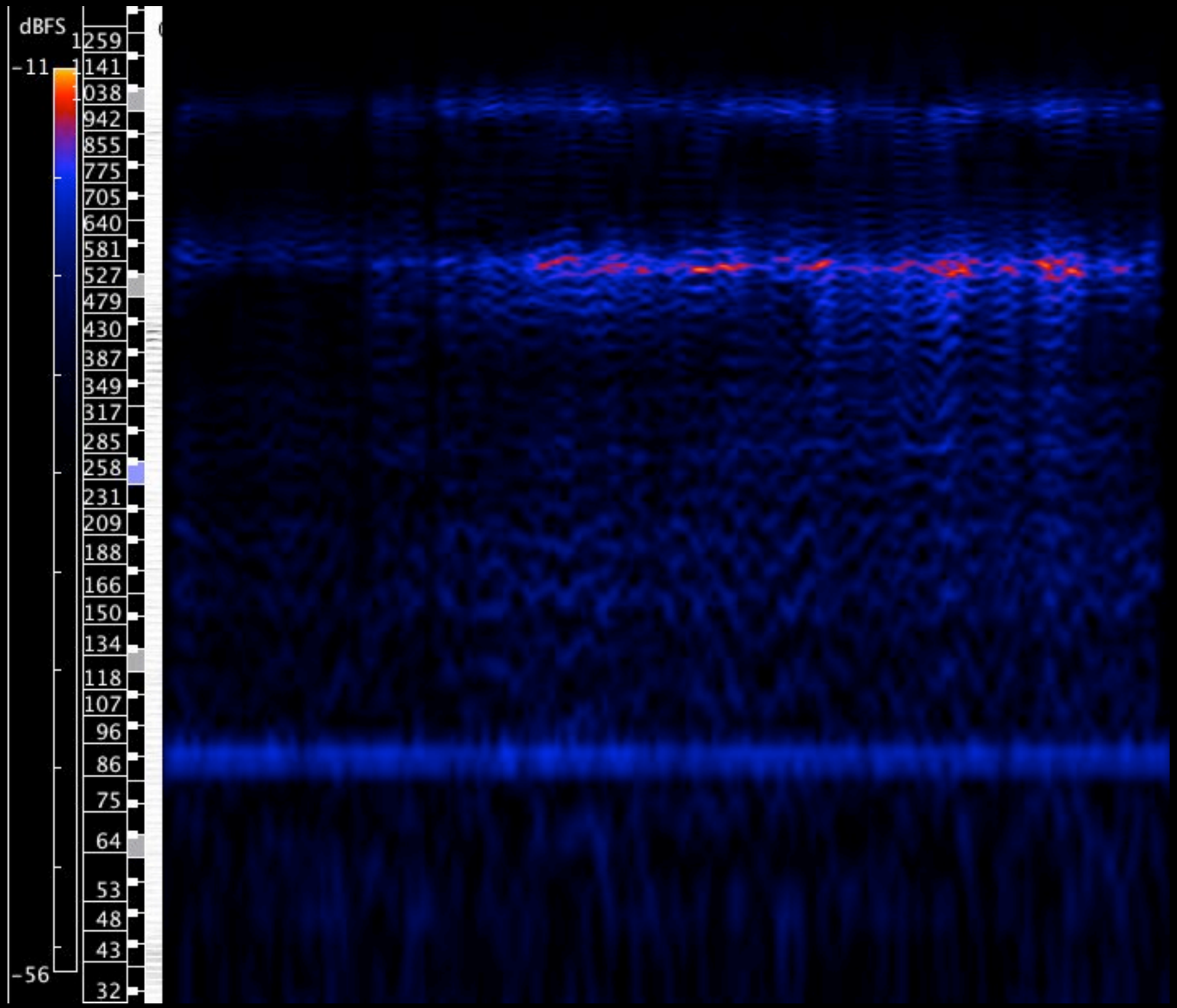


Vocal folds (stemmebånd)



Whispering

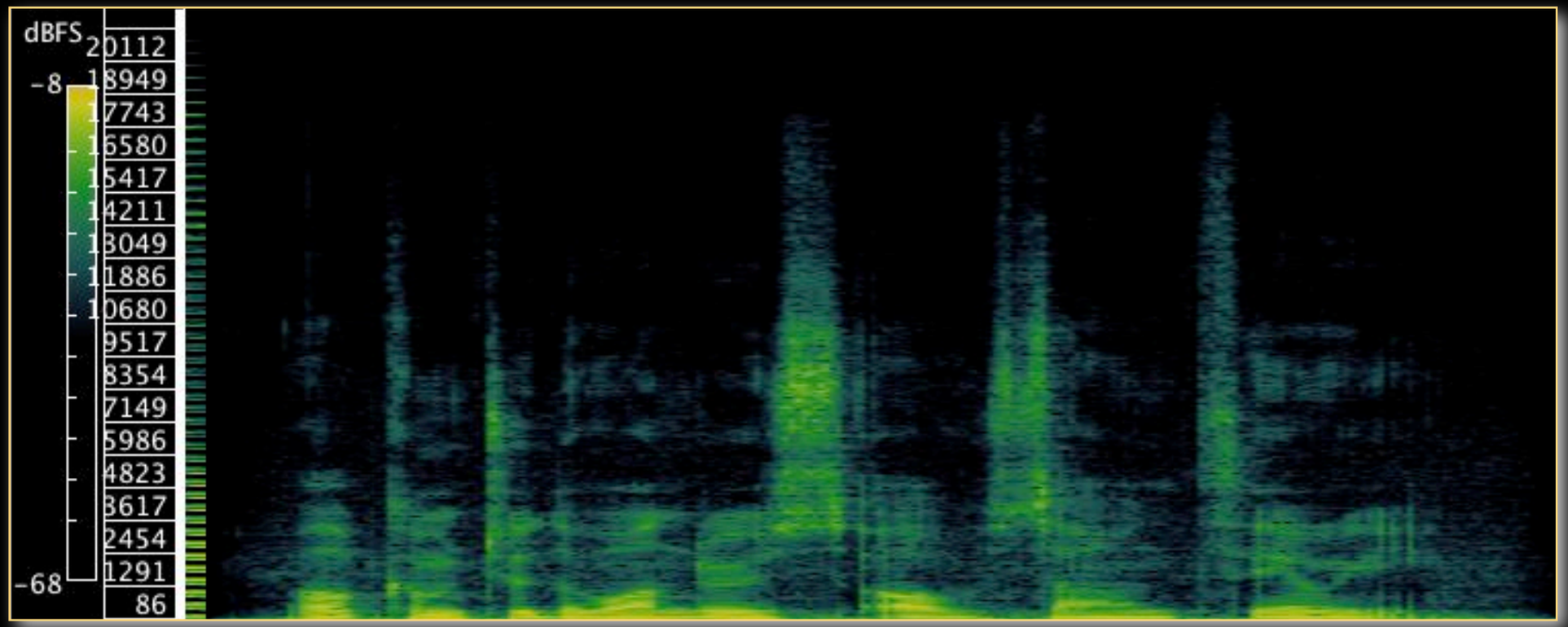
Voiced vs. unvoiced
speech vs. whispering

