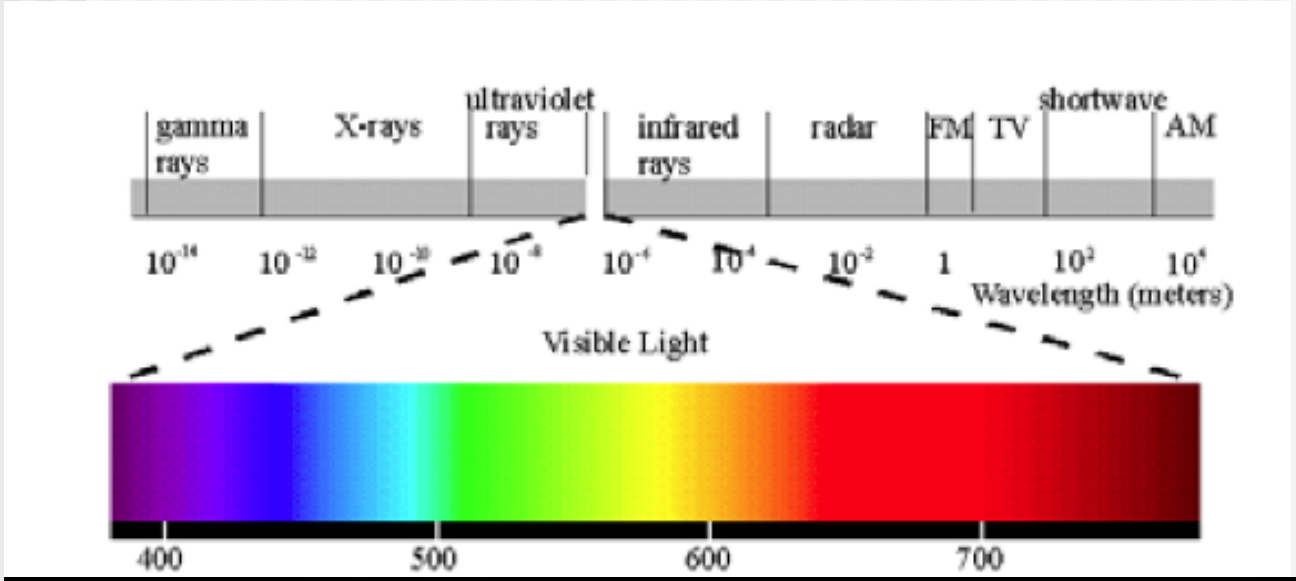
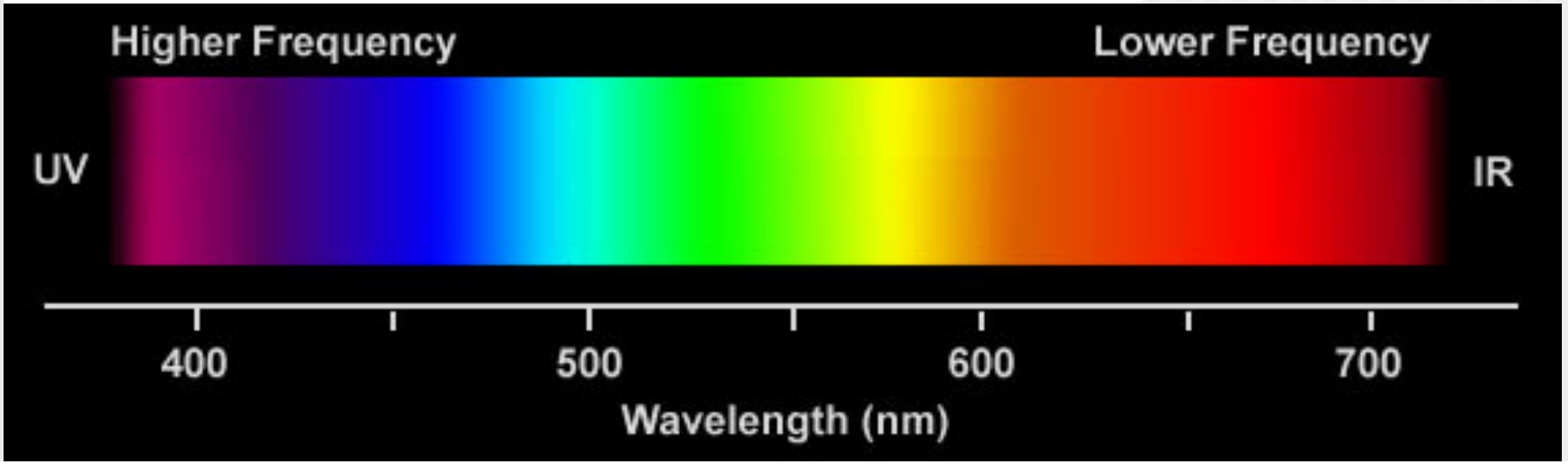
Human Acoustic Processing

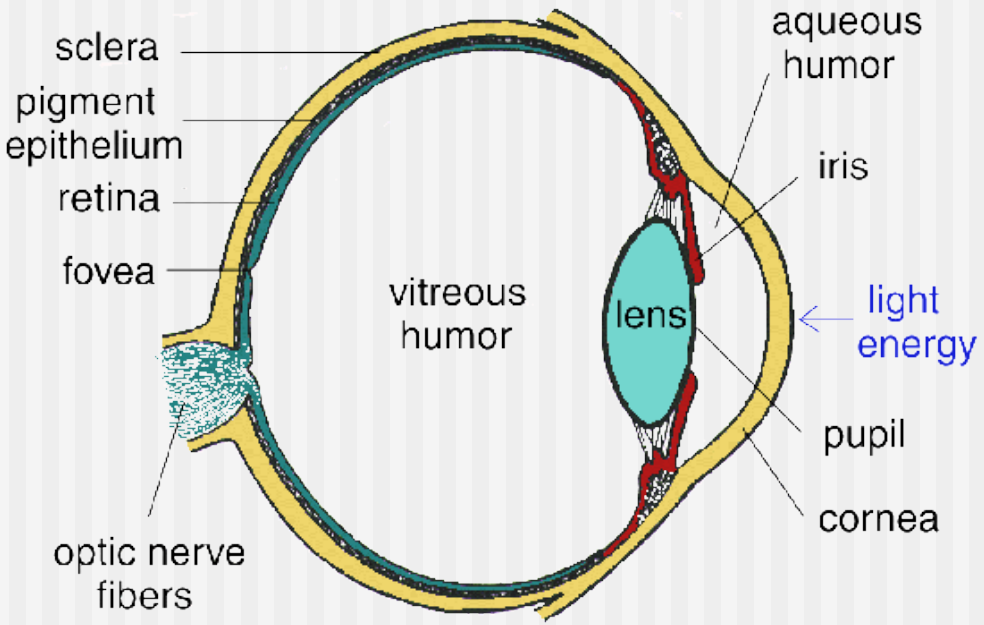
- Human Eye and Retina
  - Visible Spectrum, Color Sensitivity
- Sound and Light
- The Ear
    - Cochlea
- Auditory Pathway
- Speech Spectrogram
    - Vocal Cords
    - Formant Frequencies
- Time Warping
- Hidden Markov Models
- Signal, Time and Brain
    - Process of temporal integration, perceptual identity (Pöppel)

Visible Spectrum

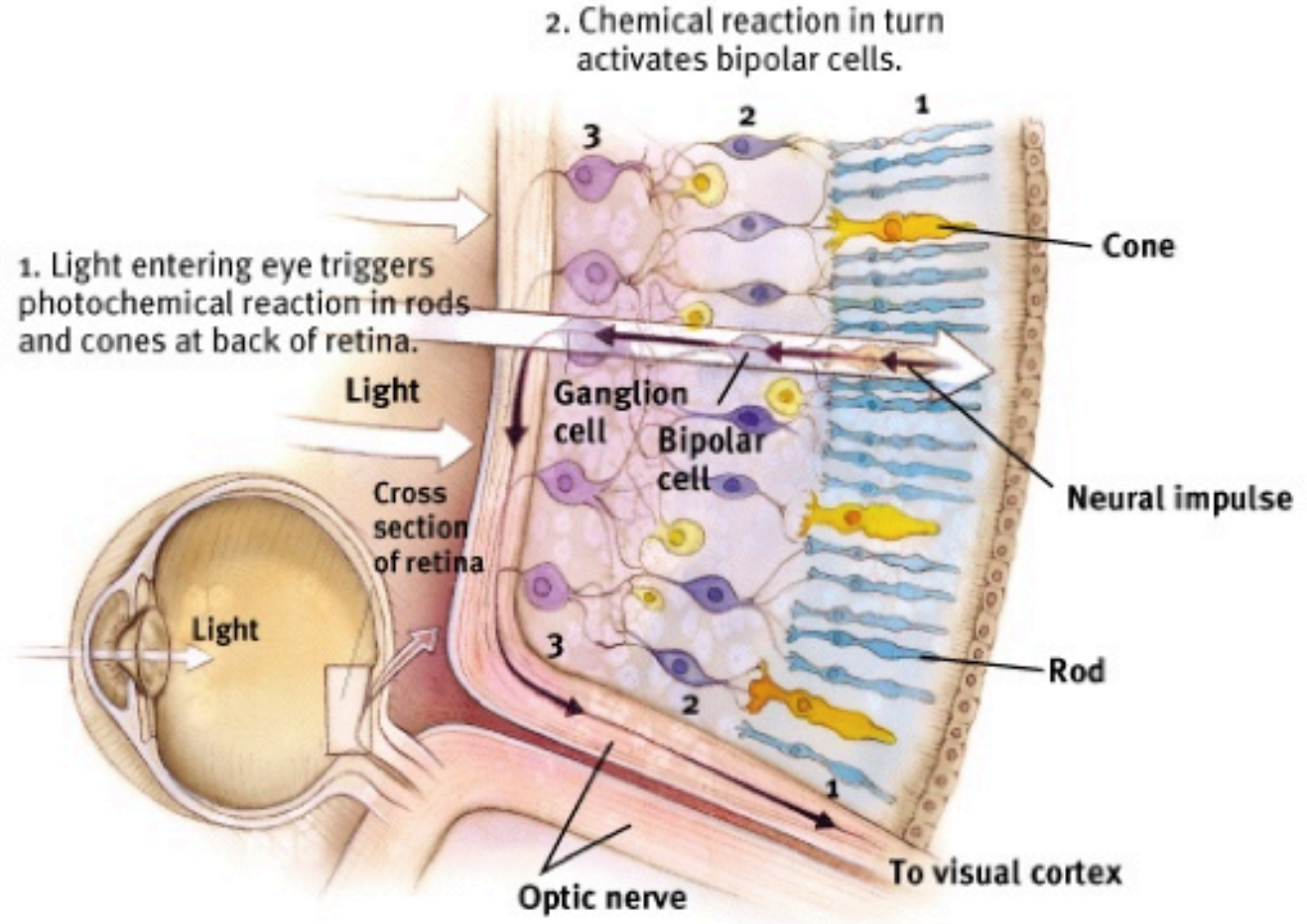




Human Eye


sclera
pigment epithelium
retina
fovea
optic nerve fibers
vitreous humor
lens
aqueous humor
iris
light energy
pupil
cornea

Eye and Retina



3. Bipolar cells then activate the ganglion cells, the axons of which converge to form the optic nerve. This nerve transmits information to the visual cortex in the brain's occipital lobe.

Signal propagation in the retina:

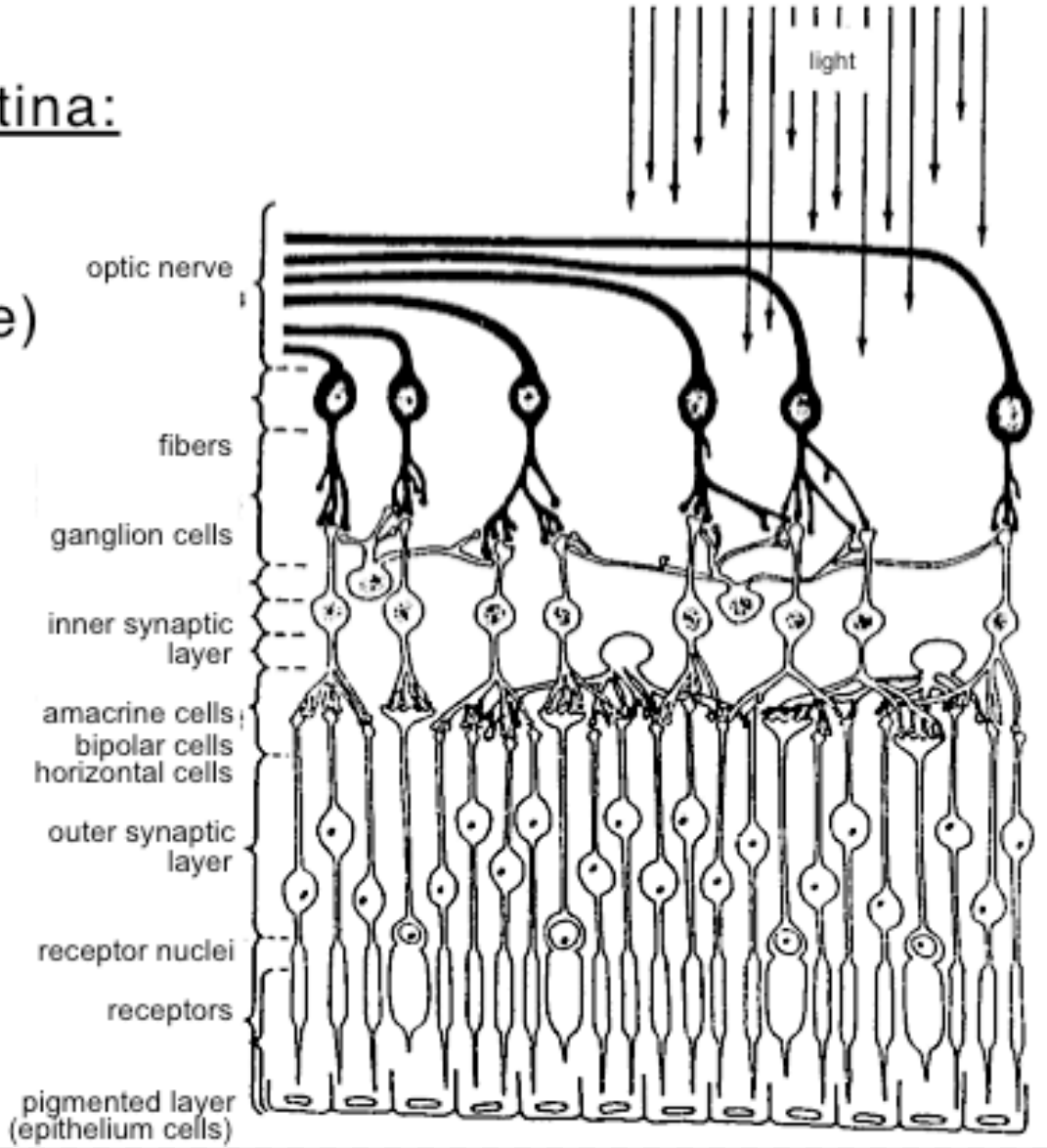
ganglion cells (=optic nerve)

↑

bipolar cells

↑

receptors

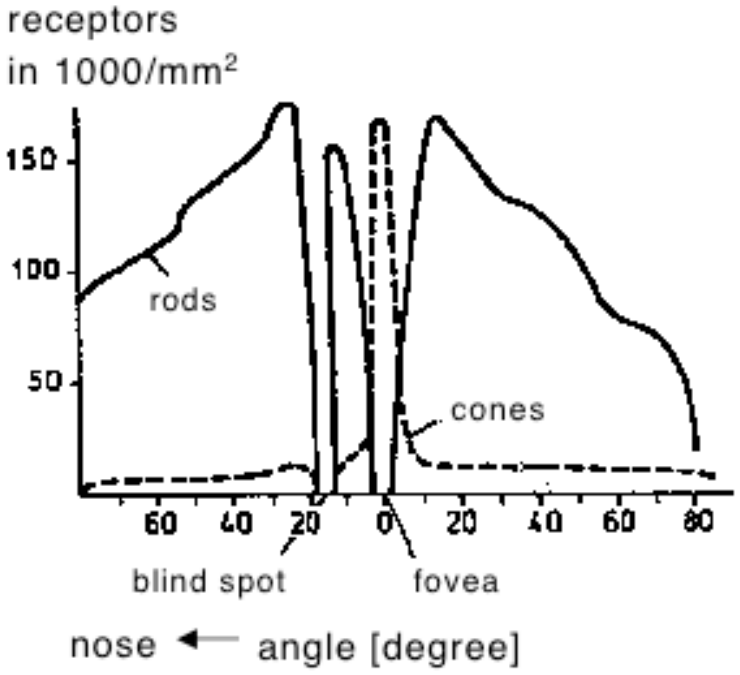


Retina

- Retina covered with light-sensitive receptors
  - rods
    - primarily for night vision & perceiving movement
    - sensitive to broad spectrum of light
    - can’t discriminate between colors
    - sense intensity or shades of gray
  - cones
    - used to sense color

Rods and Cones

Rods	Cones
high sensitivity	low sensitivity
low light vision	day light vision > 1 cd/m^2
monochrome	color
"scotopic vision"	"photopic vision"

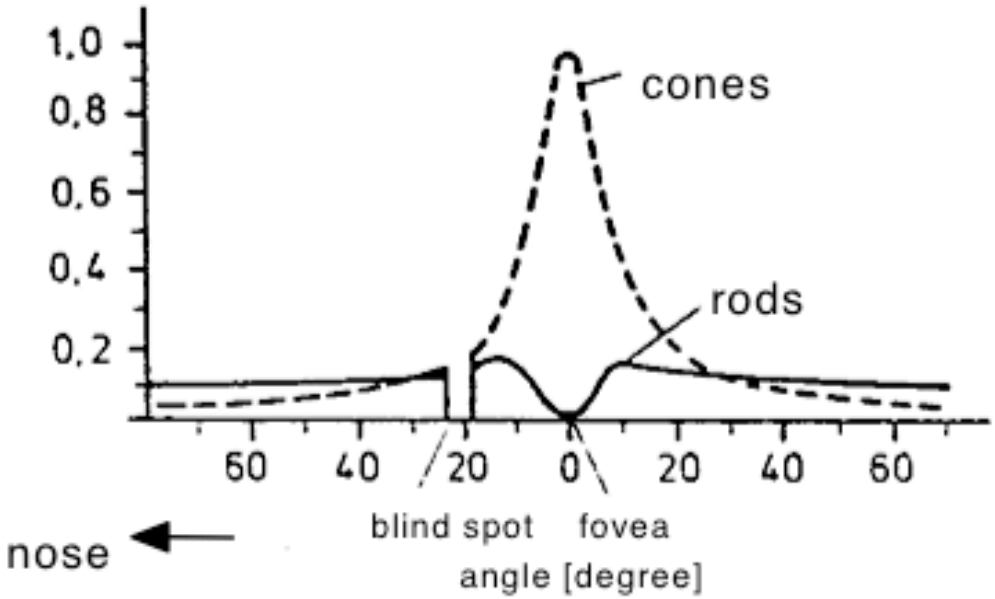


Video displays

Retina

- Center of retina has most of the cones (color)
  - allows for high acuity of objects focused at center
- Edge of retina is dominated by rods (shades of gray)
  - allows detecting motion of threats in periphery

- Spatial resolution in lines/arcmin:



Color Perception via Cones

- "Photopigments" used to sense color
- 3 types: blue, green, "red" (really yellow)
  - each sensitive to different band of spectrum
  - ratio of neural activity of the 3 $\rightarrow$ color
    - other colors are perceived by combining stimulation

