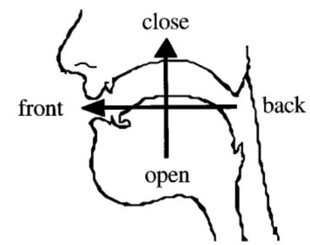
Human Voice, Hearing System and Binaural perception

Yuan-Fu Liao
National Taipei University of Technology

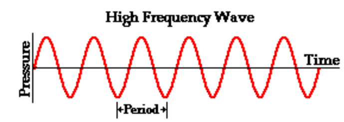
yfliao@ntut.edu.tw
http://www.ntut.edu.tw/~yfliao

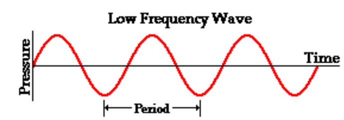
3 domains of phonetics

- articulatory phonetics

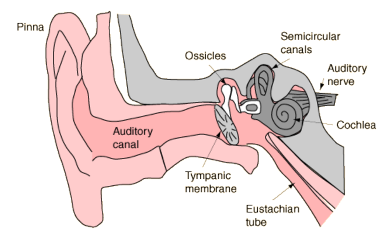


- acoustic phonetics



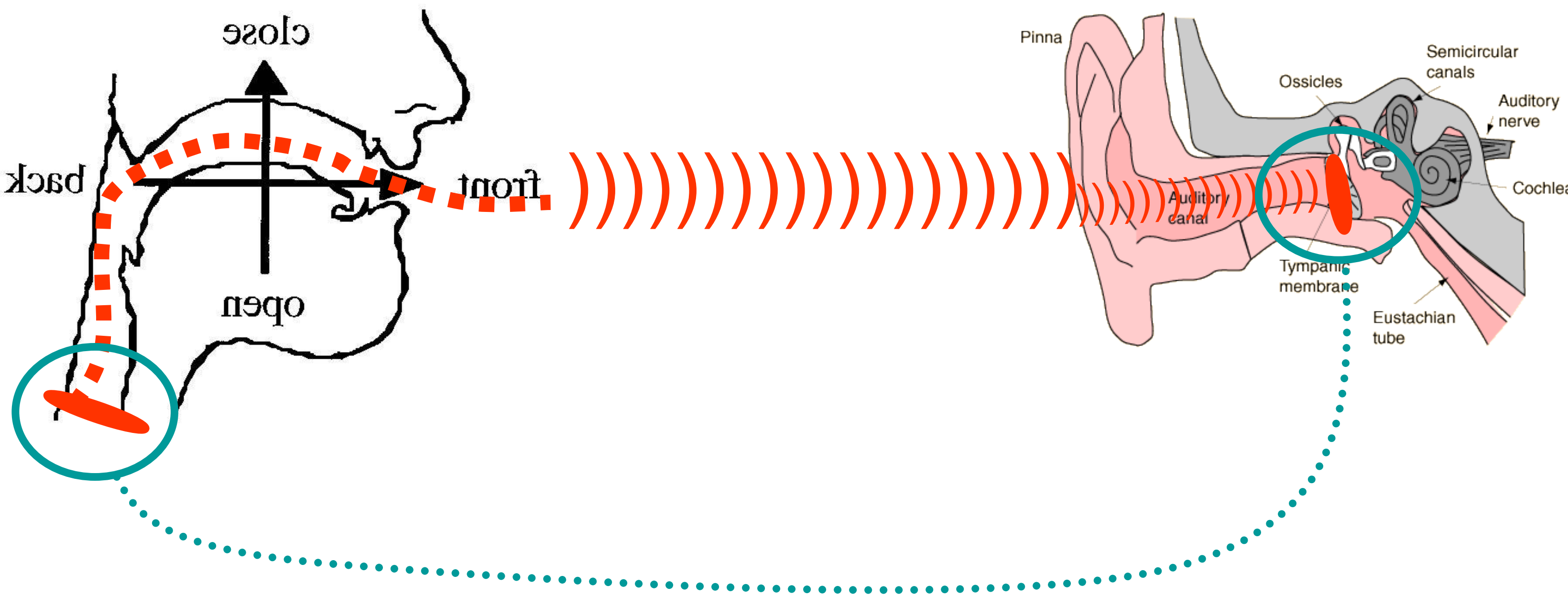


- auditory phonetics



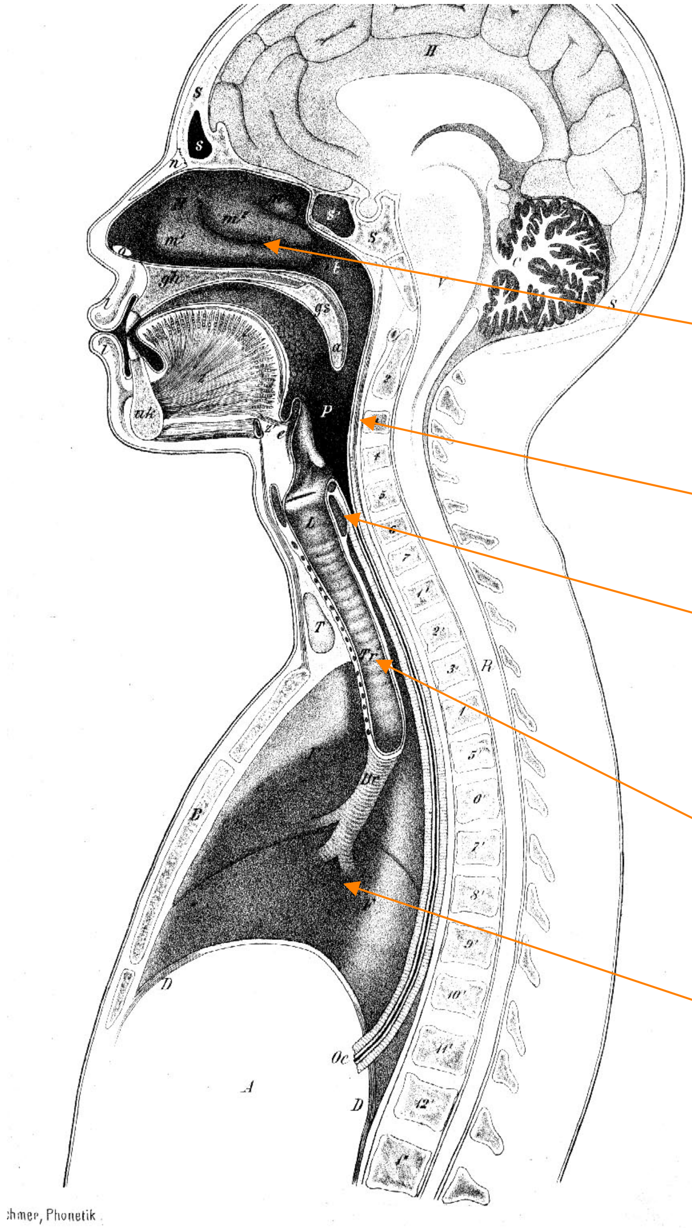
3 domains of phonetics

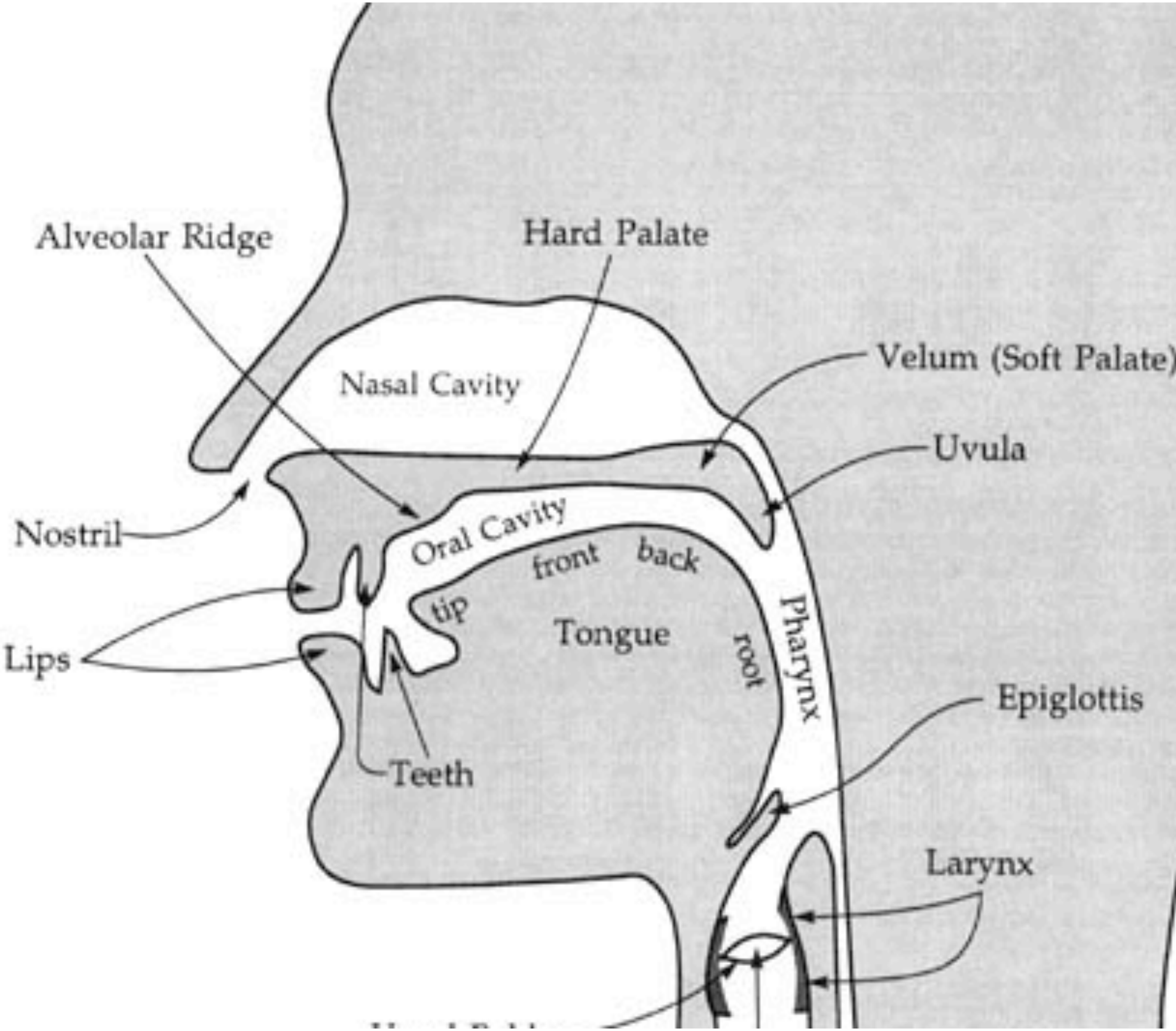
articulation acoustics audition



發聲器官結構

Sagittal section of the vocal tract
(Techmer 1880)

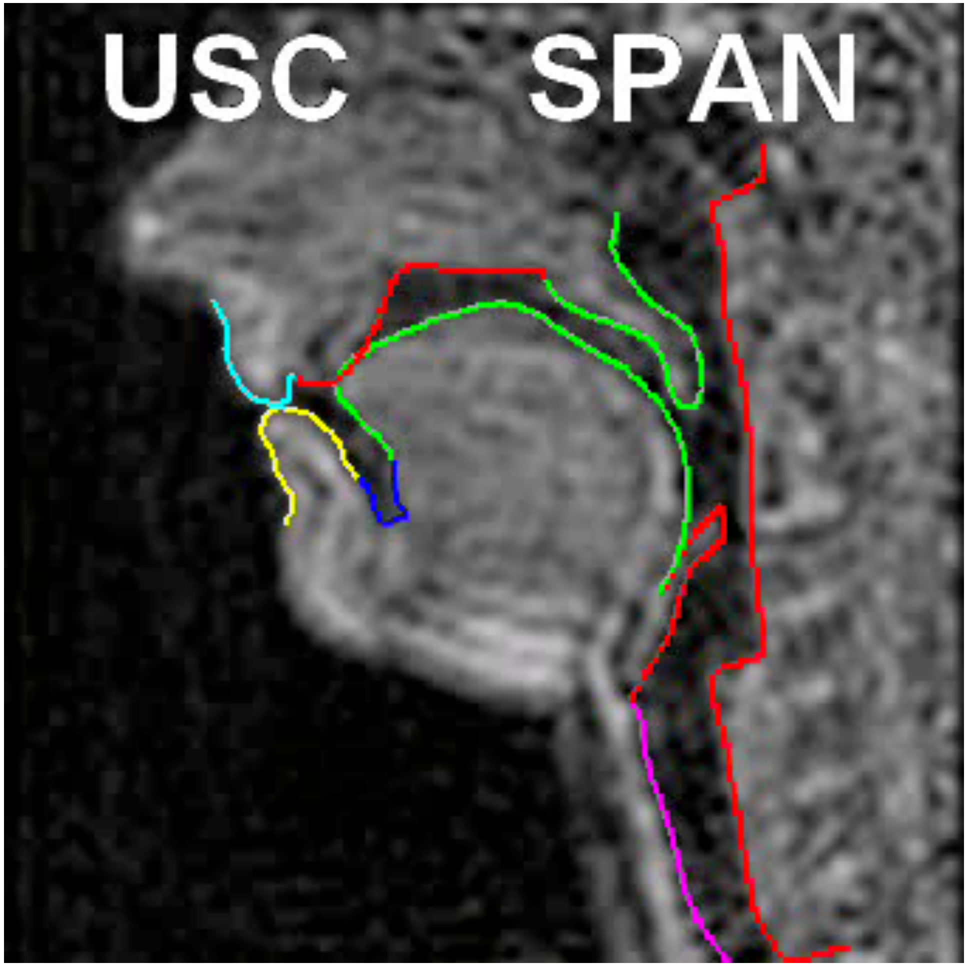




From Mark Liberman's Web Site, from Language Files

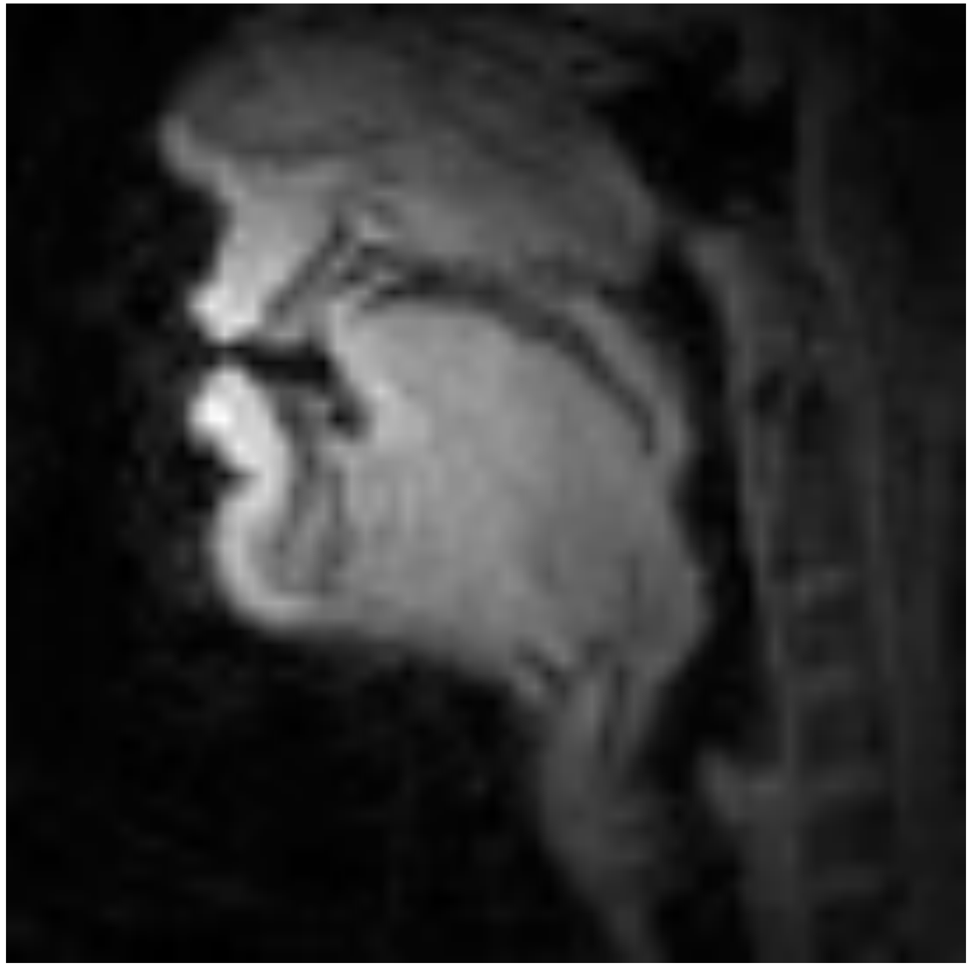
USC's SAIL Lab

Shri Narayanan



Female Singing

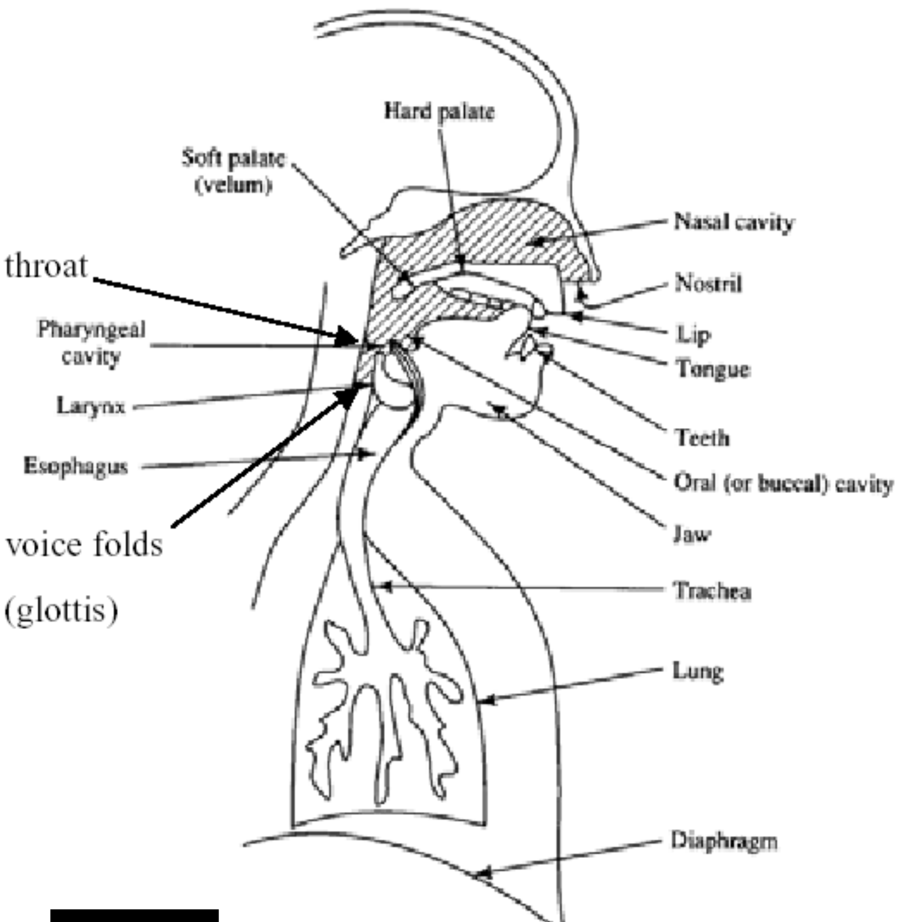
Tamil

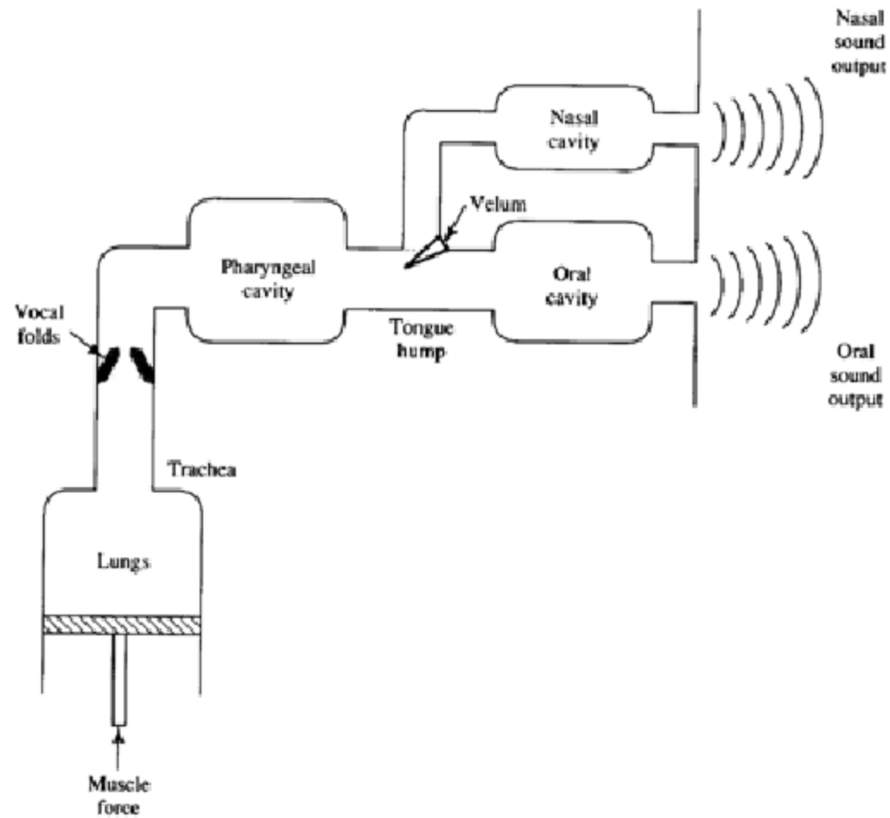


Singing in the MRI with Tyley Ross - Making the Voice Visible

- https://youtu.be/J3TwTb-T044

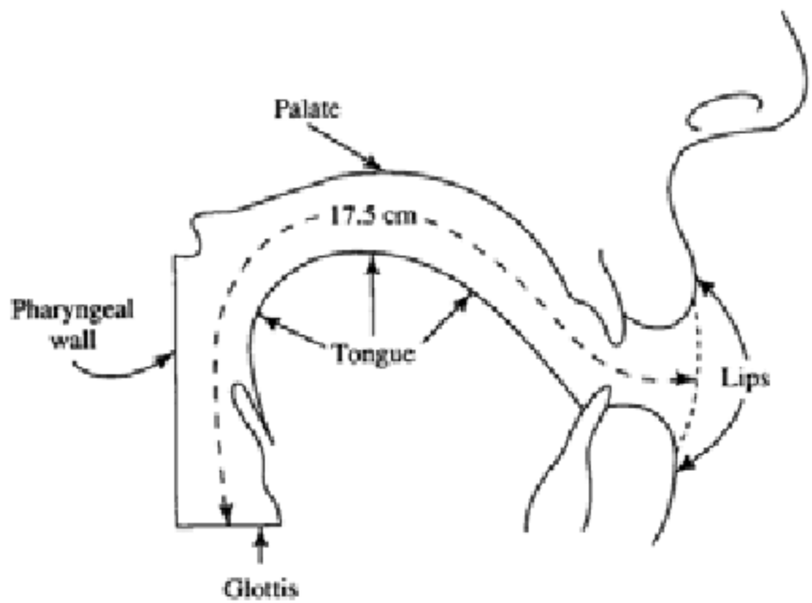
Human Speech Production

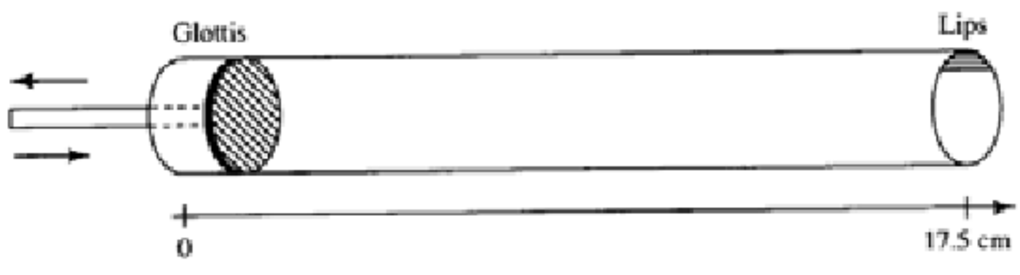




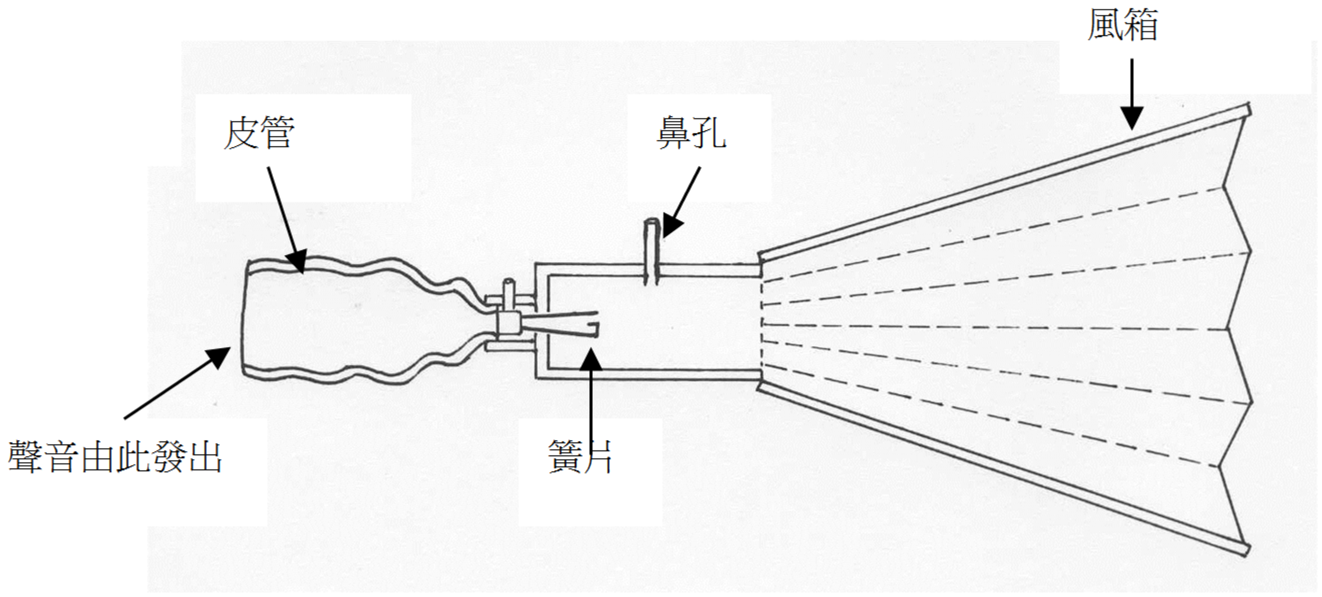
[redacted] vocal + pharyngeal cavities = vocal tract [redacted]

The Vocal Tract





Speaking Machine



Signal in Larynx

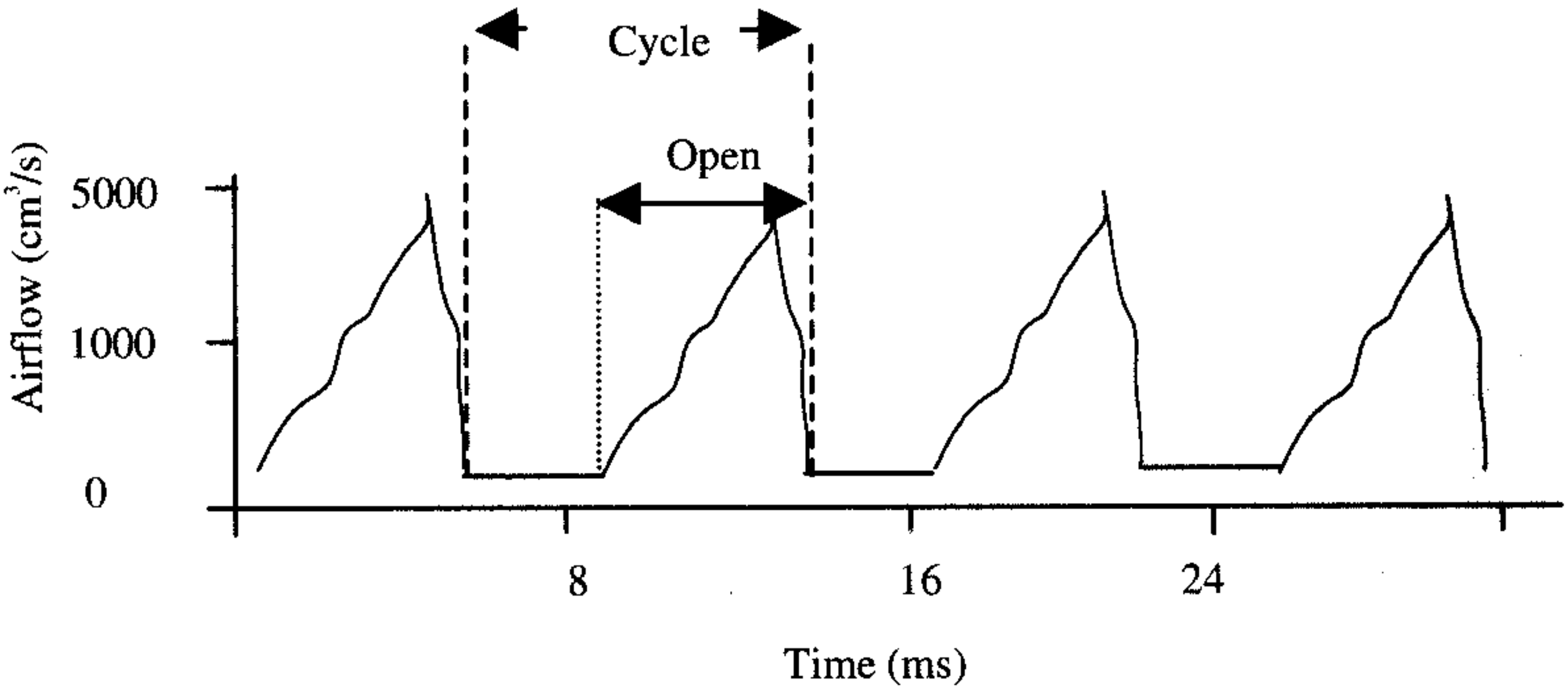
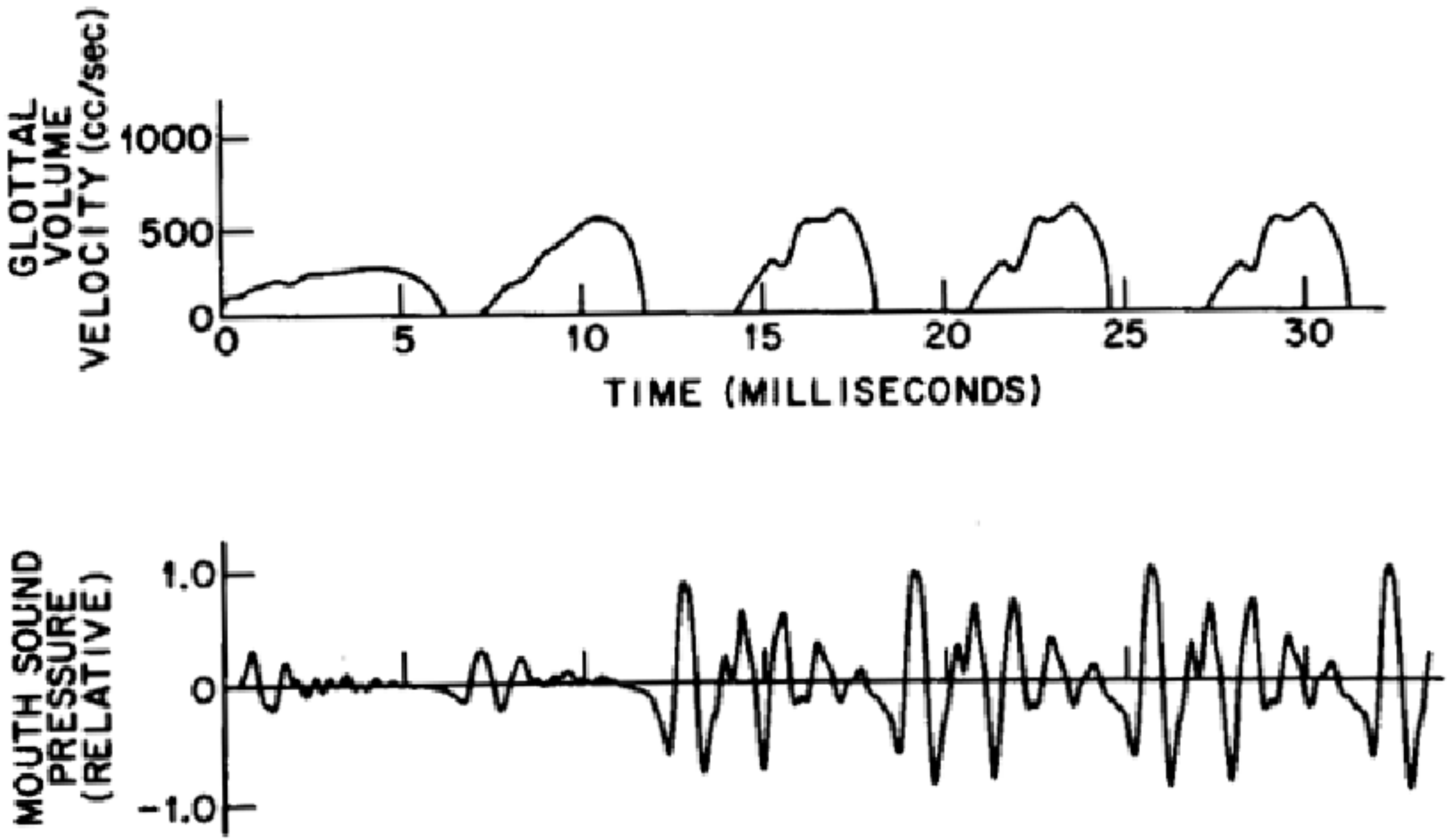


Figure 2.7 Waveform showing air flow during laryngeal cycle.