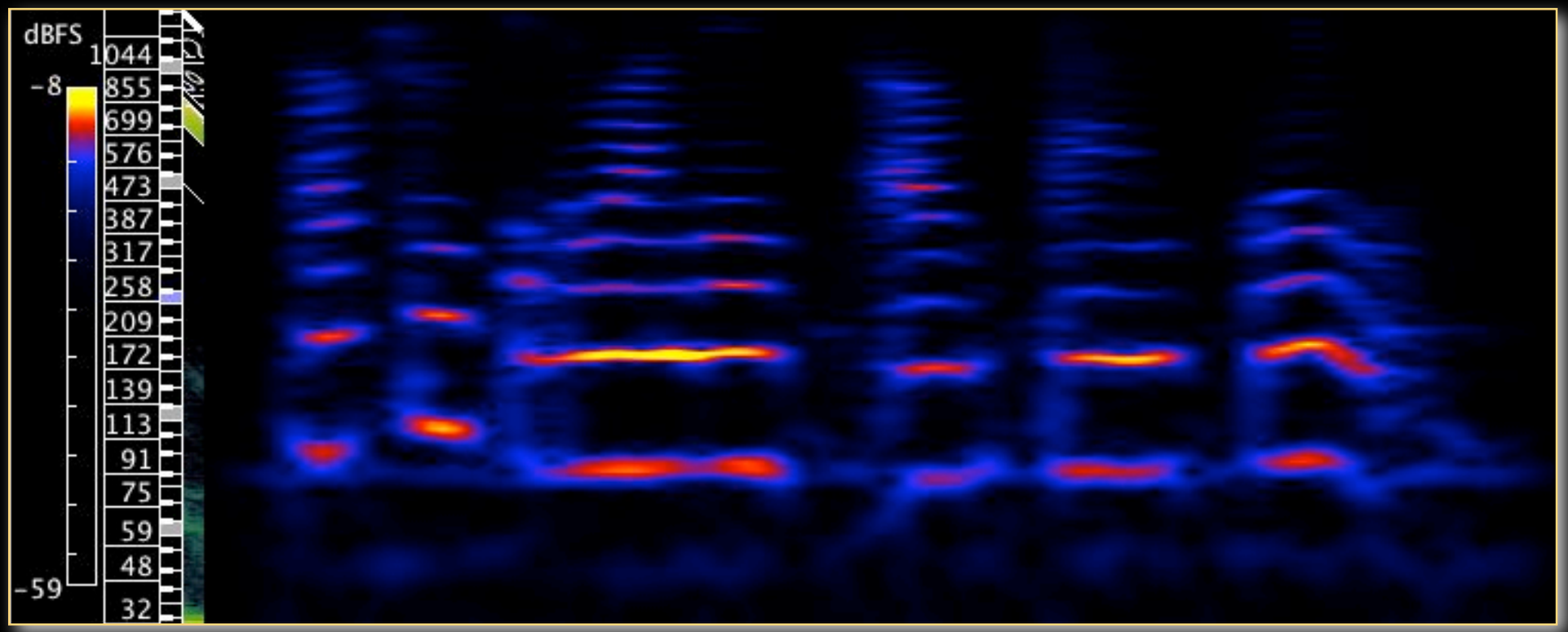

dBFS
-11
-56
1259
1141
1038
942
855
775
705
640
581
527
479
430
387
349
317
285
258
231
209
188
166
150
134
118
107
96
86
75
64
53
48
43
32