Color Sensitivity

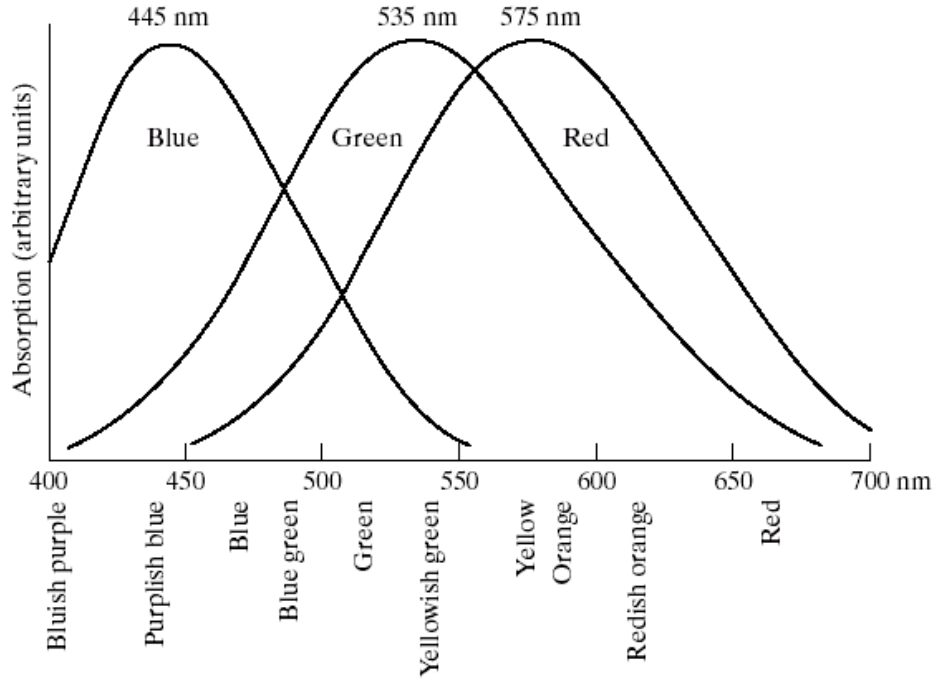
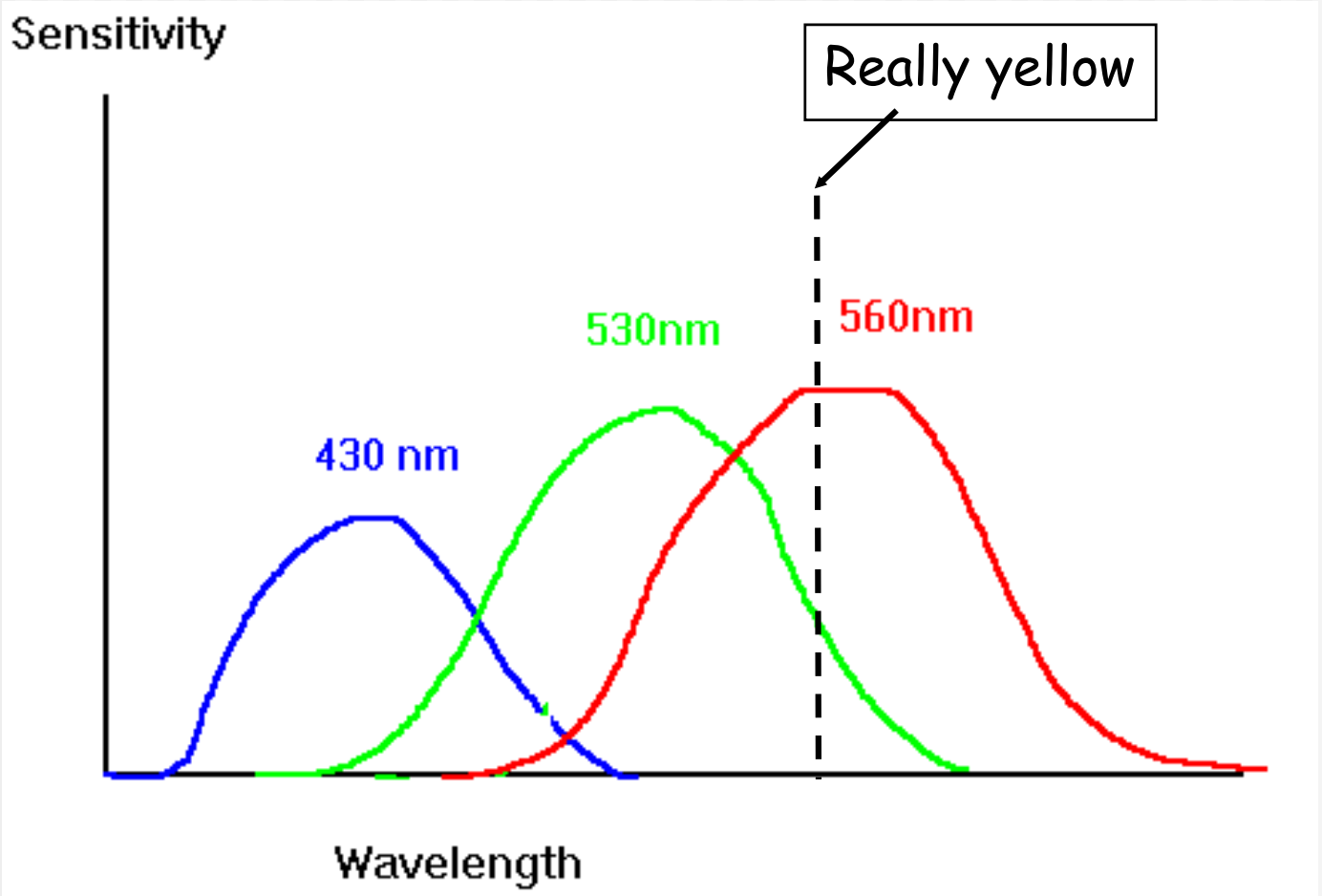


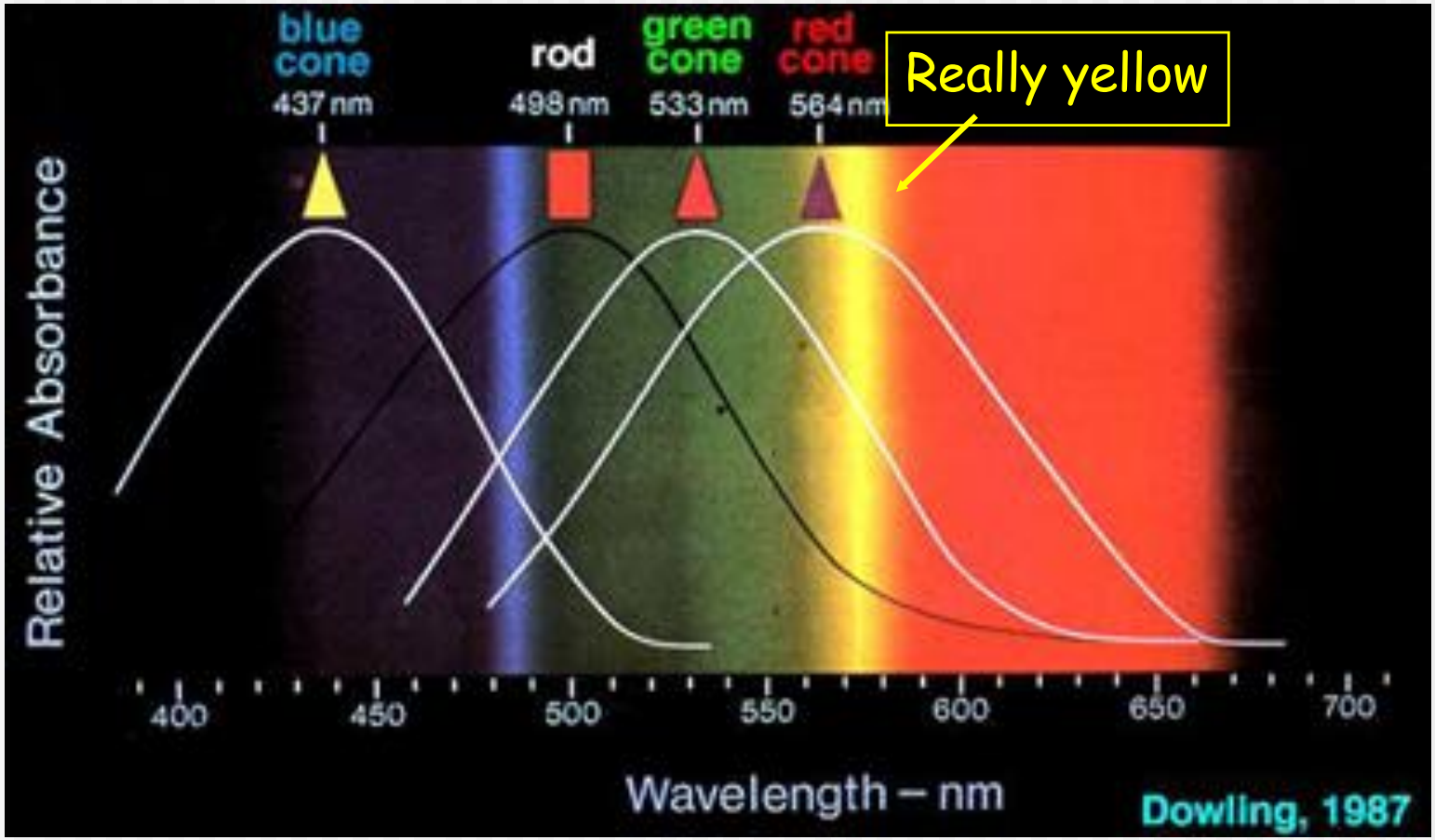
FIGURE 6.3 Absorption of light by the red, green, and blue cones in the human eye as a function of wavelength.

Color Sensitivity



from: http://www.cs.gsu.edu/classes/hypgraph/color/coloreff.htm

Color Sensitivity

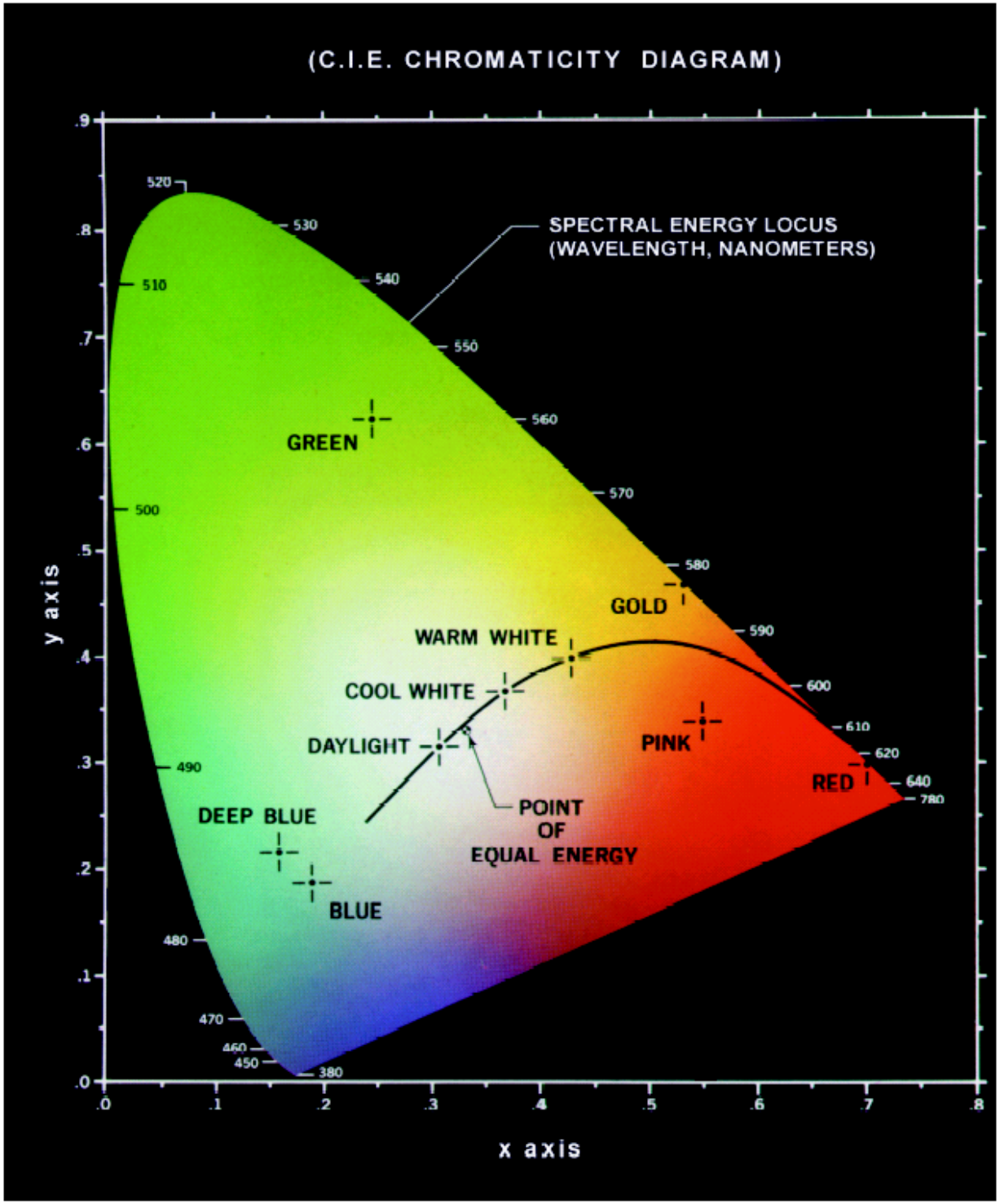


from http://insight.med.utah.edu/Webvision/index.html

Distribution of Photopigments

- Not distributed evenly
  - mainly reds (64%)
  - very few blues (4%)
    - insensitivity to short wavelengths
      - cyan to deep-blue
  - Center of retina (high acuity) has no blue cones
    - disappearance of small blue objects you fixate on

FIGURE 6.5
Chromaticity diagram. (Courtesy of the General Electric Co., Lamp Business Division.)



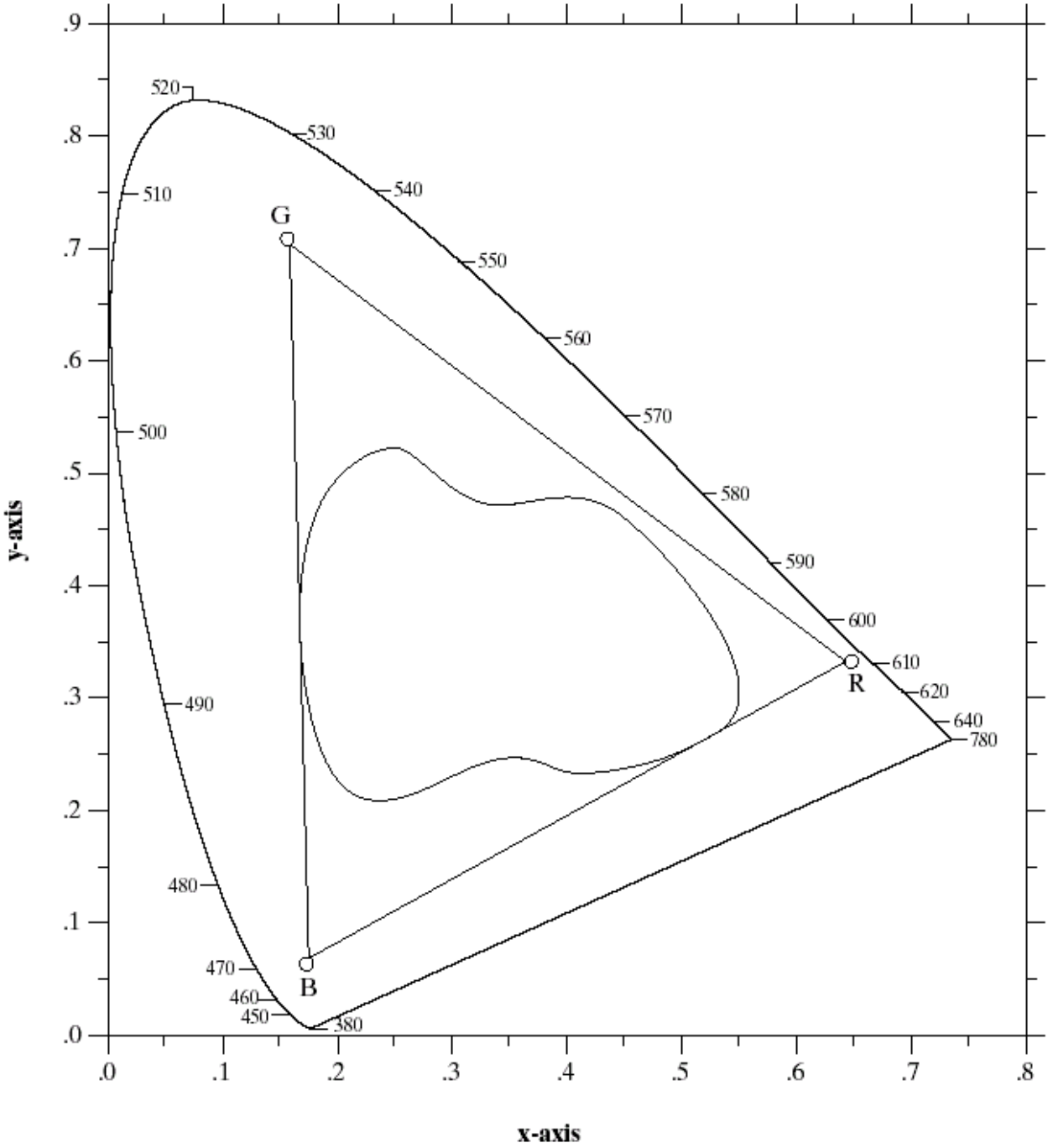
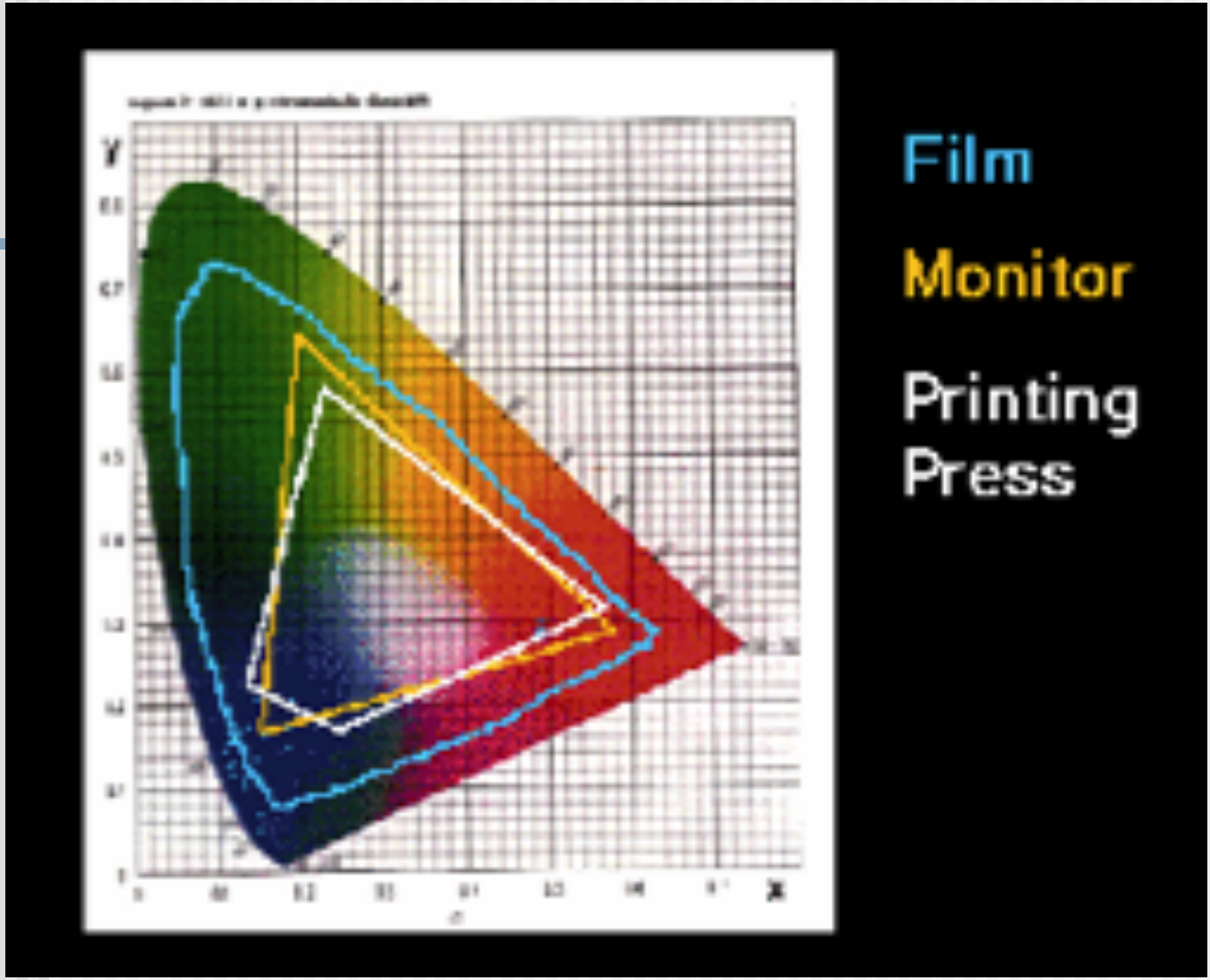
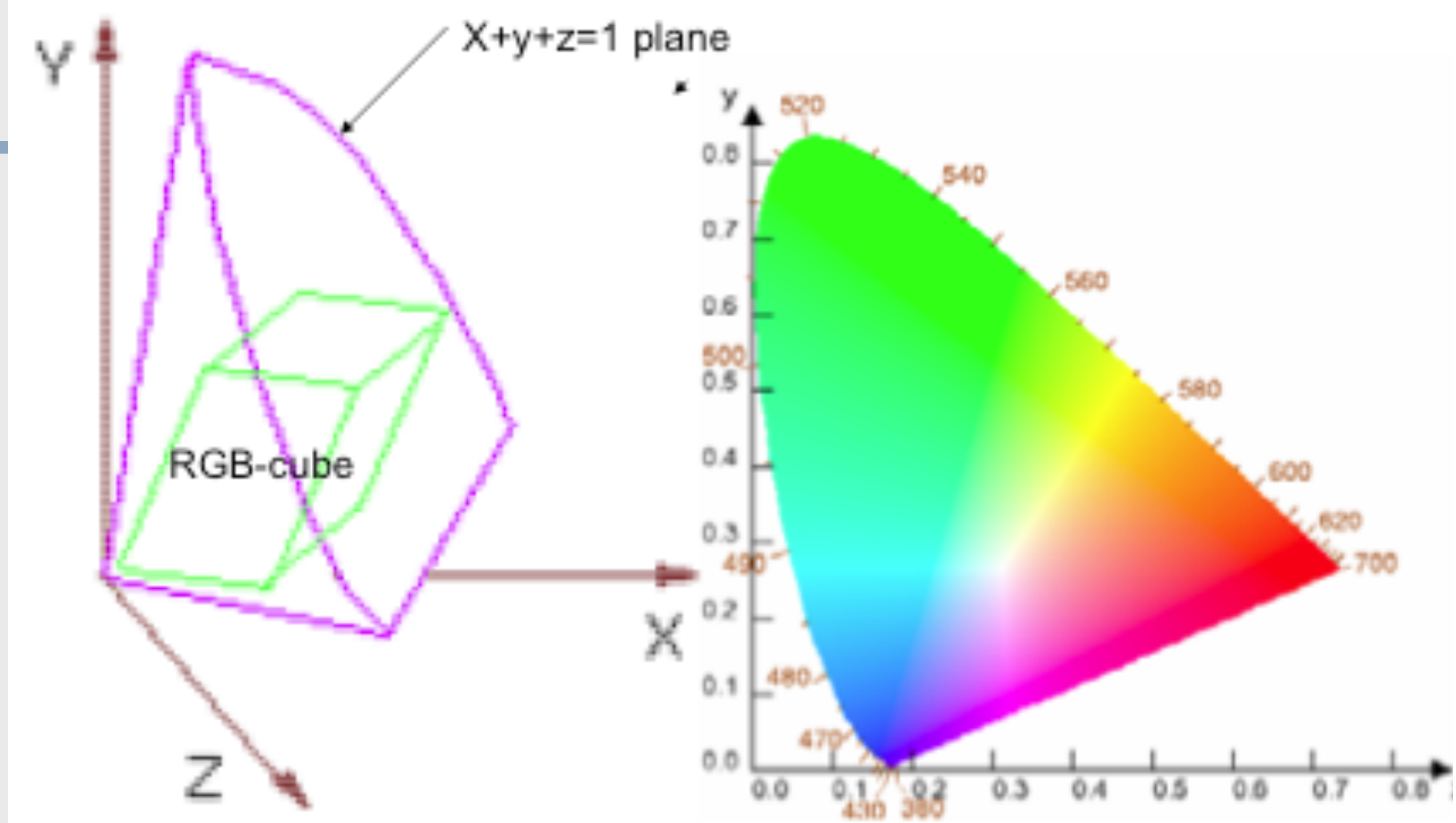


FIGURE 6.6 Typical color gamut of color monitors (triangle) and color printing devices (irregular region).