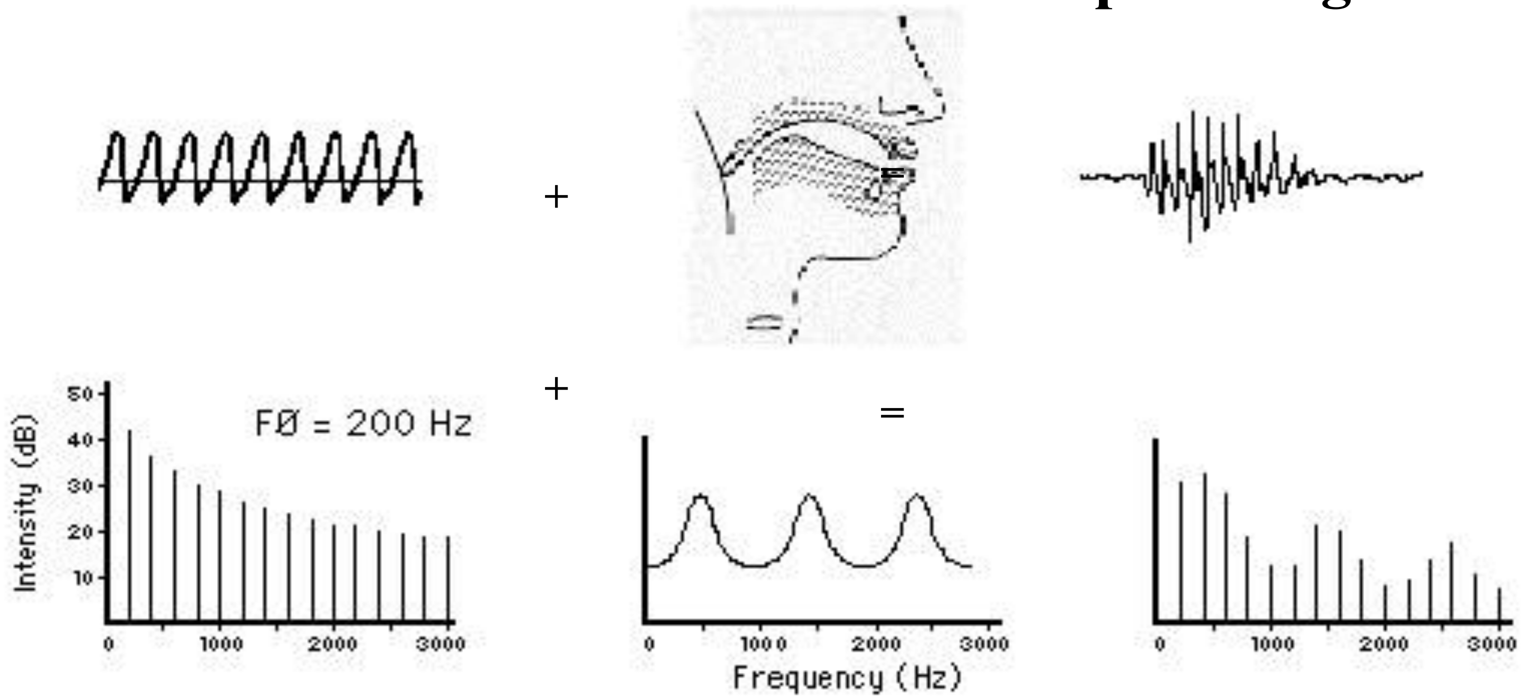
from Spoken Language Processing, Raj Reddy

Glottal Volume Velocity & Resulting Sound Pressure (Voiced)



Speech Production

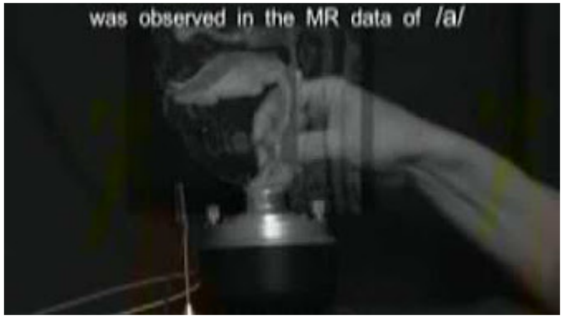
Glottal Pulses **Vocal Tract** **Speech Signal**



(a) Source Spectrum **(b) Filter Function** **(c) Output Energy Spectrum**

Vocal Tract

- Vocal tract – Real time MRI
  - https://www.youtube.com/watch?v=IdtkOyg9QIY
- Vocal Cord Stroboscopy
  - https://www.youtube.com/watch?v=hfOZxJnY4c8
- vocal tract model synthesis
  - https://www.youtube.com/watch?v=wR41CRbIjV4
  - https://www.youtube.com/watch?v=TVYYnJ_9258

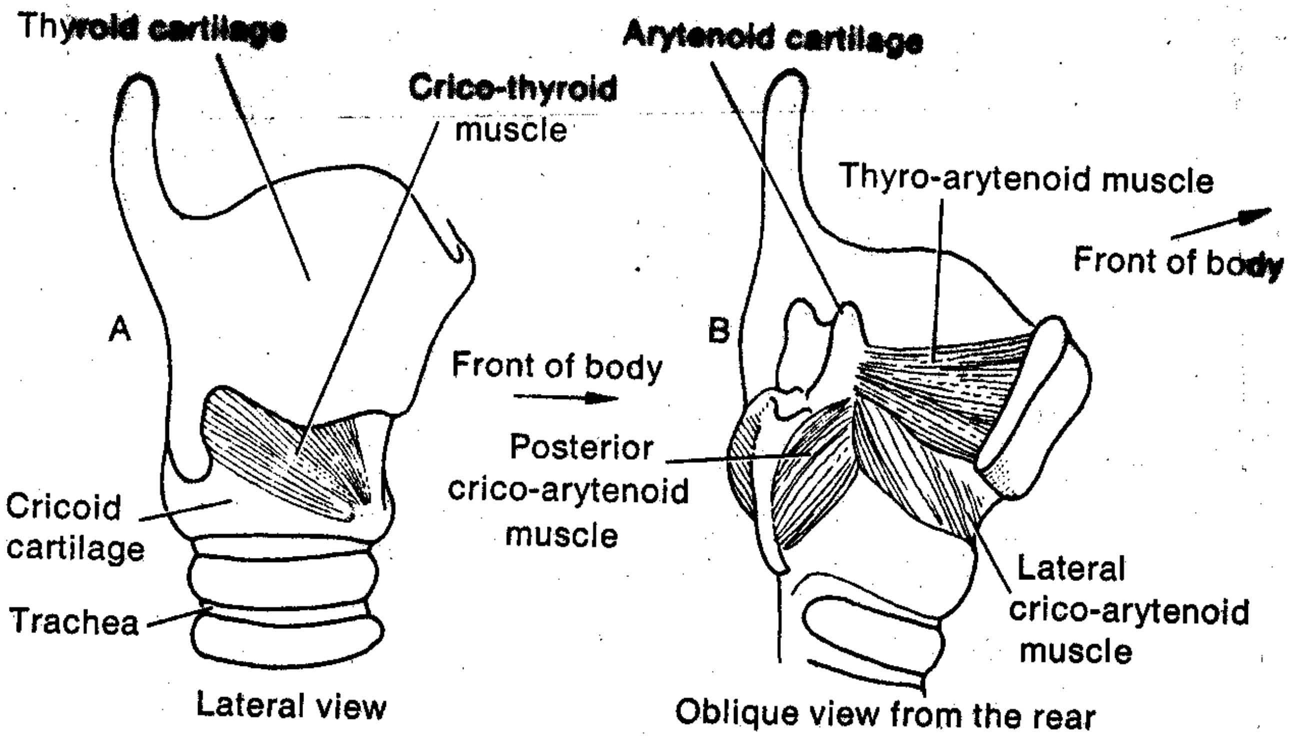


Larynx and Vocal Folds

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

Text from slides by Sharon Rose UCSD LING 11

Laryngeal setting: the Larynx



Cartilages and intrinsic muscles of the larynx.

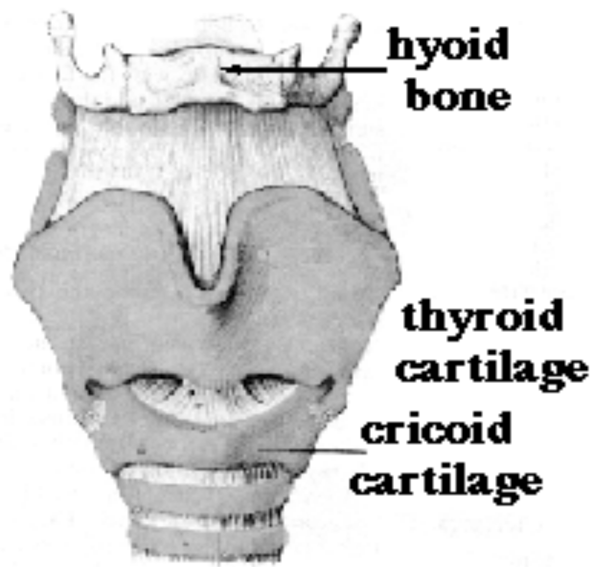


Figure thnx to John Coleman!!

Vertical slice through larynx, as seen from back

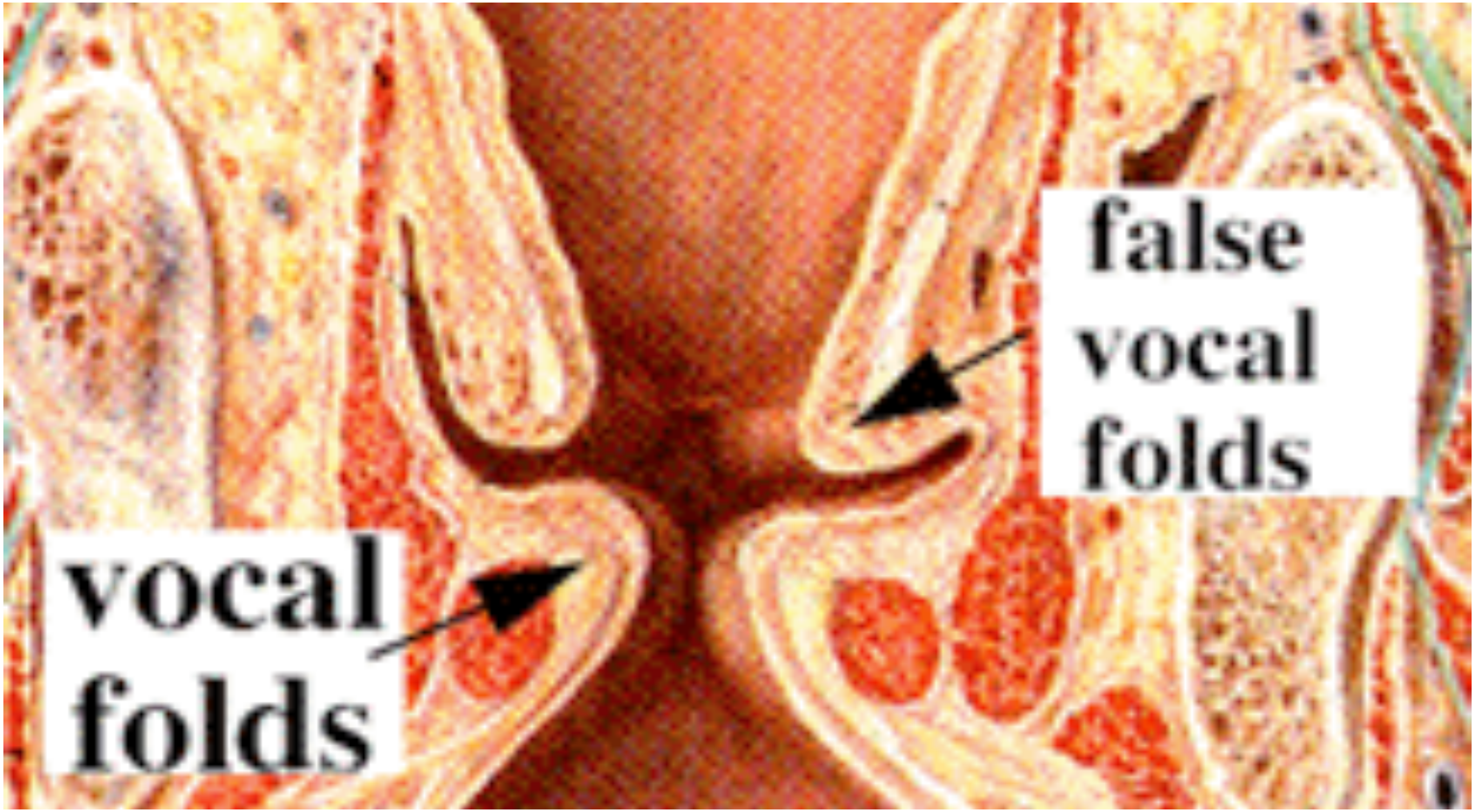
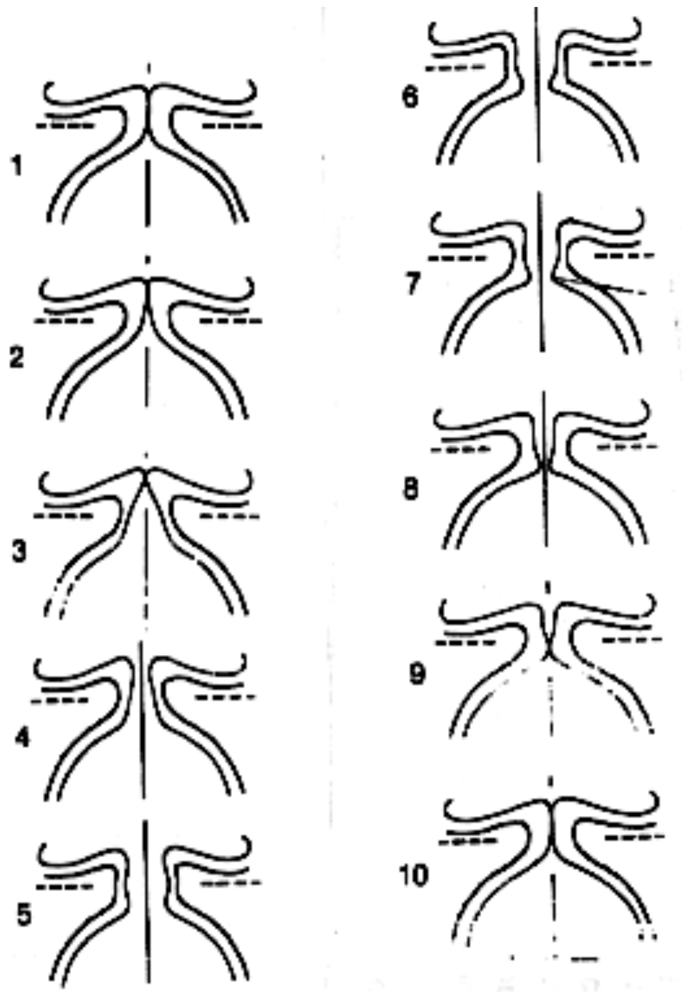


Figure thnx to John Coleman!!

Voicing:



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

Figure & text from John Coleman's web

Voicelessness

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

Consonants and Vowels

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

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

Text adapted from John Colema

Place of Articulation

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

Places of articulation

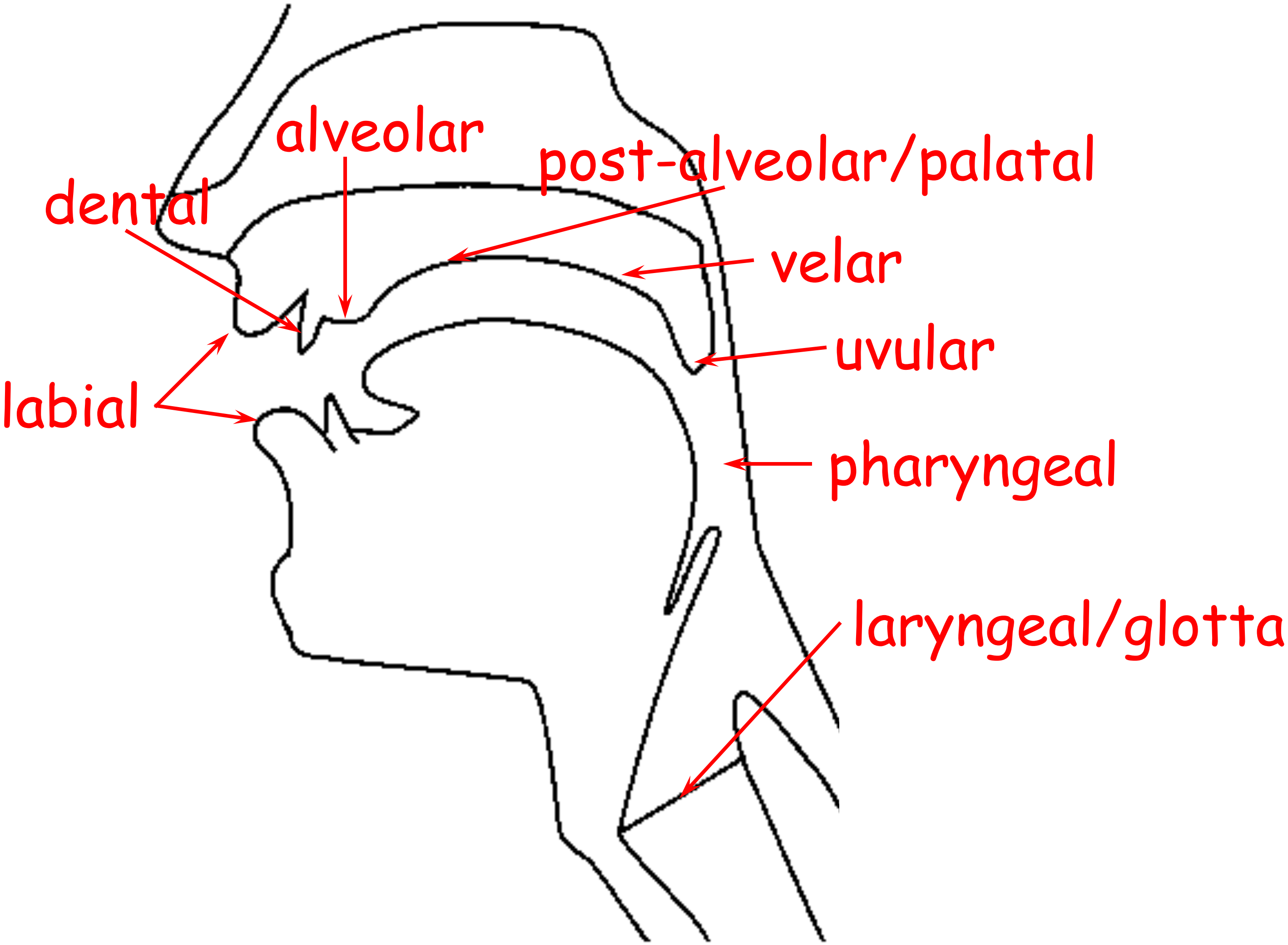
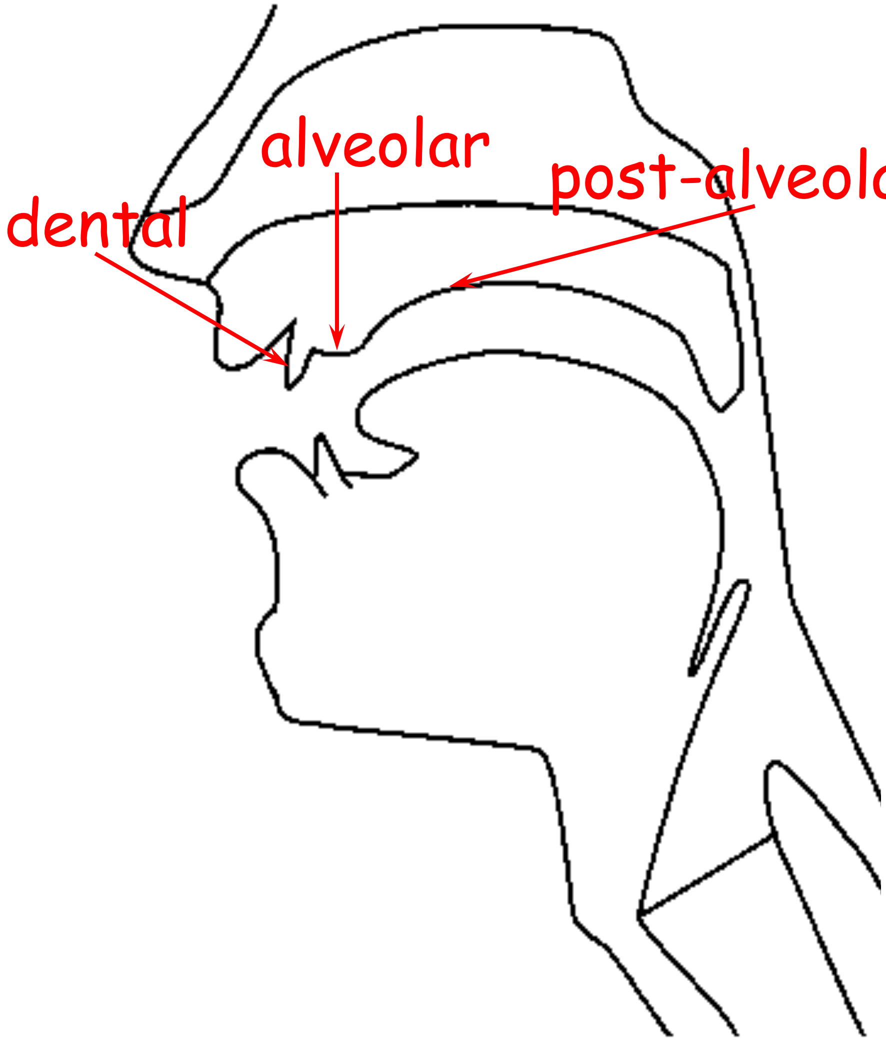


Figure thanks to Jennifer Venditti

Coronal place



Dental:

th/dh

Alveolar:

t/d/s/z/l

Post:

sh/zh/y

Figure thanks to Jennifer Vend

Dorsal Place

Velar:

k/g/ng

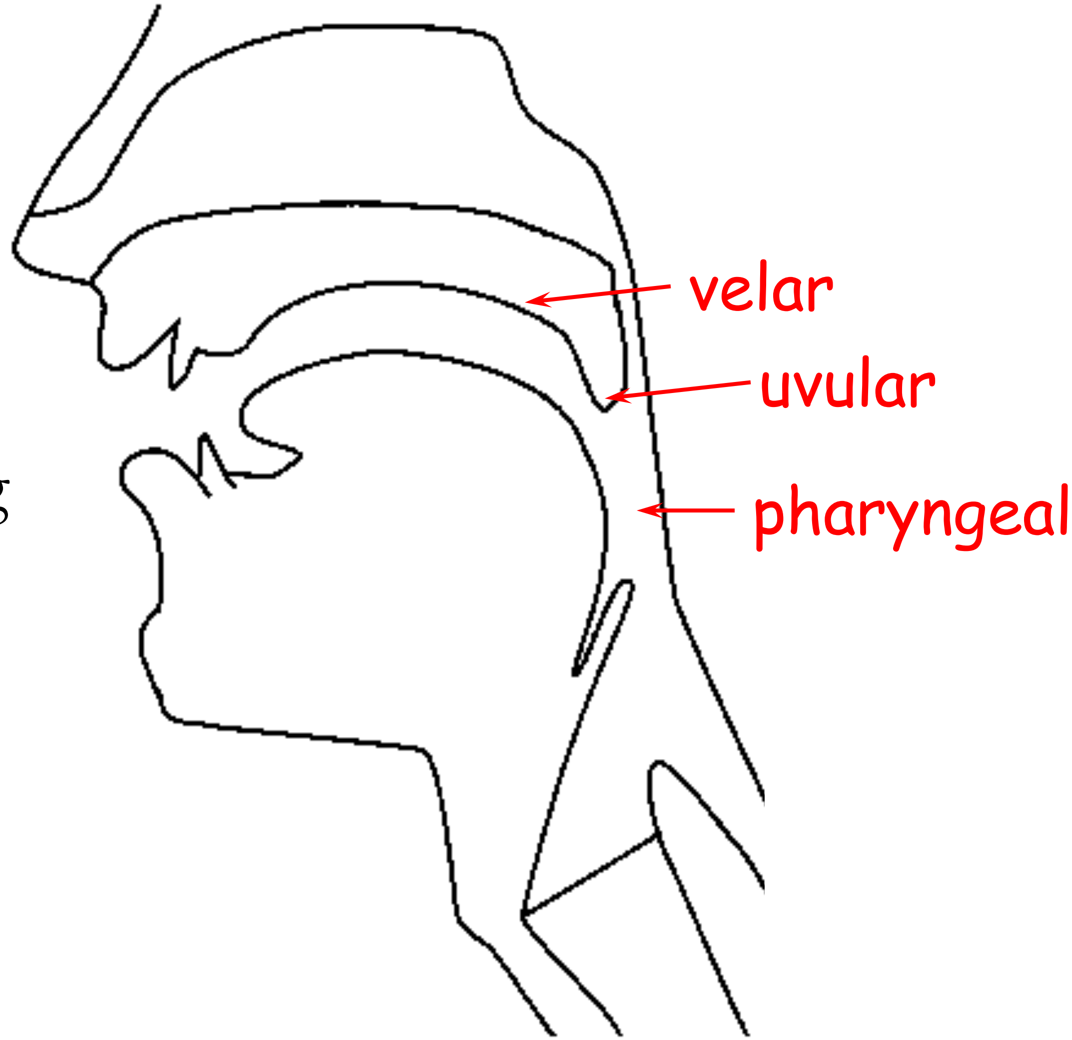
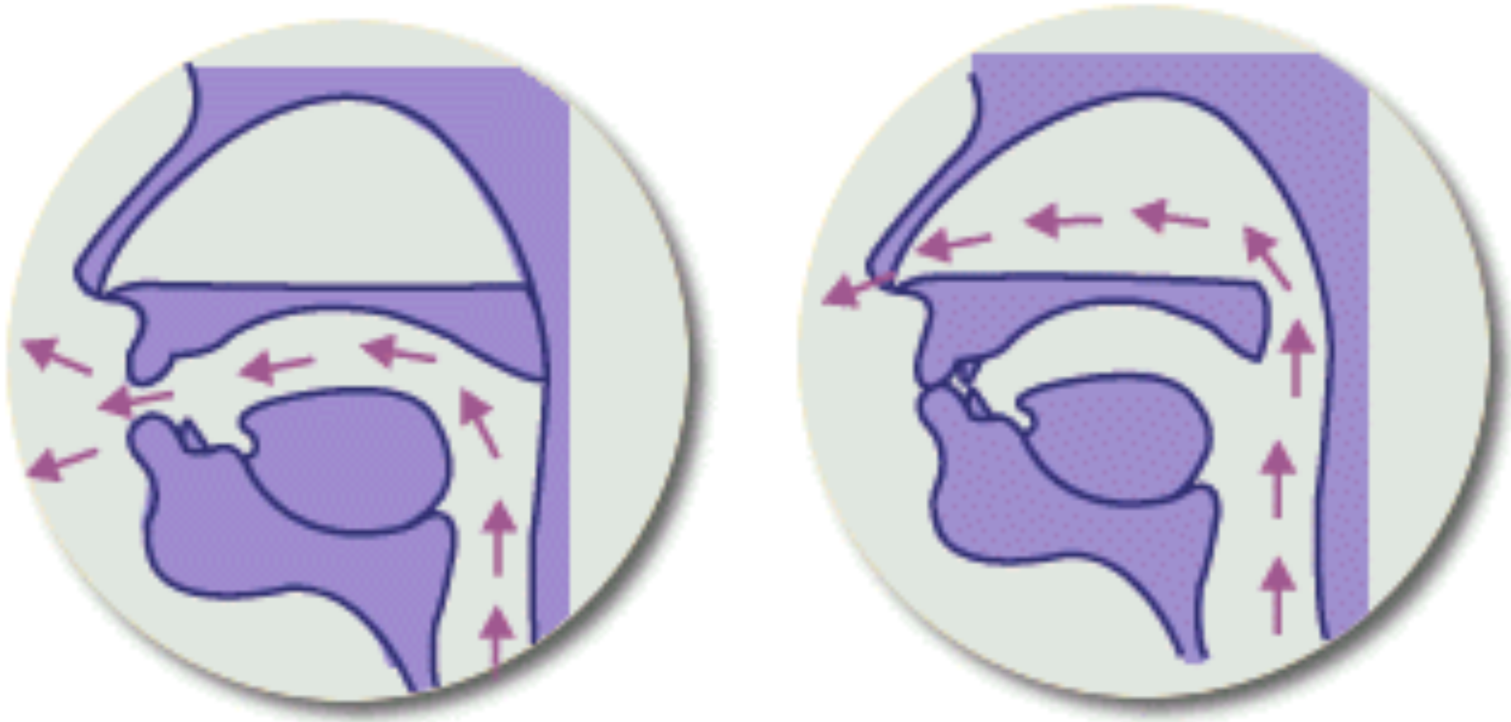