Child: 300 Hz
Female: 220 Hz
Male: 110 Hz

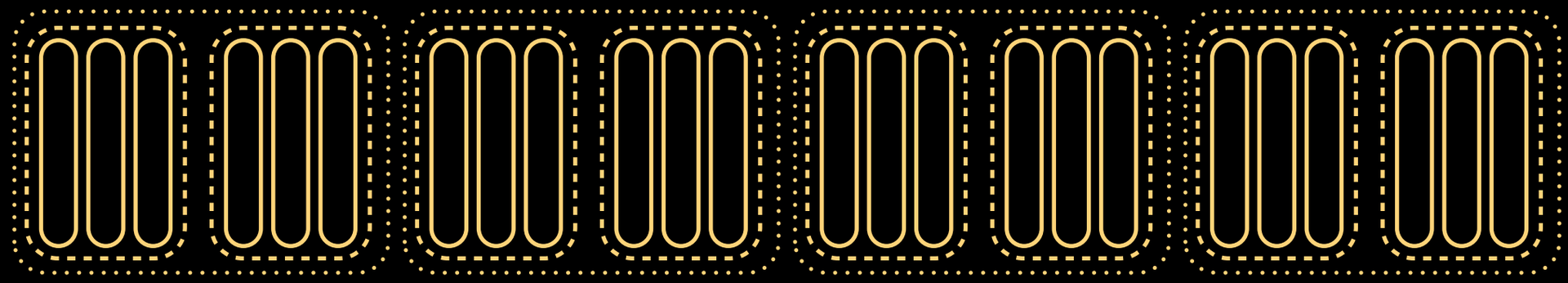
Vocal tract

nasal cavity (nose)
oral cavity (mouth)
tongue
lips


dBFS
-8
-68
20112
18949
17743
16580
15417
14211
13049
11886
10680
9517
8354
7149
5986
4823
3617
2454
1291
86


dBFS
-8
-59
1044
855
699
576
473
387
317
258
209
172
139
113
91
75
59
48
32

Phoneme
Syllable
Word



Phoneme

Vowels

Consonants

THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)

	LABIAL		CORONAL				DORSAL			RADICAL	
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Plosive	p b	p̪ b̪		t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʡ
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ʜ ʢ
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Trill	ʙ			r					ʀ		
Tap, Flap		ⱱ		ɾ		ɽ					
Lateral fricative				ɬ ɮ		ꞎ	ʎ̝	ʟ̝			
Lateral approximant				l		ɭ	ʎ	ʟ			
Lateral flap				ɺ		ɭ̆					

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured ɦ. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

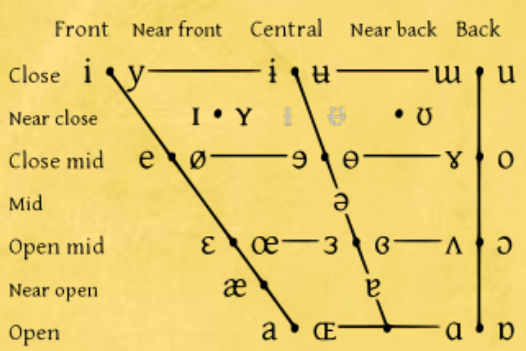
CONSONANTS (NON-PULMONIC)

Anterior click releases (require posterior stops)	Voiced implosives	Ejectives
ʘ Bilabial fricated	ɓ Bilabial	ʼ *Examples:*
ǀ Laminal alveolar fricated ("dental")	ɗ Dental or alveolar	pʼ Bilabial
ǃ Apical (post)alveolar abrupt ("retroflex")	ʄ Palatal	tʼ Dental or alveolar
ǂ Laminal postalveolar abrupt ("palatal")	ɠ Velar	kʼ Velar
ǁ Lateral alveolar fricated ("lateral")	ʛ Uvular	sʼ Alveolar fricative

CONSONANTS (CO-ARTICULATED)

- ʍ Voiceless labialized velar approximant
- w Voiced labialized velar approximant
- ɥ Voiced labialized palatal approximant
- ɕ Voiceless palatalized postalveolar (alveolo-palat
- ʑ Voiced palatalized postalveolar (alveolo-palatal
- ɧ Simultaneous x and ʃ (disputed)
- k͡p t͡s Affricates and double articulations may be joine

VOWELS



Vowels at right & left of bullets are rounded & unrounded.

SUPRASEGMENTALS

ˈ Primary stress	ˈˈ Extra stress		
ˌ Secondary stress	[ˌfoʊnəˈtɪʃən]		
eː Long	eˑ Half-long		
e Short	ĕ Extra-short		
. Syllable break	‿ Linking (no break)		

INTONATION

| Minor (foot) break	
‖ Major (intonation) break	
↗ Global rise	↘ Global fall

TONE

Level tones		*Conto*
e̋ ˥ Top		ě
é ˦ High		ê
ē ˧ Mid		e᷄
è ˨ Low		e᷅
ȅ ˩ Bottom		e᷈
Tone terracing		e᷉
ꜛ Upstep		
ꜜ Downstep		