Film
Monitor
Printing
Press

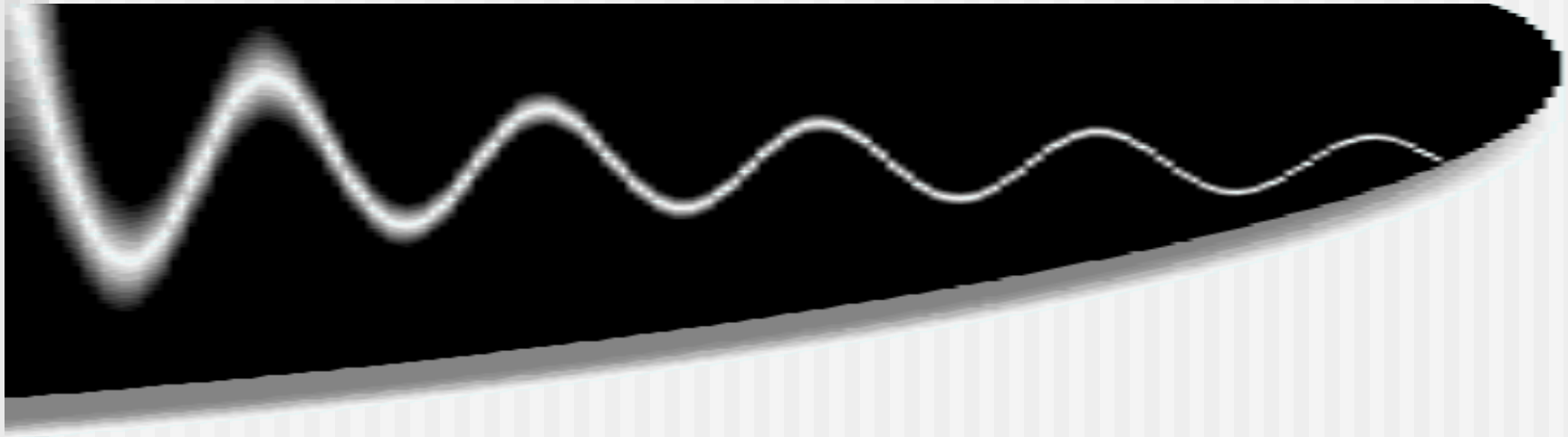

Y
X+y+z=1 plane
RGB-cube
Z
X
y
520
540
560
580
600
620
700
500
490
480
470
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
430
380

Color Deficiency (also known as “color blindness”)

- Trouble discriminating colors
  - besets about 9% of population
  - two major types
- Different photopigment response
  - reduces capability to discern small color diffs
    - particularly those of low brightness
  - most common
- Red-green deficiency is best known
  - lack of either green or red photopigment $\rightarrow$?
    - can’t discriminate colors dependent on R & G

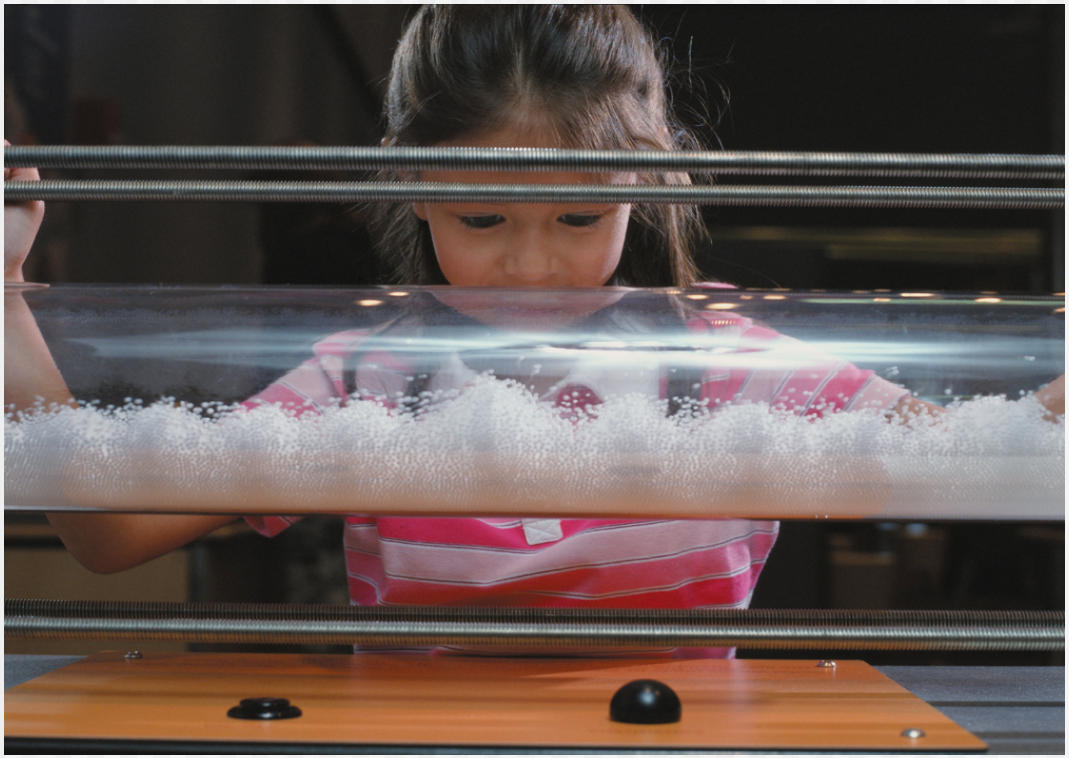
SOUND vs. LIGHT

- Sound travels through space as waves, but these waves are much longer that light waves
- Sound is only a wave which is transmitted by a medium like air, water, etc..
- Sound can bend around corners, unlike light which travels in straight lines



SOUND: Properties

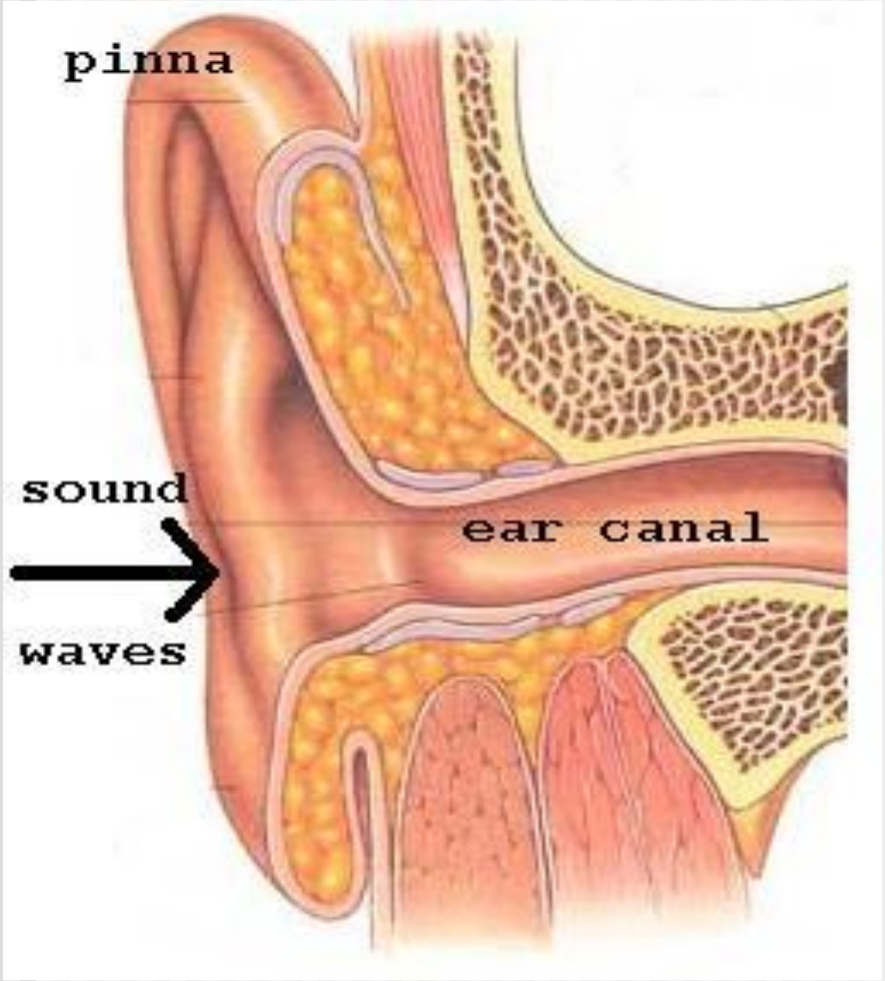
- Purely mechanical – it is produced by the vibration of an object
- Presence of a sound wave changes local air pressure
- Speed of travel depends on medium



THE EAR: Structure

- THE OUTER EAR: Gathers sound and directs it to the ear drum
- THE MIDDLE EAR: Transforms sound energy for inner ear
- THE INNER EAR: Location of auditory receptors

THE EAR: The outer ear

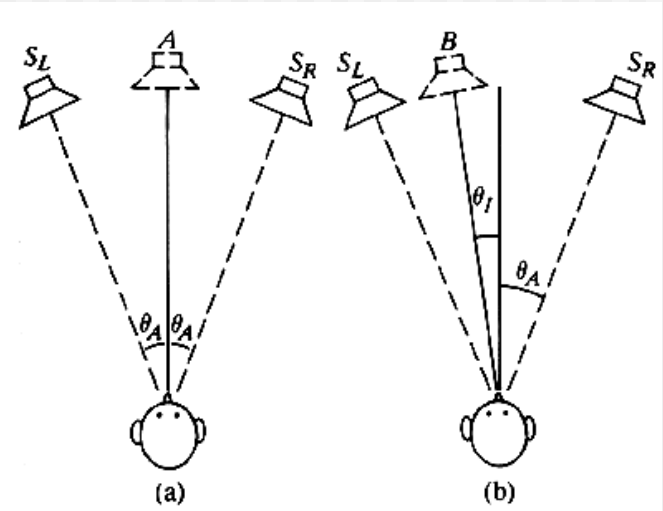


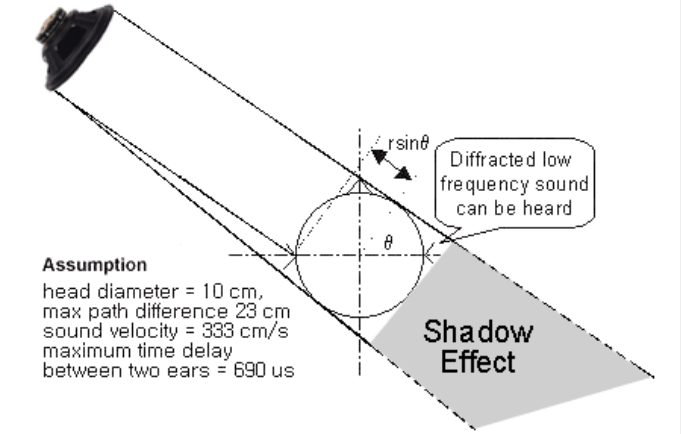
- PINNA: Important for sound gathering and localization of sound
- EAR CANAL or AUDITORY MEATUS: important for sound selection
- EARDRUM or TYMPANIC MEMBRANE:

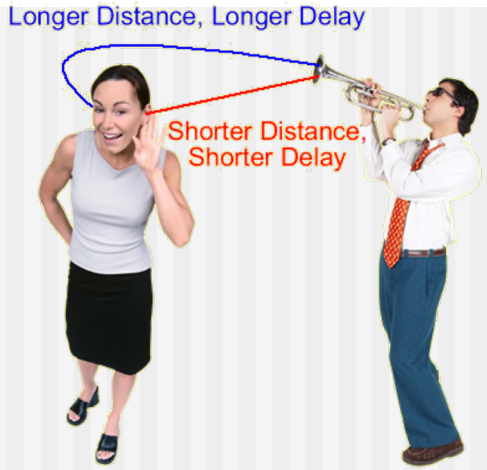
  vibrates in response to sound/pressure change

Pinna

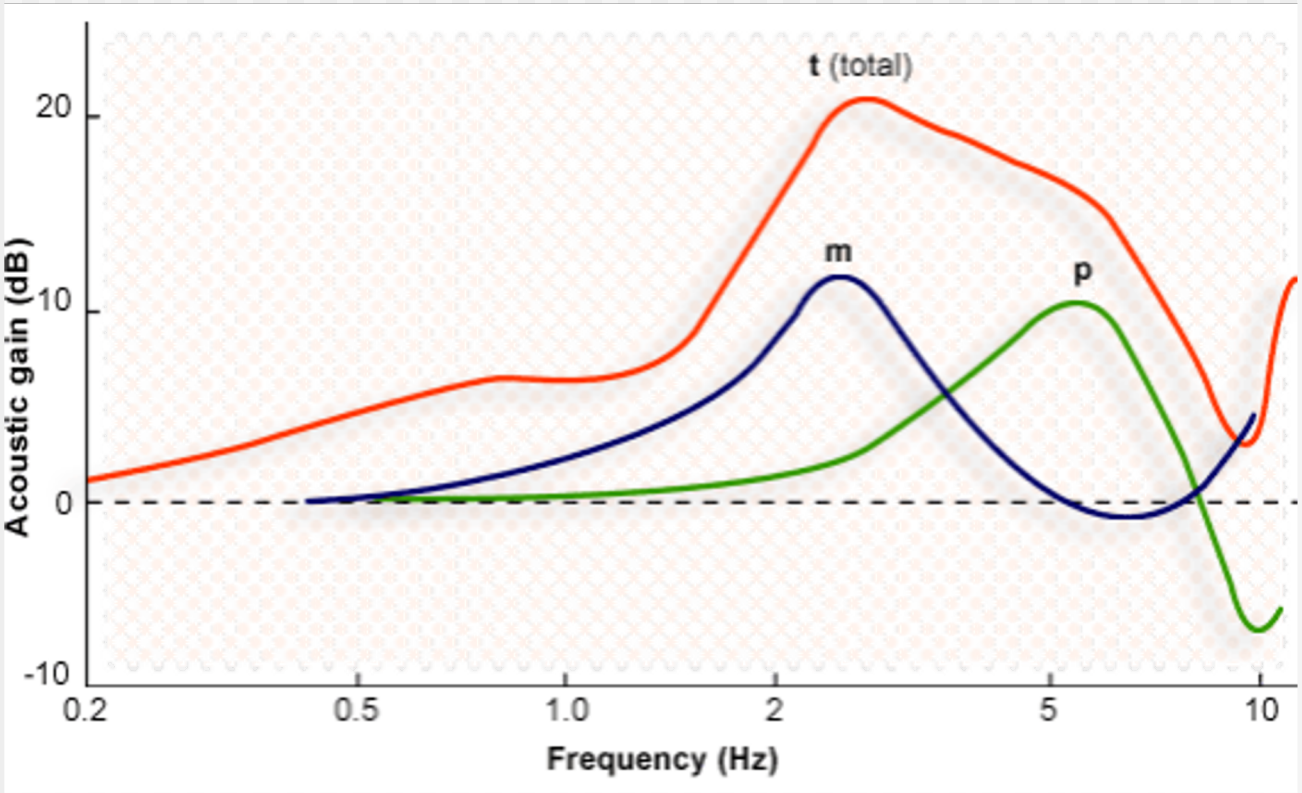
- The visible portion that is commonly referred to as "the ear"
- Helps localize sound sources
- Directs sound into the ear
- Each individual's pinna creates a distinctive imprint on the acoustic wave traveling into the auditory canal





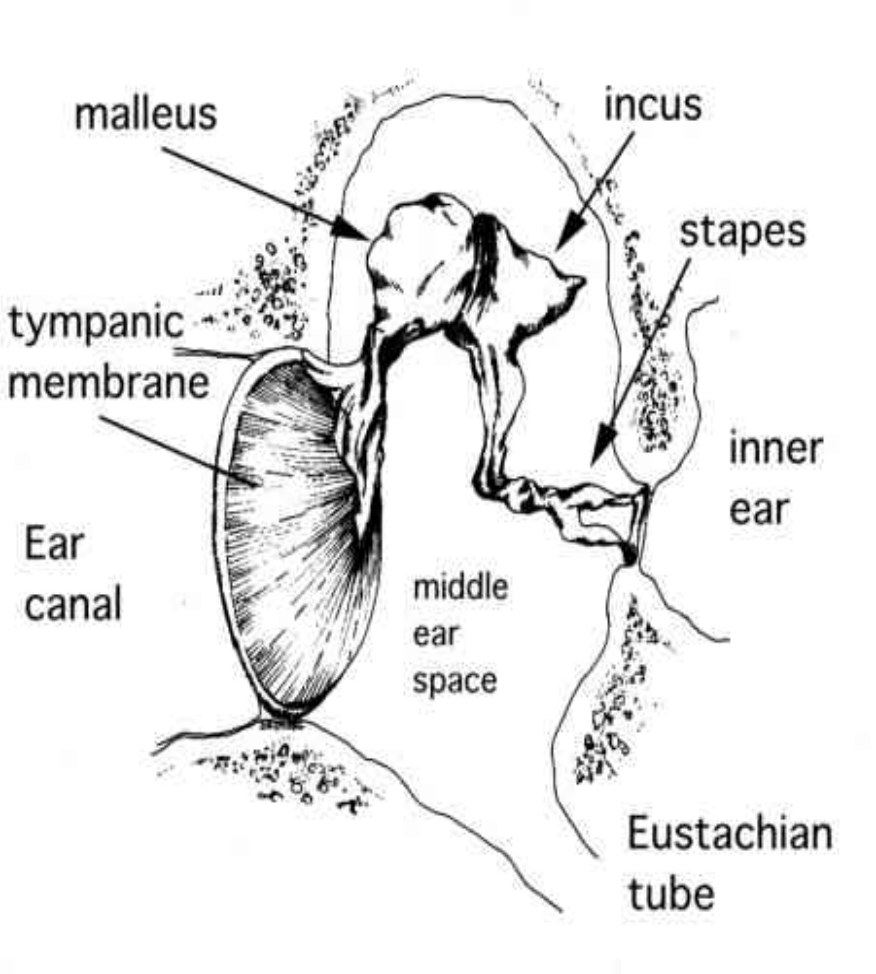


Outer Ear Resonance



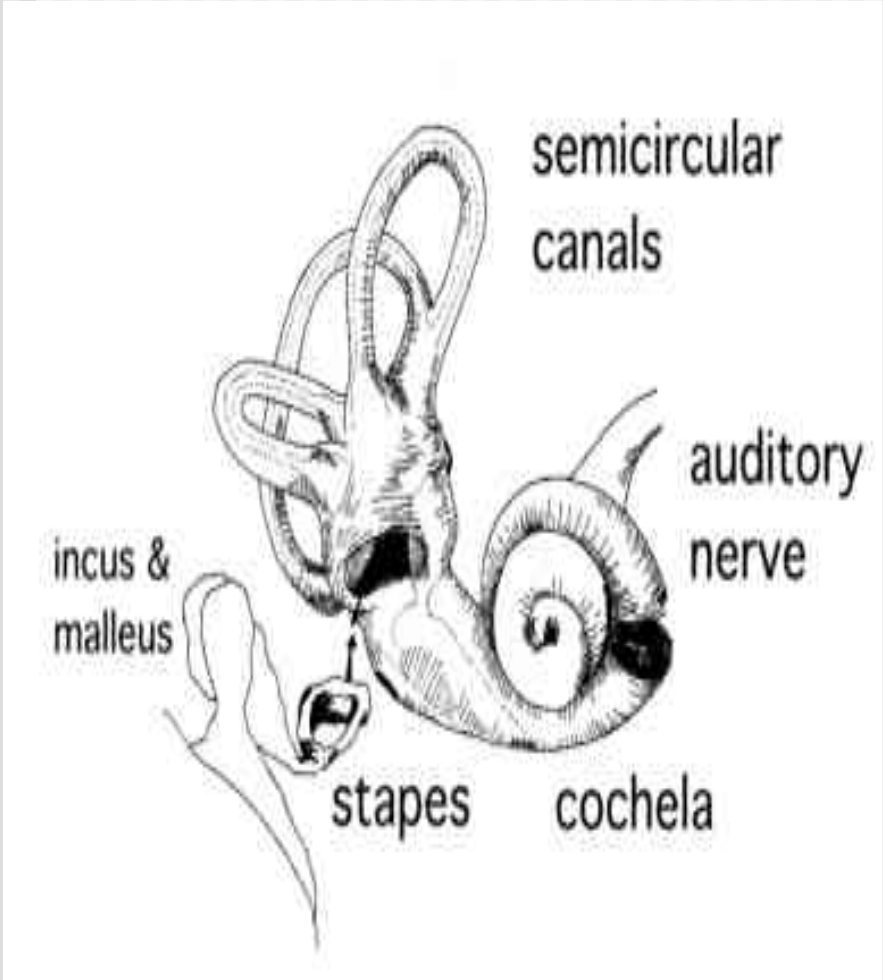
- Influence of pinna (p), Influence of ear canal (m)
- Combine influence (t)