Figure thanks to Jennifer Vend

Manner of Articulation

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

Oral vs. Nasal Sounds



Thanks to Jong-bok Kim for this f

More on Manner of articulation of consonants

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

Text from Ladefoged "A Course in Ph

More on manner of articulation of consonants

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

Articulatory parameters for English consonants (in ARPAbet)

PLACE OF ARTICULATION

MANNER OF ARTICULATION

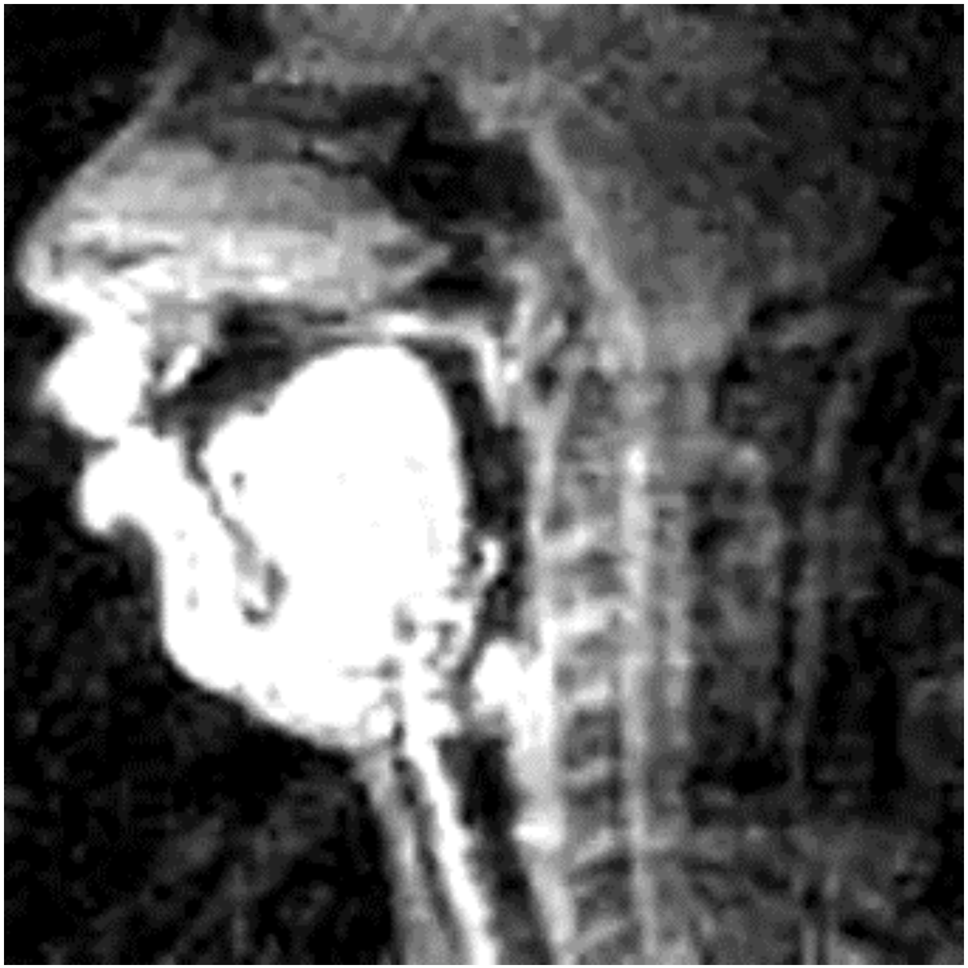
	bilabial		labio-dental		inter-dental		alveolar		palatal		velar		glottal	
stop	p	b					t	d			k	g	q	
fric.			f	v	th	dh	s	z	sh	zh			h	
affric.									ch	jh				
nasal		m						n				ng		
approx		w						l/r		y				
flap							dx							

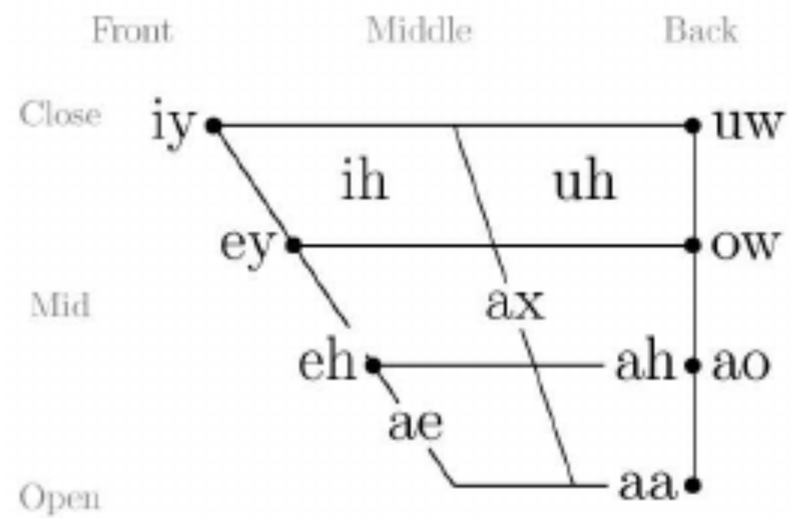
VOICING:

voiceless	voiced

Table from Jennifer Venditti

Tongue position for vowels





Vowels

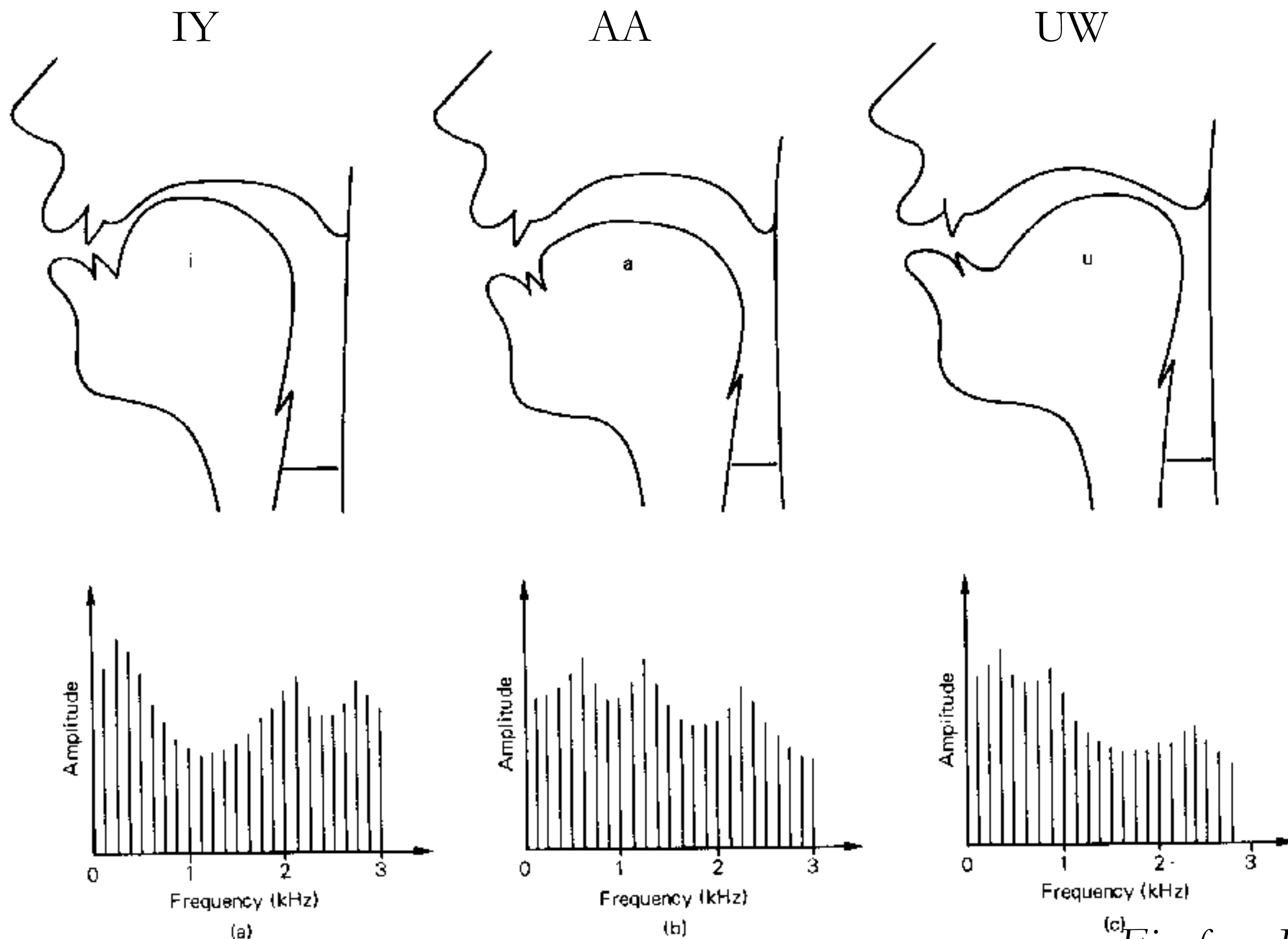
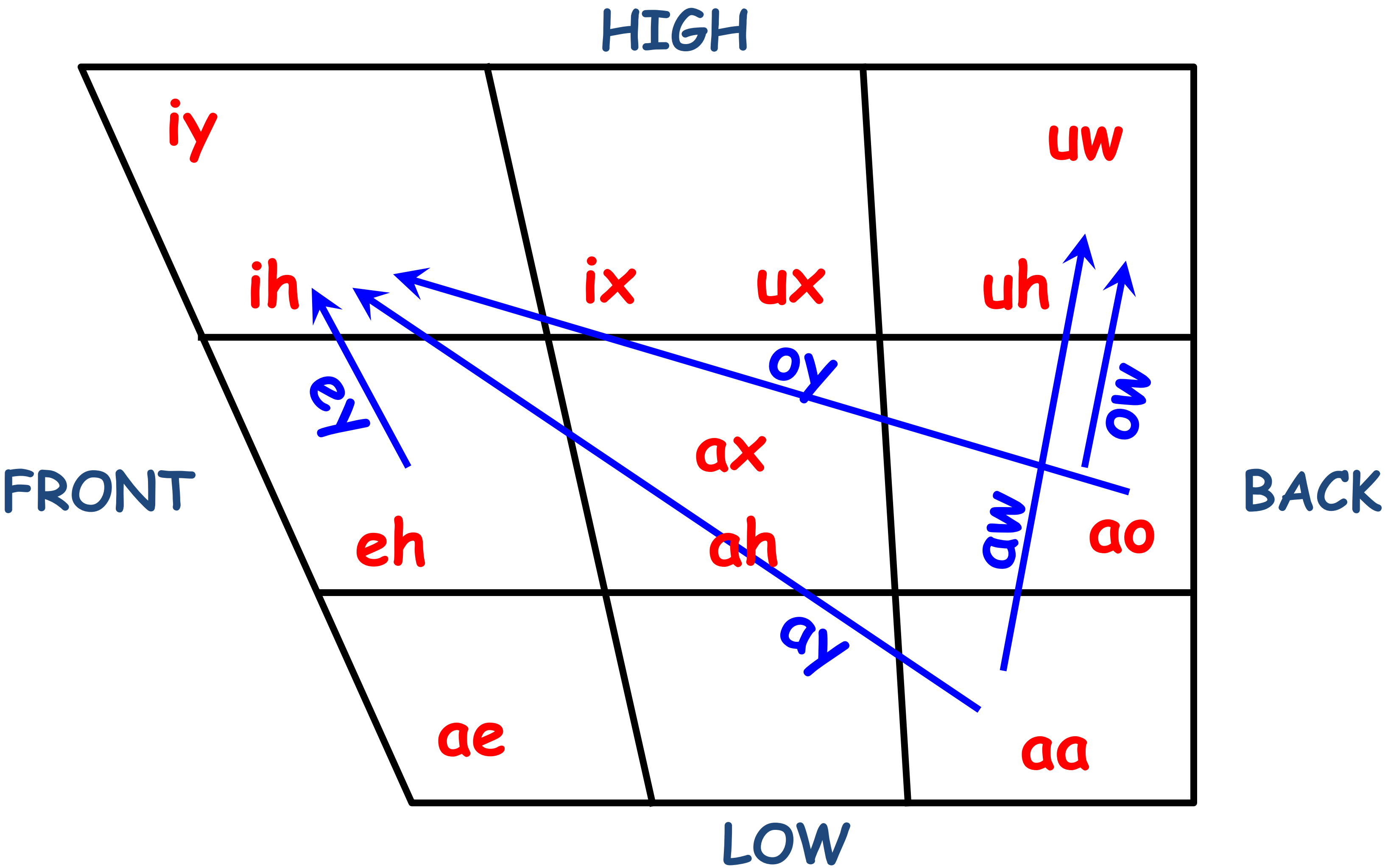


Fig. from Eric Keller

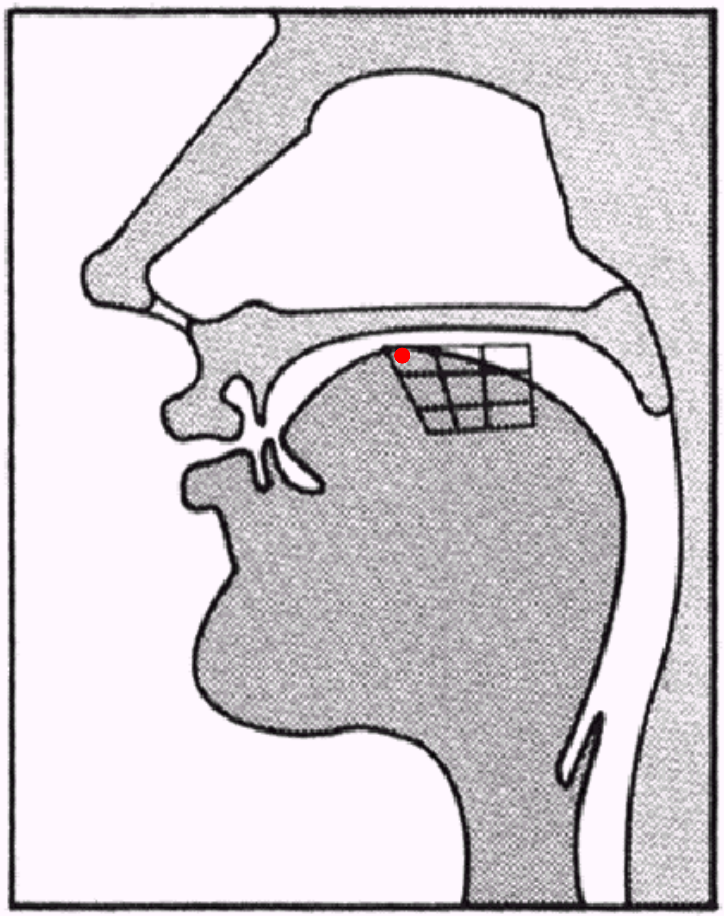
American English Vowel Space



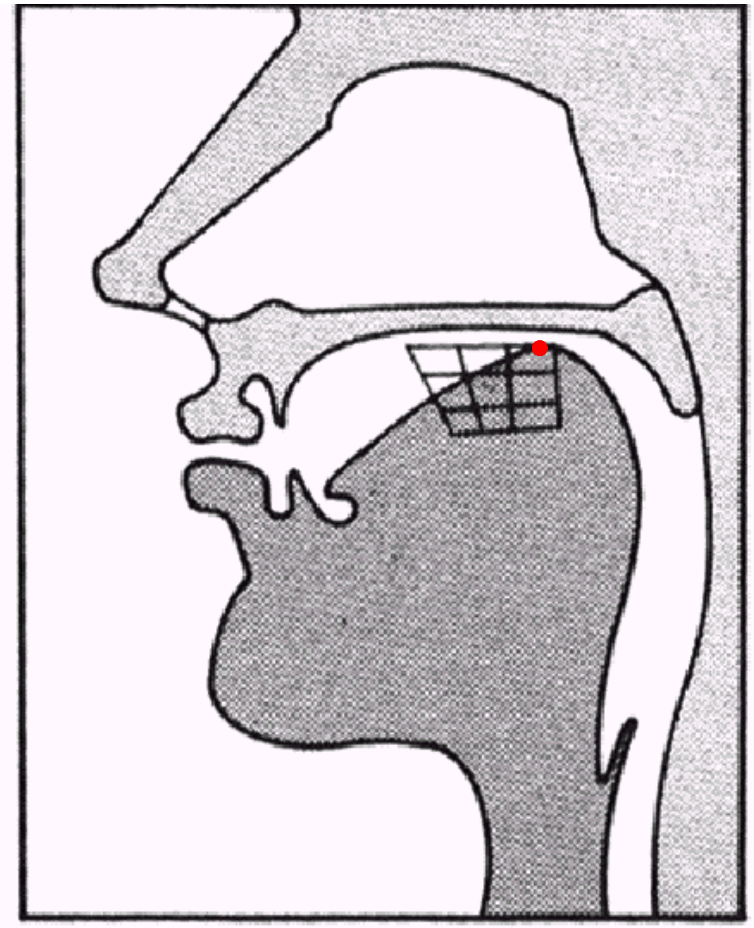
Red: Vowels, Blue: Dipthongs

Figure from Jennifer Venditti

[iy] vs. [uw]



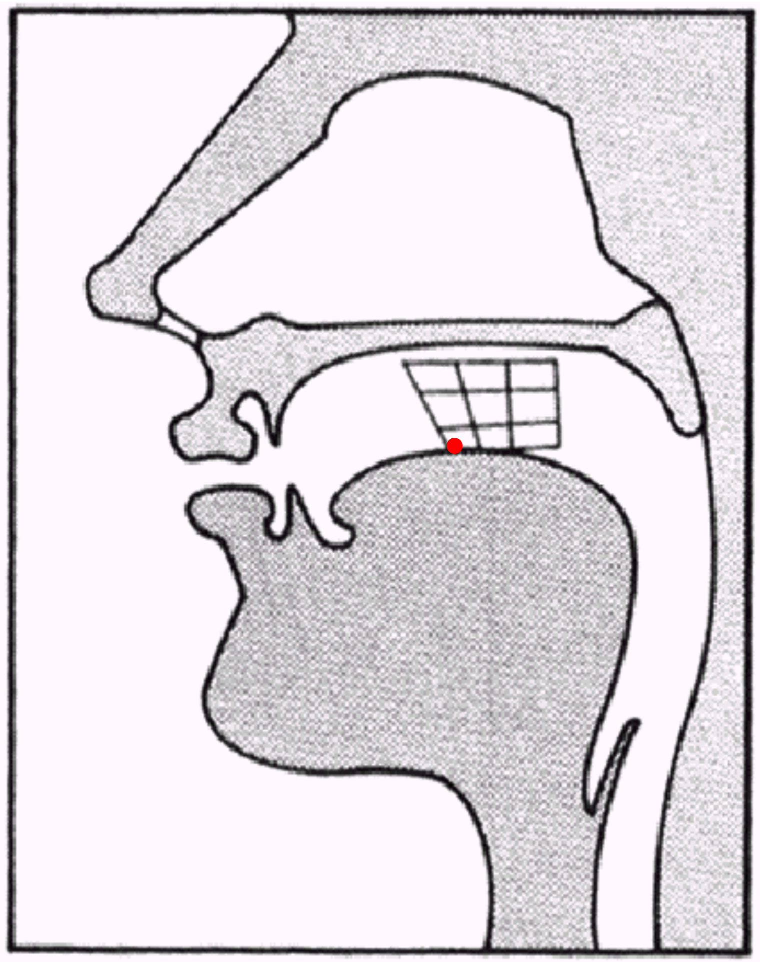
/i/



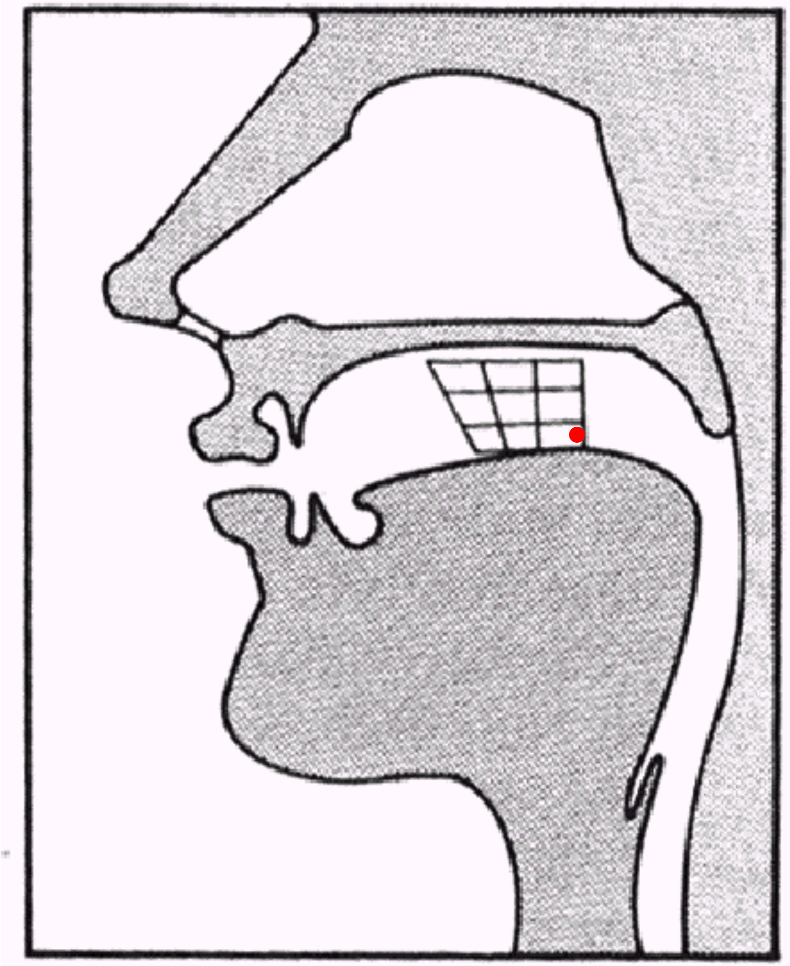
/u/

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

[ae] vs. [aa]



/æ/



/ɑ/

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

IPA

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2018)

CONSONANTS (PULMONIC)

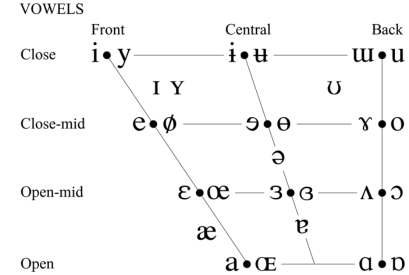
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant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

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

CONSONANTS (NON-PULMONIC)

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

VOWELS



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

OTHER SYMBOLS

- ʍ Voiceless labial-velar fricative
- w Voiced labial-velar approximant
- ɥ Voiced labial-palatal approximant
- ʜ Voiceless epiglottal fricative
- ʢ Voiced epiglottal fricative
- ʡ Epiglottal plosive
- ɕ ʑ Alveolo-palatal fricatives
- ɺ Voiced alveolar lateral flap
- ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. t͡s k͡p

DIACRITICS Some diacritics may be placed above a symbol with a descender, e.g. ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼	◌̻	Laminal	t̻ d̻
◌̹	More rounded	ɔ̹	ʷ	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̜	Less rounded	ɔ̜	ʲ	Palatalized	tʲ dʲ	ⁿ	Nasal release	dⁿ
◌̟	Advanced	u̟	ˠ	Velarized	tˠ dˠ	ˡ	Lateral release	dˡ
◌̠	Retracted	e̠	ˤ	Pharyngealized	tˤ dˤ	˺	No audible release	d̚
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ			
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝ (ɹ̝ = voiced alveolar fricative)			
◌̩	Syllabic	n̩	◌̞	Lowered	e̞ (β̞ = voiced bilabial approximant)			
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘			
˞	Rhoticity	ɚ a˞	◌̙	Retracted Tongue Root	e̙			

SUPRASEGMENTALS

- ˈ Primary stress ˌfoʊnəˈtɪʃən
- ˌ Secondary stress
- ː Long eː
- ˑ Half-long eˑ
- ◌̆ Extra-short ĕ
- | Minor (foot) group
- ‖ Major (intonation) group
- . Syllable break ɹi.ækt
- ‿ Linking (absence of a break)

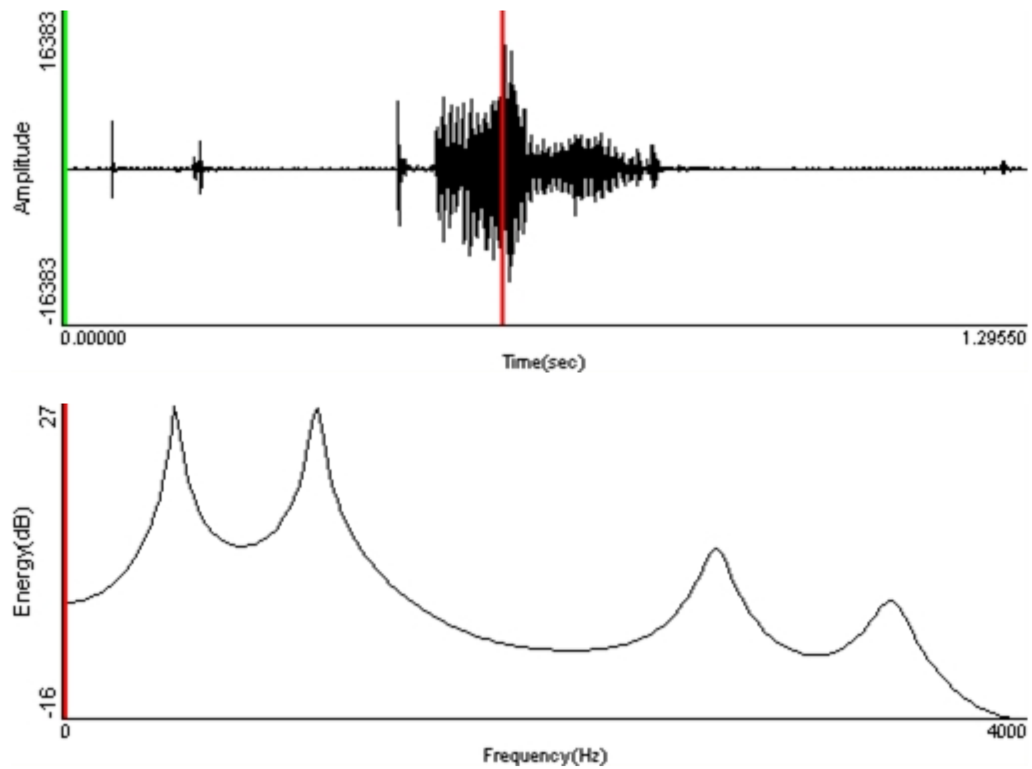
TONES AND WORD ACCENTS

LEVEL			CONTOUR		
e̋ or	˥	Extra high	ě or	˩˥	Rising
é	˦	High	ê	˥˩	Falling
ē	˧	Mid	e᷄	˧˥	High rising
è	˨	Low	e᷅	˩˧	Low rising
ȅ	˩	Extra low	e᷈	˧˦˧	Rising-falling
ꜜ		Downstep	↗		Global rise
ꜛ		Upstep	↘		Global fall