DIACRITICS

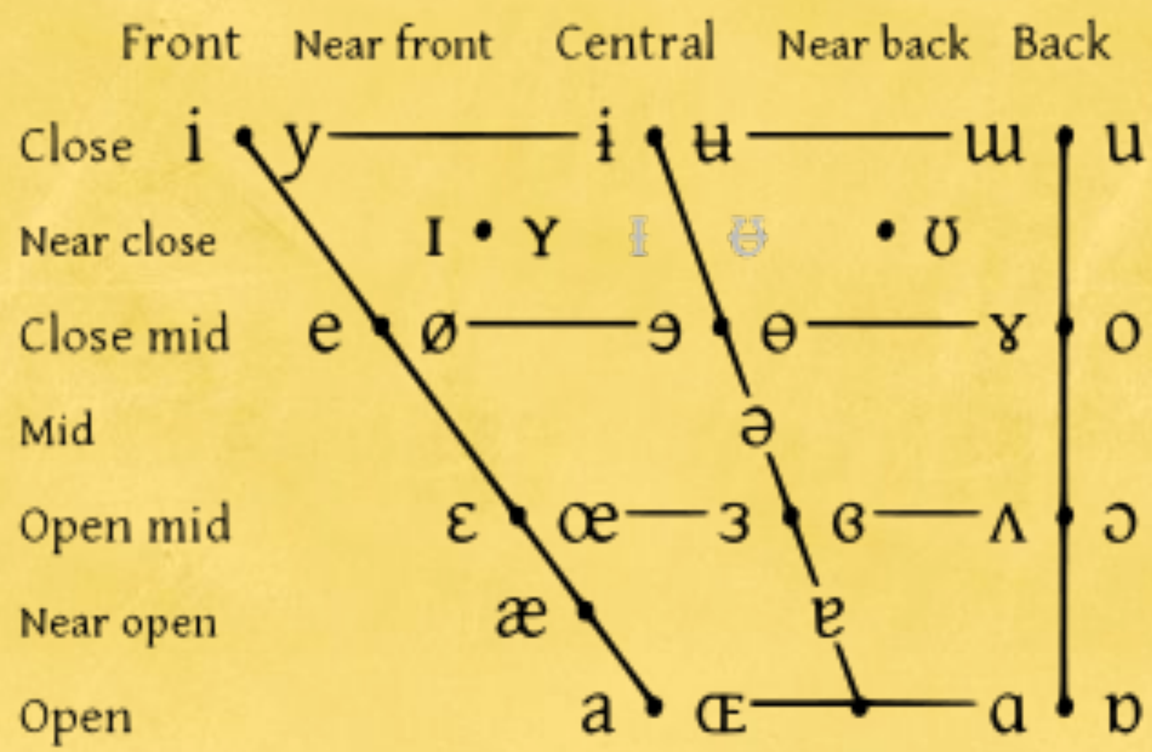
Diacritics may be placed above a symbol with a descender, as ŋ̊. Other IPA symbols may appear as diacri
phonetic detail: tᶠ (fricative release), bʱ (breathy voice), ˀa (glottal onset), ᵊ (epenthetic schwa), oᵘ (diph

SYLLABICITY & RELEASES		PHONATION		PRIMARY ARTICULATION		SECONDARY ARTICULATION			
n̩ ɹ̩	Syllabic	n̥ d̥	Voiceless or Slack voice	t̪ b̪	Dental	tʷ dʷ	Labialized	ɔ̹ x	
e̯ ʊ̯	Non-syllabic	s̬ d̬	Modal voice or Stiff voice	t̺ d̺	Apical	tʲ dʲ	Palatalized	ɔ̜ xʷ	
tʰ ʰt	(Pre)aspirated	n̤ a̤	Breathy voice	t̻ d̻	Laminal	tˠ dˠ	Velarized	ẽ z̃	
dⁿ	Nasal release	n̰ a̰	Creaky voice	u̟ t̟	Advanced	tˤ dˤ	Pharyngealized	ɚ ɝ	
dˡ	Lateral release	n͈ a͈	Strident	i̠ t̠	Retracted	ɫ ᵶ	Velarized or pharyngealized	e̘ o̘	
t̚	No audible release	n̼ d̼	Linguolabial	ä ï	Centralized	ŭ̽	Mid-centralized	e̙ o̙	Retracted tongue root
e̞ β̞	Lowered (β̞ is a bilabial approximant)			e̝ ɹ̝	Raised (ɹ̝ is a voiced alveolar non-sibilant fricative, ɾ̞ a fricative trill)				

CONSONANTS (NON-PULMONIC)

Anterior click releases (require posterior stops)	Voiced implosives	Ejectives
ʘ Bilabial fricated	ɓ Bilabial	ʼ *Examples:*
ǀ Laminal alveolar fricated ("dental")	ɗ Dental or alveolar	pʼ Bilabial
ǃ Apical (post)alveolar abrupt ("retroflex")	ʄ Palatal	tʼ Dental or alveolar
ǂ Laminal postalveolar abrupt ("palatal")	ɠ Velar	kʼ Velar
ǁ Lateral alveolar fricated ("lateral")	ʛ Uvular	sʼ Alveolar fricative

VOWELS



Vowels at right & left of bullets are rounded & unrounded.