THE EAR: The Middle Ear



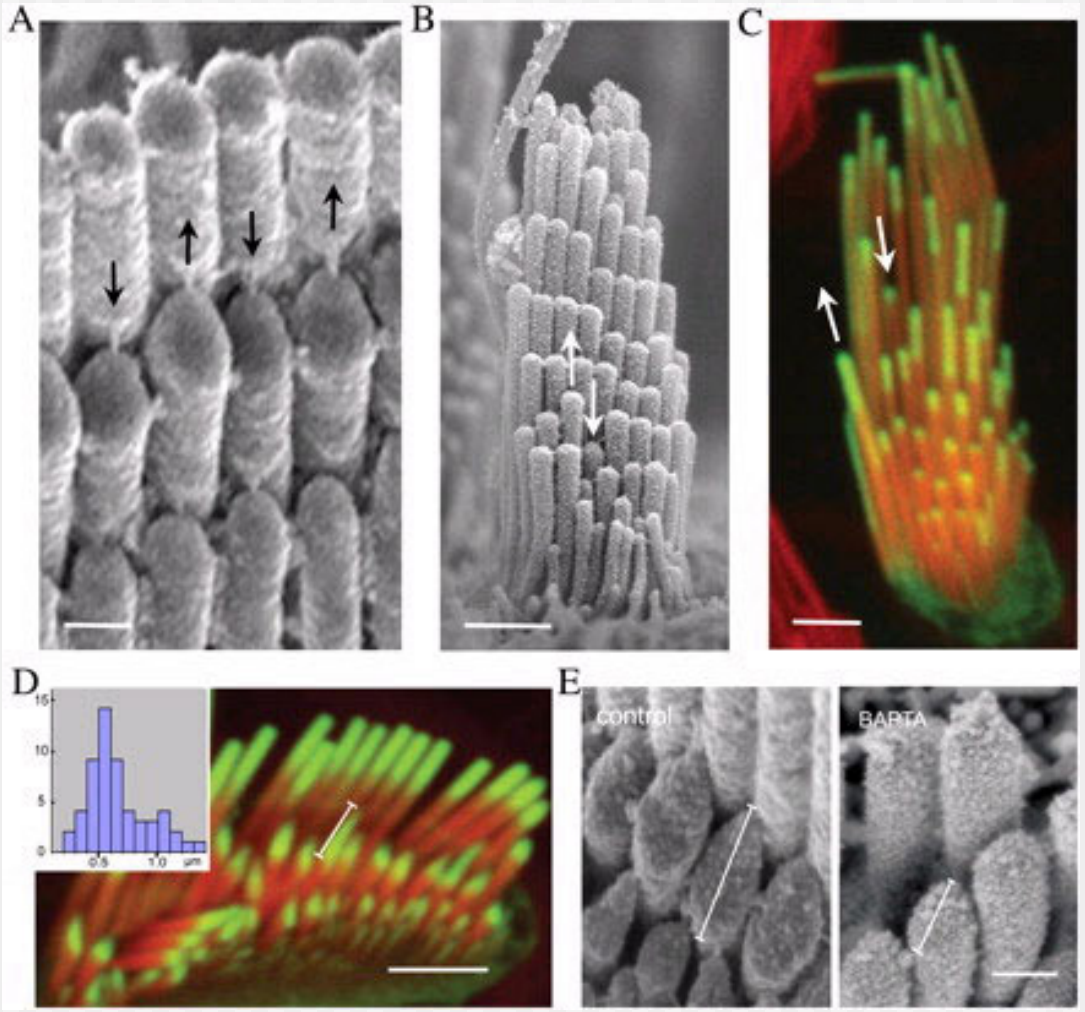
- OSSICLES: malleus, incus, and stapes; the stapes is the output signal of middle ear
- Important for sound transformation and reduction of potentially harmful sounds

THE EAR: The Inner Ear

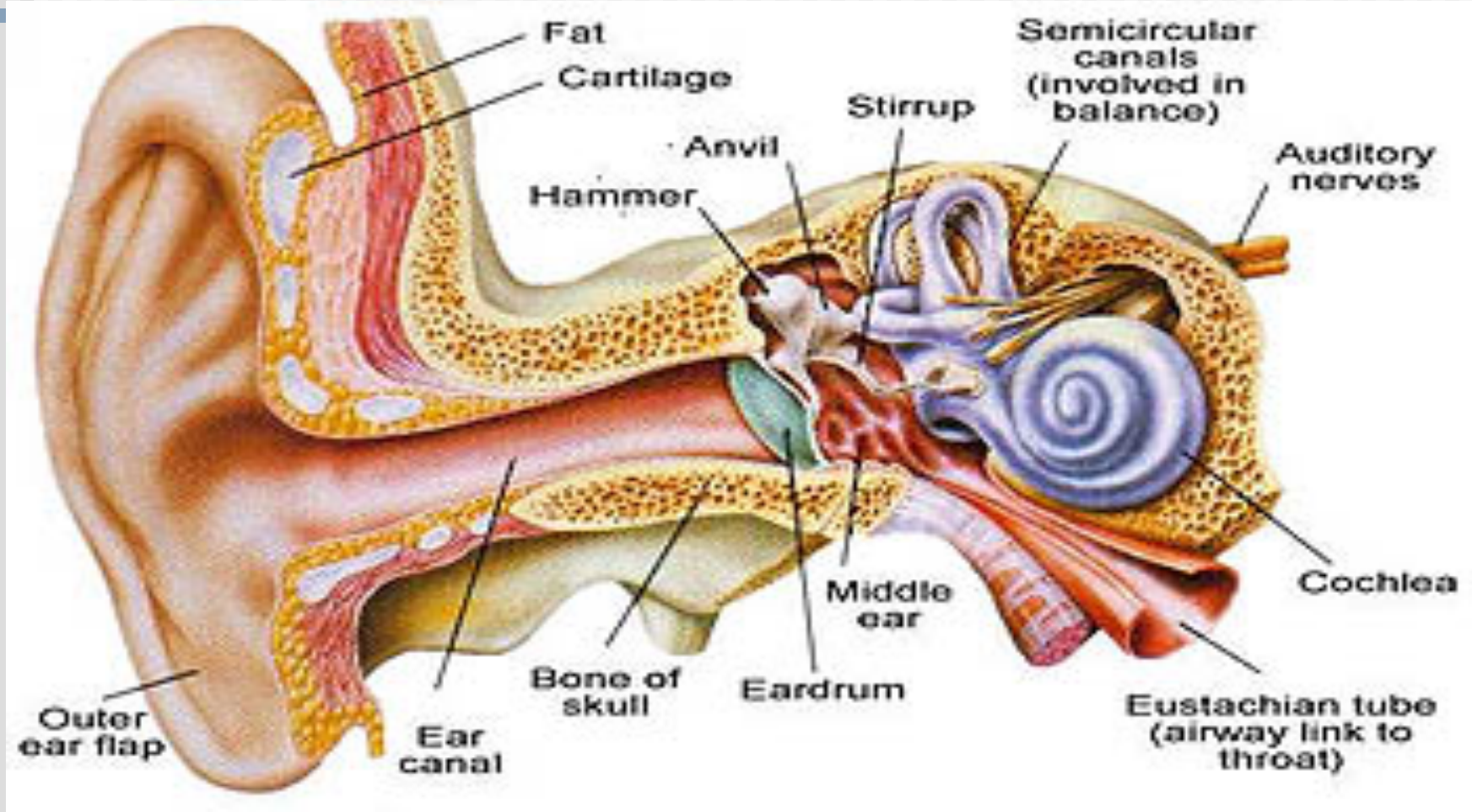


- SEMICIRCULAR CANALS: fluid-filled tubes important for balance
- COCHLEA: fluid-filled, spiral-shaped structure that contains **auditory receptors (hair cells)**

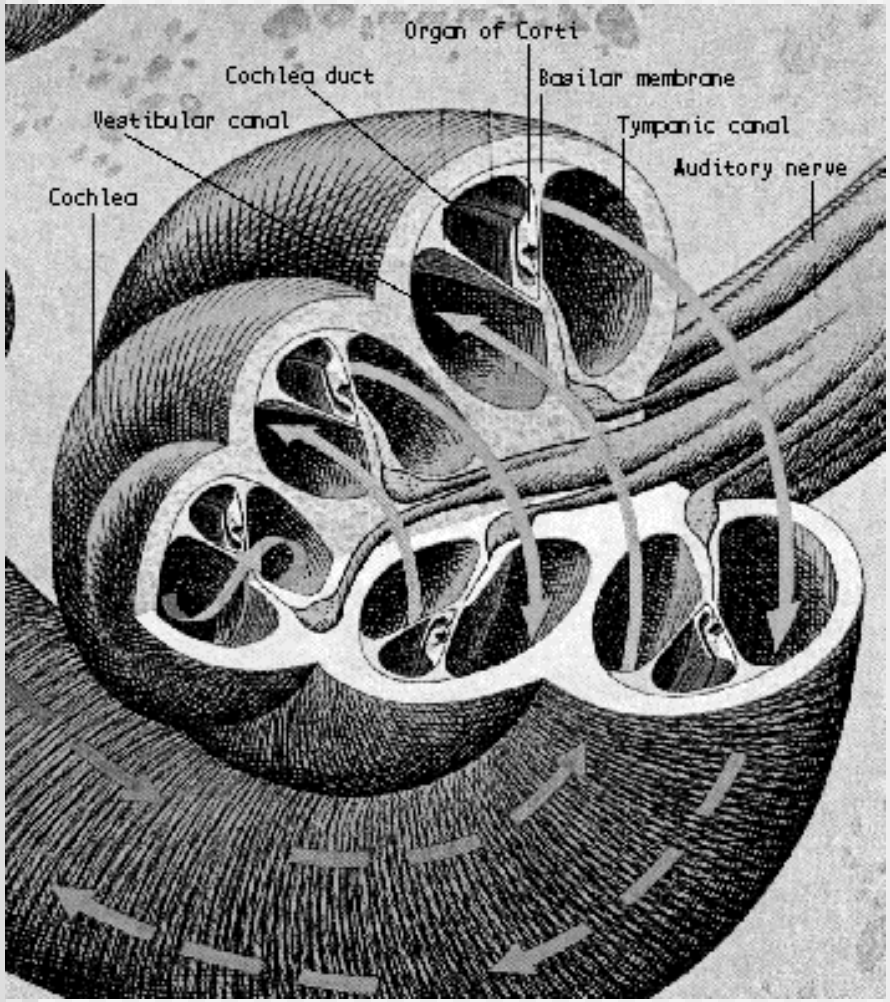
Auditory receptors (hair cells)



THE EAR: Diagram

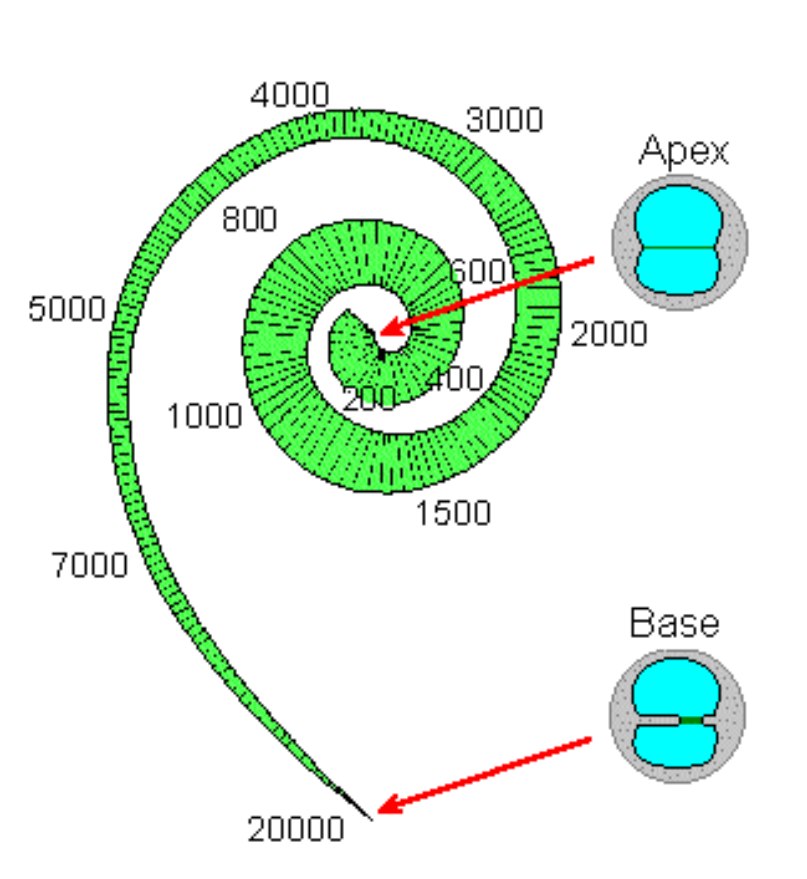


Cochlear section

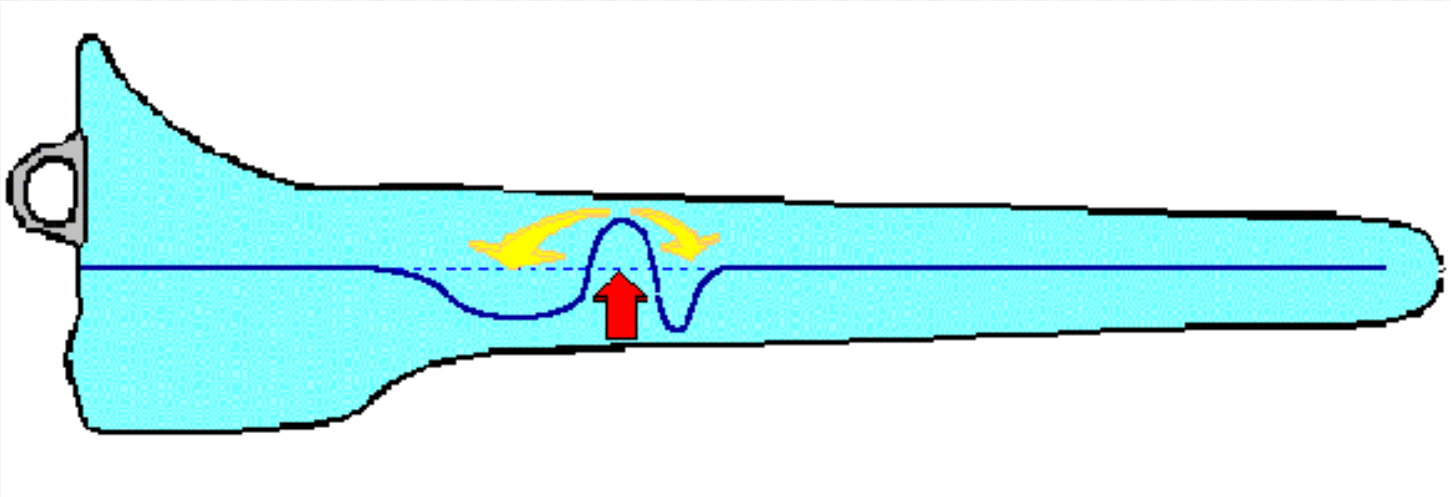


- Cochlea, or inner ear, has a spiral form:
  - vestibular canal
  - basilar membrane
  - tympanic canal
  - auditory nerve

Cochlea

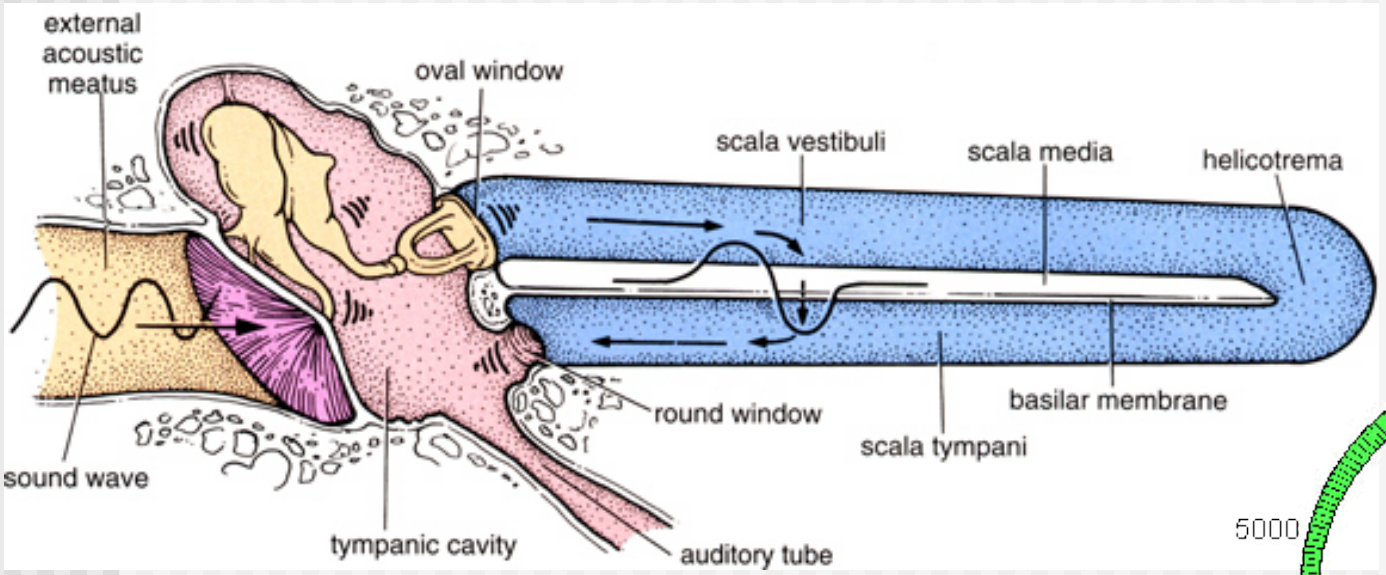

4000
3000
Apex
800
600
5000
2000
400
200
1000
1500
7000
Base
20000

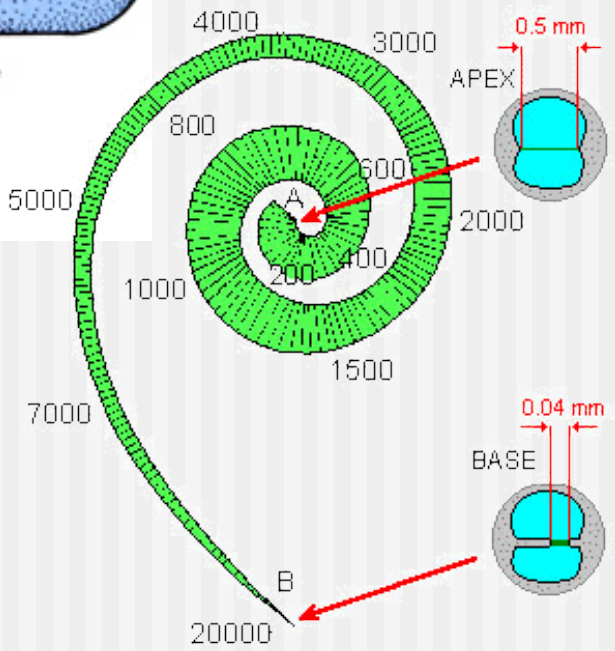
Basilar membrane



- sound enters at the *stapes*
- travels along the basilar membrane
- vibrates at matching position
- activates auditory nerves

Traveling Waves



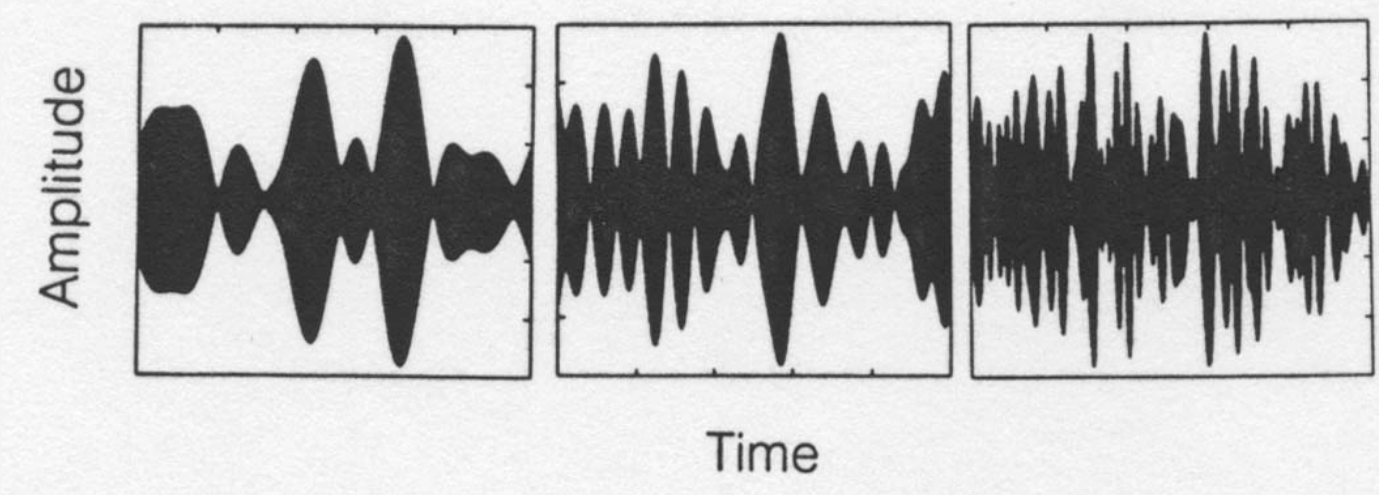


- Traveling wave
- Basilar membrane
- Traveling Wave info

How Sound Travels Through the Ear...