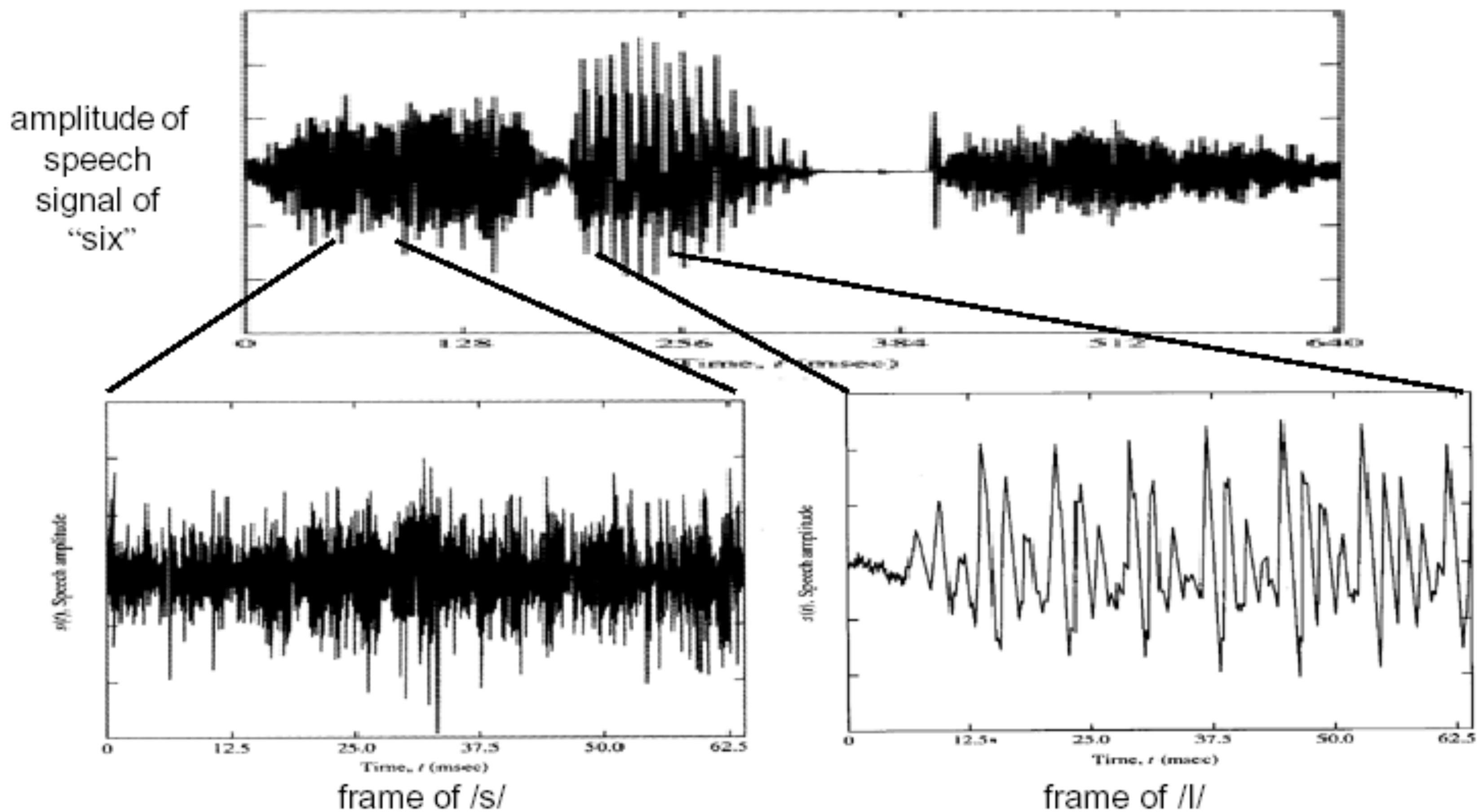
Typefaces: Doulos SIL (metatext); Doulos SIL, IPA Kiel, IPA LS Uni (symbols)

聲學語音學

聲音波形及頻譜



Acoustical Analysis
(speech signal of “six”)



Spectrogram of the Speech Waveform "sees"

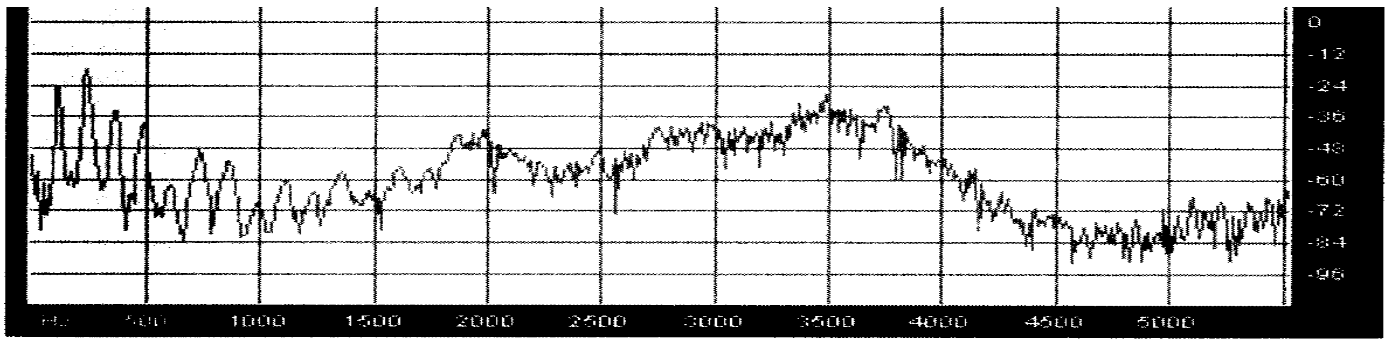
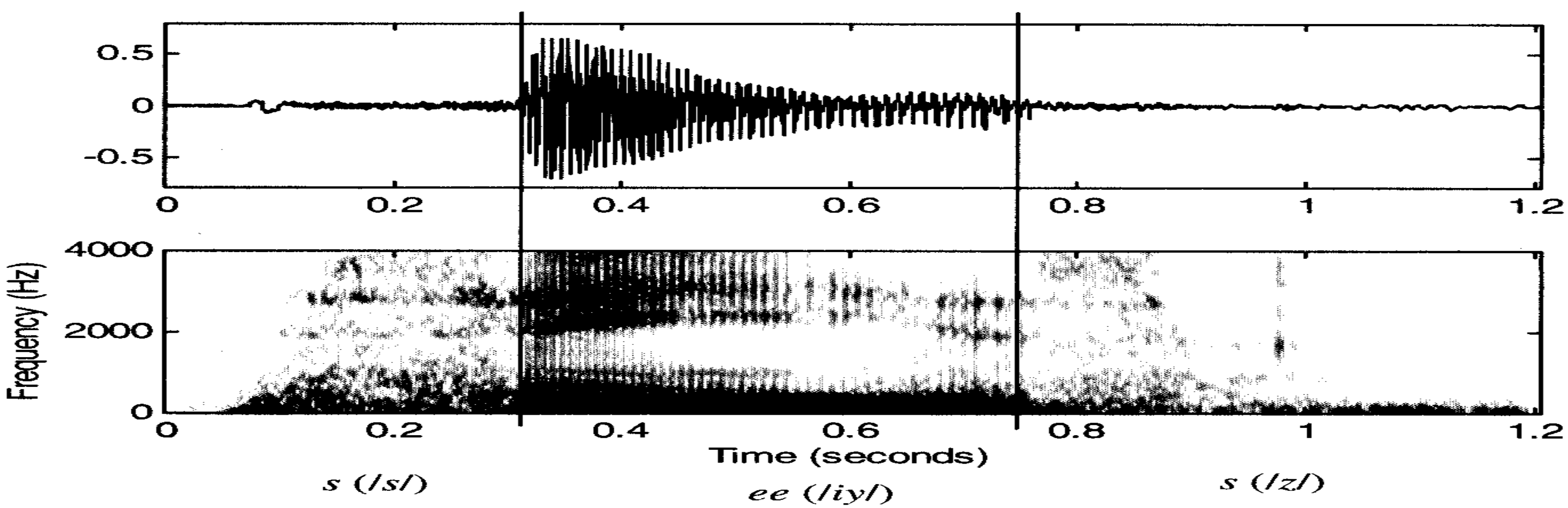
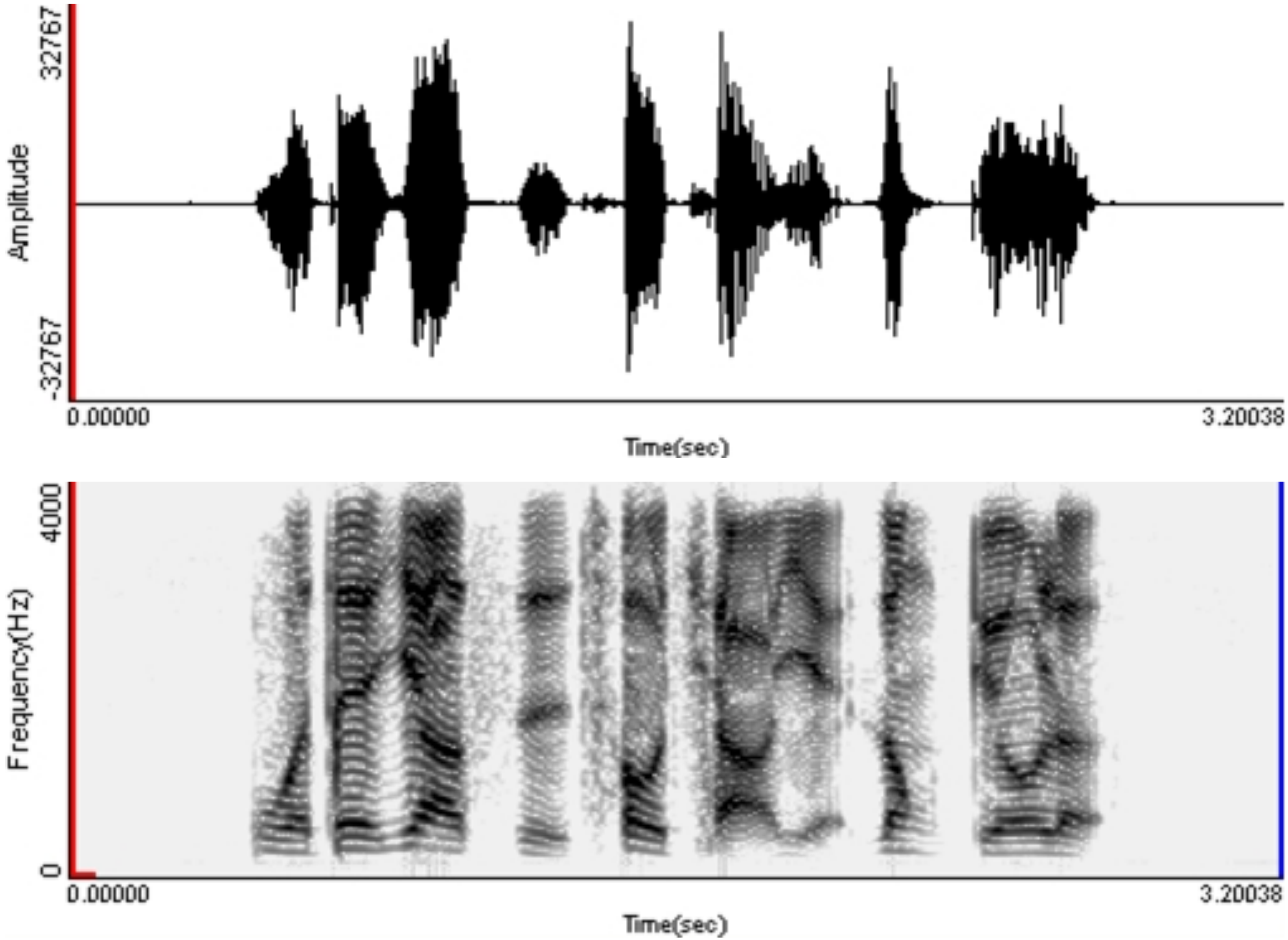


Figure 2.8 **A spectral analysis of the vowel** ***/iy/*****, showing characteristically uneven distribution of energy at different frequencies.**

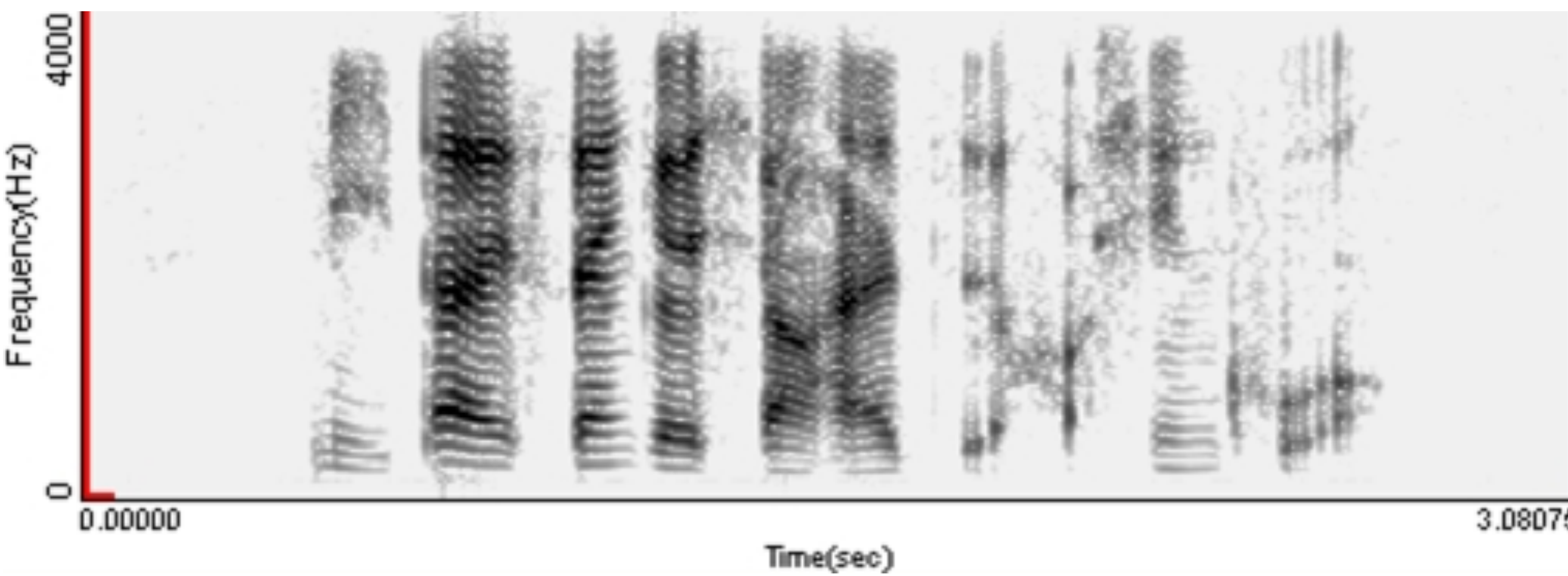


聲音的聲譜圖


32767
Amplitude
-32767
0.00000
3.20038
Time(sec)
4000
Frequency(Hz)
0
0.00000
3.20038
Time(sec)

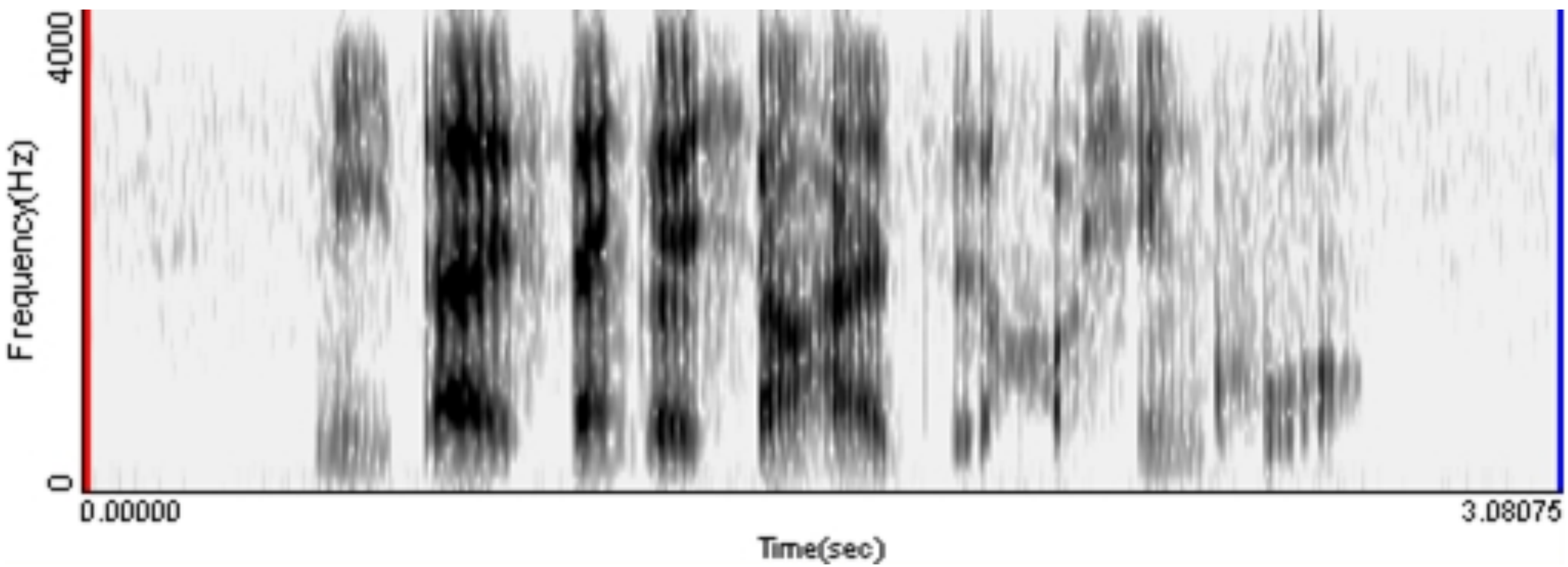
窄頻帶的聲譜圖(narrowband spectrograms)

聲譜圖的產生是用富利葉轉換(Fourier transform)，當我們用較長的分析視窗(analysis windows)，約20ms，對應頻寬約為45 Hz，得到的頻率解析度較高，頻譜上可以看到諧振的成分。在聲譜圖上呈現等距的黑白相間橫線條，其間距就是基頻(F0) 。

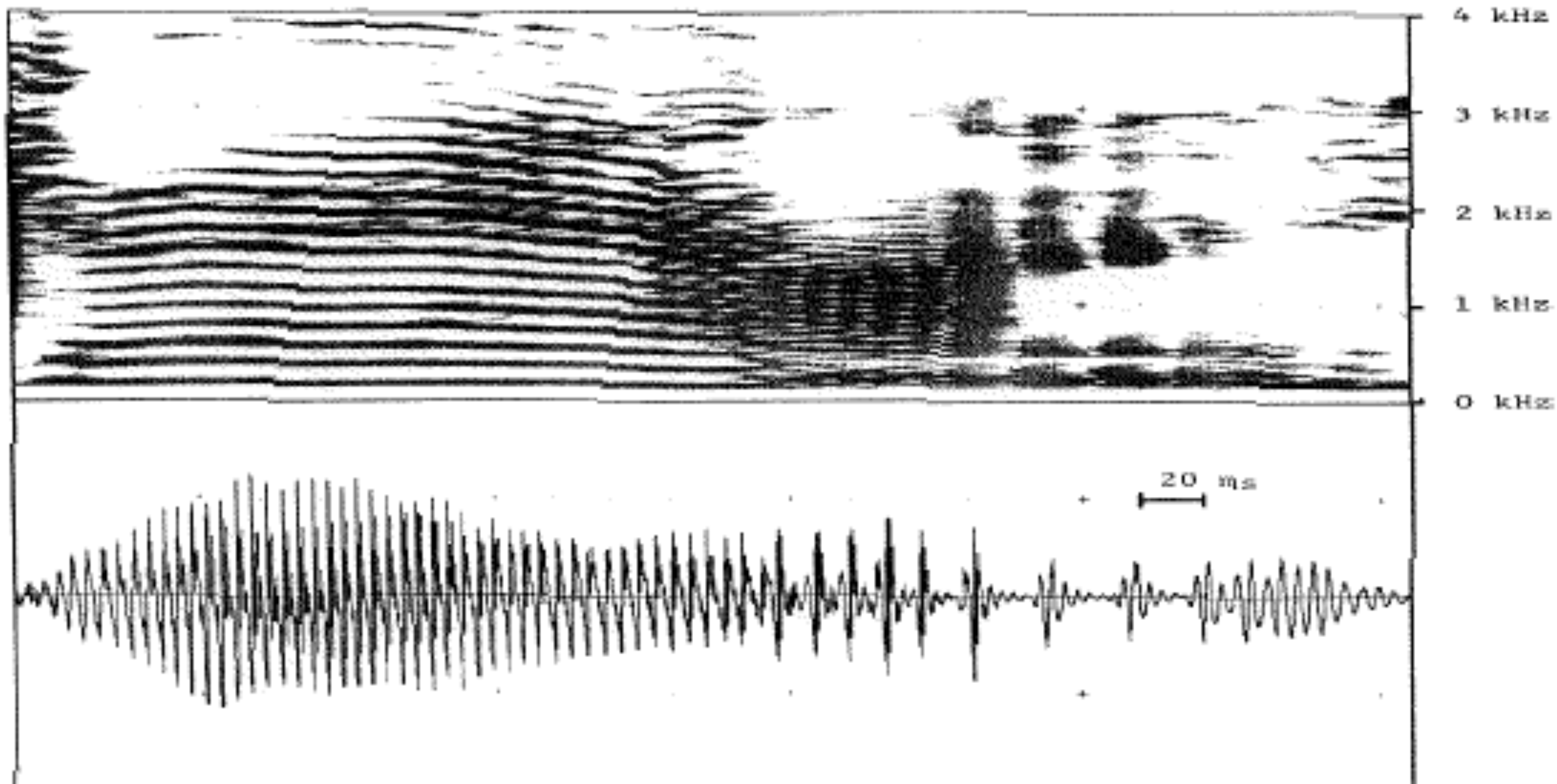


寬頻帶的聲譜圖(wideband spectrograms)

若是在轉換演算時用較少的取樣點，分析視窗大約3ms，對應頻寬約300 Hz，則頻譜上看不到諧振成分，在聲譜圖上看不到等距的黑白相間。頻率解析度較低，反而是時軸上的解析度較高，看到明顯的垂直線條。

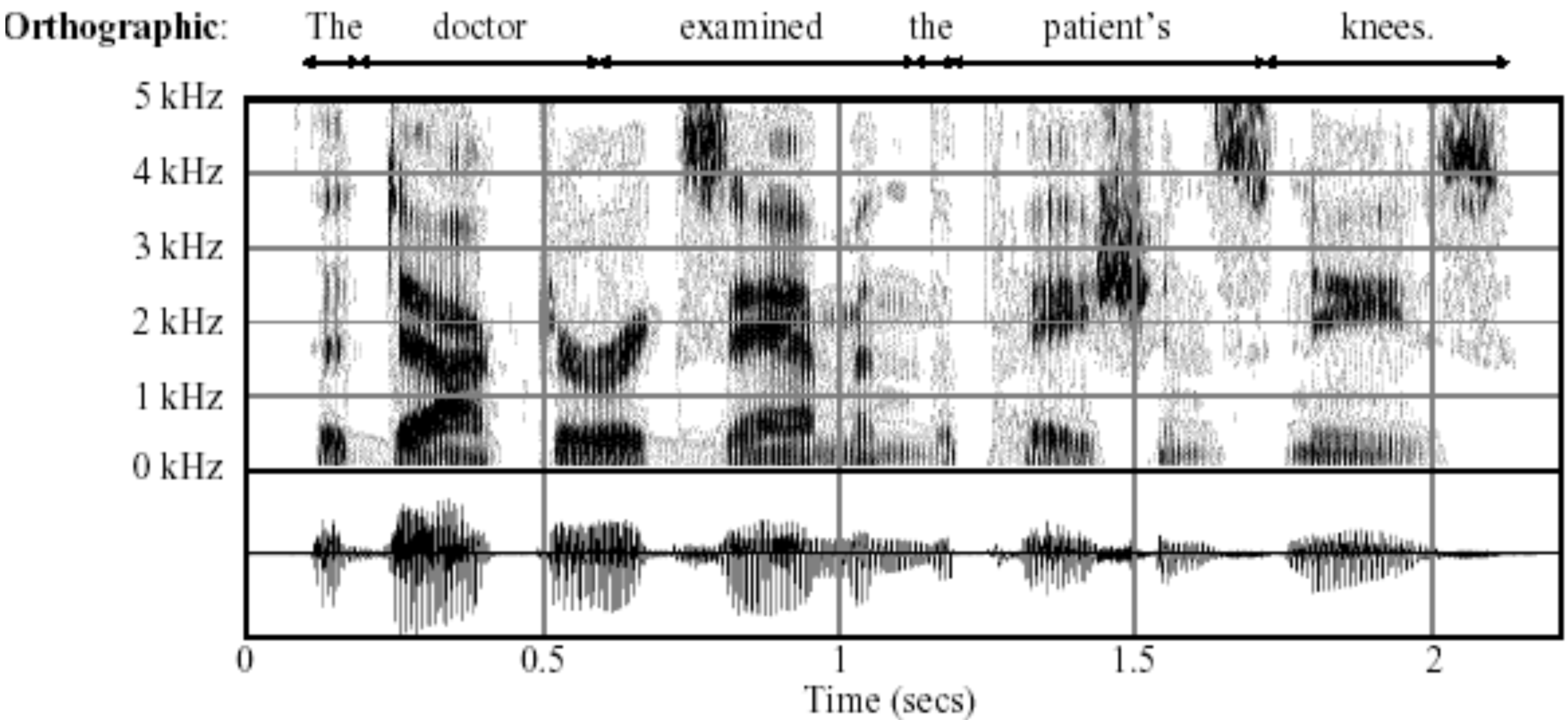


Narrowband Spectrogram



Wideband Spectrogram

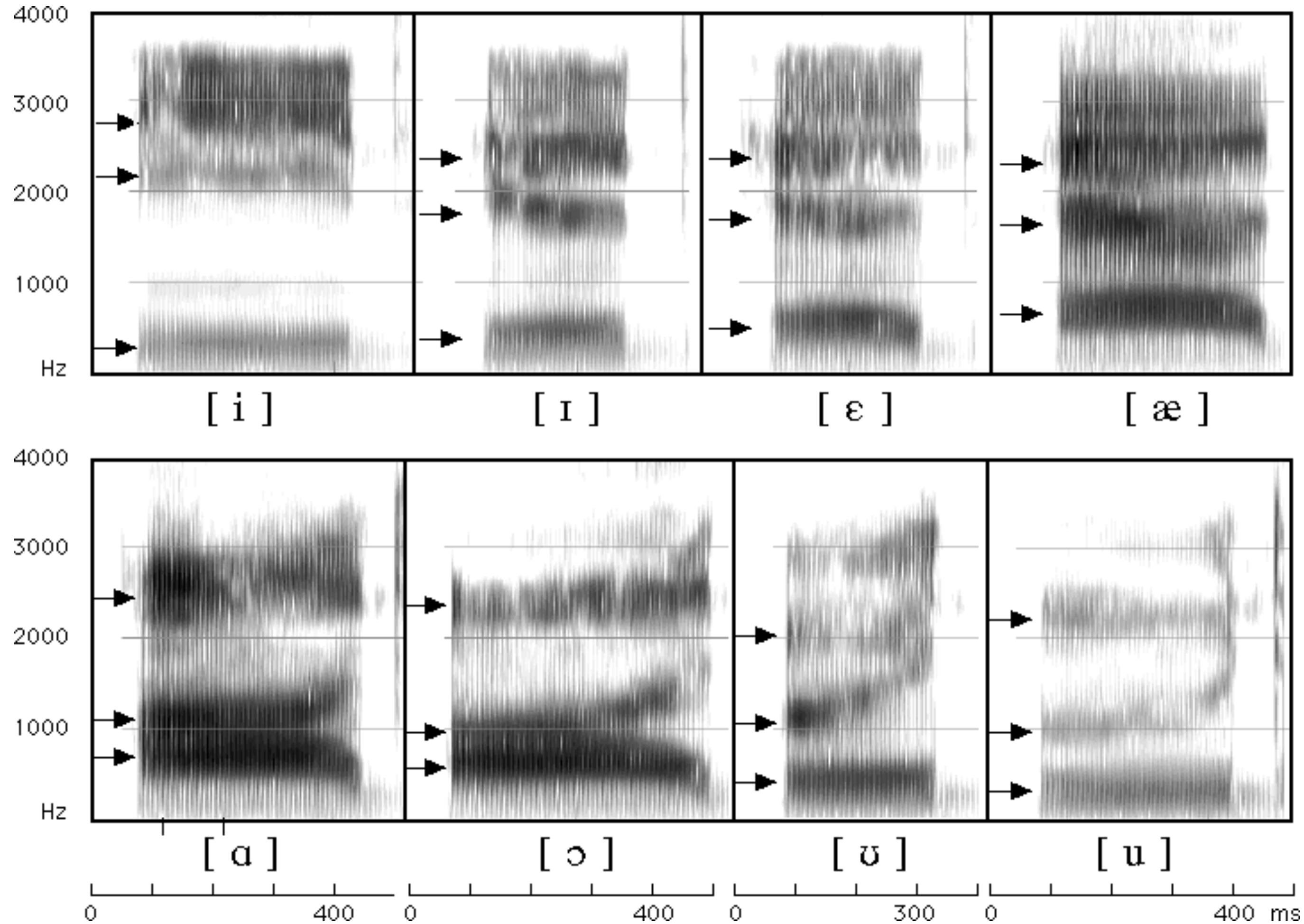
Standard wideband spectrogram ($f_s = 10\ kHz$, $T_w = 6\ ms$):