Syllables

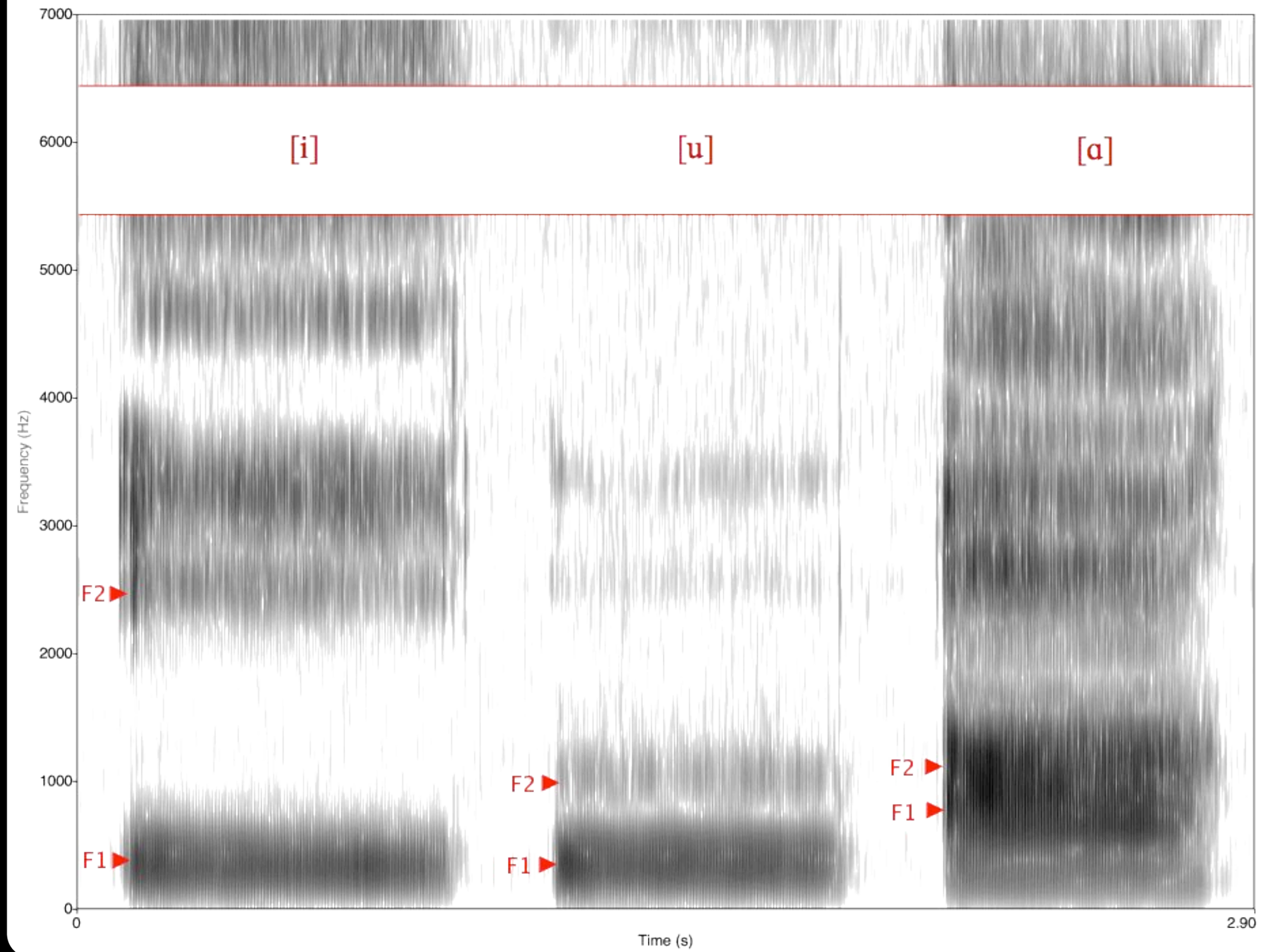
schtroumpfed
/ˈʃtruːmpft/

squirrelled
/ˈskwɜrld/

broughammed
/ˈbruːmd/

http://en.wikipedia.org/wiki/List_of_the_longest_English_words_with_one_syllable

Formants


7000
6000
5000
4000
3000
2000
1000
0
Frequency (Hz)
[i]
[u]
[ɑ]
F2
F1
F2
F1
F2
F1
0
2.90
Time (s)

Vowel formant centers

Vowel	IPA	Formant f_1	Formant f_2
u	u	**320 Hz**	800 Hz
o	o	**500 Hz**	1000 Hz
ɑ	ɑ	**700 Hz**	1150 Hz
a	a	**1000 Hz**	1400 Hz
ø	ø	500 Hz	**1500 Hz**
y	y	320 Hz	**1650 Hz**
æ	ɛ	700 Hz	**1800 Hz**
e	e	500 Hz	**2300 Hz**
i	i	320 Hz	**2500 Hz**

Vowel formants

Vowel	Main formant region
u	200–400 Hz
o	400–600 Hz
a	800–1200 Hz
e	400–600 and 2200–2600 Hz
i	200–400 and 3000–3500 Hz

Prosody

Speech analysis and synthesis

Speech analysis

Speech recognition

Speech synthesis

Singing

Change vowel sounds

Singing vs. speaking

Second formant is lower in frequency

Singer's formant (~2500 - 3000 Hz)

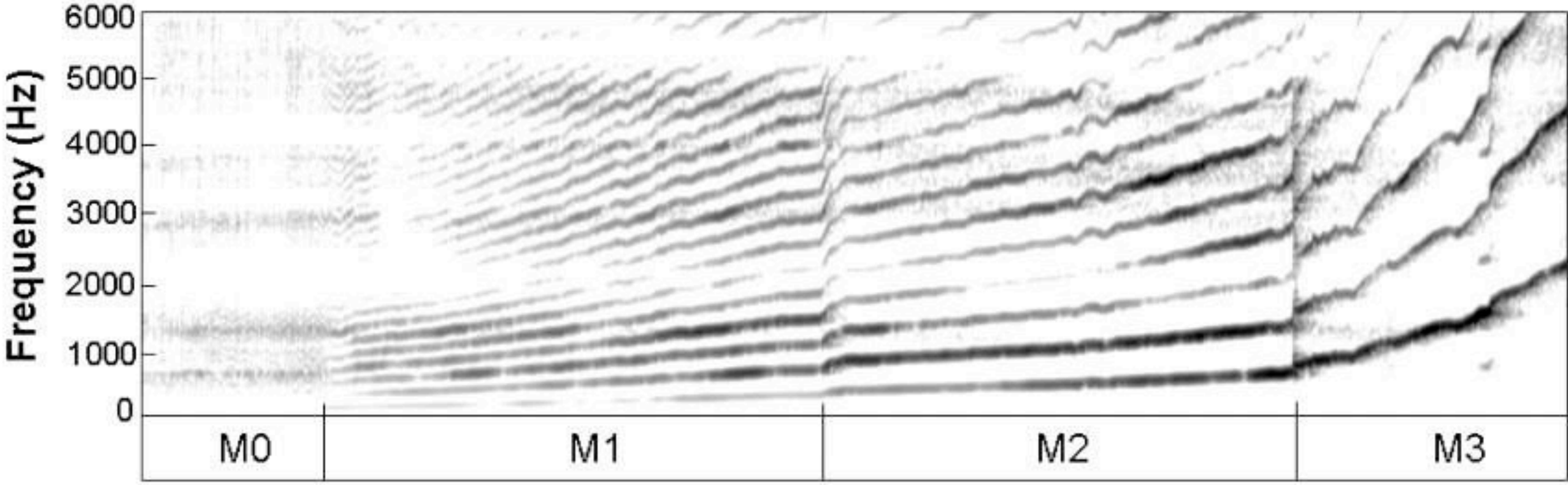
Registers

Mechanism 0 (M0): ‘vocal fry’