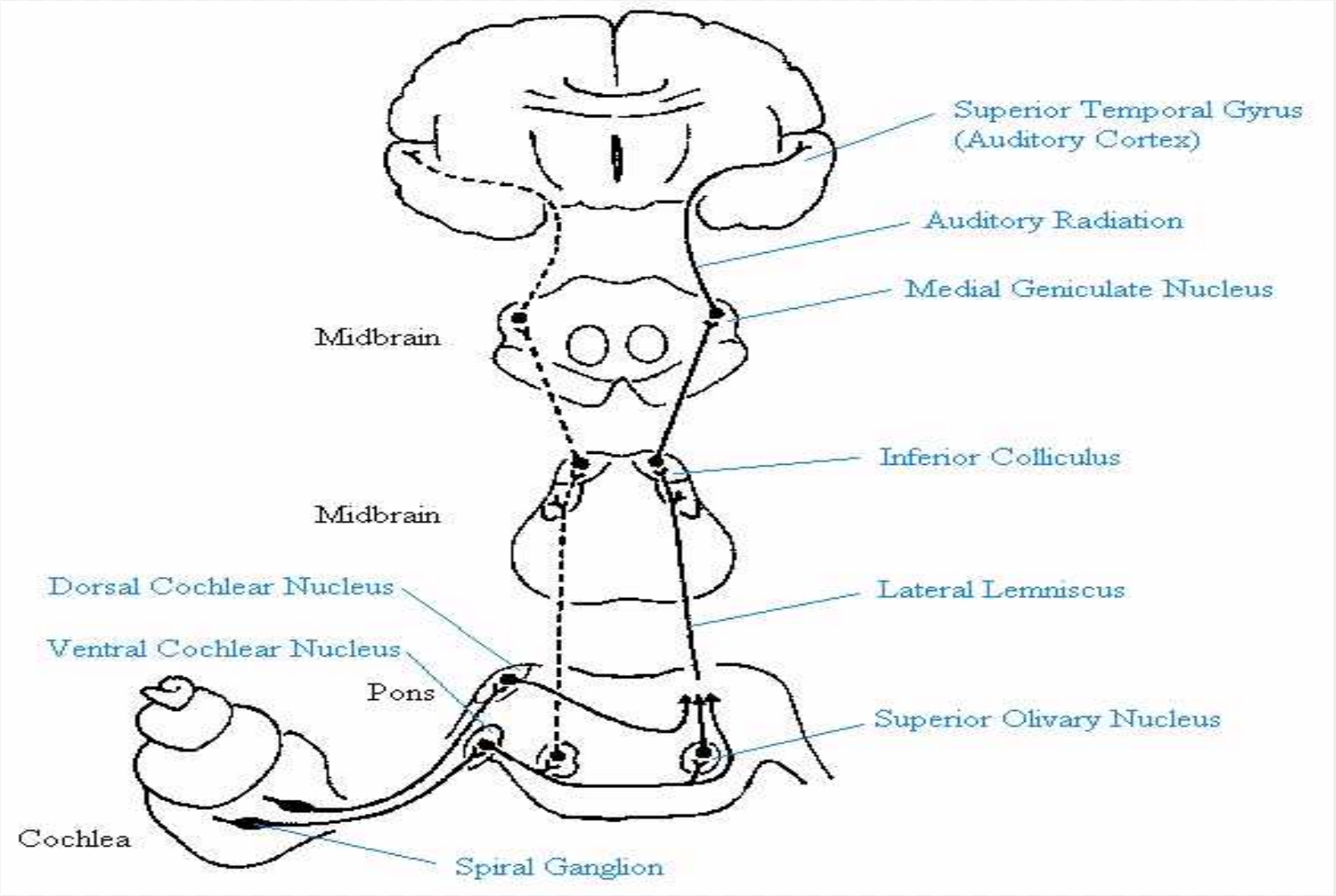
- *Acoustic energy*, in the form of sound waves, is channeled into the ear canal by the **pinna**
  - Sound waves strike the **tympanic membrane,** causing it to vibrate like a drum, and changing it into *mechanical energy*
  - The **malleus**, which is attached to the tympanic membrane, starts the **ossicles** into motion. (The middle ear components mechanically amplify sound)
- The **stapes** moves in and out of the **oval window** of the **cochlea** creating a fluid motion
- The fluid movement within the cochlea causes membranes in the **Organ of Corti** to shear against the **hair cells**
  - This creates an *electrical signal* which is sent via the **Auditory Nerve** to the brain, where sound is interpreted

Temporal resolution



- Auditory nerve fibers do not fire at the instant at which sounds begin or end
- Auditory nerve fibers do not fire on every cycle of sound
- Adaptation occurs to longer duration sounds
  - Spontaneous activity occurs when no sound is present

THE AUDITORY PATHWAY: Diagram



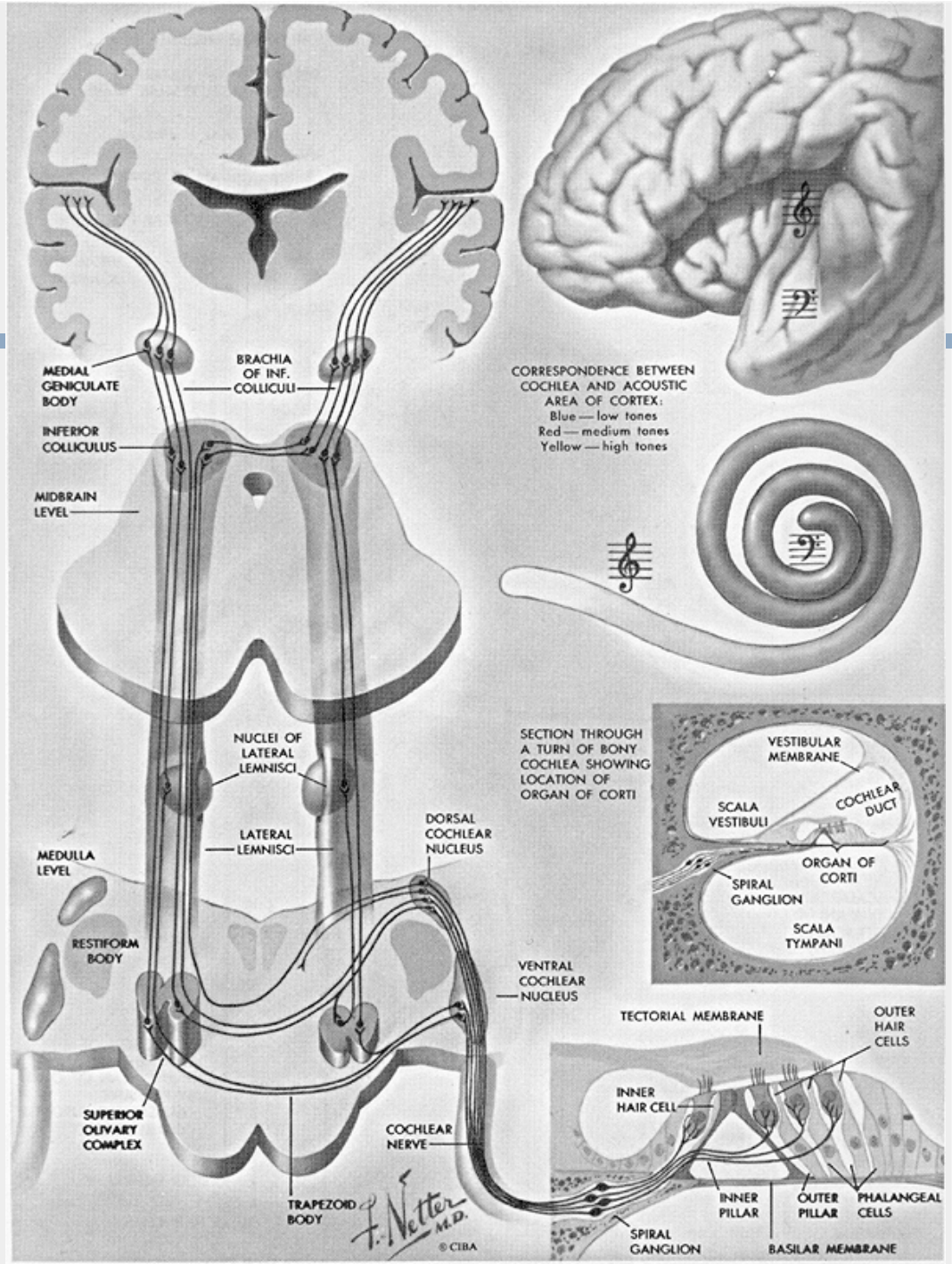
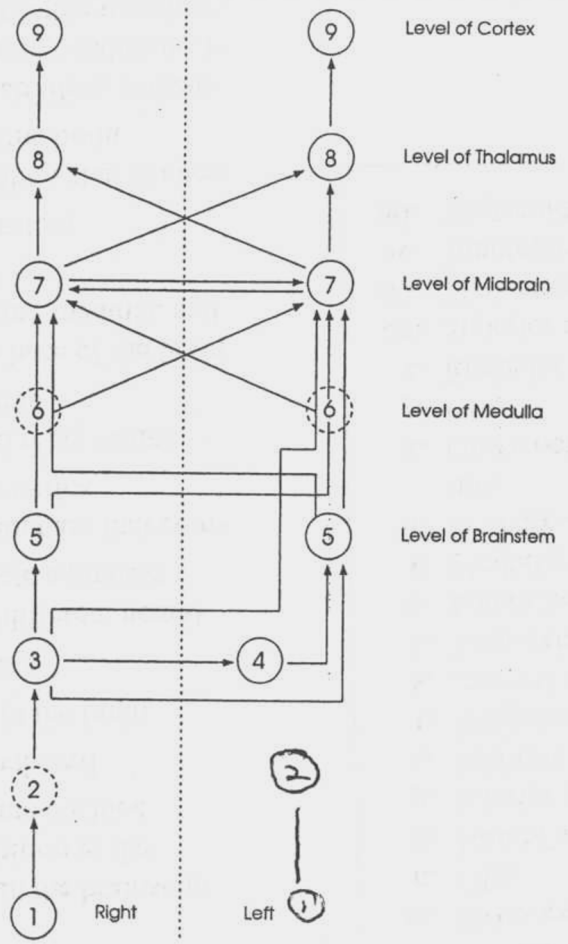
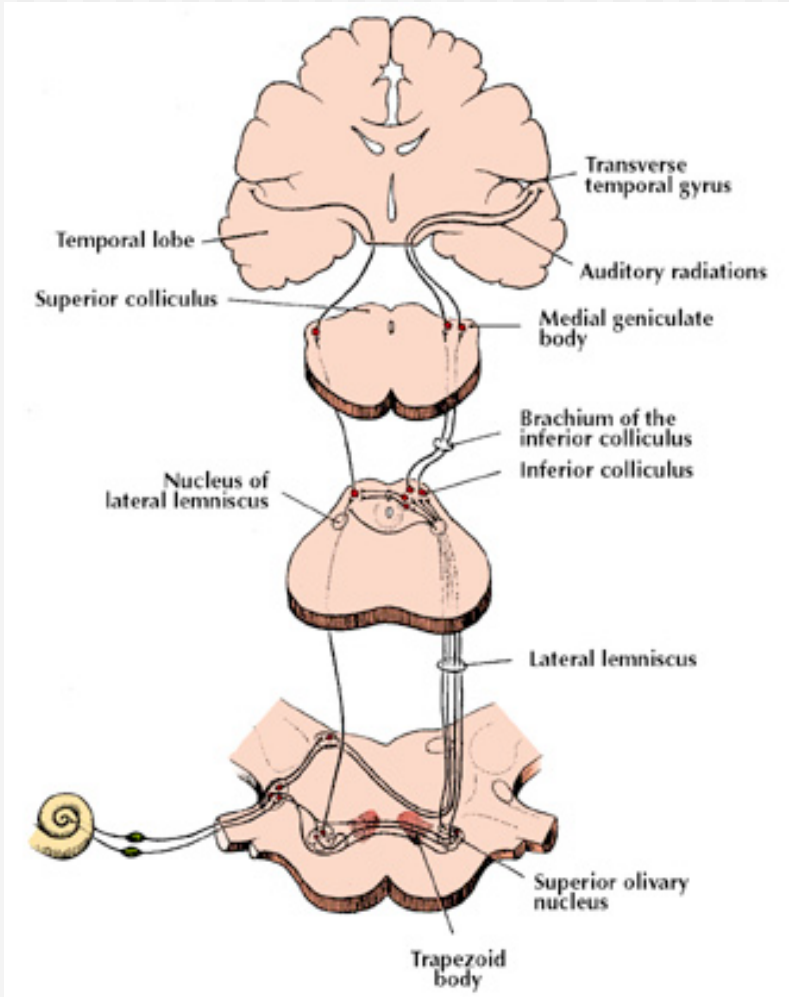


Figure 2.3. Central pathways of hearing.

Central Auditory Path





Short-term spectrum (STFT)

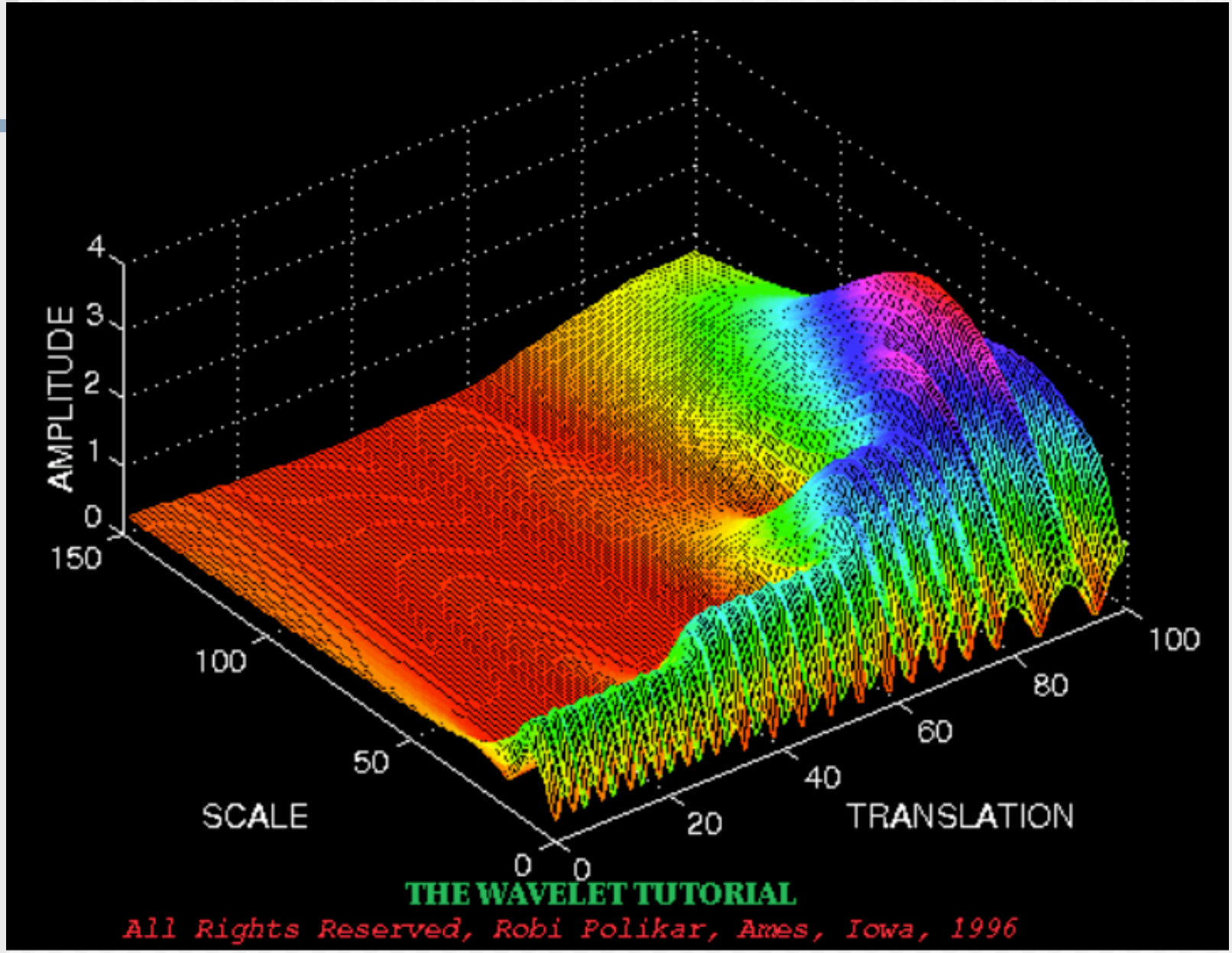
- Represents the distribution of **power** with respect to **frequency** over a time interval centred at time, *t*, like a vertical slice through the spectrogram
- From a **source-filter** perspective, it gives us some information about the shape of the vocal tract at time *t*
- From a **human speech perception** view, it provides similar information to that sent from the cochlea to the auditory nerve

THE SHORT TERM FOURIER TRANSFORM (Remember?)