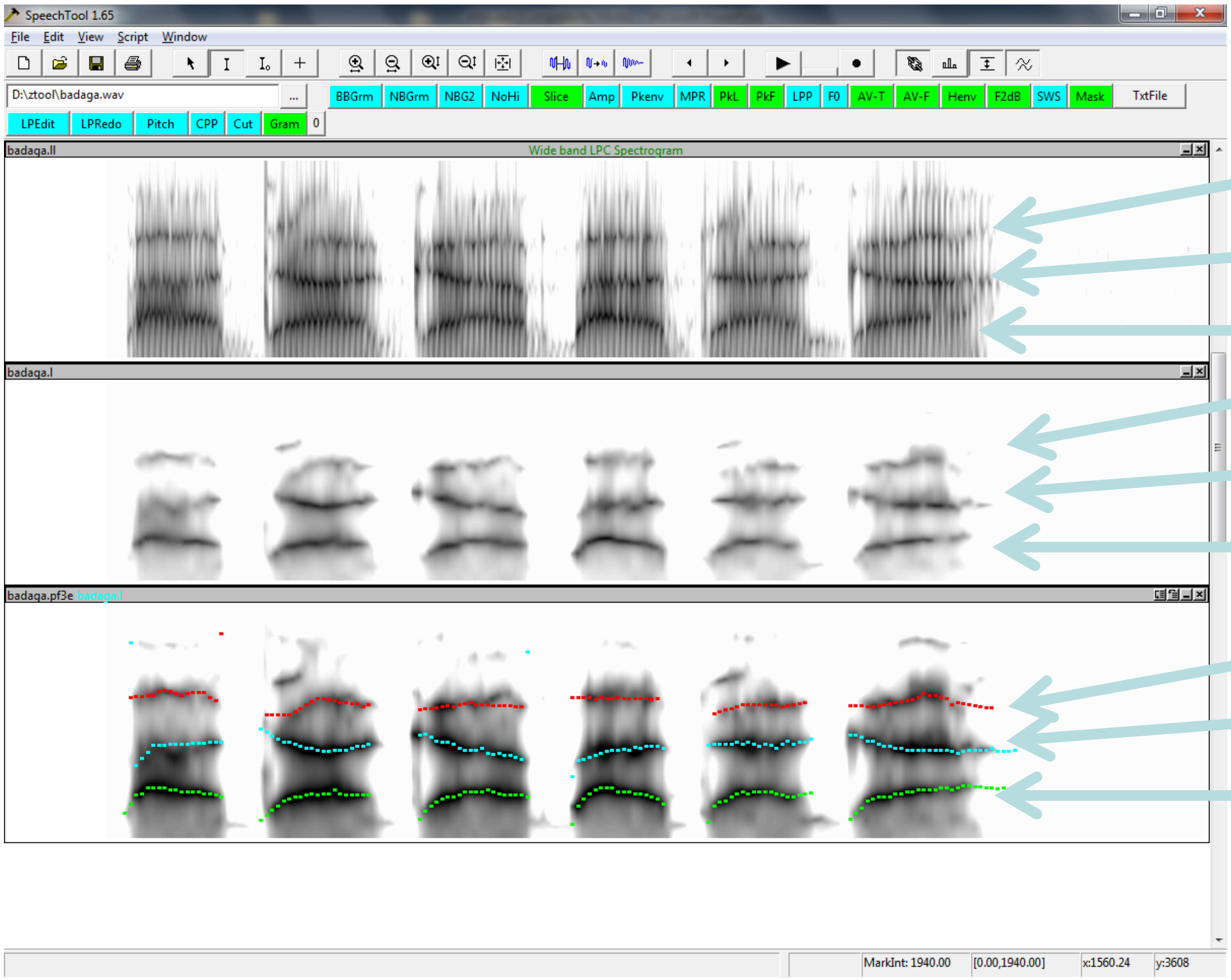
Formants

- Vowels largely distinguished by 2 characteristic pitches.
- One of them (the higher of the two) goes downward throughout the series iy ih eh ae aa ao ou u
- The other goes up for the first four vowels and then down for the next four.
- These are called "formants" of the vowels, lower is 1st formant, higher is 2nd formant.

Seeing formants: the spectrogram



What happens to the sound pattern when the vocal tract opens?



3rd formant (F_3)

2nd formant (F_2)

1st formant (F_1)

3rd formant (F_3)

2nd formant (F_2)

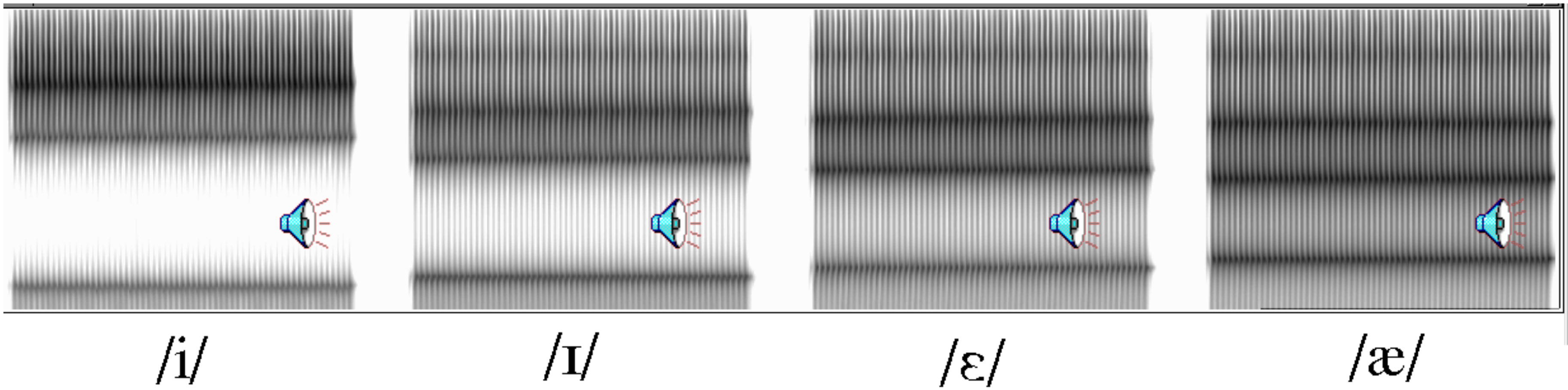
1st formant (F_1)

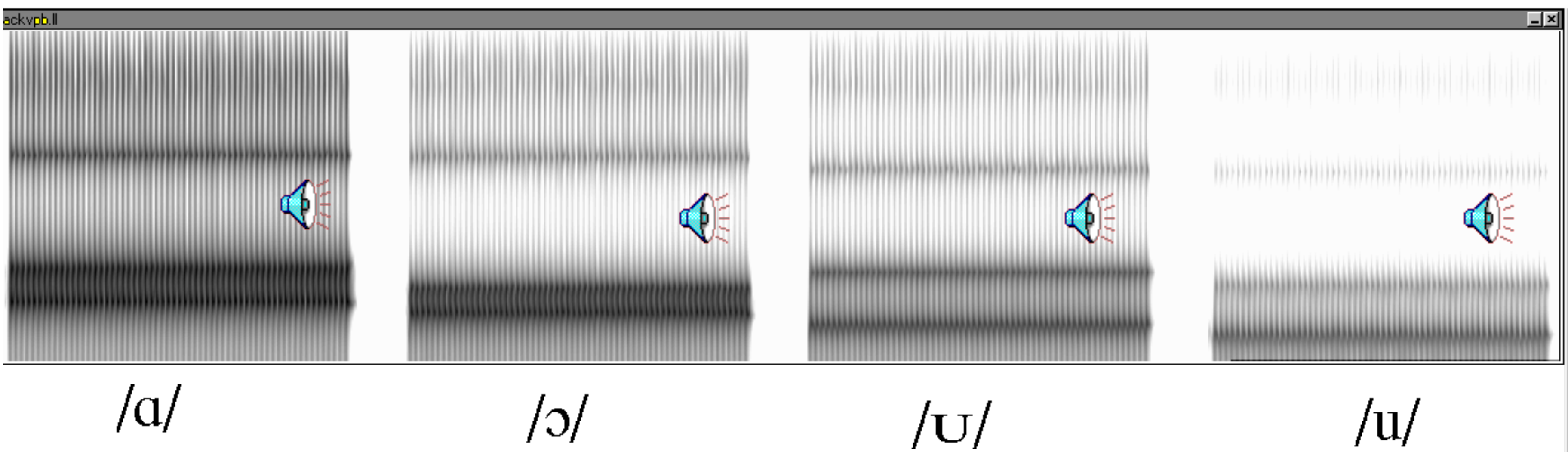
3rd formant (F_3)

2nd formant (F_2)

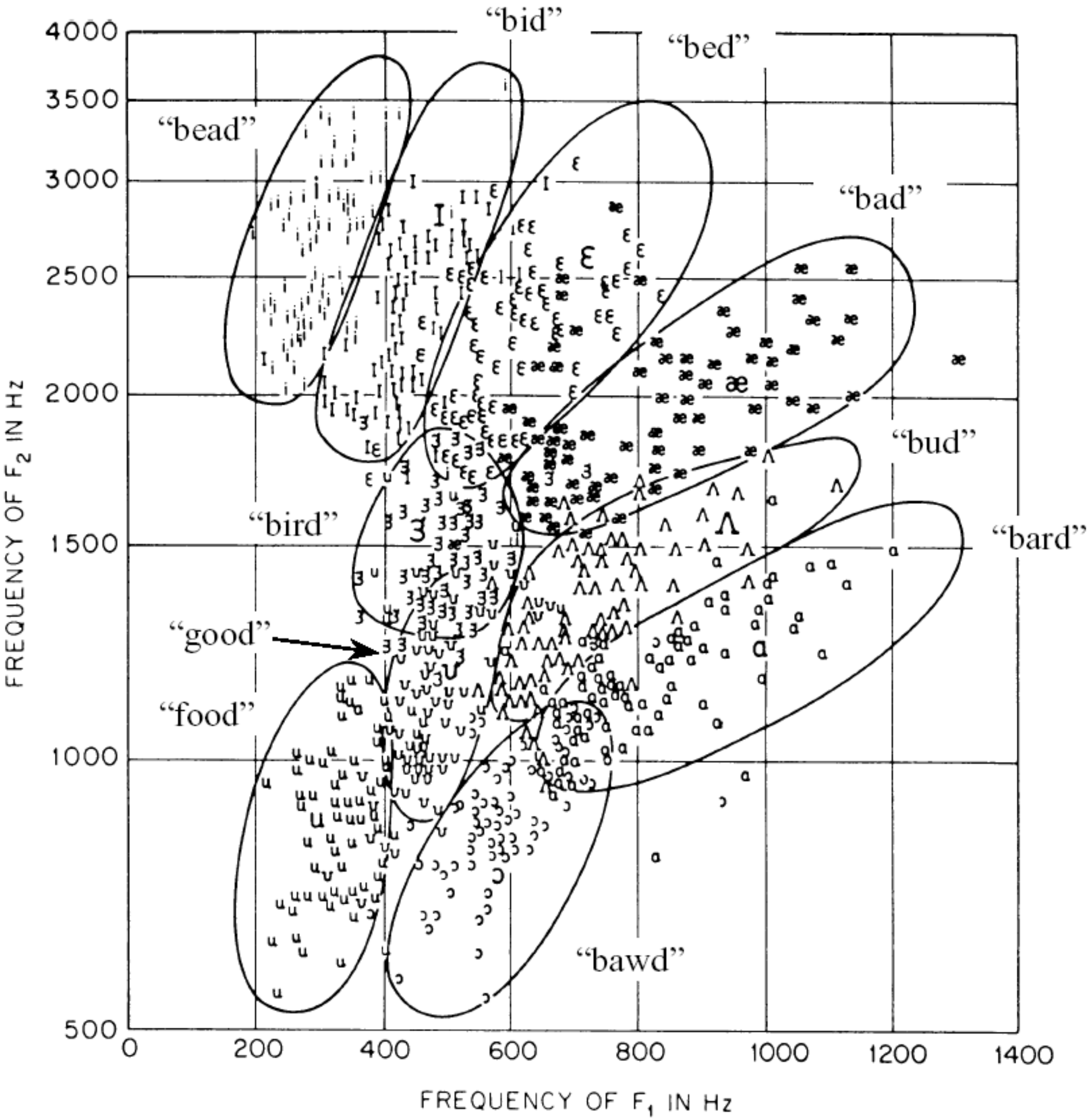
1st formant (F_1)

Acoustic patterns for some American English Vowels
Dark horizontal bars are called *formants*. Each vowel seems to have a distinctive distribution of formants.



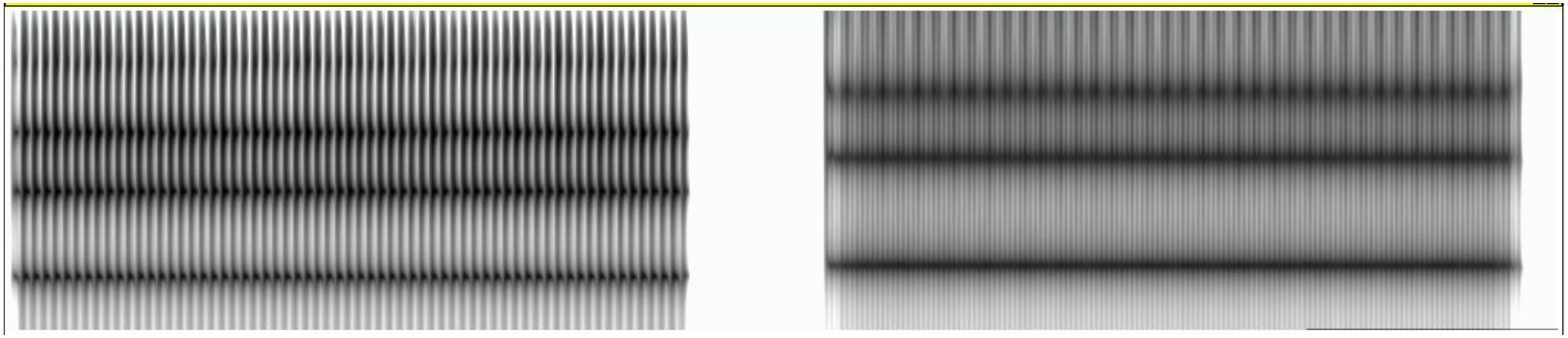


American Male speakers:



Formant Frequency Variability

/A/ (as in “cat”) spoken by a man and a child



Adult Male **Child**

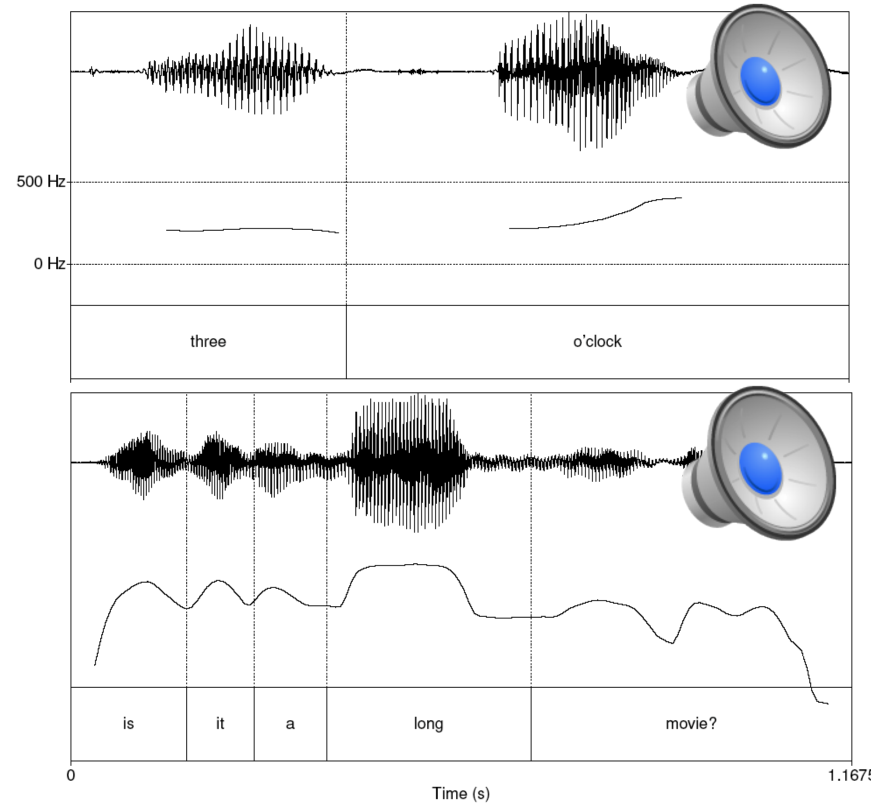
Notice that the formant frequencies are quite different, yet they are both heard as good instances of /A/.

超音段特徵(suprasegmental feature)

當我們說出完整的一句話時，是包含了一連串的發音過程，發音器官進行一連串的動作，將一個一個音節的發音產生出來。

在這過程中，會有音高的改變，音節時間長短的變化，重音的不同，以及呼吸中斷的停頓等現象，我們將這些現象所引起的聲調與節奏(rhythm)因素，稱為說話的韻律特徵(prosodic feature)。這些特徵超越音位的音段，所以稱為超音段特徵。

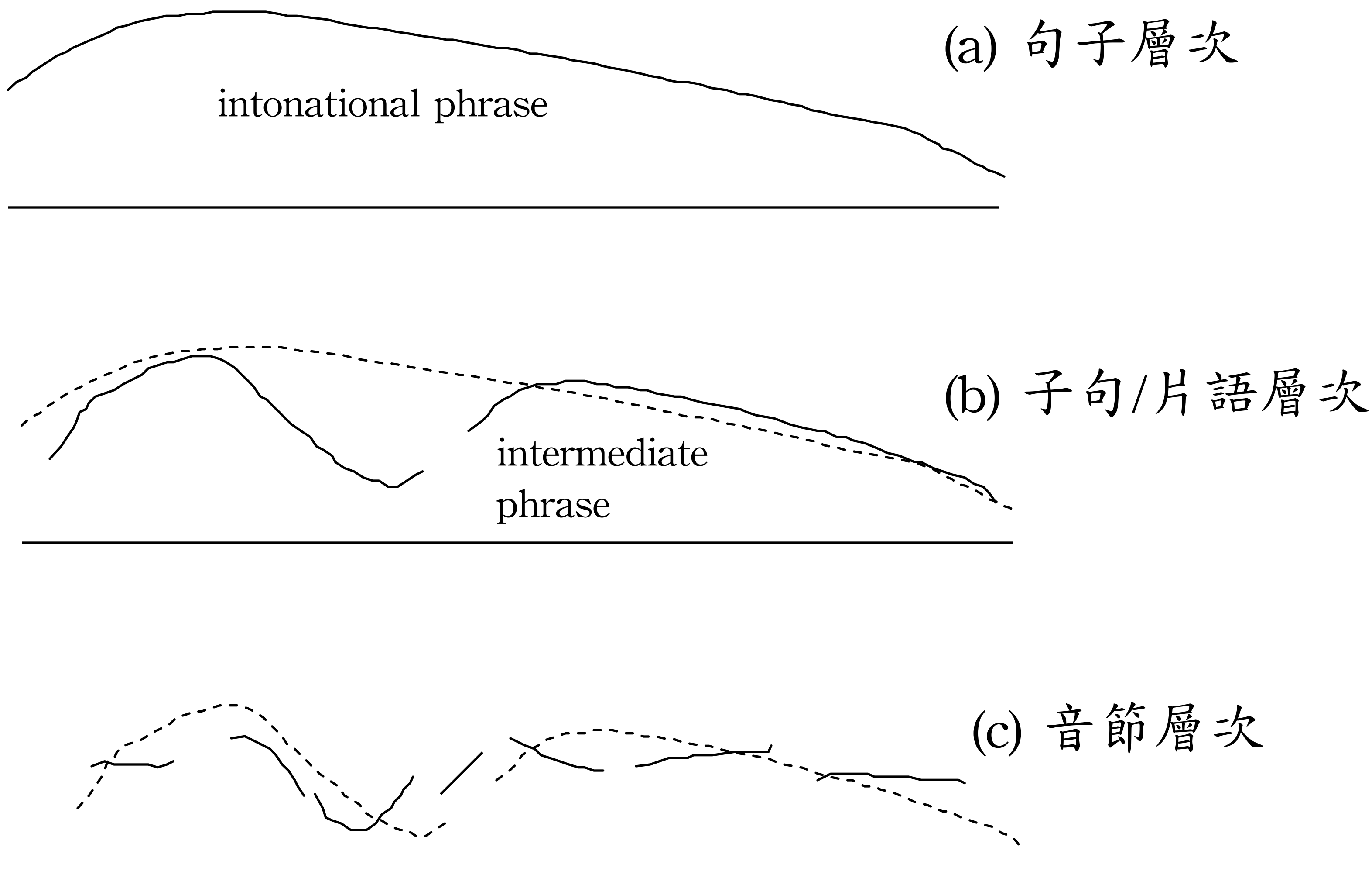
- F0
- Duration
- Intensity (energy)
- Break



French [bread and cheese]

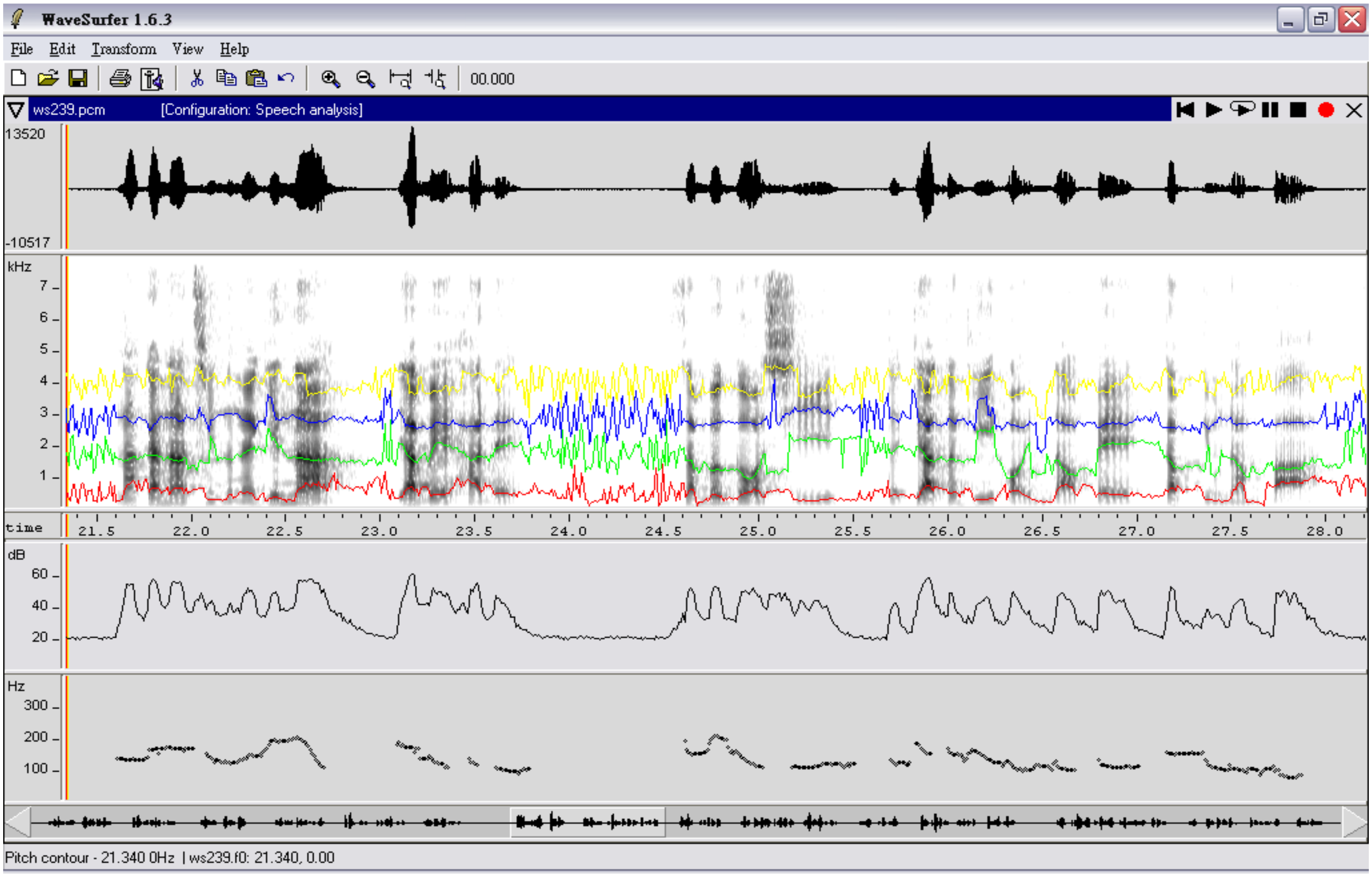
[French bread] and [cheese]

韻律片語結構



(a) 句子層次
intonational phrase
(b) 子句/片語層次
intermediate
phrase
(c) 音節層次

範例：台語韻律



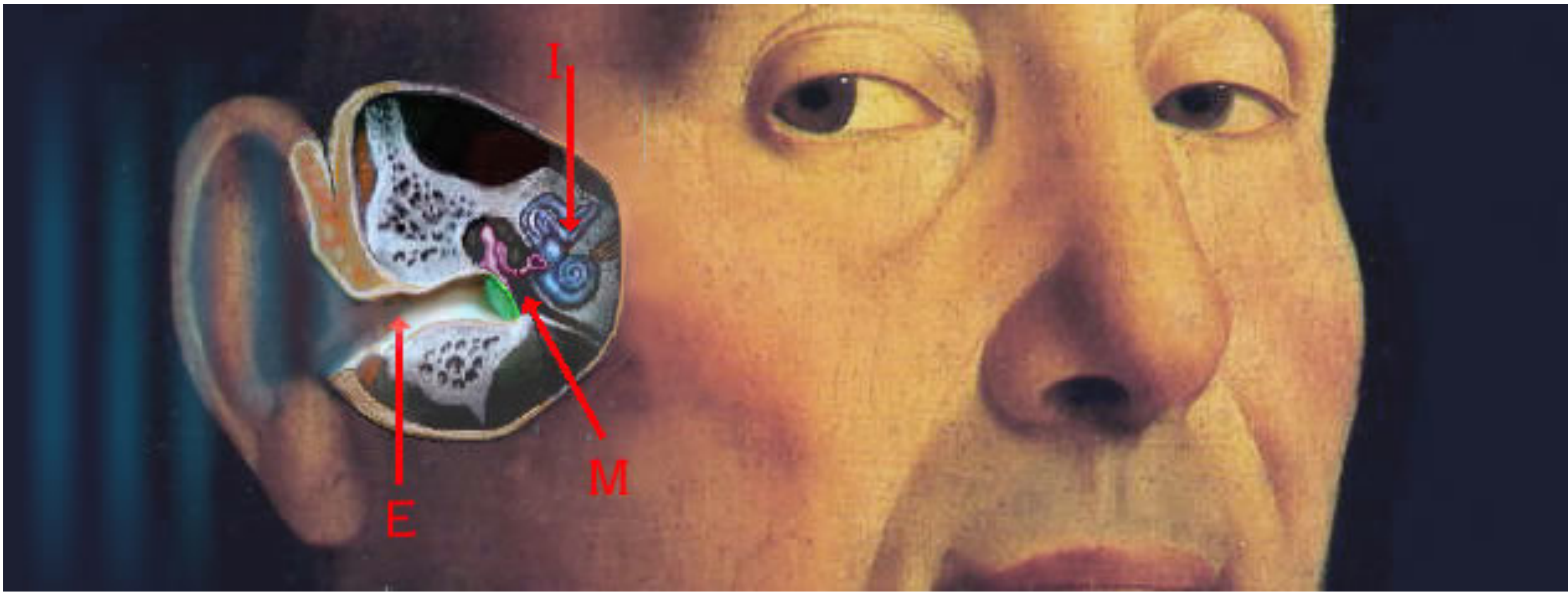
WaveSurfer 1.6.3
File Edit Transform View Help
00.000
ws239.pcm [Configuration: Speech analysis]
13520
-10517
kHz
7
6
5
4
3
2
1
time 21.5 22.0 22.5 23.0 23.5 24.0 24.5 25.0 25.5 26.0 26.5 27.0 27.5 28.0
dB
60
40
20
Hz
300
200
100
Pitch contour - 21.340 0Hz | ws239.f0: 21.340, 0.00