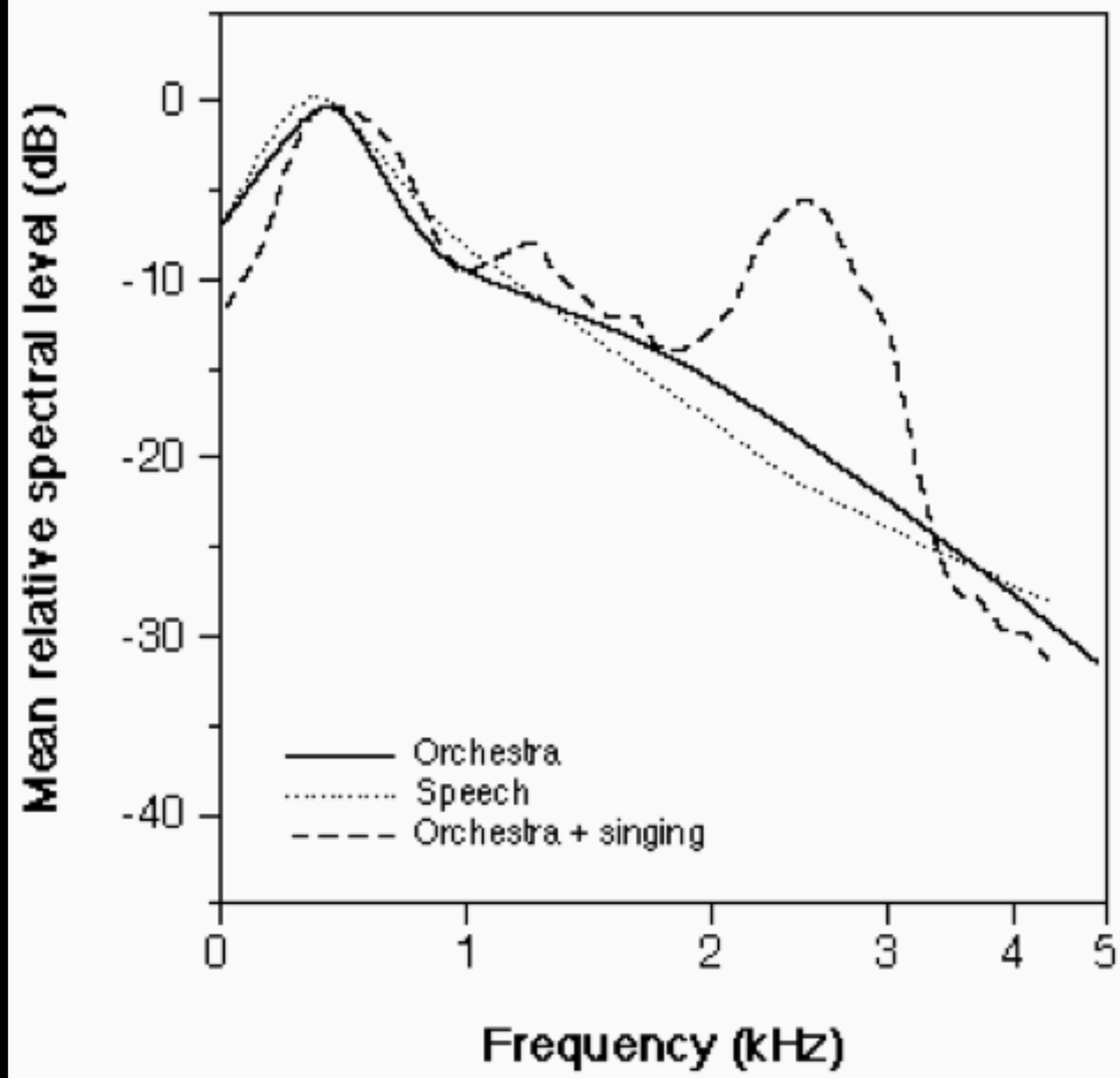
Mechanism 1 (M1): ‘chest’