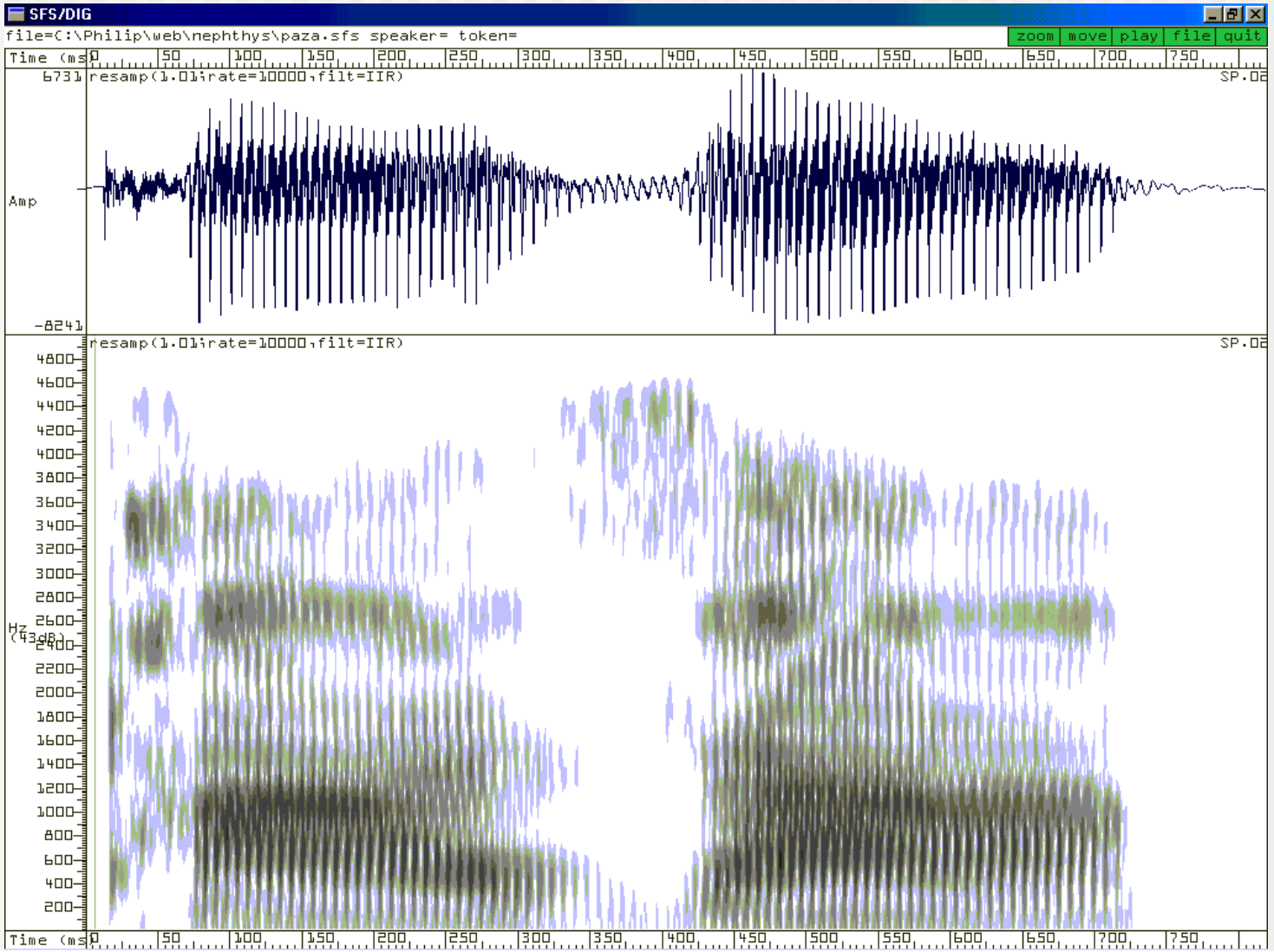
- If this region where the signal can be assumed to be stationary small...
  - we look at that signal from narrow windows, narrow enough that the portion of the signal seen from these windows are indeed stationary
  - This approach of researchers ended up with a revised version of the Fourier transform, so-called : The Short Time Fourier Transform (STFT)

- The problem with the STFT has to do with the width of the window function that is used
- Narrow window ➔ good time resolution, poor frequency resolution
- Wide window ➔ good frequency resolution, poor time resolution

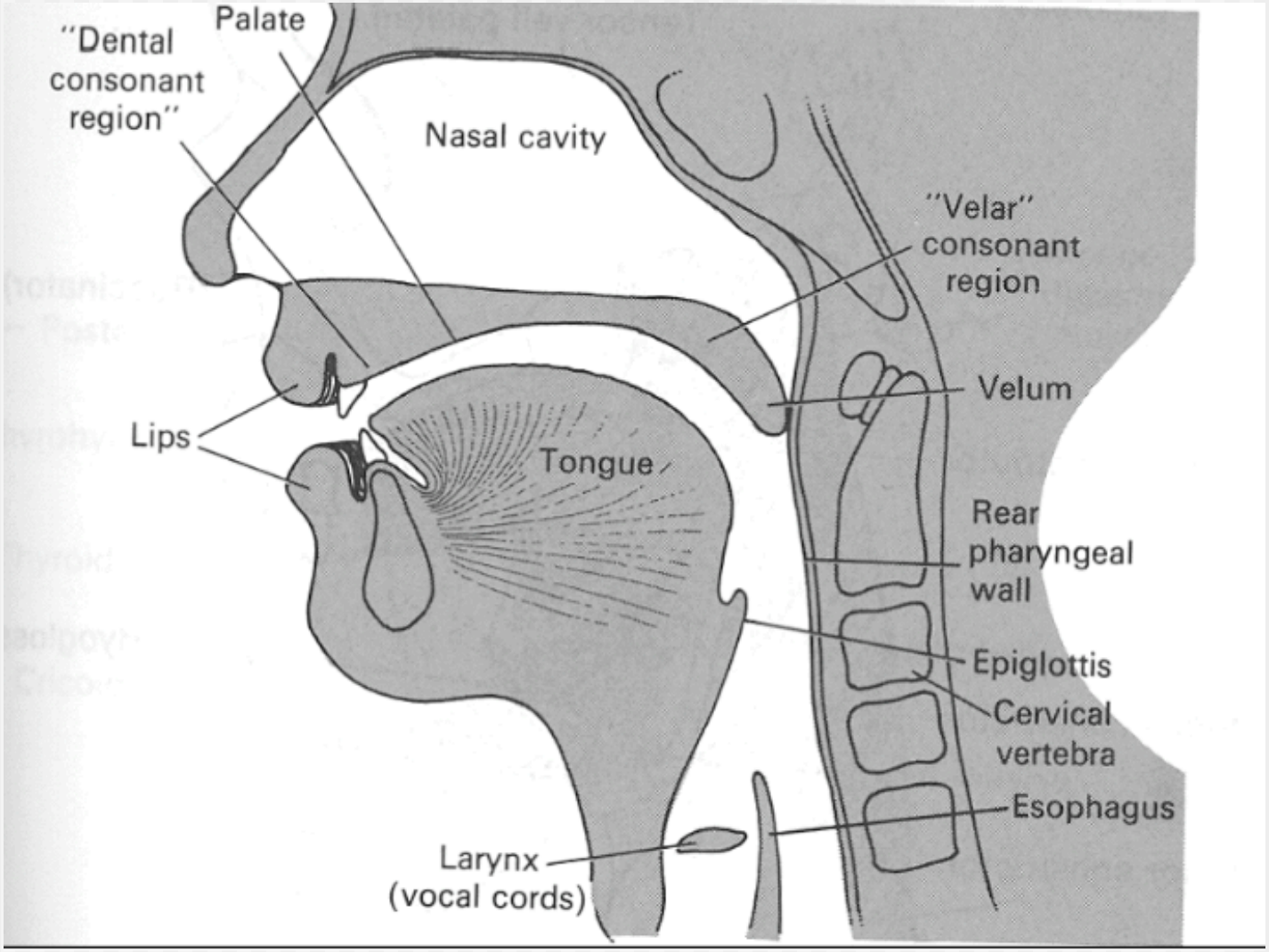
CWT (Better!!)


AMPLITUDE
4
3
2
1
0
150
100
50
SCALE
0
0
20
40
60
80
100
TRANSLATION
THE WAVELET TUTORIAL
All Rights Reserved, Robi Polikar, Ames, Iowa, 1996

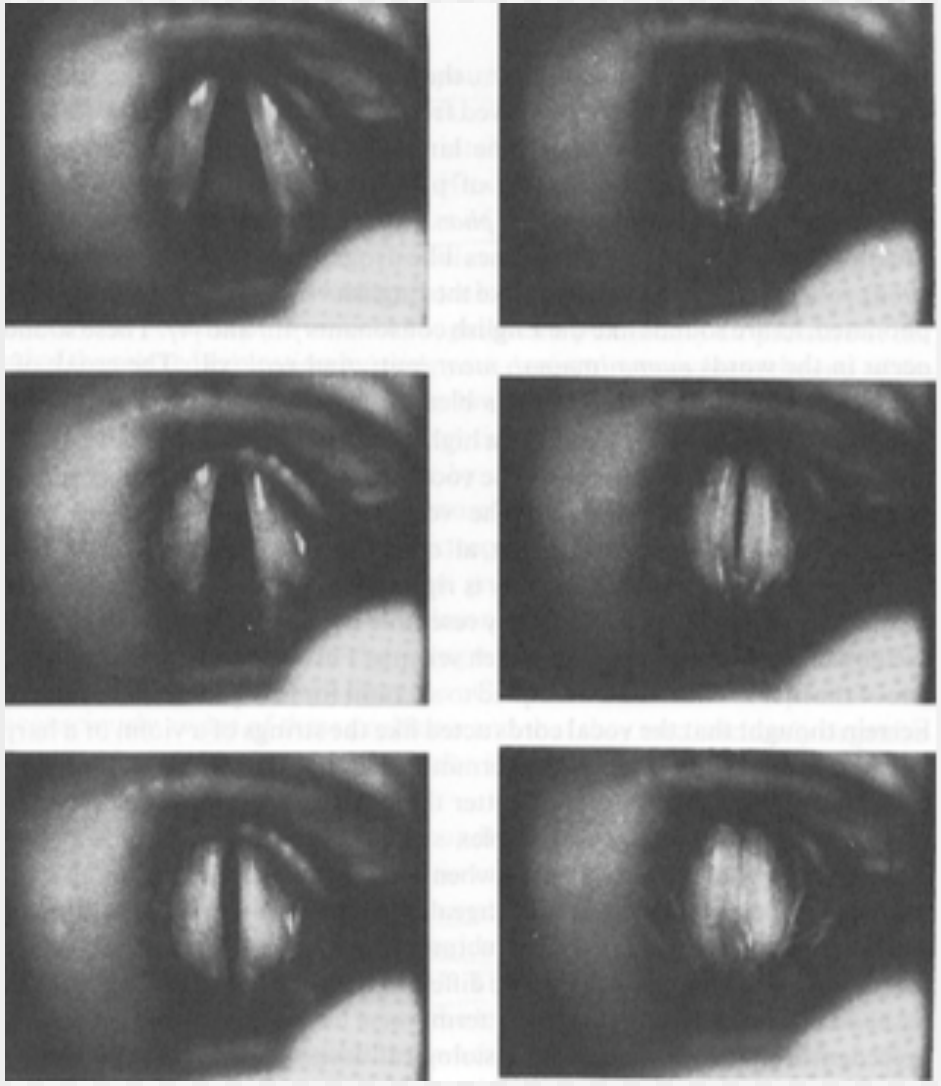
Speech spectrogram
CWT still to new…



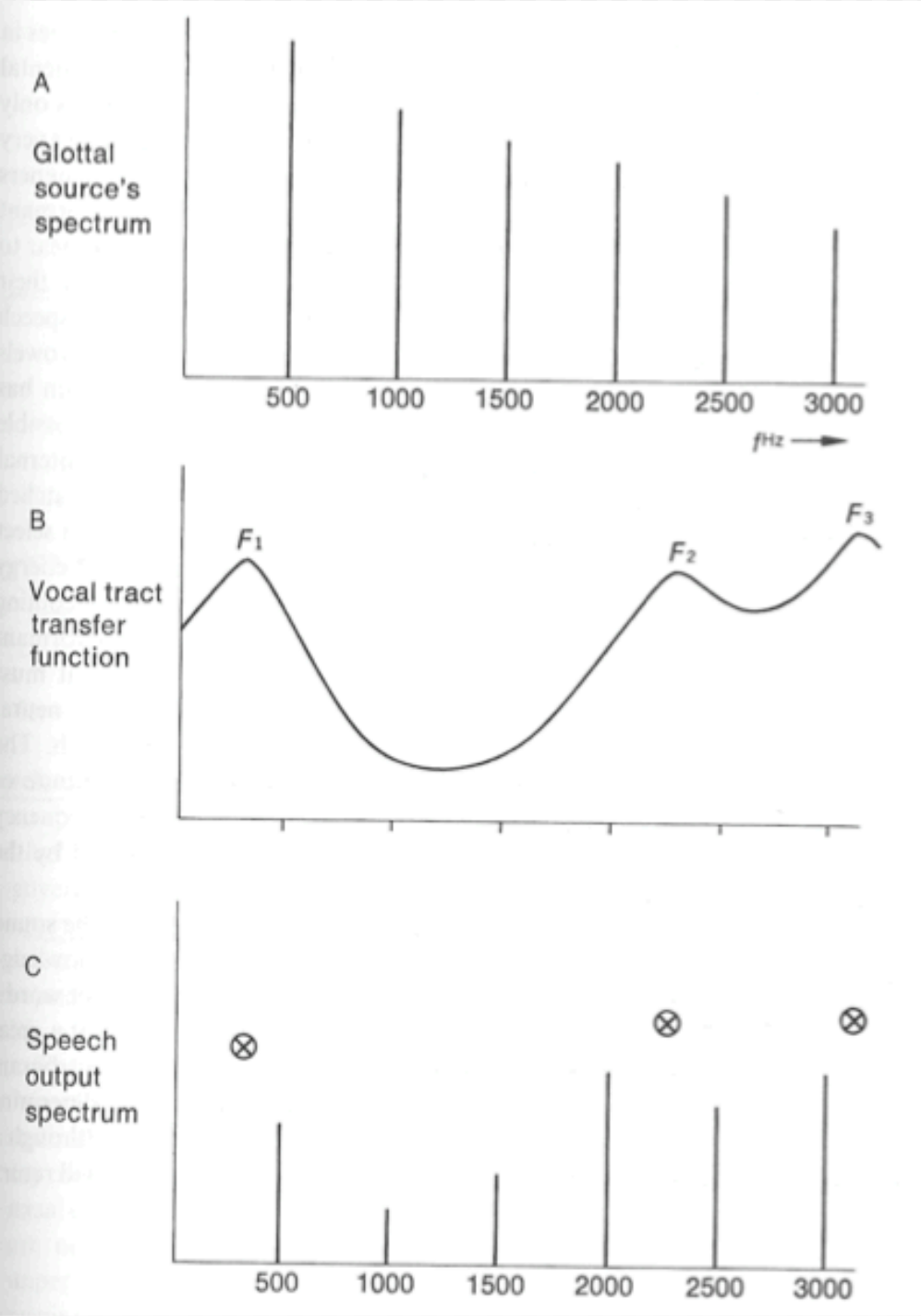
Glottal refers to the opening between the vocal cords of the larynx