聽覺

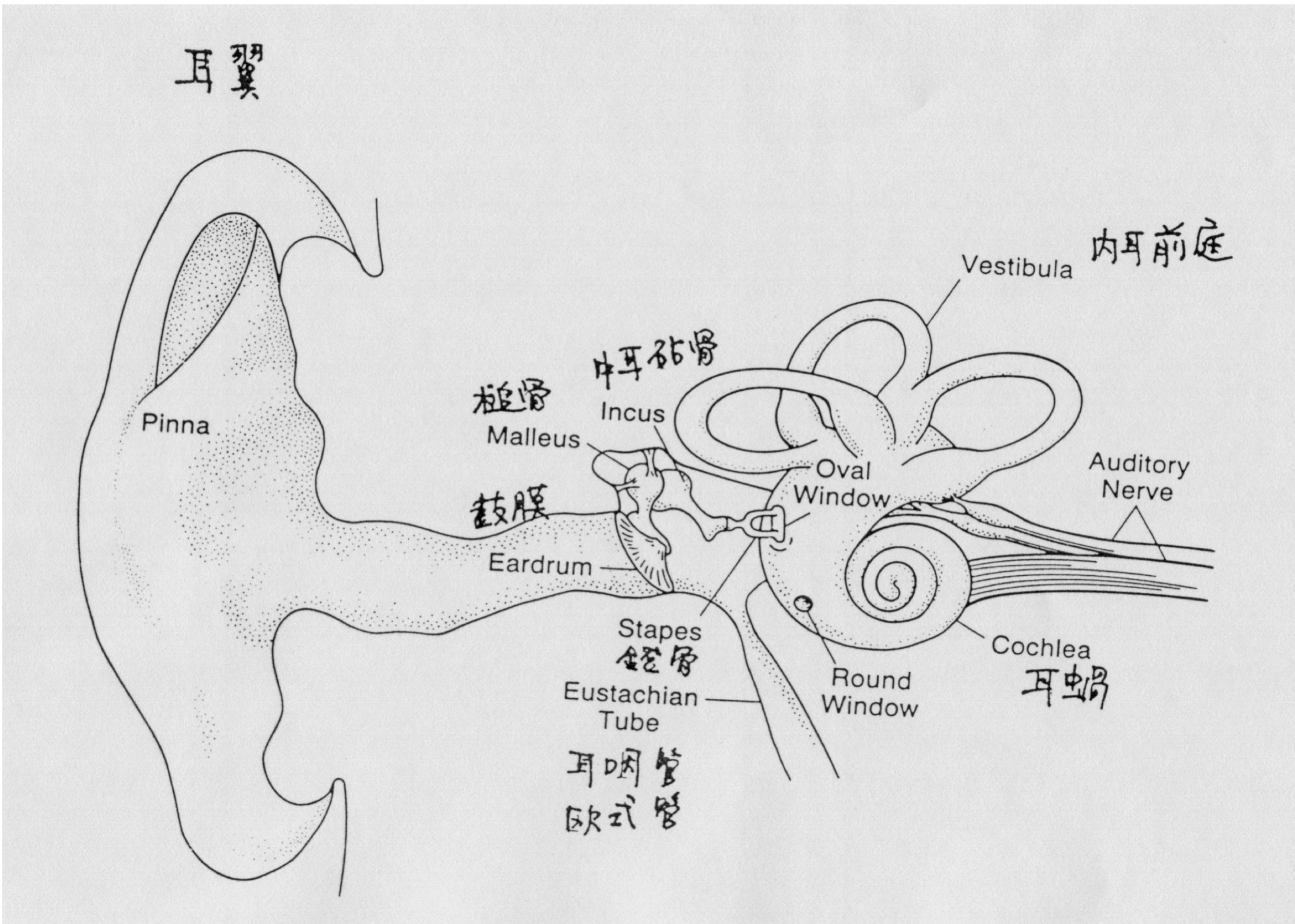
用語言作溝通，就是說話的人能讓對方聽懂他在說些什麼，雙方用說與聽來達成訊息交換與感情交流。而語言的學習過程，更是不斷地聽，然後學習會說，所以語言與聽覺有密切的關聯性。

本章將說明人類聽覺器官的構造，以及如何聽聲音。並且對於人耳在聽覺上的兩個重要特性，遮蔽效應與臨界頻帶，做了詳細敘述，這是語音處理中常常會用到的聽覺特性。

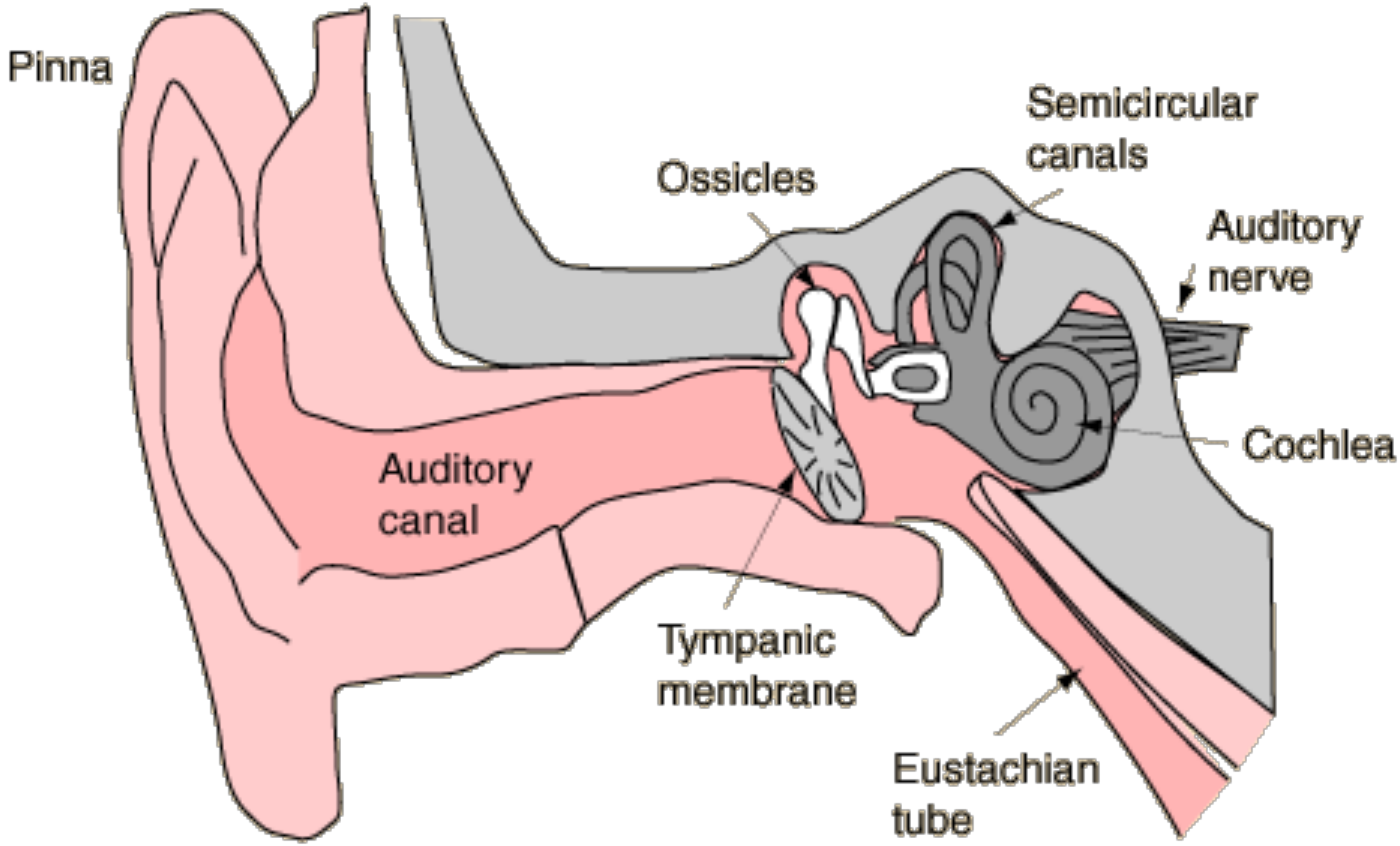
人耳結構

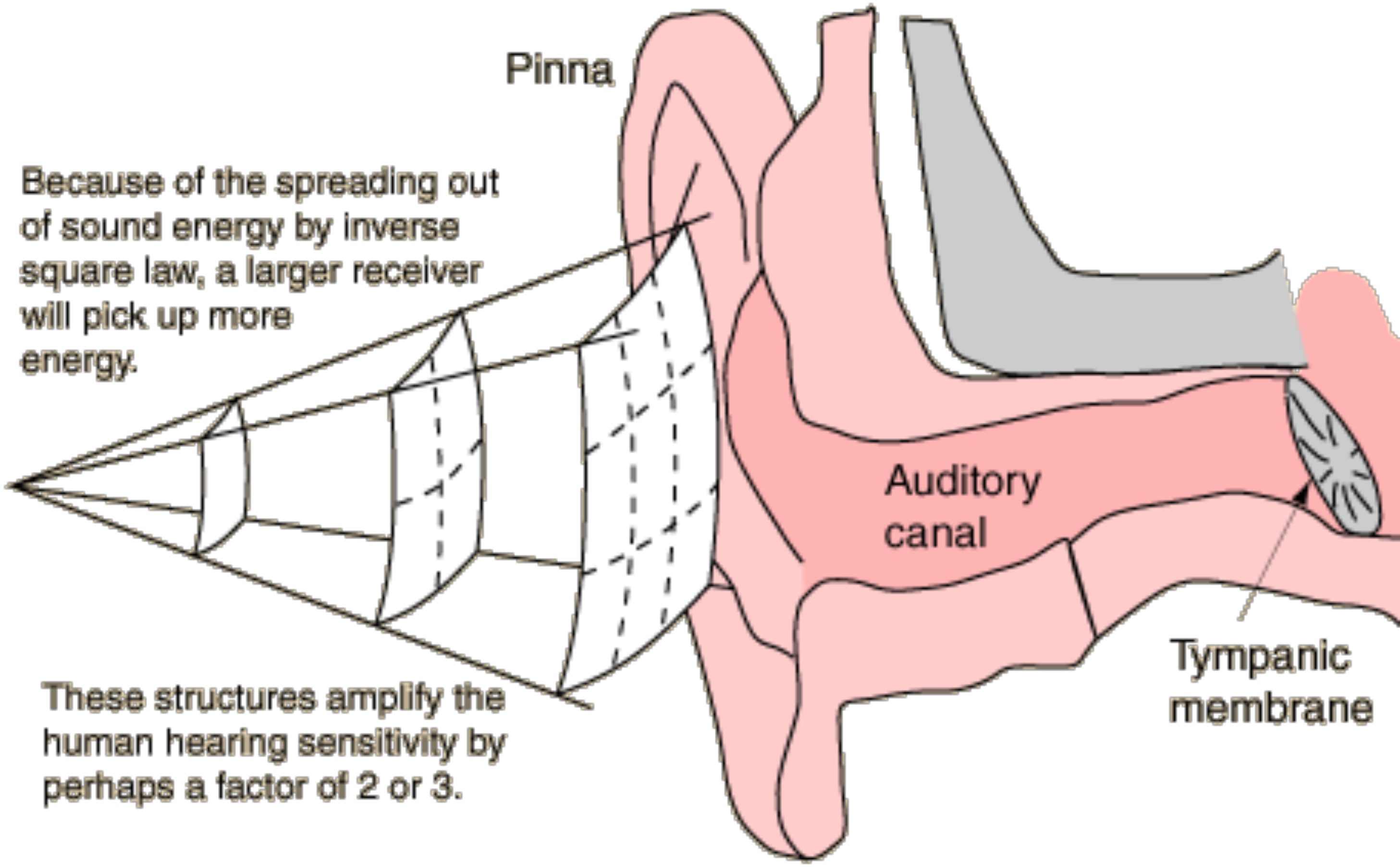


人耳結構



耳翼
Pinna
內耳前庭
Vestibula
中耳砧骨
Incus
槌骨
Malleus
鼓膜
Eardrum
Oval Window
Auditory Nerve
Stapes
鐙骨
Eustachian Tube
耳咽管
歐式管
Round Window
Cochlea
耳蝸

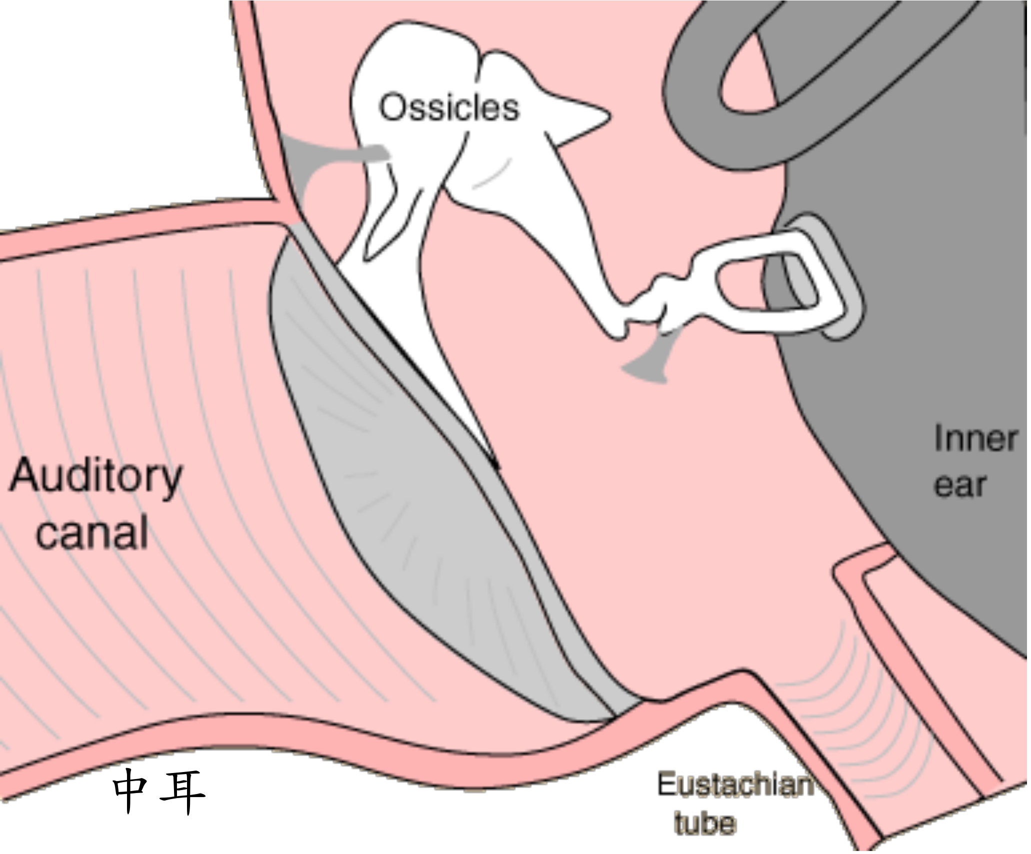

Pinna
Semicircular canals
Ossicles
Auditory nerve
Cochlea
Auditory canal
Tympanic membrane
Eustachian tube

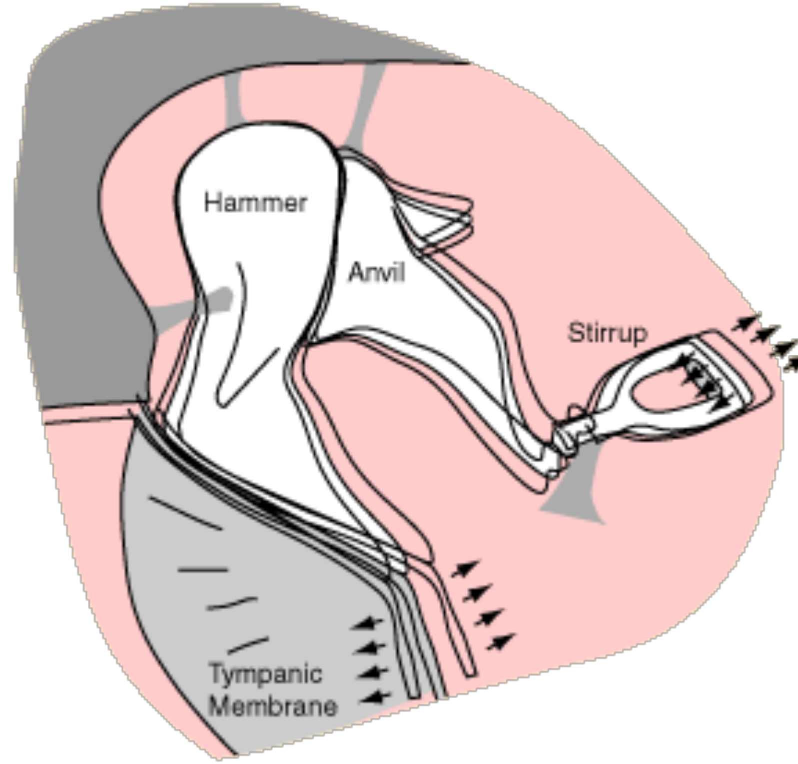


外耳

從耳翼(pinnas)到鼓膜(eardrum)這一段叫做外耳。耳翼的功能在幫助判斷聲音的來源方向，它對於聽者前方來的聲音比較敏感。

耳翼到鼓膜之間的通道叫做耳道(meatus)，這是一個長約2.7公分直徑約0.7公分的通道。對於聲波的傳播而言，這條通道等於是一個四分之一波長的共振腔，它的第一個共振頻率大約是3 kHz，這個共振作用將3～5 kHz的聲波放大15 dB左右，因此使得我們的聽覺對於3～5 kHz的頻率範圍比較敏感。

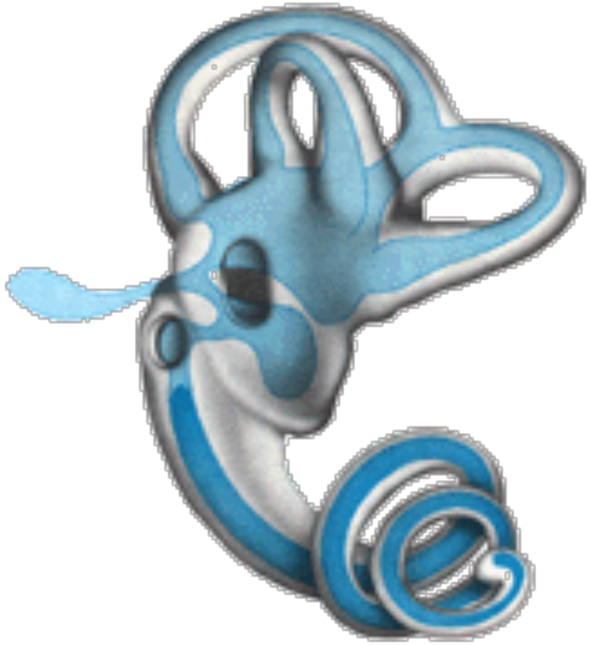


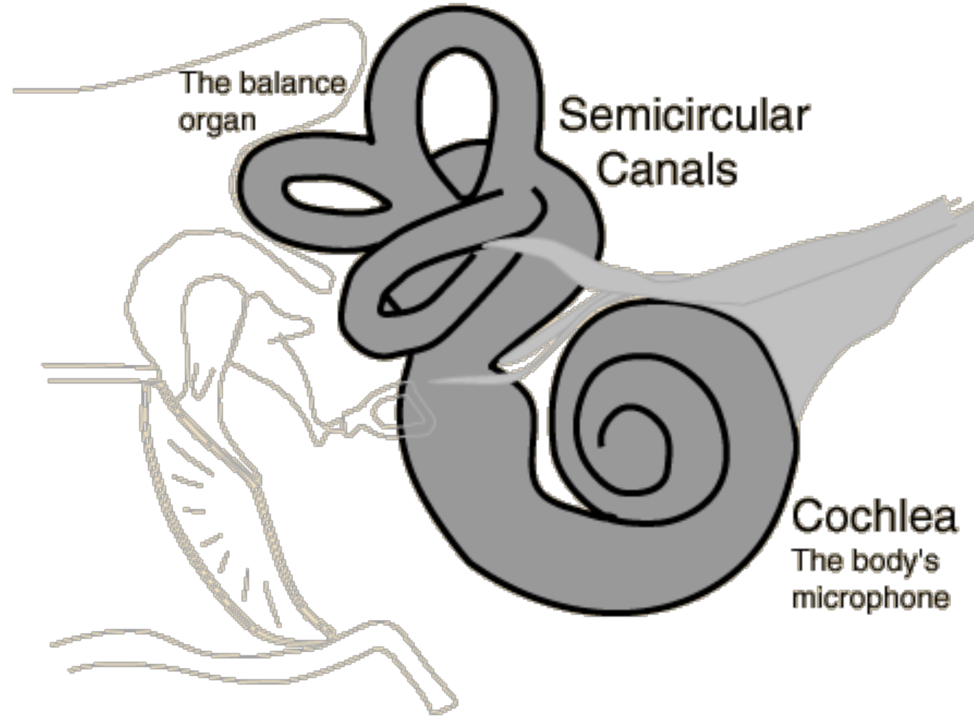


中耳

在鼓膜與耳蝸之間有一個大約6立方公分的小空間，稱之為中耳。在這個小空間內，有三塊小骨頭，分別是槌骨(hammer，或稱為malleus)， 砧骨(anvil或稱為incus)與鐙骨(stapes，或稱為stirrup)。

槌骨黏接在鼓膜上，聲波造成的空氣振動會使鼓膜振動，進而推動槌骨的振動，鐙骨貼在耳蝸的橢圓形窗(oval window)上，槌骨的振動經過砧骨與鐙骨的傳遞，將振動訊號傳到耳蝸內的淋巴液，所以中耳可以看成是聲波到淋巴液的訊號轉換，在此轉換過程中，對於1 kHz以上的訊號做了 -15 dB/oct的衰減，可以看成是一個低通濾波器(low pass filter)。





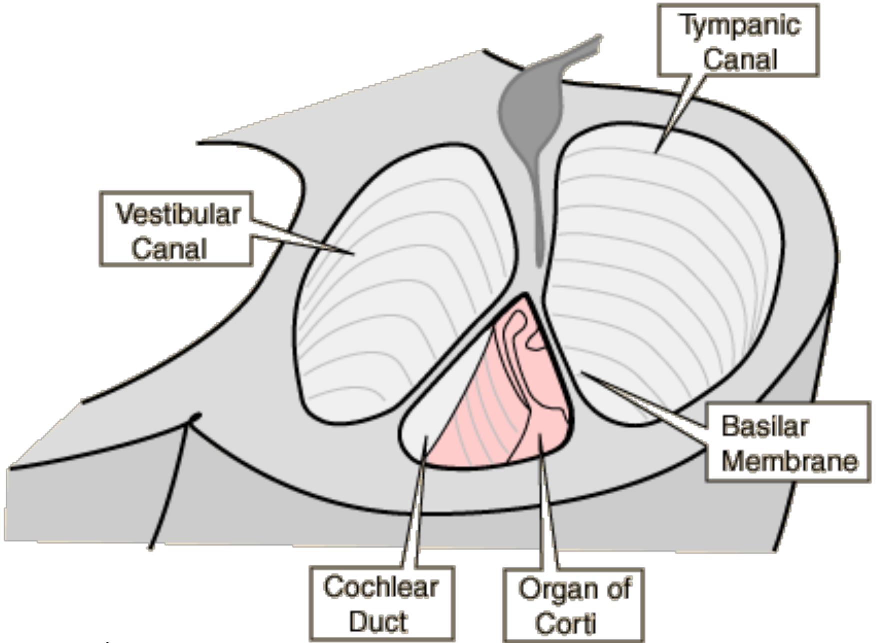
內耳

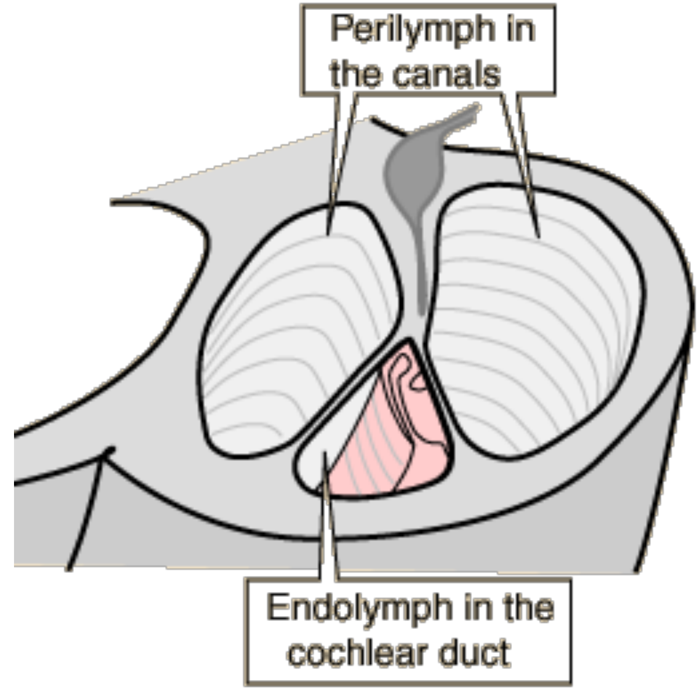
內耳指的就是耳蝸(cochlea)，它是一條充滿凝膠狀淋巴液的管子，卷曲約2.5圈成蝸殼狀。

鐙骨貼在耳蝸的橢圓形窗上，從橢圓形窗傳來的機械振動轉換成神經纖維的電性刺激訊號。

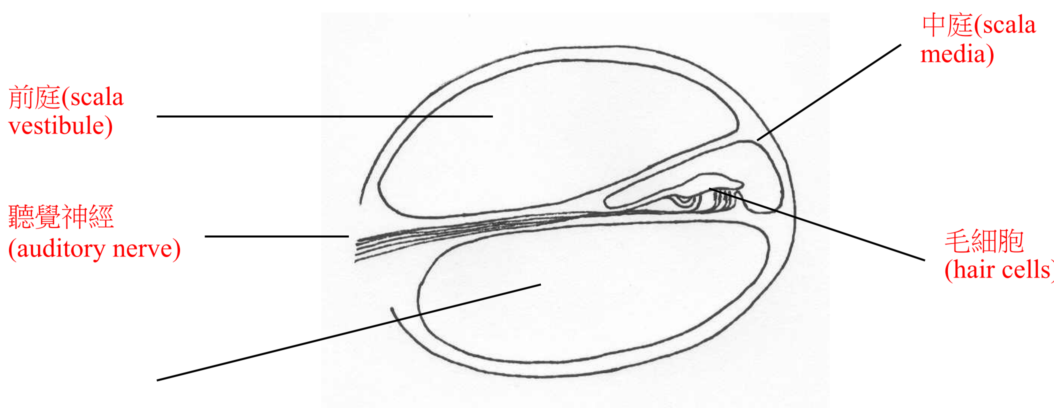
另外有一個圓窗(round window)，是一層薄膜，讓耳蝸內的淋巴液受到壓力時，有一個紓解的地方。

這個卷曲的管子長約3.5公分，內部有兩個薄膜，將管子分隔成三個充滿淋巴液的空間，分別稱為前庭(scala vestibuli)、鼓室(scala tympani)與中室(scala media)。



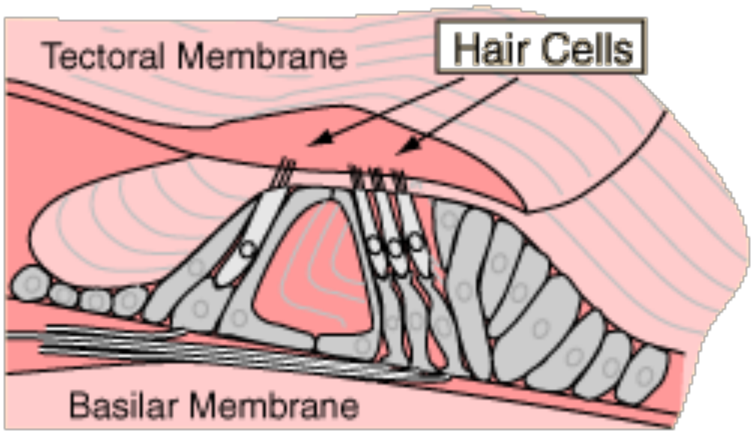


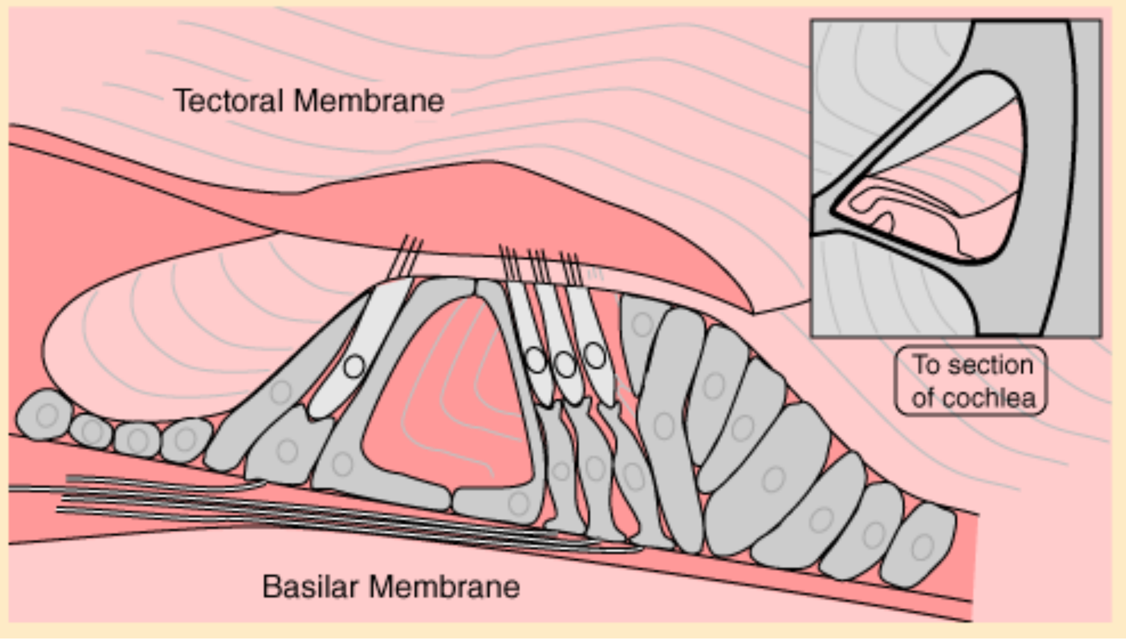
內耳耳蝸的橫切面



鼓室與中室之間的隔膜稱為基膜(basilar membrane)，長約3.2～3.5公分，靠近鐙骨處的厚度約0.04mm，越往內伸越厚，到達頂端時厚度約0.5mm。