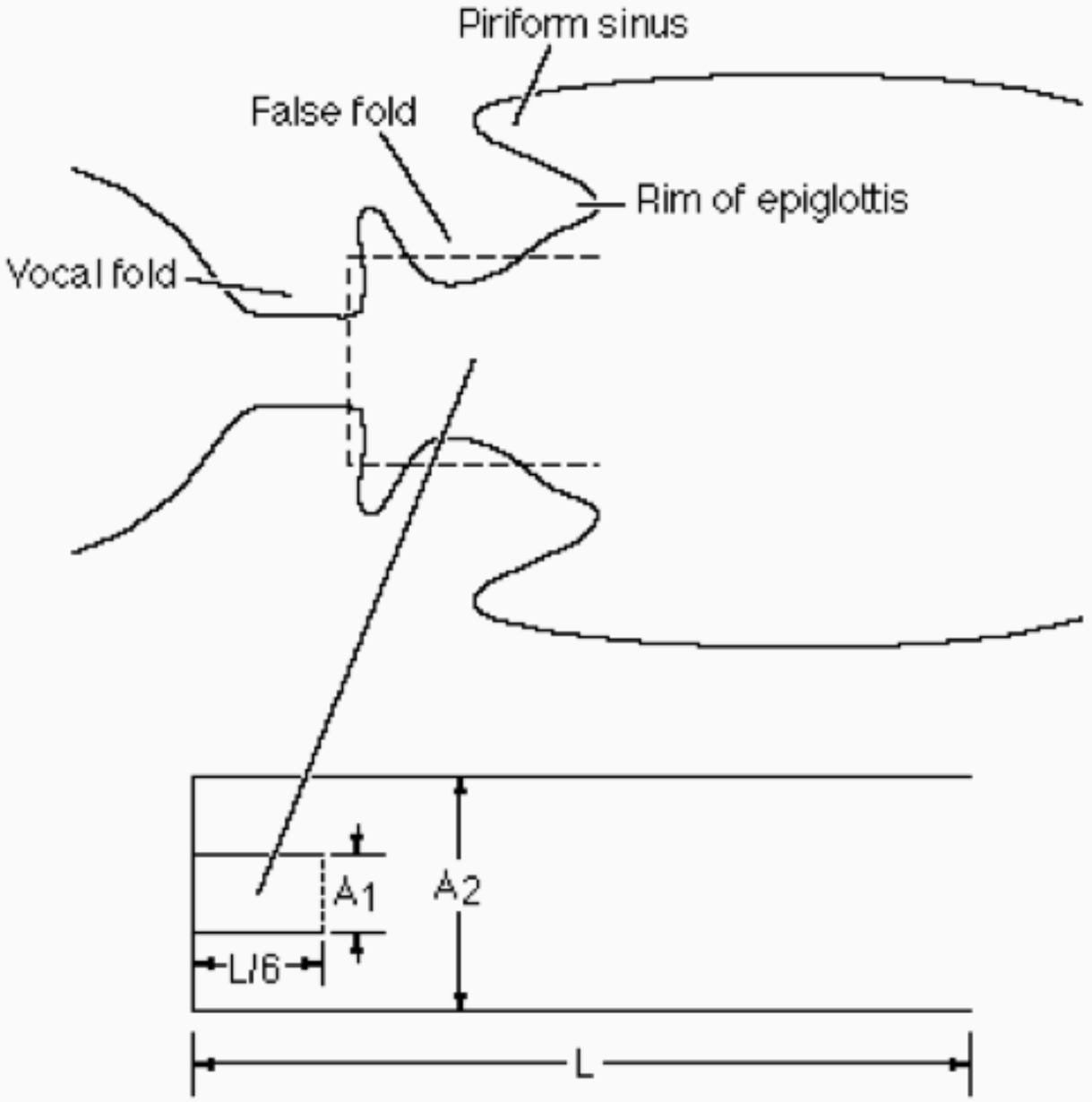
Mechanism 2 (M2): head/falsetto

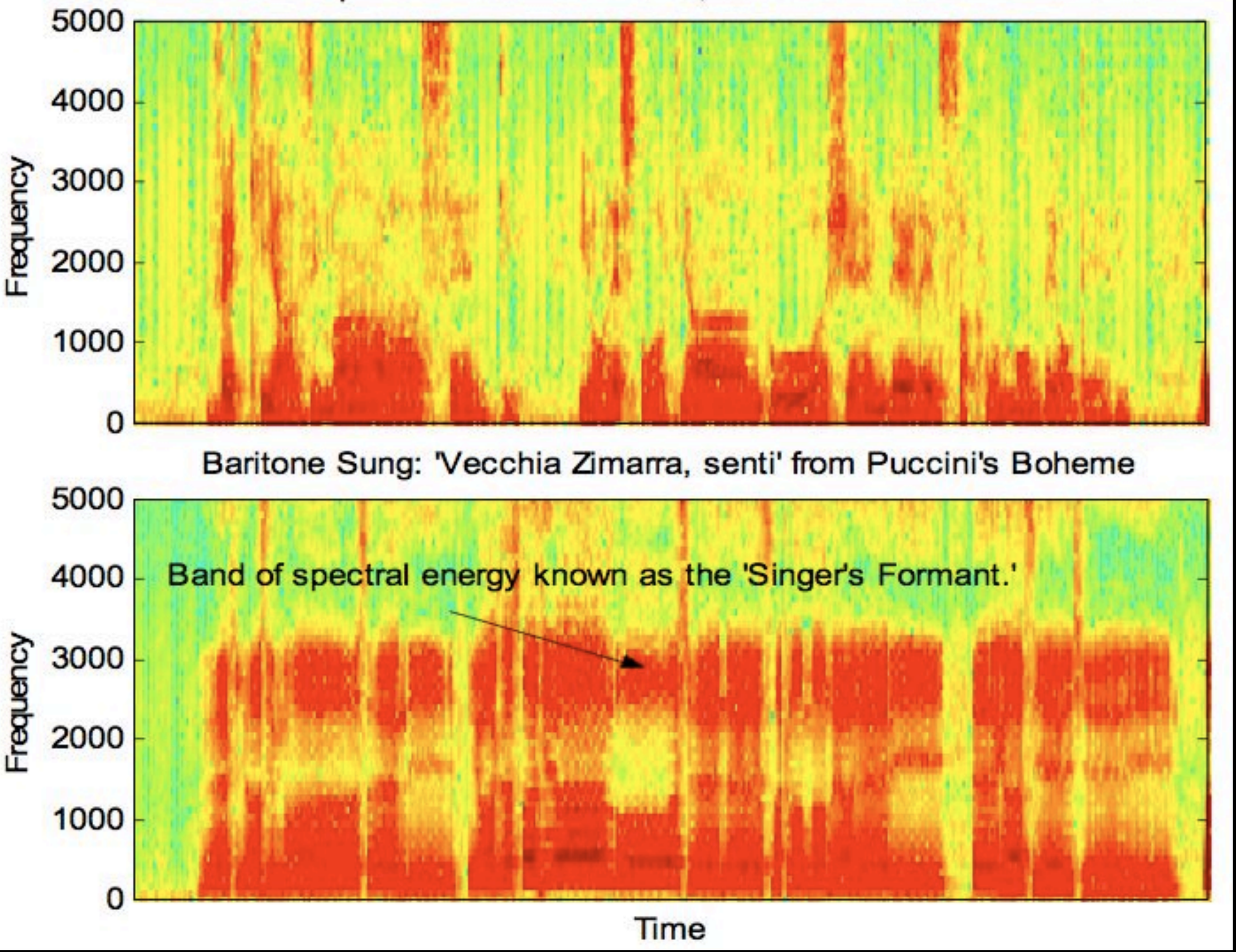
Mechanism 3 (M3): ‘flageolet’