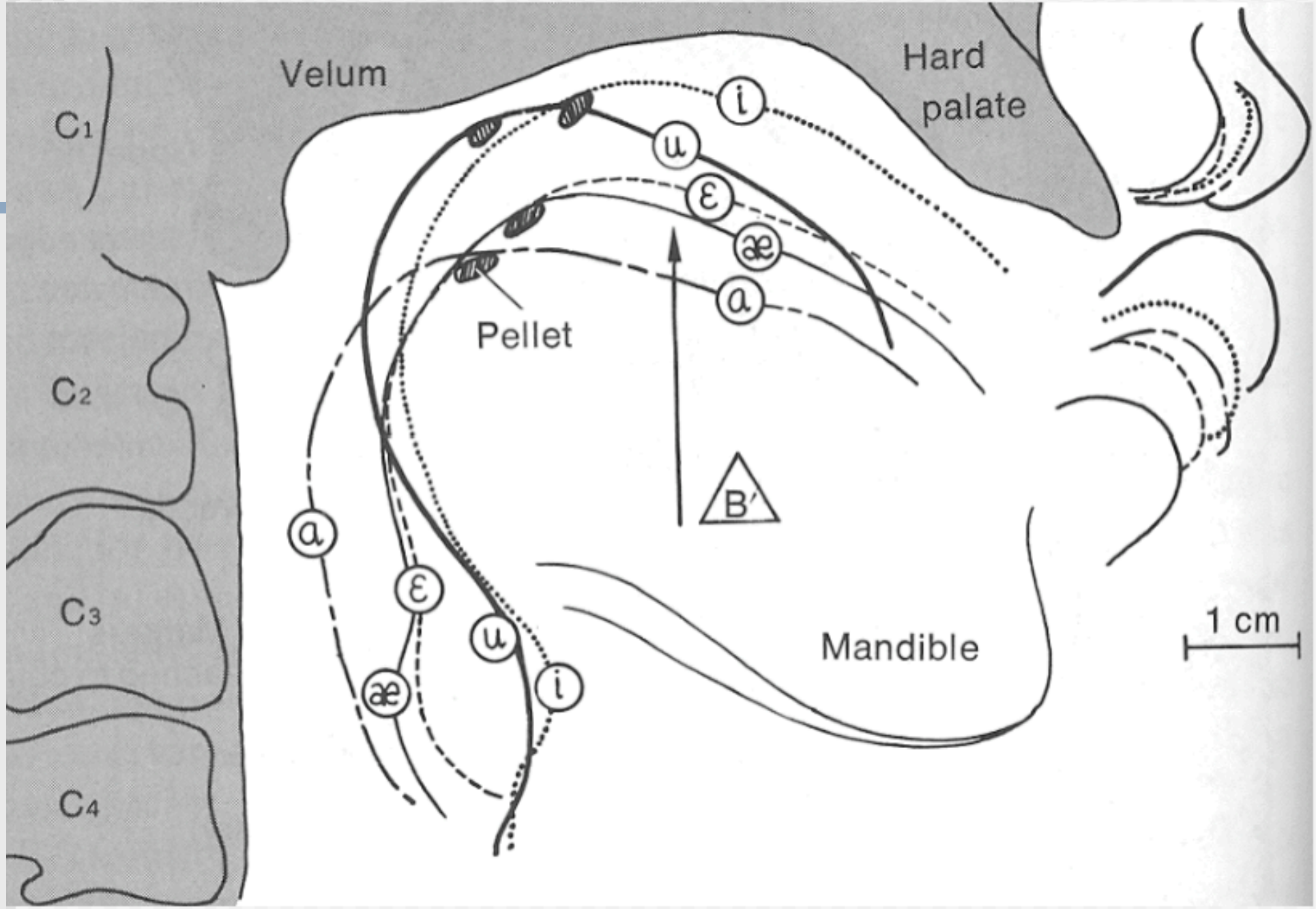


Vocal cords

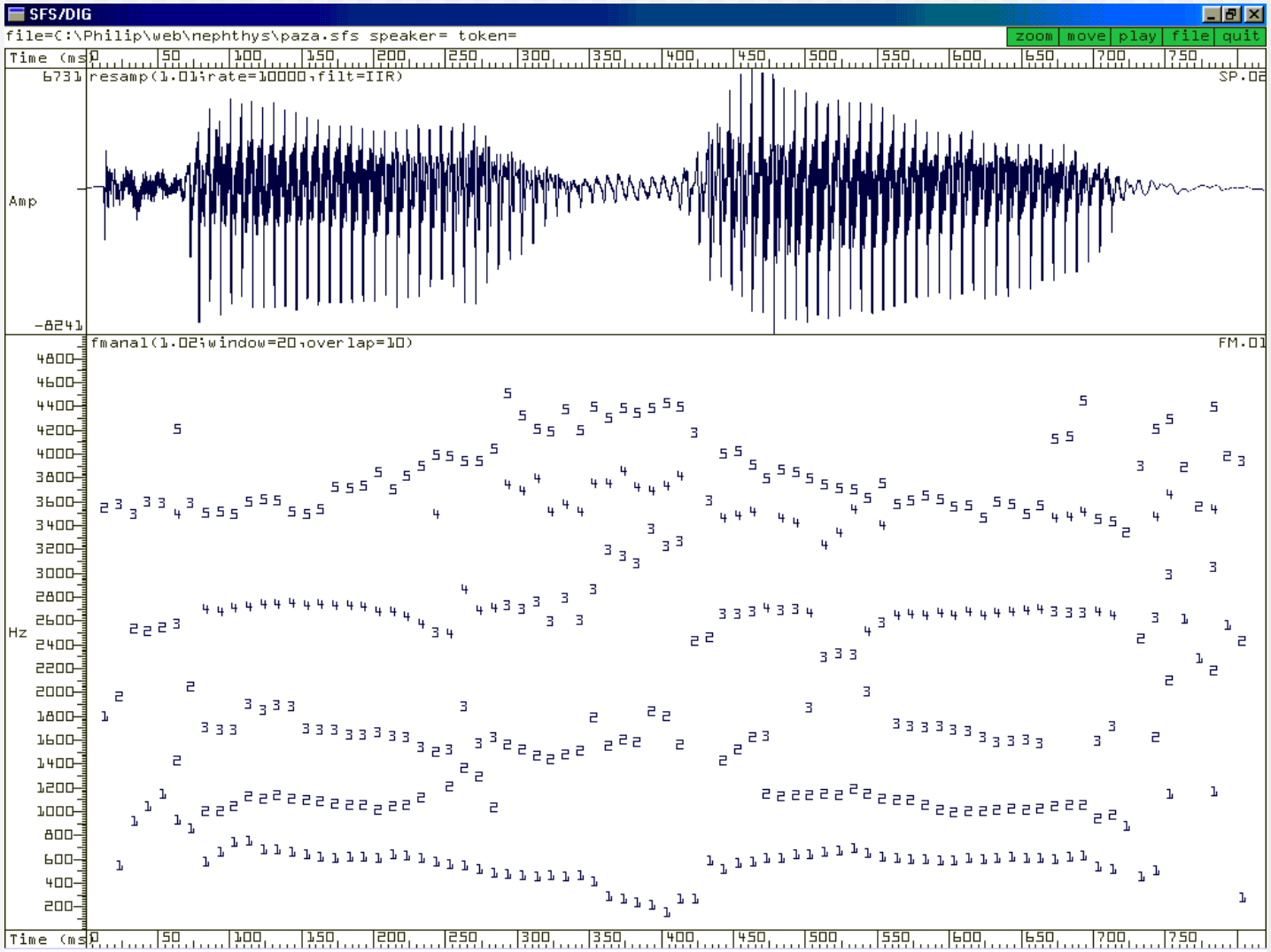


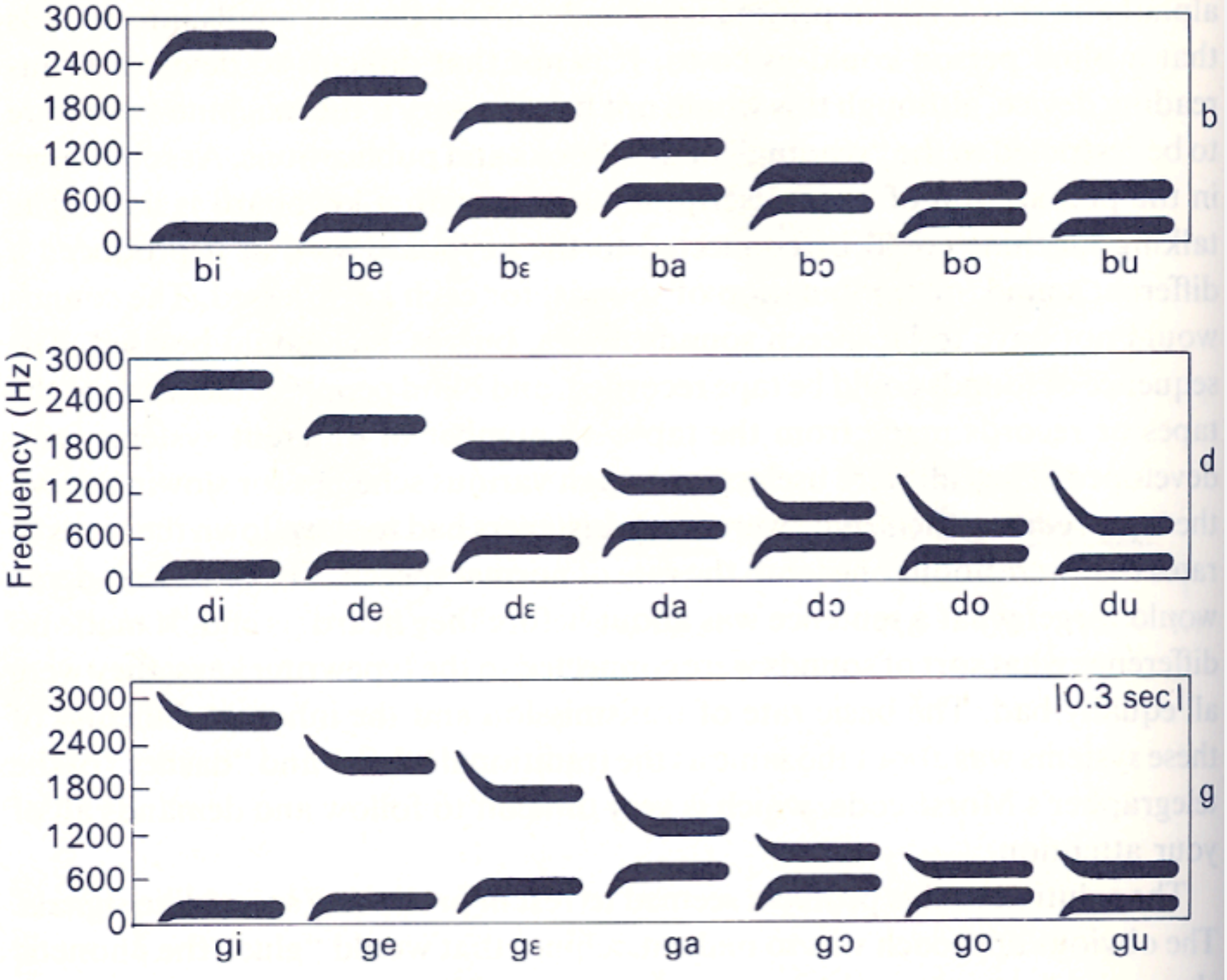
- A) and B)
  - The vocal tract transfers the source (glotattal spectrum) into speech
- B)
  - Formant Frequencies (F1, F2, F3..)
- C)
  - Formant Frequencies are not directly manifested in the speech signal

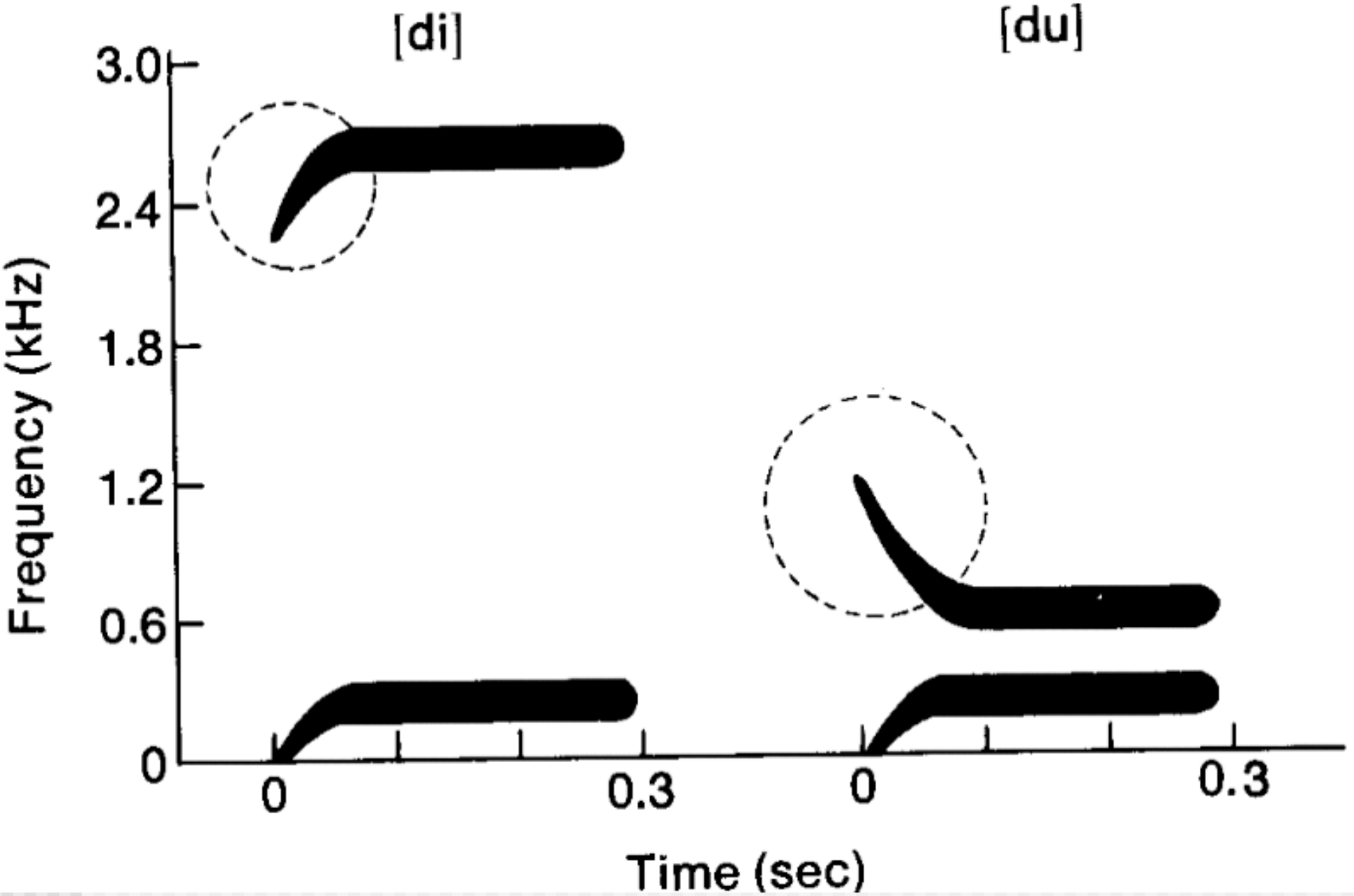


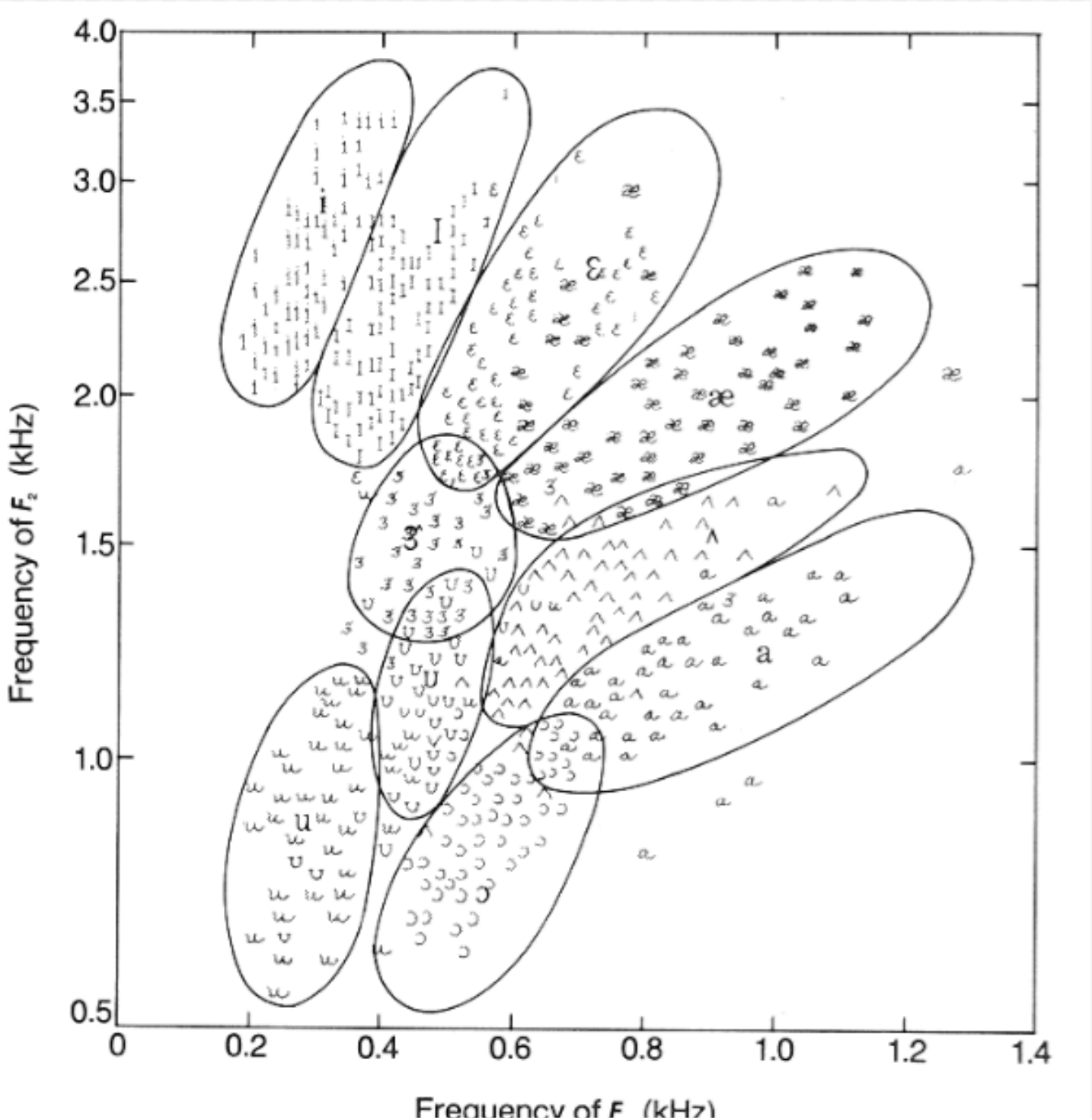

Velum
Hard palate
C1
C2
C3
C4
i
u
ɛ
æ
ɑ
Pellet
B′
ɑ
ɛ
u
æ
i
Mandible
1 cm

Derived formant tracks




Frequency (Hz)
3000
2400
1800
1200
600
0
b
bi
be
bɛ
ba
bɔ
bo
bu
d
di
de
dɛ
da
dɔ
do
du
g
gi
ge
gɛ
ga
gɔ
go
gu
|0.3 sec|


[di]
[du]
3.0
2.4
1.8
1.2
0.6
0
Frequency (kHz)
0
0.3
0
0.3
Time (sec)


4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Frequency of F_2 (kHz)
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Frequency of F_1 (kHz)

Time Warping

Consider two time series:

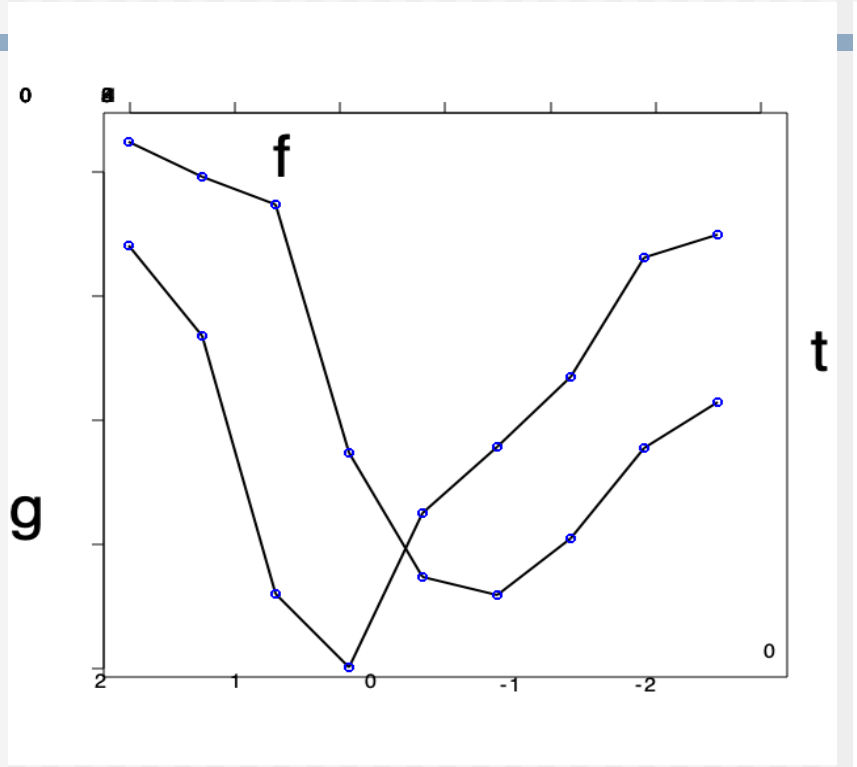
$$F = \{f(t_i),\ i = 1,\ldots,N\} \quad and \quad G = \{g(t_j),\ j = 1,\ldots,M\}$$

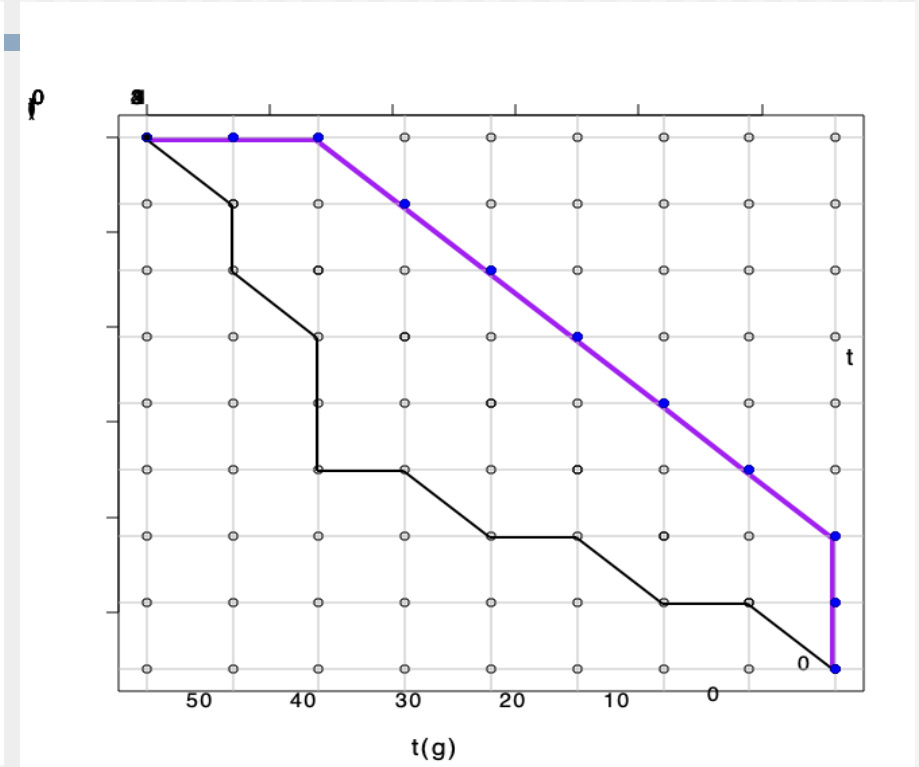
To find the best match between F and G we minimize a measure of discrepancy

$$C(F,G,w) = \sum_{k=1}^{K} d(f(t_{i(k)}), g(t_{j(k)}))r(k)$$

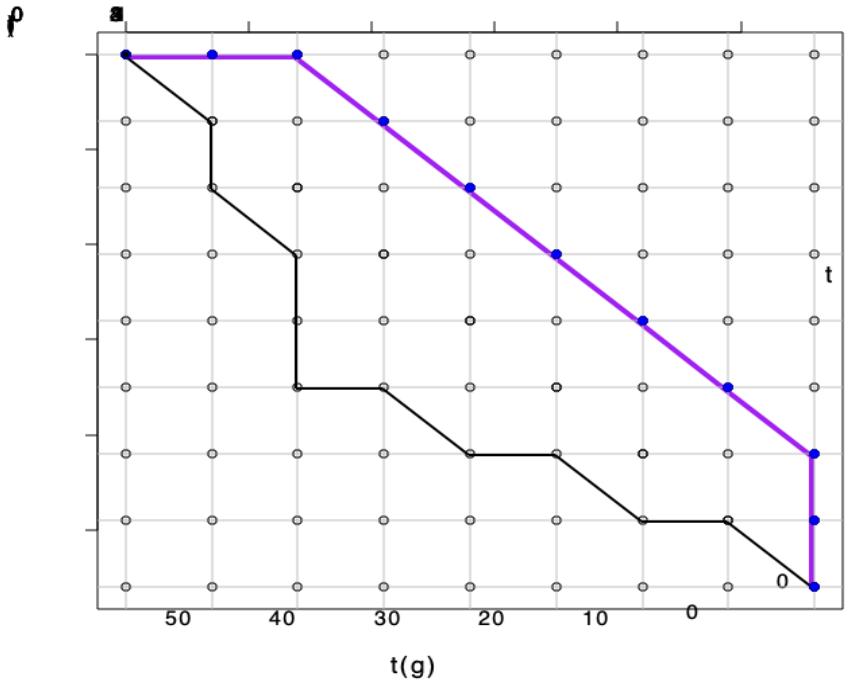
over a warping path w = (i, j), connecting (1,1) and (N,M).