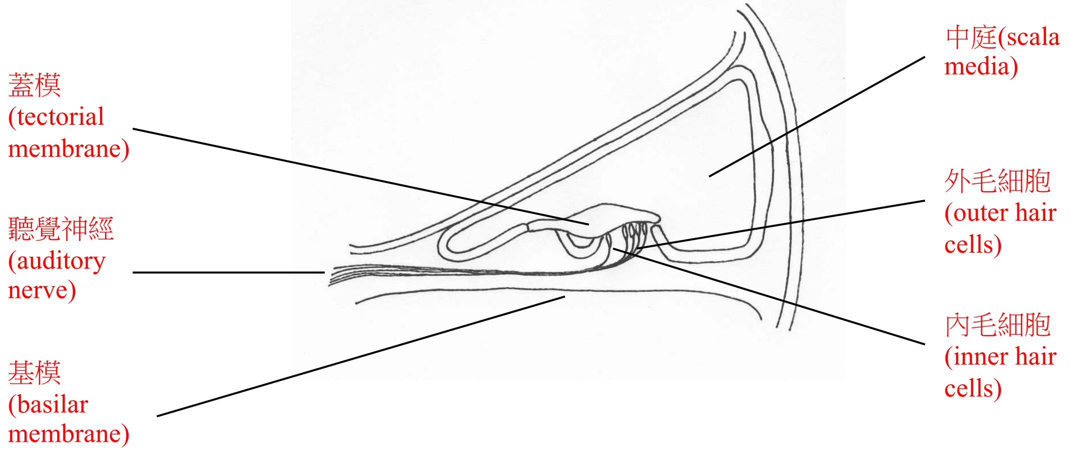
在基膜上有一個皮層組織，內含約30,000個感測毛細胞(sensory hair cells)，沿著管長方向排列成若干行。

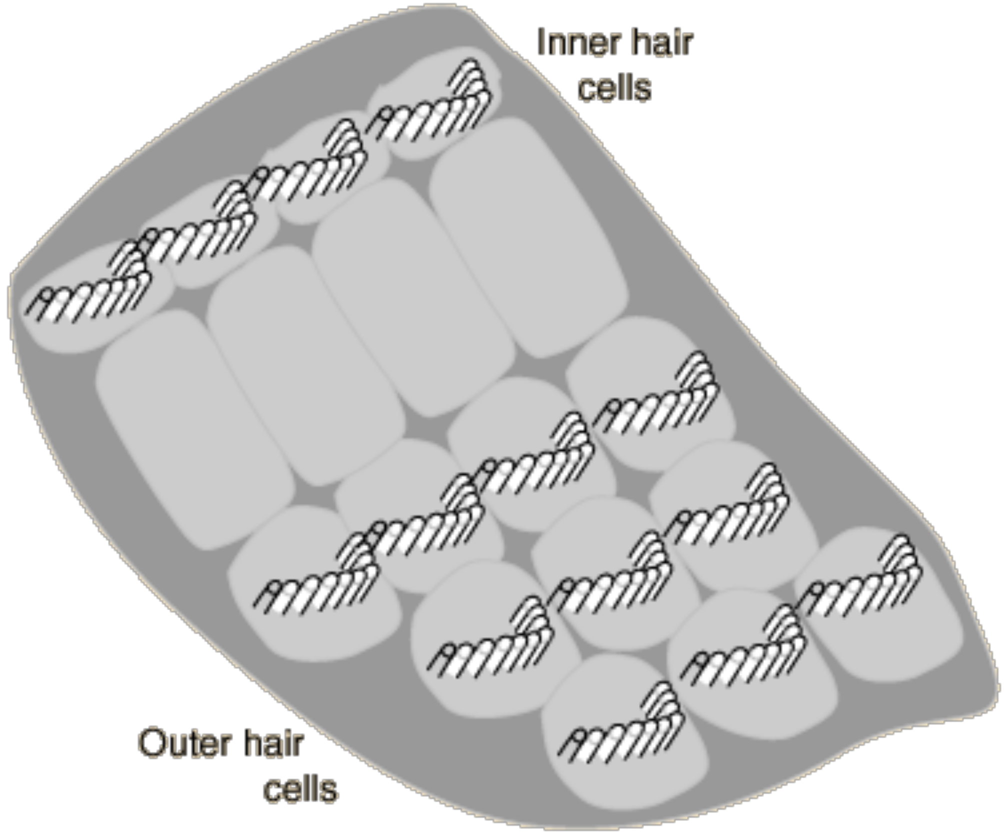




聽覺神經的終端就是這些毛細胞，每根聽覺神經約有40～140個終端的毛細胞。毛細胞上端頂著蓋膜(tectorial membrane)，蓋膜與基膜之相對運動就會刺激毛細胞，發射電性訊號，經由聽覺神經傳回大腦。

中室內的毛細胞構造

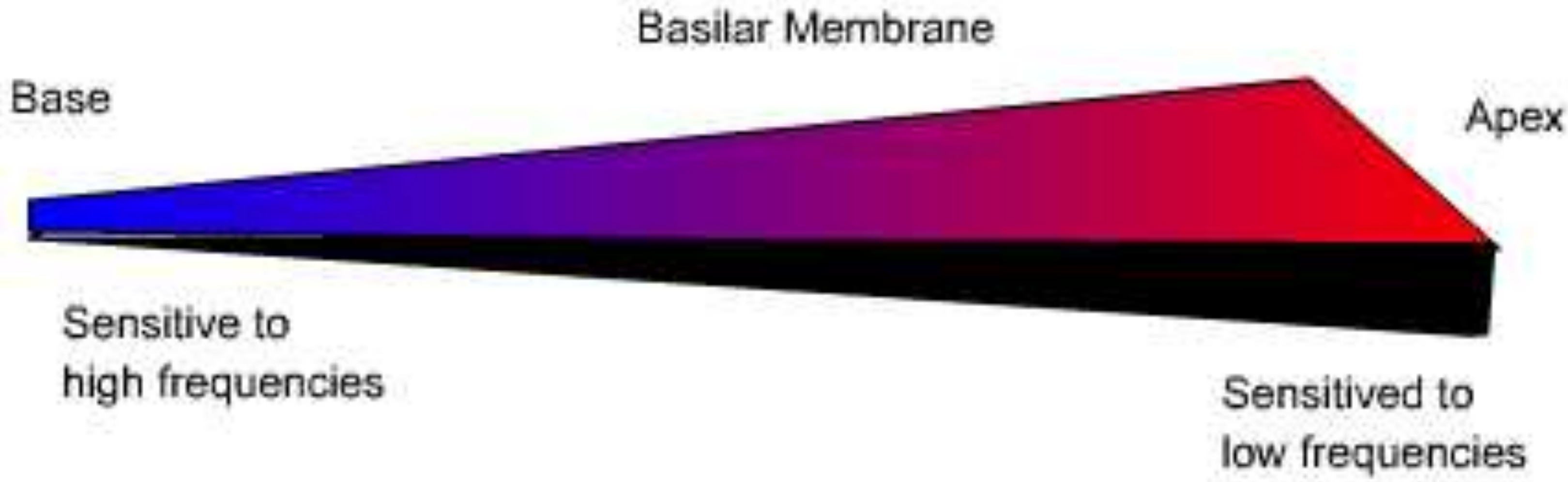




內毛細胞(inner hair cells)呈單行，約有5000個細胞，佔大約95%的神經纖維終端。

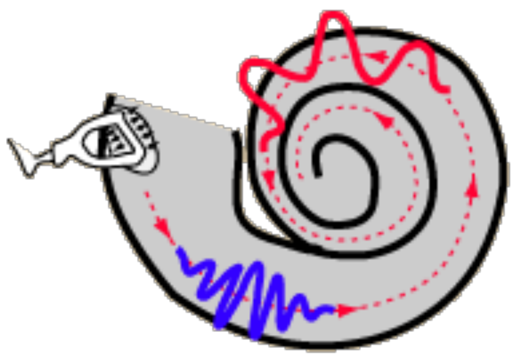
外毛細胞(outer hair cells)有若干行，數量較多，但是佔大約5%的神經纖維終端。

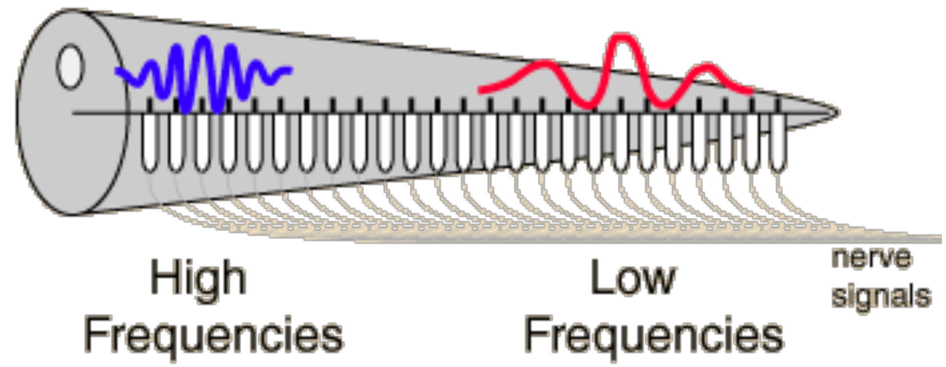
內毛細胞與外毛細胞具有不同的反應特性，外毛細胞主要是由基膜帶動而產生刺激，內毛細胞是反應基膜的運動速度，外毛細胞對於基膜的彎曲會有感應，而內毛細胞則是感應沿基膜方向的運動。



基膜的拉緊程度(tautness)與形狀是沿著耳蝸軸線改變，因此它對於頻率的反應也隨著改變。靠近橢圓形窗的基部是較薄較硬，在尖端處則較厚重，沿著基膜的每一點有其特性頻率(characteristic frequency)，對於聲波的頻率響應曲線呈現帶通濾波器(band pass filter)的特性。

它有大約固定的Q值，也就是固定的中央頻率與頻寬的比值，因此頻率上的解析度是在低頻比較高，在高頻則比較低。連接高頻特性頻率處的毛細胞所反應的頻帶較寬，而連接低頻特性頻率處的毛細胞所反應的頻帶就較窄。

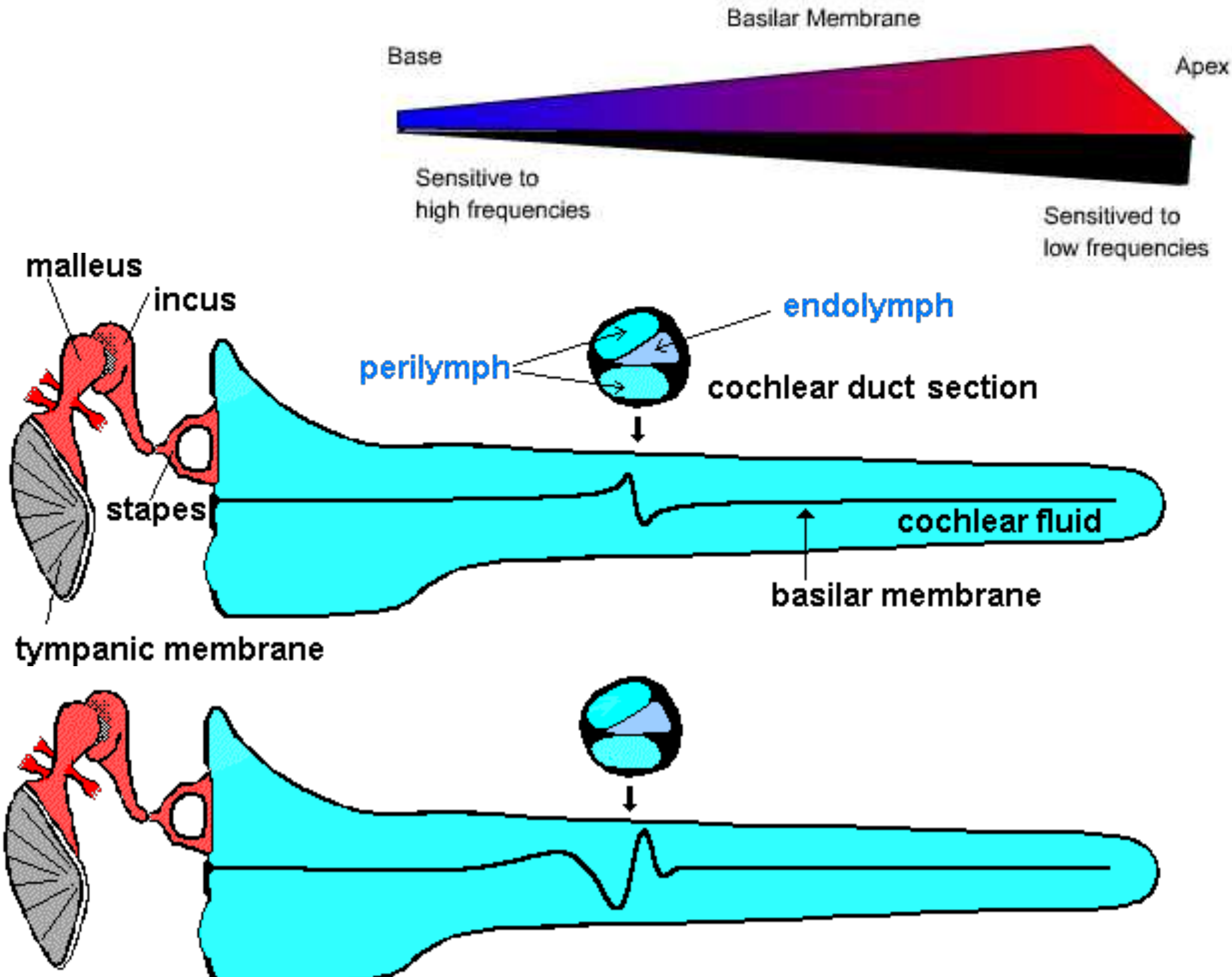




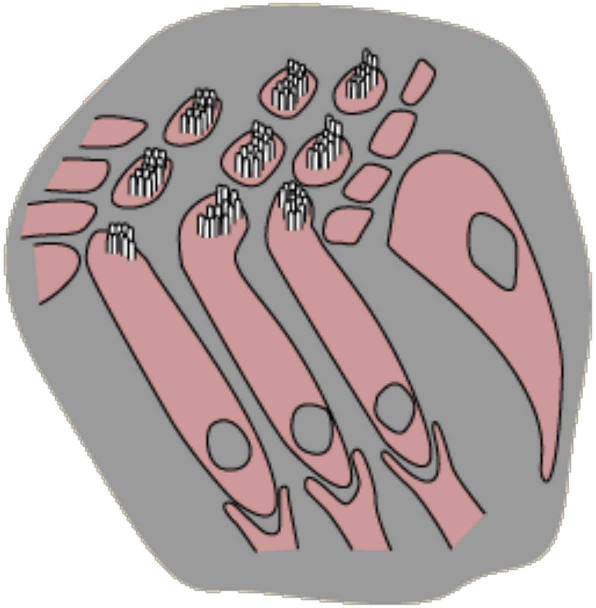
當有一固定頻率的聲波傳入到橢圓形窗時，這個振動的壓力會立刻傳遍耳蝸，使得基膜跟著這個頻率振動。

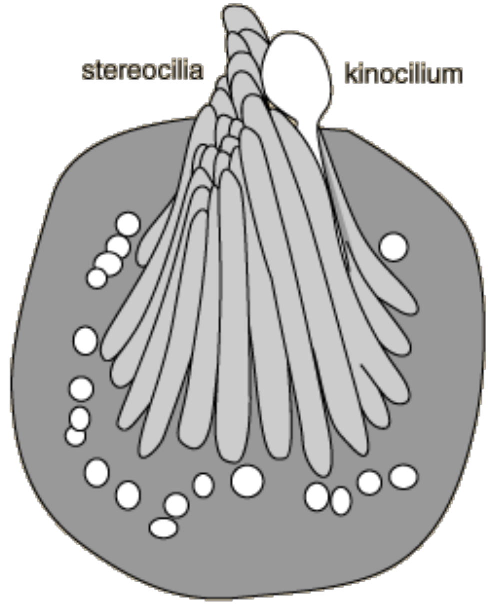
因為聲波在耳蝸液體內的傳播速度為1600m/s，對於長度只有3.5公分的耳蝸來說，沿著基膜是看不出相對的延遲，事實上聲音從頭殼或耳道傳達到耳蝸，基膜的運動是一樣的，因此，人耳對於聲音訊號的相位差異不太有感覺。

基膜的運動會從基部向尖部傳遞，這個波的傳送速度沿著基膜遞減，這個傳導波(traveling wave)在到達與聲波頻率相近的特性頻率處時，振幅會最大。對於一個瞬間脈衝訊號(impulse)的聲音，如卡搭聲(chick)，在基膜上不同特性頻率位置上會反映它自己的頻率，在基部消失比較快，在尖部則消失較慢。

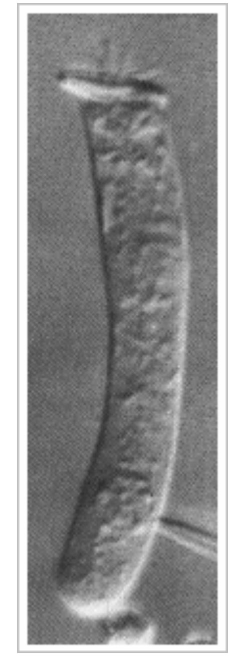


Basilar Membrane
Base
Apex
Sensitive to
high frequencies
Sensitived to
low frequencies
malleus
incus
endolymph
perilymph
cochlear duct section
stapes
cochlear fluid
basilar membrane
tympanic membrane





Adapted from electron-scanning micrograph at 16,800x . A. J. Hudspeth, R. Jacobs, Science News, Oct 20, 1984.



毛細胞是經由聽覺神經連接到大腦，當毛細胞受到彎曲的壓力與剪力時，神經細胞會發出長約0.5～1.0ms的脈衝訊號。毛細胞的張力會改變它的導電性而影響其化學物質的釋放，造成神經細胞發射脈衝訊號。

動物實驗中證實，在卡搭聲與單頻率聲音的刺激下，產生的發射型態不同，沒有聲音刺激時，神經細胞的發射訊號是平均每秒10～50次的隨機發射，當有聲音刺激時，神經細胞的發射訊號會因為音量大小與頻率而不同，聲音越大則發射的次數越多越密。

在基膜上特性頻率與刺激訊號頻率相同時，該處的神經細胞發射次數會最多，附近的神經細胞發射次數就隨距離遞減。

Auditory

- How Hearing Works?
  - https://www.youtube.com/watch?annotation_id=annotation_1840566121&feature=iv&src_vid=MXt_gX2Srgo&v=T8lKKlnnC6M
- Journey Into the Ear
  - https://www.youtube.com/watch?v=MIKvReMGu5Q
- Dancing hair cell (ear)
  - https://www.youtube.com/watch?v=Xo9bwQuYrRo

聲音感知

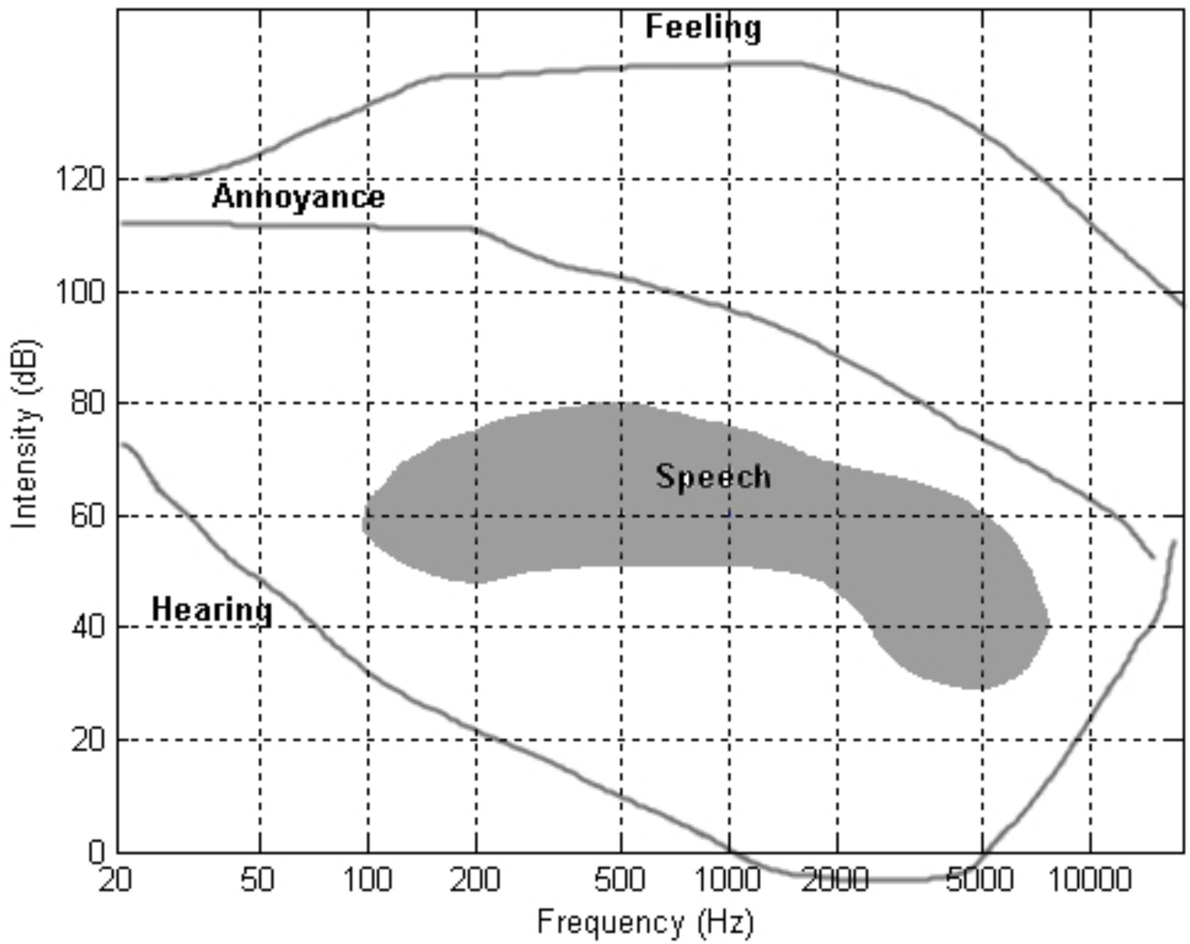
人類在分辨聲音時，是取決於頻域上的解析度，聽覺器官等於是將時域的聲音波形轉換成在頻域的頻譜，能否聽到聲音或分辨聲音，是取決於音強(intensity)與頻譜(spectrum)。

一般人的聽覺器官可以感知頻率範圍16 Hz到18 kHz的聲音，動態範圍約1000倍。

超出人類聽覺感知範圍，如20 kHz以上的聲波，我們稱之為超音波(ultrasonic)。

人耳對於1 kHz到5 kHz的聲音最為敏感，但是對於1 kHz以下與5 kHz以上的聲音，就要較大的音強才聽得到。

人類聽覺的
範圍



Feeling
Annoyance
Speech
Hearing
Intensity (dB)
Frequency (Hz)
0
20
40
60
80
100
120
20
50
100
200
500
1000
2000
5000
10000

音強以聲壓位階(sound pressure level，SPL)為量測單位，它的參考基準是1 kHz振動頻率下振動空氣壓力為0.0002dyn/cm2，這是正好可以聽到的音強。

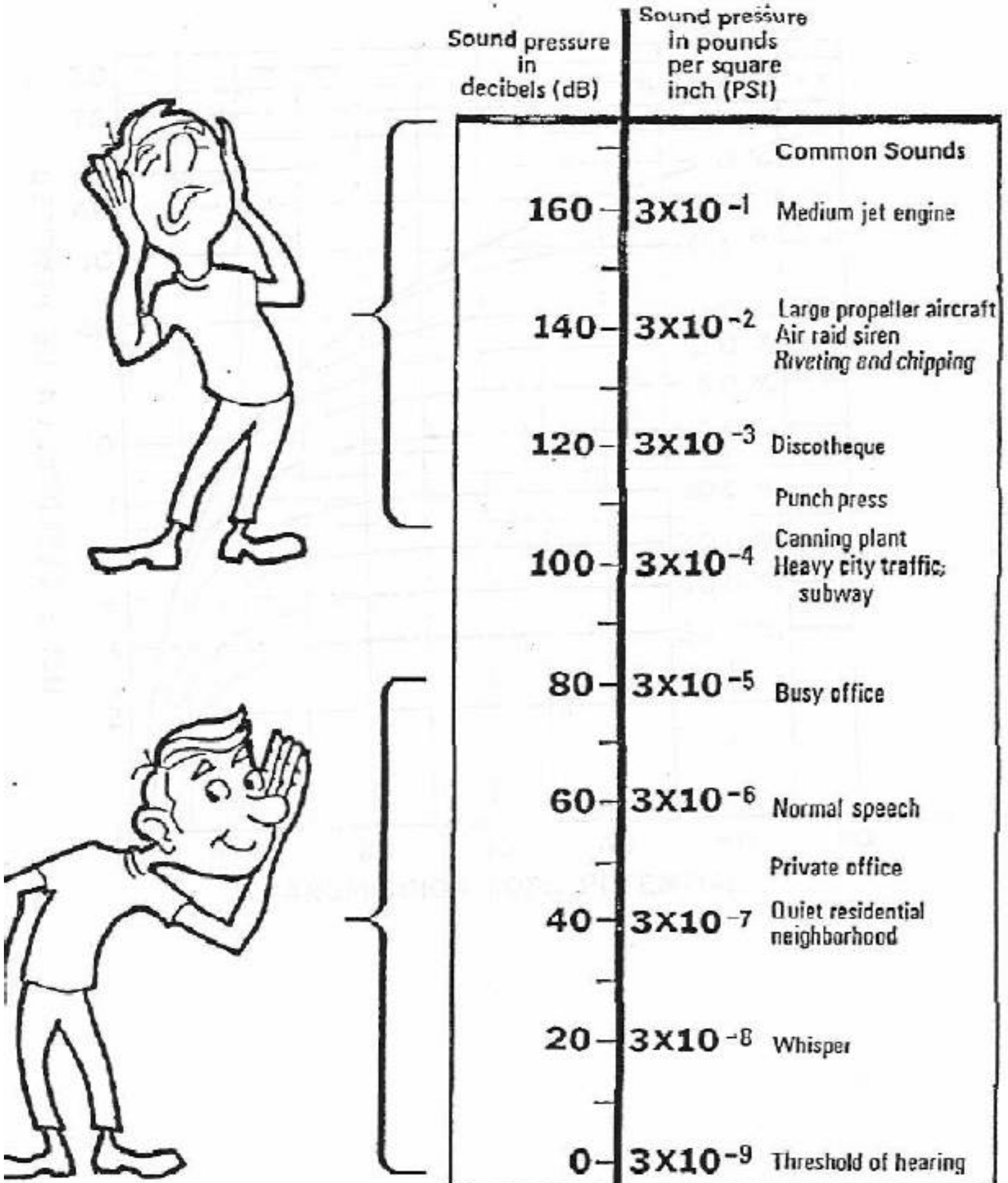
以聲音在1 kHz振動時的空氣壓力為參考基準，相對的音強倍數取dB值，就是其聲壓位階。

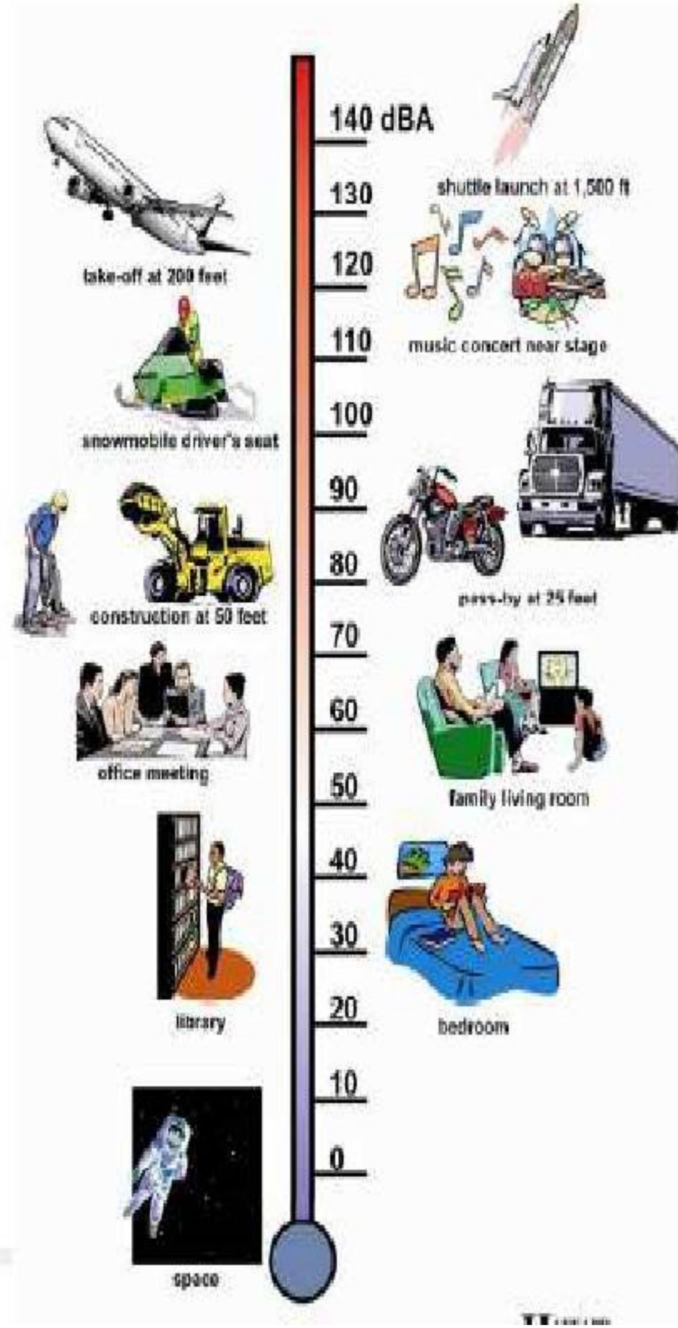
1 kHz到5 kHz之間是聽覺最敏感處，小於0 dB音強的聲音還可以聽得到，小於1 kHz或大於5 kHz的聲音，音強要增大才能聽得到。