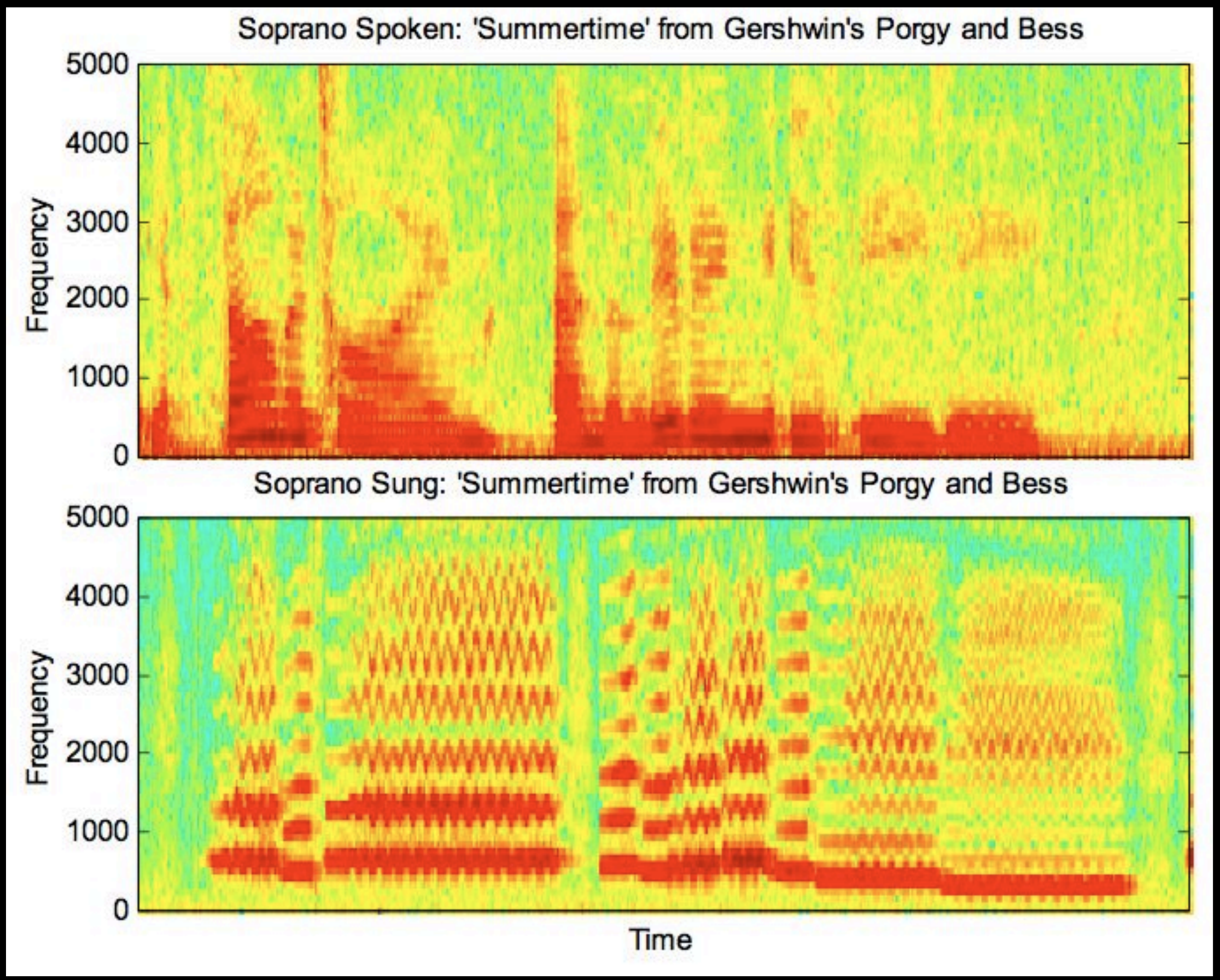

Frequency (Hz)
6000
5000
4000
3000
2000
1000
0
M0
M1
M2
M3

Singers' formant


Mean relative spectral level (dB)
0
-10
-20
-30
-40
0
1
2
3
4
5
Frequency (kHz)
Orchestra
Speech
Orchestra + singing


Piriform sinus
False fold
Rim of epiglottis
Vocal fold
A1
A2
L/6
L


Baritone Spoken: 'Vecchia Zimarra, senti' from Puccini's Boheme
5000
4000
3000
2000
1000
0
Frequency
Baritone Sung: 'Vecchia Zimarra, senti' from Puccini's Boheme
5000
4000
3000
2000
1000
0
Frequency
Band of spectral energy known as the 'Singer's Formant.'
Time


Soprano Spoken: 'Summertime' from Gershwin's Porgy and Bess
5000
4000
3000
2000
1000
0
Frequency
Soprano Sung: 'Summertime' from Gershwin's Porgy and Bess
5000
4000
3000
2000
1000
0
Frequency
Time

Overtone/harmonic singing