這條聽得到的臨界曲線就是聽覺閾(hearing threshold)。

若干聲音的聲壓位階

聲音	SPL (dB)
噴射引擎 (10公尺距離)	150
疼痛的聽覺界限	130
超音速飛機起飛 (500公尺外)	120
火車鳴笛 (100呎外)	117
搖滾樂	110
鏈鋸 (1公尺距離)	100
割草機 (1.5公尺距離)	90
噴射客機起飛 (坐在客艙內靠窗位置)	85
汽車 (15呎距離)	70
公路上汽車 (15公尺距離)	70
談話 (1公尺距離)	50 – 70
電腦 (2呎距離)	54
市郊住宅區	40 – 50
錄音室	20
輕聲 (1.5公尺距離)	20
輕聲呼吸	10
聽覺閾	0

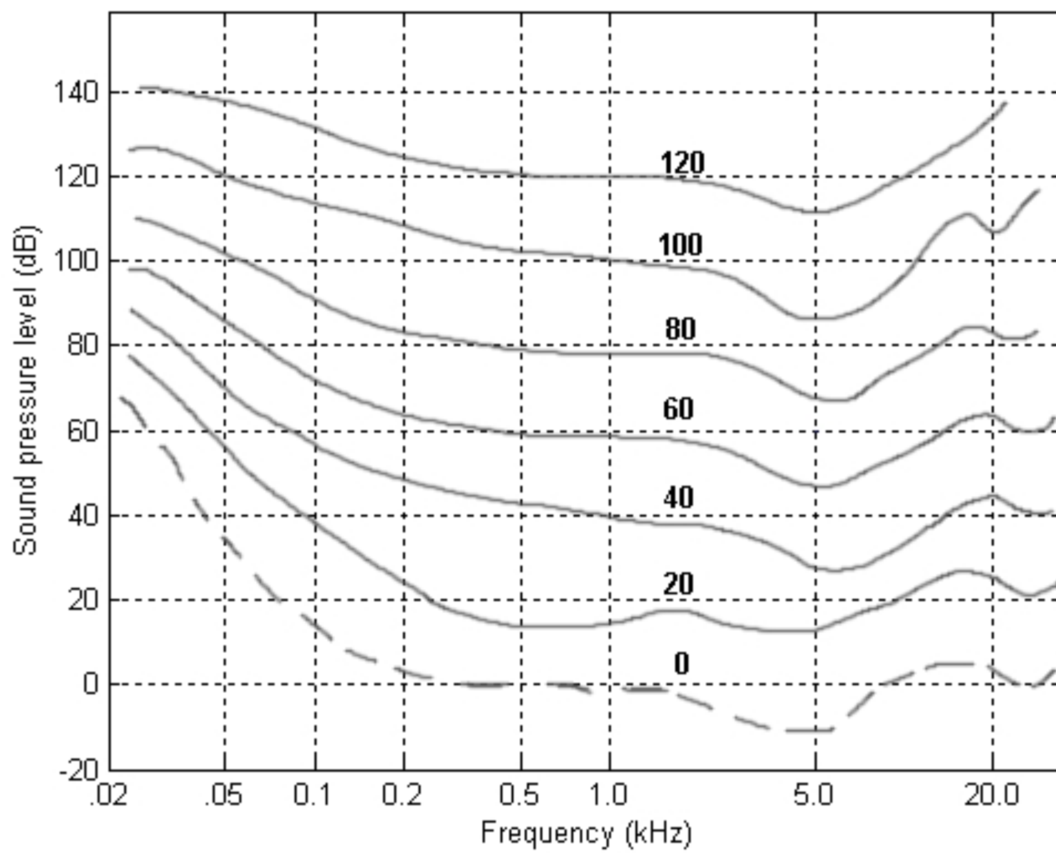

Sound pressure in decibels (dB)
Sound pressure in pounds per square inch (PSI)
Common Sounds
160 3X10 -1 Medium jet engine
140 3X10 -2 Large propeller aircraft
Air raid siren
Riveting and chipping
120 3X10 -3 Discotheque
Punch press
Canning plant
100 3X10 -4 Heavy city traffic; subway
80 3X10 -5 Busy office
60 3X10 -6 Normal speech
Private office
40 3X10 -7 Quiet residential neighborhood
20 3X10 -8 Whisper
0 3X10 -9 Threshold of hearing


140 dBA
130
120
110
100
90
80
70
60
50
40
30
20
10
0
take-off at 200 feet
snowmobile driver's seat
construction at 50 feet
office meeting
library
space
shuttle launch at 1,500 ft
music concert near stage
pass-by at 25 feet
family living room
bedroom

在衡量聲音大小時，有一個單位叫做響度(loudness)，它以1 kHz的單頻率聲音為參考基準，在不同頻率下，聽覺上的響度與1 kHz時的響度一樣時，對應的聲壓位階連成一曲線，這就是等響度曲線(equal-loudness contours)。

響度單位用Phon，等同於1 kHz聲音時的聲壓值，各個不同的響度曲線，與聽覺臨界曲線幾乎是平行的。

等響度曲線



140
120
100
80
60
40
20
0
-20
Sound pressure level (dB)
.02
.05
0.1
0.2
0.5
1.0
5.0
20.0
Frequency (kHz)
120
100
80
60
40
20
0

在聽覺感知上，常常以兩個相近的聲音來測試聽者的分辨能力，看看兩者是否聽起來不一樣。

在頻率方面，兩個音強相等的單頻率聲音(tones)，若是在1 kHz以下，相差1～3 Hz就可以分辨其不同。但在高頻時，頻率差要更大些才分辨出不同，如在8 kHz的聲音，要100 Hz的差別才分得出來。

在整個可以聽到聲音的頻率範圍內，大約可以分辨出1600個不同的頻率。在音強方面，可區別的音強差是在0.3 dB～1 dB之間，如在1 kHz 60 dB時，可分辨的差異是0.3 dB。

在聽覺範圍內可以有350個可分辨的音強，連同1600個可分辨的頻率，全部有大約300,000個可分辨的聲音。

另一個現象，是聲音要夠長才能夠聽辨，太短的聲音在聽覺上並不敏感。如音長超過250 ms，可以分辨850個不同的頻率，但是10 ms的音長就只能分辨120個不同的頻率。

有週期性波形的聲音，會有其聲調，通常分辨兩個聲調不同的聲音，是看它們的音高，音高與聲帶的振動週期相關。

一個聲音具有某一種音高(pitch)，我們可以拿一個正弦波的聲音，調整其頻率，使到聽起來音高相同。

如果一個聲音沒有週期性，如擦音，它們的差別就是看音色(timbre)，高頻成分多者聽起來音色較亮，所以音色反應的是頻譜的包絡線圖(envelope)。

遮蔽效應與臨界頻帶

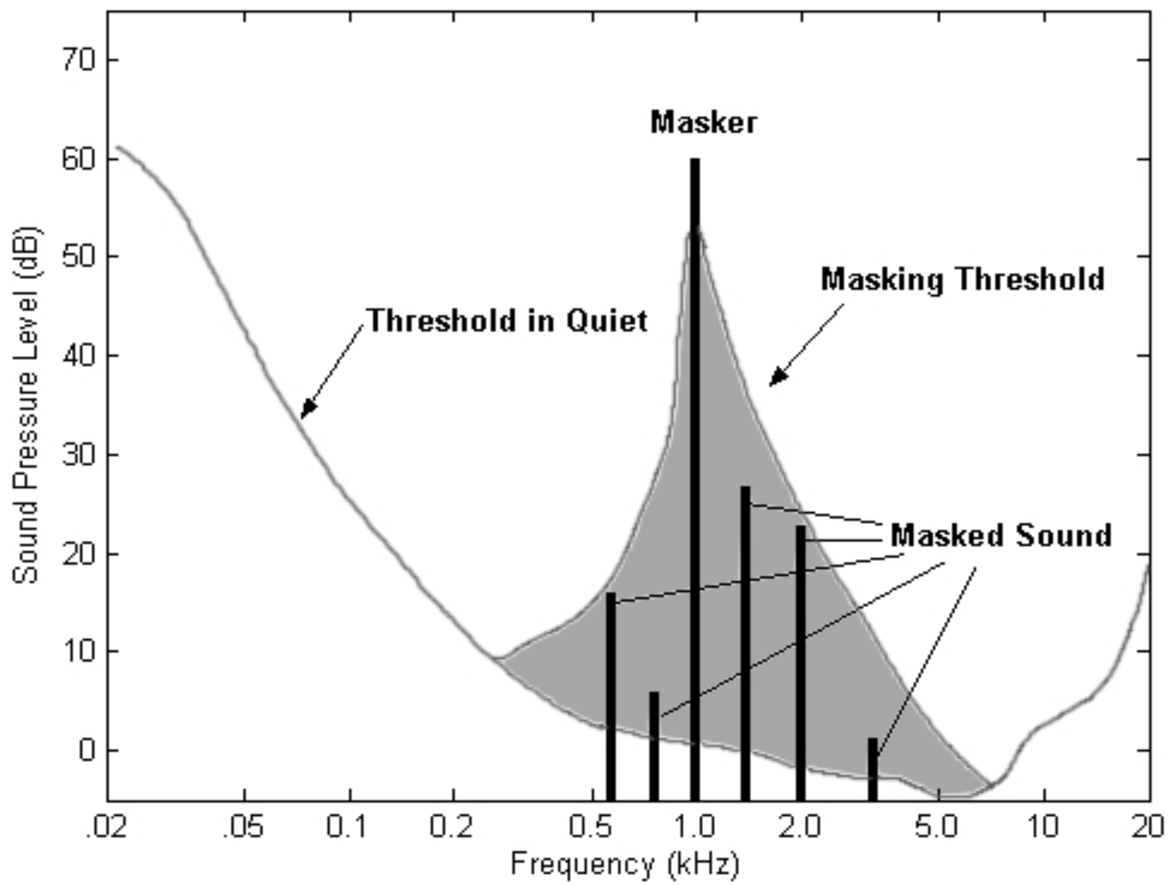
遮蔽效應(masking effect)

當一個聲音存在時，它會將另一個聲音的感知臨界值提高，也就是說當某一頻率的聲音，有一特定音強存在時，另一個不同頻率的聲音要將音強提高才會被聽到，這就是遮蔽效應。

同時存在的聲音，常常是低頻的聲音傾向於遮蔽掉高頻的聲音，這是一種頻率遮蔽(frequency masking)的現象。

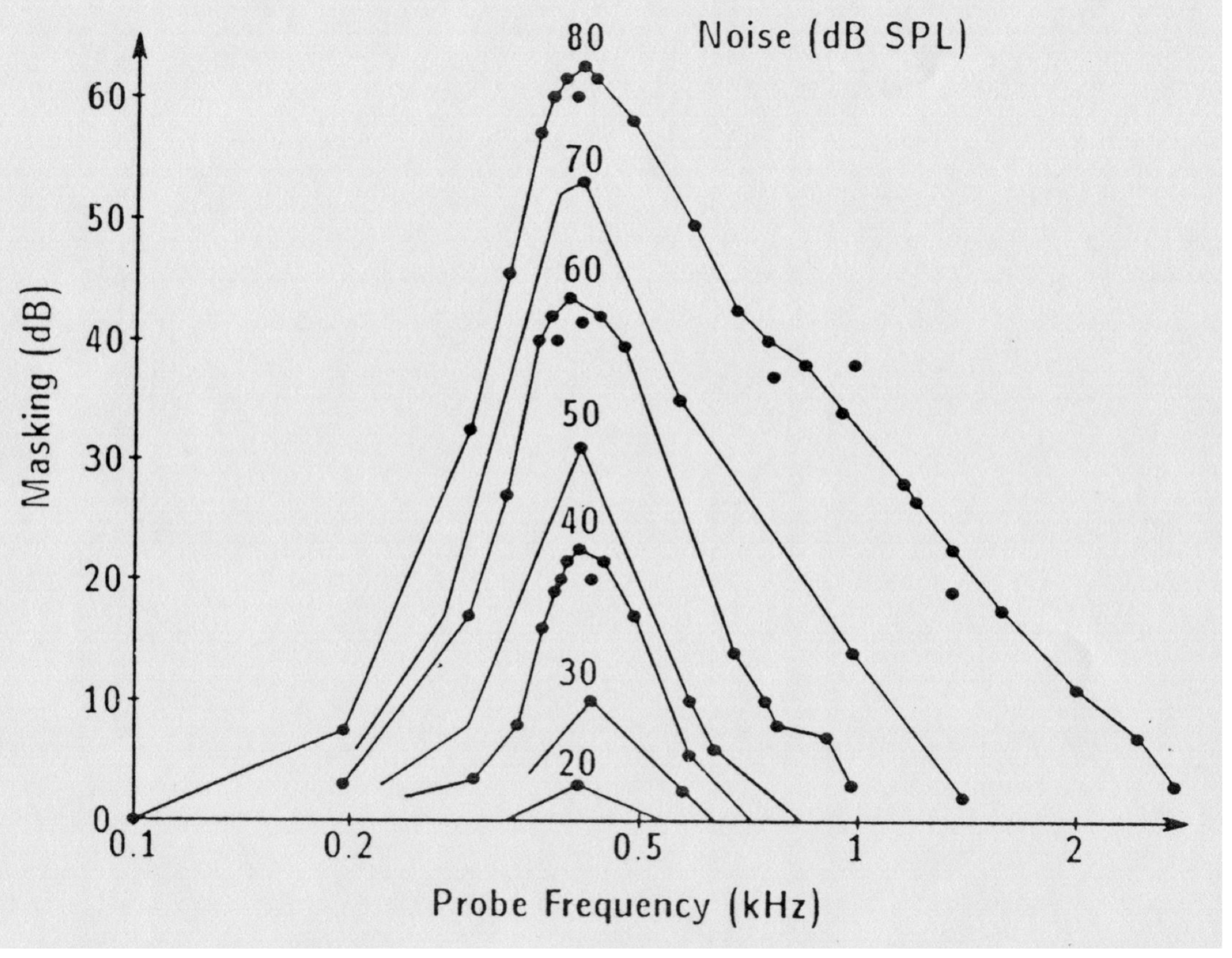
某一段時間存在一個聲音，在此聲音剛結束的一小段時間內，其他的聲音會聽不見這種遮蔽現象叫做時間遮蔽(temporal masking)。 這個現象也會發生在此存在聲音之前，這是由於聲音在聽覺器官中傳遞時延遲所造成的。

頻率遮蔽的
情形

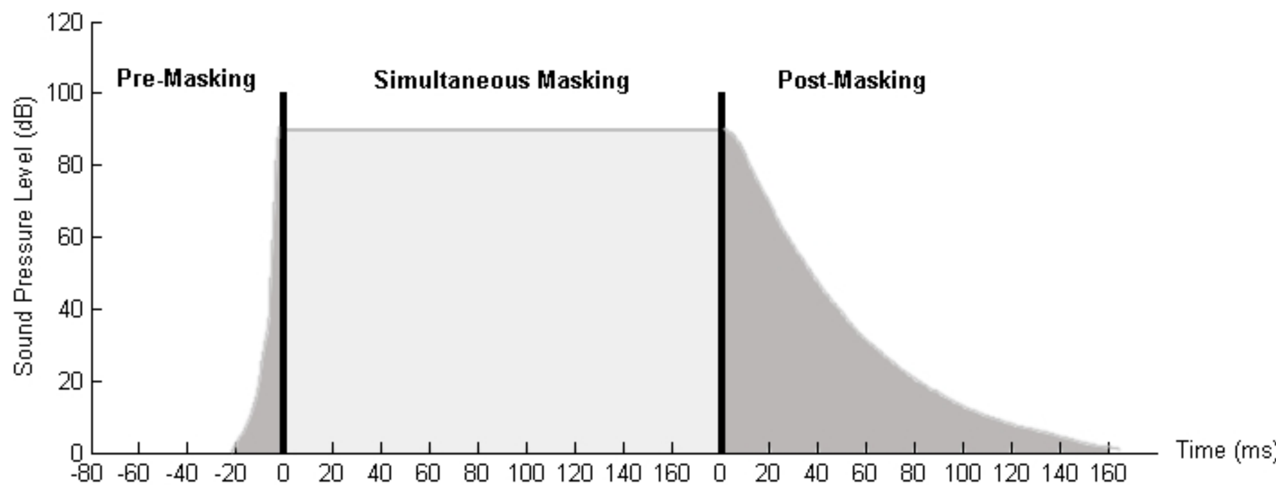


Masker
Masking Threshold
Threshold in Quiet
Masked Sound
Sound Pressure Level (dB)
Frequency (kHz)
70
60
50
40
30
20
10
0
.02
.05
0.1
0.2
0.5
1.0
2.0
5.0
10
20

遮蔽效應之 Pattern



時間遮蔽的情形



通常前遮蔽的時間很短，約20ms，而後遮蔽時間較長，約200ms。

臨界頻帶(critical band)

當我們改變窄帶聲音刺激(narrowband sound stimulus)時，若其聲音成分跨越某一頻率，在聽覺上就會感到有差異，而在一個頻率範圍內時，則不感覺到差異，這個頻率範圍就叫做臨界頻帶。

當兩個聲音訊號同在一個臨界頻帶範圍內時，能量高的聲音就把另一個聲音遮蔽掉。

如果有一個帶通的噪音維持了固定的音強，當其頻寬增加，聽到的響度仍是一樣，一直到頻寬跨過臨界頻帶時，聽起來就不一樣，覺得噪音響度變大，這是因為鄰近的神經元(neurons)也發射訊號。

Critical Ratio Experiment

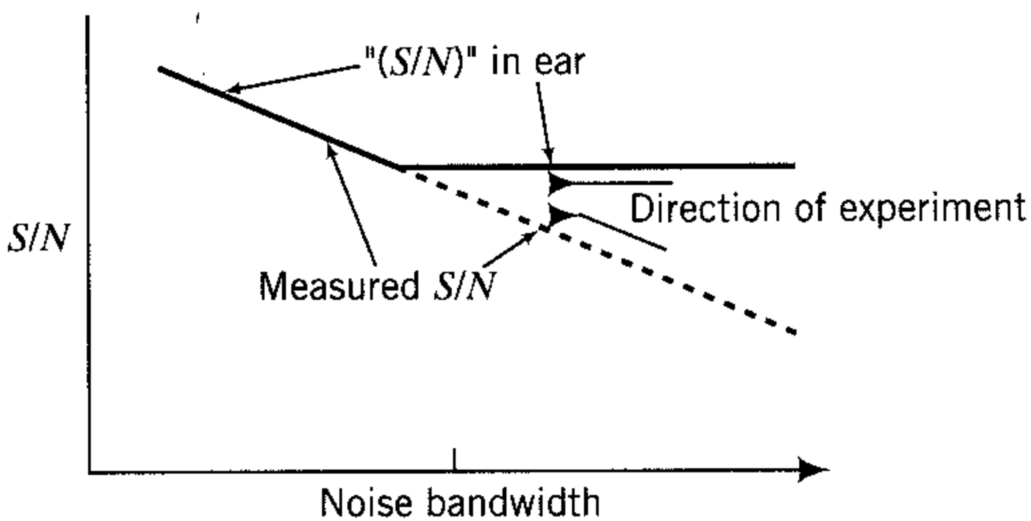


FIGURE 15.3 Critical ratio experiment. The solid line shows the apparent signal-to-noise ratio of the psychological filter; the dashed line shows the signal-to-noise ratio of the stimulus.

臨界頻帶濾波器(critical band filters)

臨界頻帶濾波器的形狀，可以用寬帶的低通與高通噪音去遮蔽一個頻率聲音而得到。

假設在一個臨界頻帶中，其中心頻率處有一個單頻率聲音，我們產生一個低通的噪音(low passed noise)，逐漸增加這個低通噪音的截止頻率(cutoff frequency)，使其超過中心頻率，這時候就會有更多的噪音能量進入此臨界頻帶。

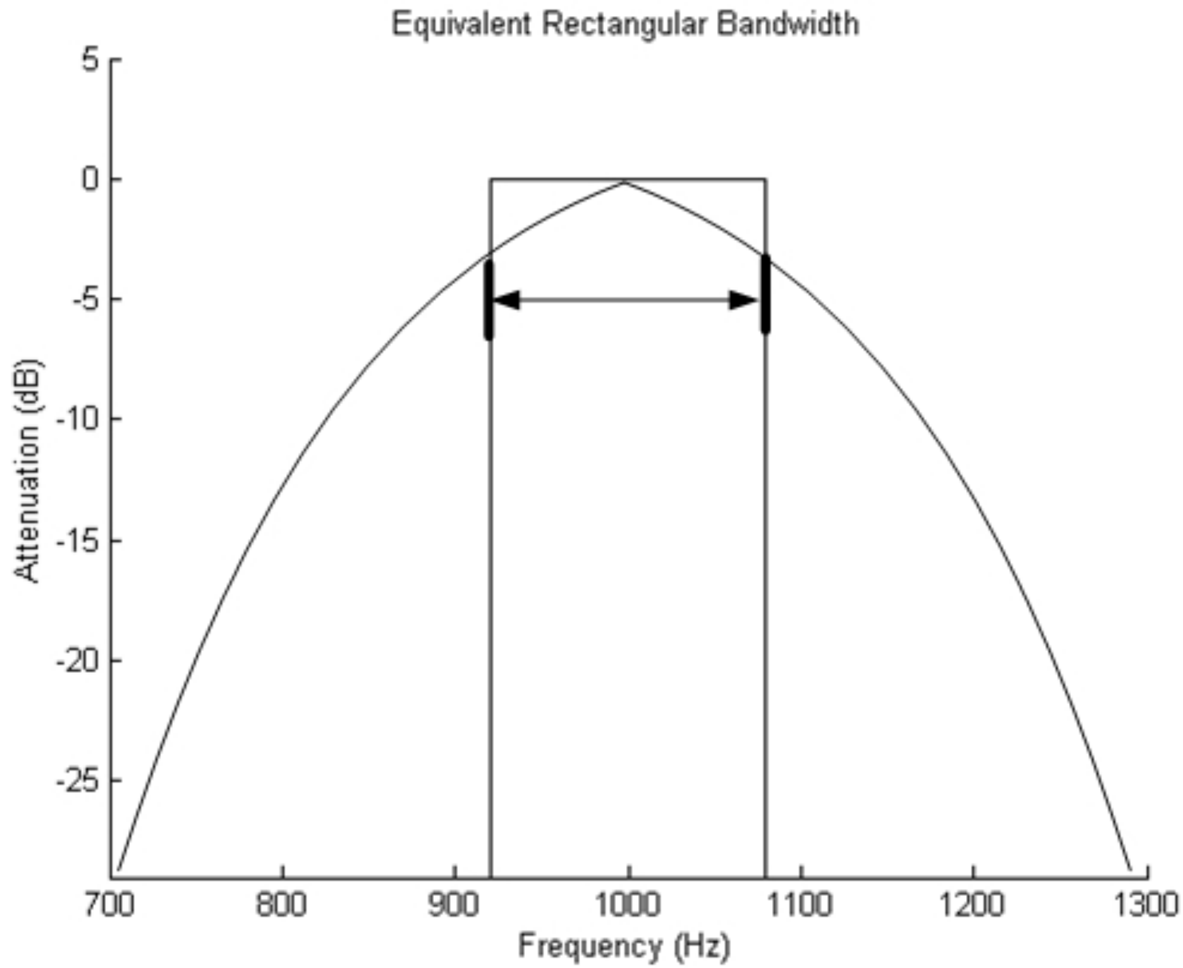
改變噪音與單頻率聲音之相對振幅，就可以找到正好遮蔽掉單頻率聲音的臨界值，這個值會是截止頻率的函數。

同樣的，用高通噪音(high passed noise)逐漸減低這個高通噪音的截止頻率，使其跨過中心頻率，一樣的去找出遮蔽掉單頻率聲音的臨界值，這個臨界值也是截止頻率的函數。

將前述兩個臨界值曲線繪出，就構成臨界頻帶濾波器的形狀。這個濾波器在線性頻率刻度上幾乎是對稱的，有很陡的裙邊，近似於一個三角形。

在500 Hz的臨界頻帶時，裙邊的陡降是65 dB/oct，在8 kHz的臨界頻帶，陡降是100 dB/oct。在500 Hz以下時，臨界頻帶的頻寬大約100 Hz，在較高頻時，頻寬逐漸增加，在大於1 kHz時，頻寬的增加呈對數比例。到了4 kHz時，頻寬為700 Hz。這些臨界頻帶濾波器有常數的Q值，也就是頻寬正比於頻率。

臨界頻帶濾波器的形狀



Equivalent Rectangular Bandwidth
5
0
-5
-10
-15
-20
-25
Attenuation (dB)
700
800
900
1000
1100
1200
1300
Frequency (Hz)

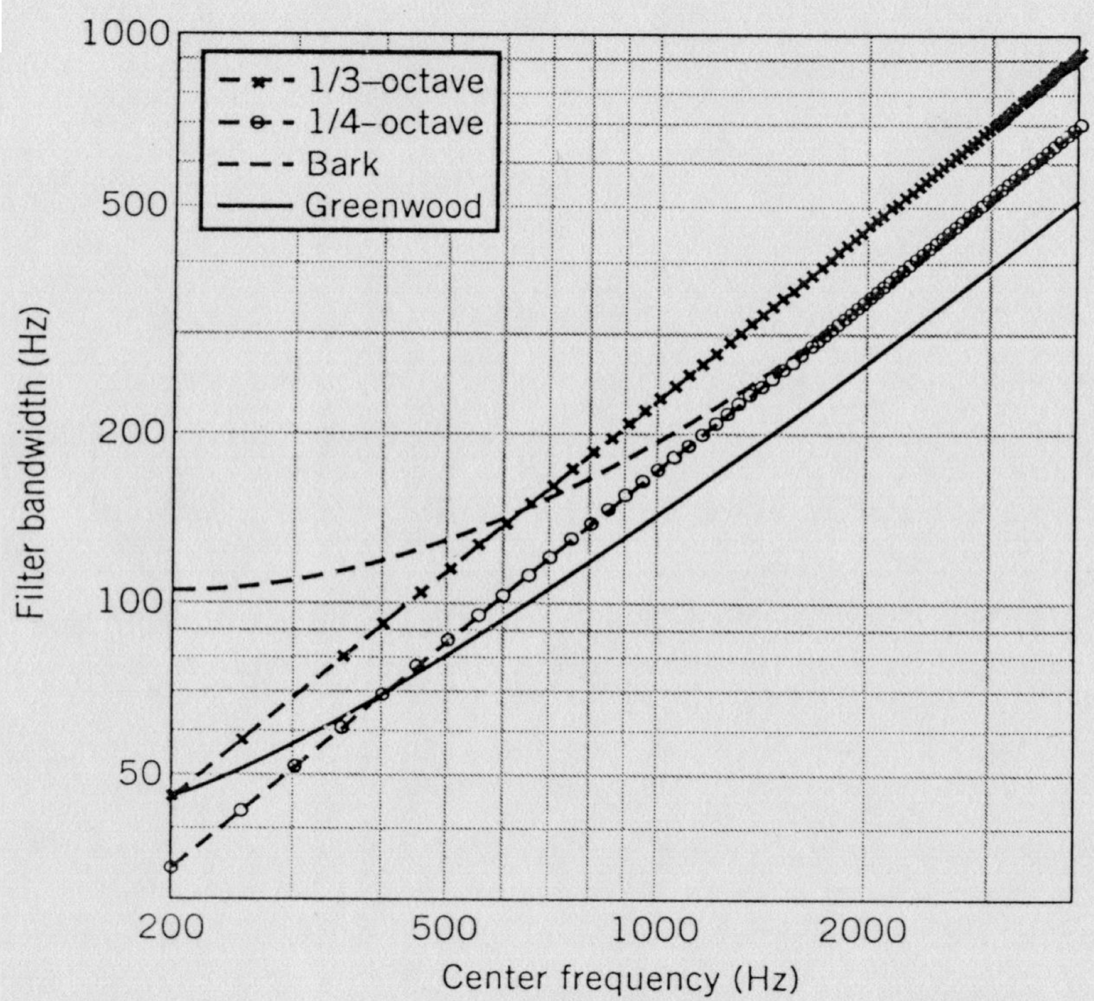
在人耳聽覺範圍內，可以分成24個臨界頻帶，對應在耳蝸的基膜上，大約是每1.5 mm一個臨界頻帶，其中含有1200條神經纖維。

24個臨界頻帶

頻帶編號	下限頻率	中心頻率	上限頻率
1	0	50	100
2	100	150	200
3	200	250	300
4	300	350	400
5	400	450	510
6	510	570	630
7	630	700	770
8	770	840	920
9	920	1000	1080
10	1080	1170	1270
11	1270	1370	1480
12	1480	1600	1720

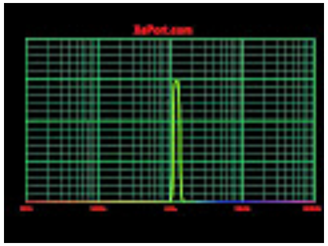
頻帶編號	下限頻率	中心頻率	上限頻率
13	1720	1850	2000
14	2000	2150	2320
15	2320	2500	2700
16	2700	2900	3150
17	3150	3400	3700
18	3700	4000	4400
19	4400	4800	5300
20	5300	5800	6400
21	6400	7000	7700
22	7700	8500	9500
23	9500	10500	12000
24	12000	13500	15500

中心頻率和頻寬之關係



frequency f
on, with two

- Earphone test - from bass to upper limit of human ear
  - https://www.youtube.com/watch?v=cvBtQmY2B5I



- How Old Are Your Ears? (Hearing Test)
  - https://www.youtube.com/watch?v=VxcbppCX6Rk&t=3s

- Hearing Test - Can You Hear the 1 kHz Tone?
  - https://www.youtube.com/watch?v=k6DVywW5NR4&index=62&list=PLQjOyrNQ5tImVbYxybZQQwKCtPIUtyE-t

- Can You Trust Your Ears? (Audio Illusions)
  - https://www.youtube.com/watch?v=kzo45hWXRWU