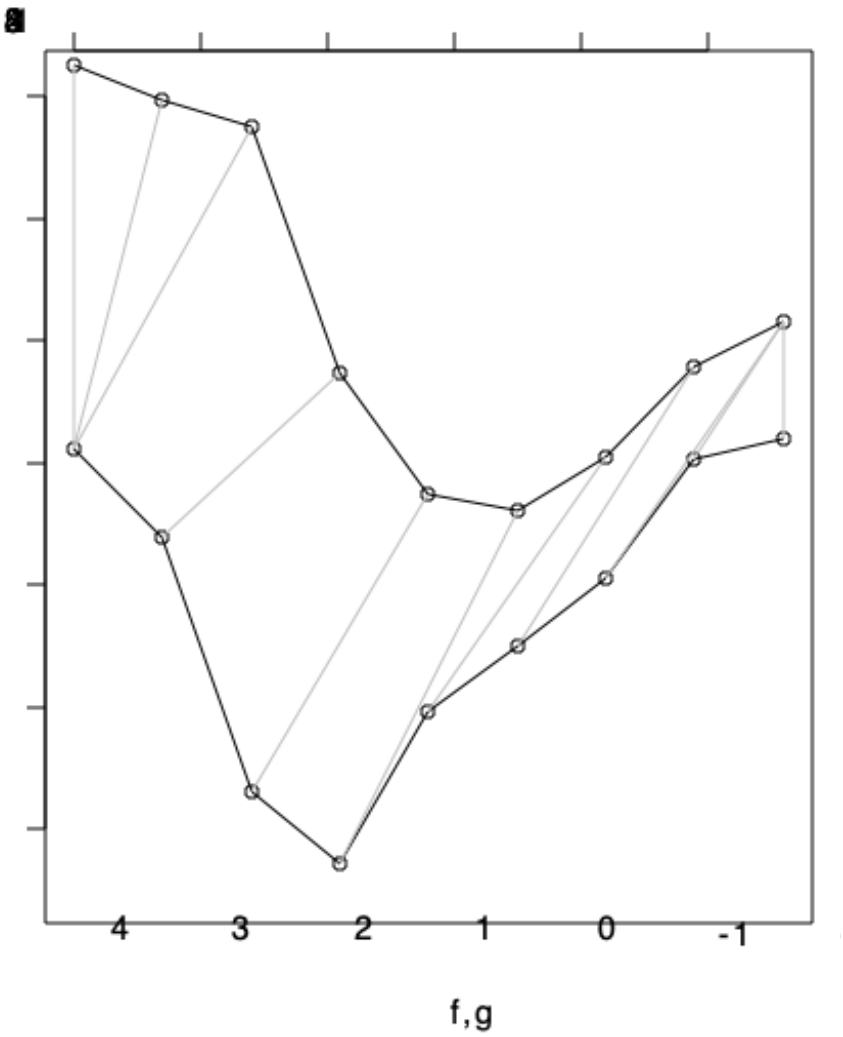
Warping Path



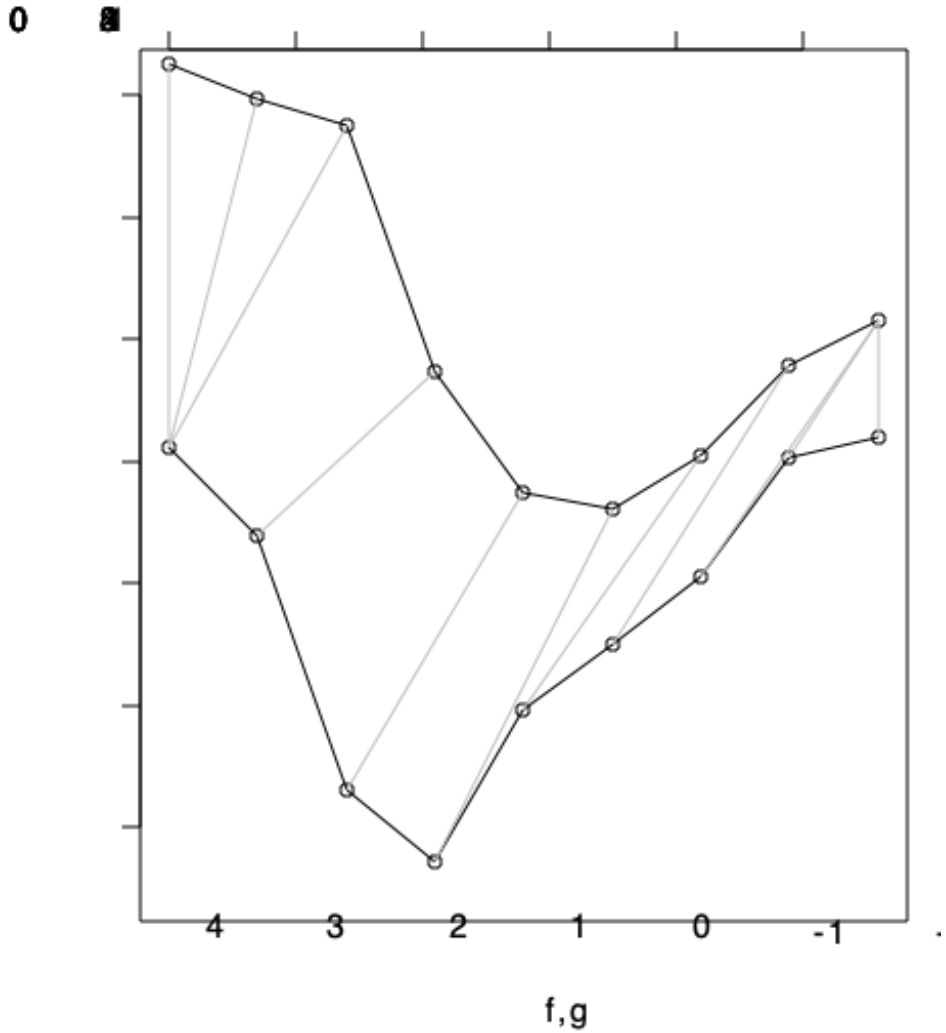


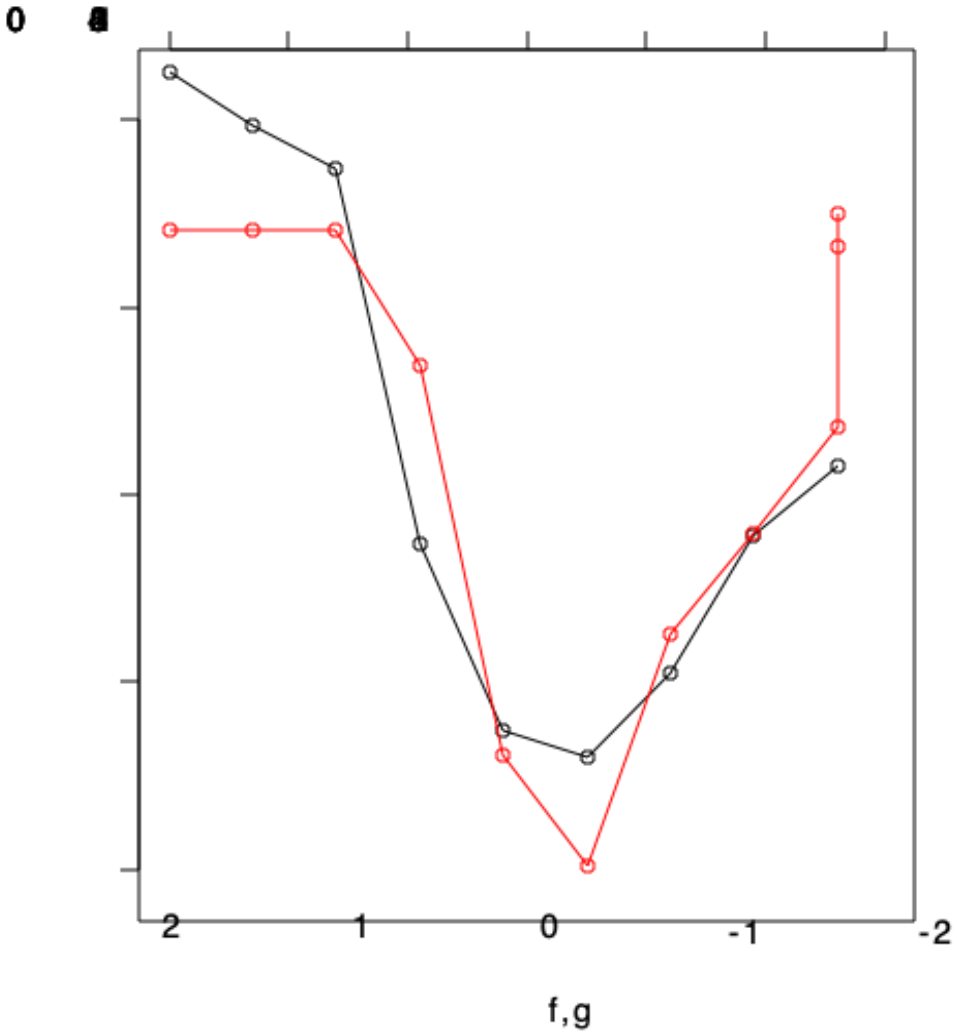
Every path on the time lattice maps the time points of f to the time points of g defining a normalized time scale for the two series.


50
40
30
20
10
0
t
0
t(g)

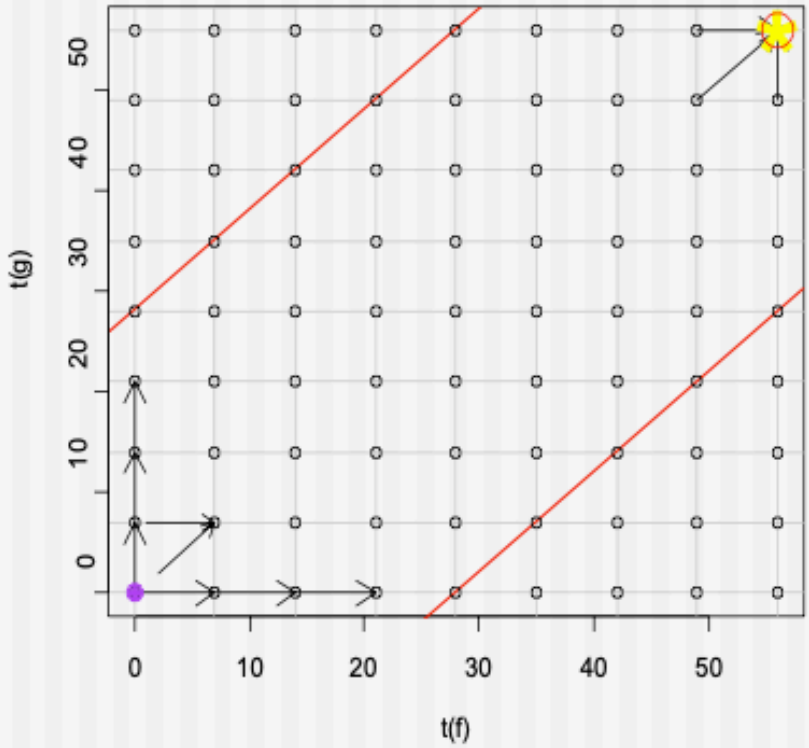

4
3
2
1
0
-1
f,g

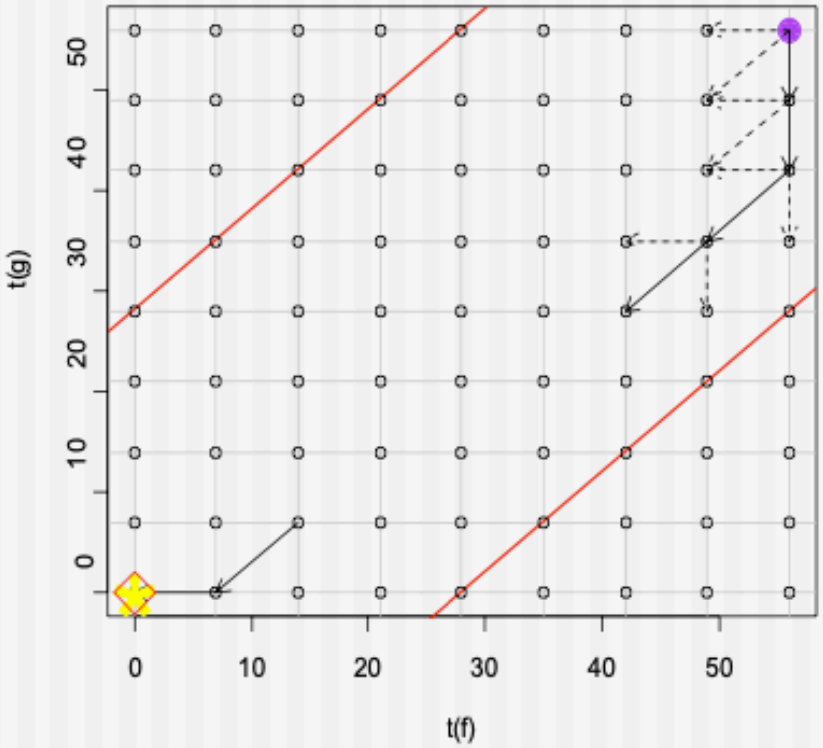
Warping effect



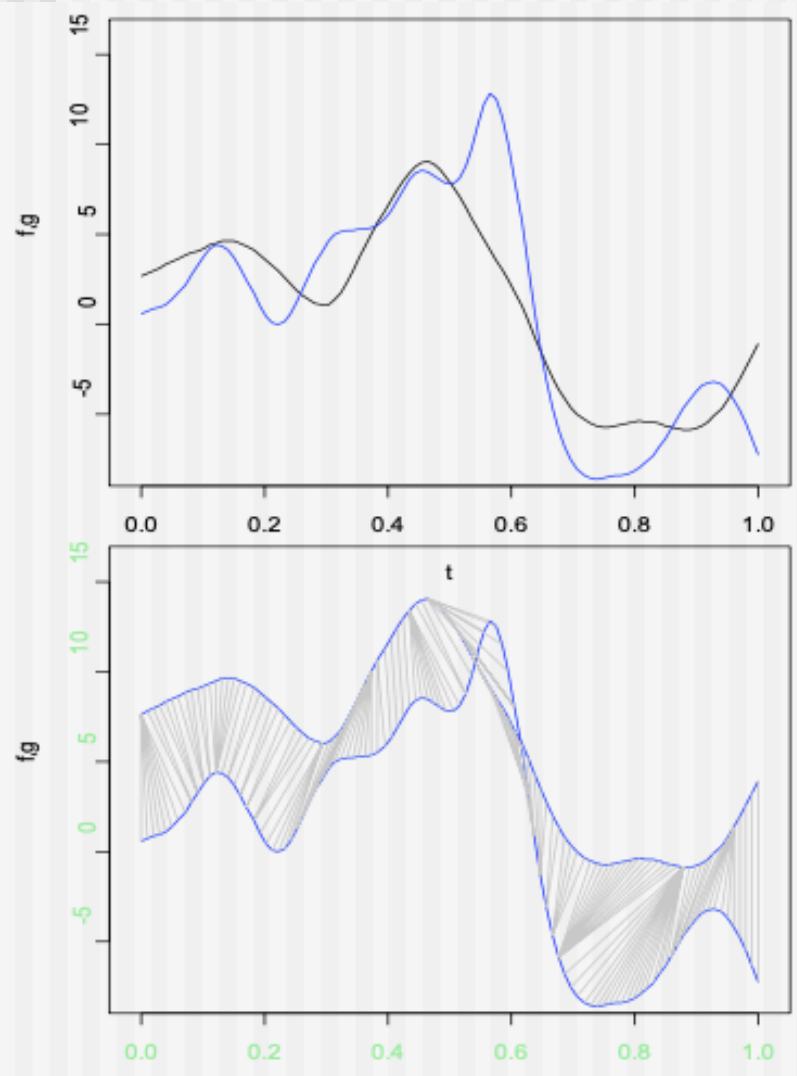


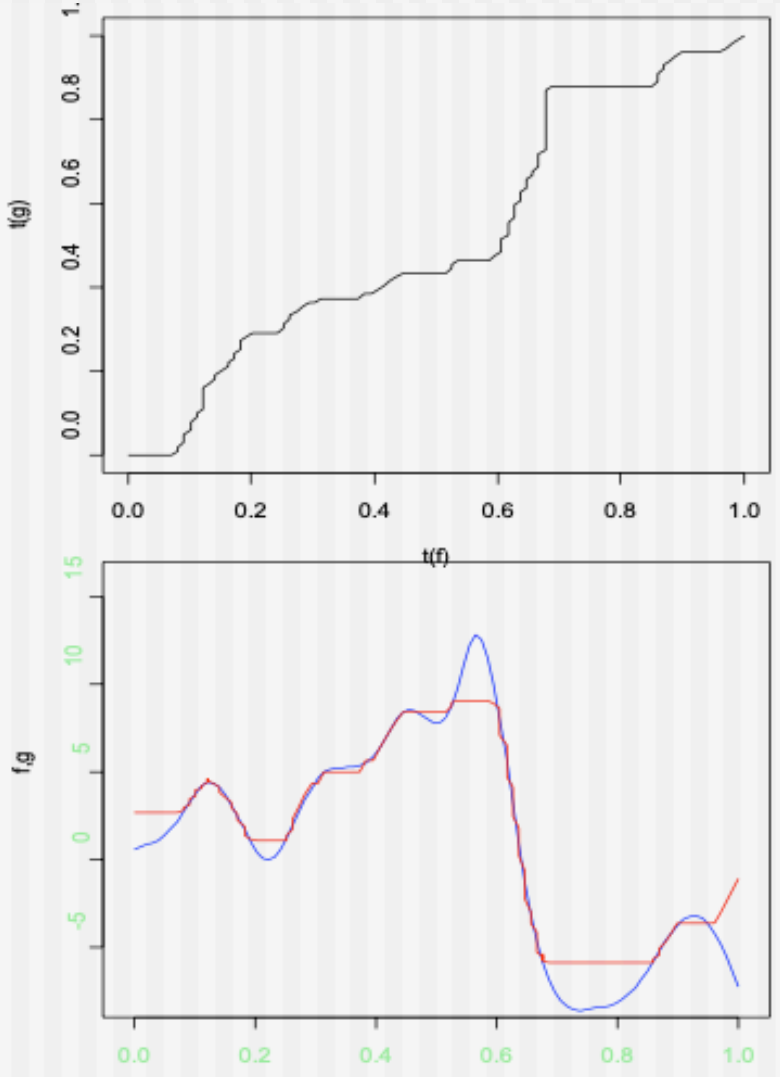
DP Implementation (Backwards Induction)



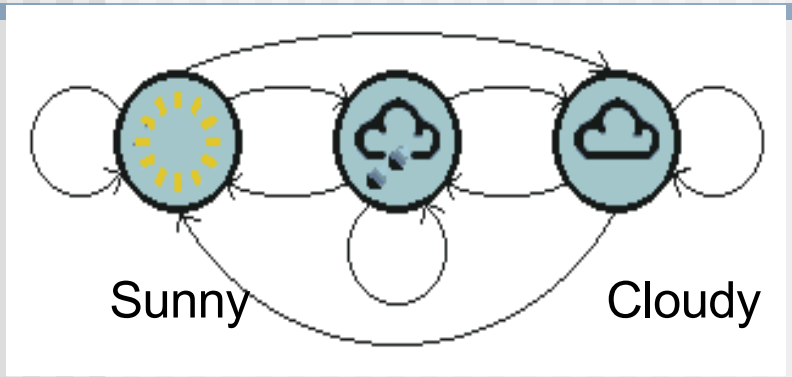


Alignment of shape invariant shifts





Markov Chains



weather yesterday \ weather today	Sun	Cloud	Rain
Sun	0.5	0.25	0.25
Cloud	0.375	0.125	0.375
Rain	0.125	0.625	0.375

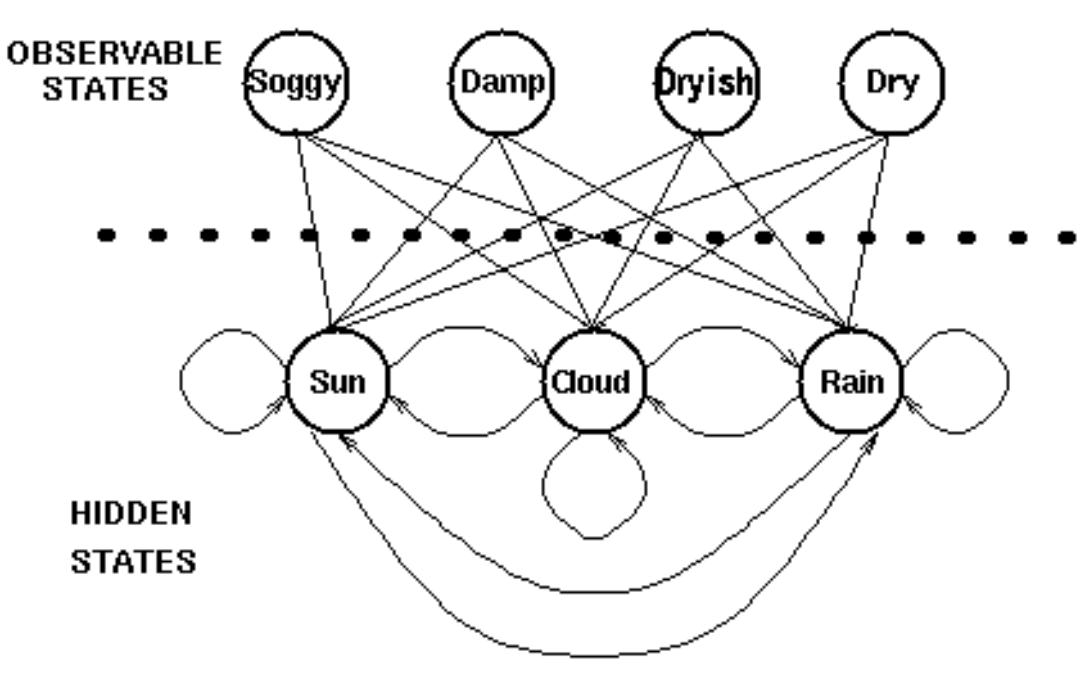
States : Three states - sunny, cloudy, rainy.

State transition matrix : The probability of the weather given the previous day's weather.

Sun	Cloud	Rain
1.0	0.0	0.0

Initial Distribution : Defining the probability of the system being in each of the states at time 0.

Hidden Markov Models

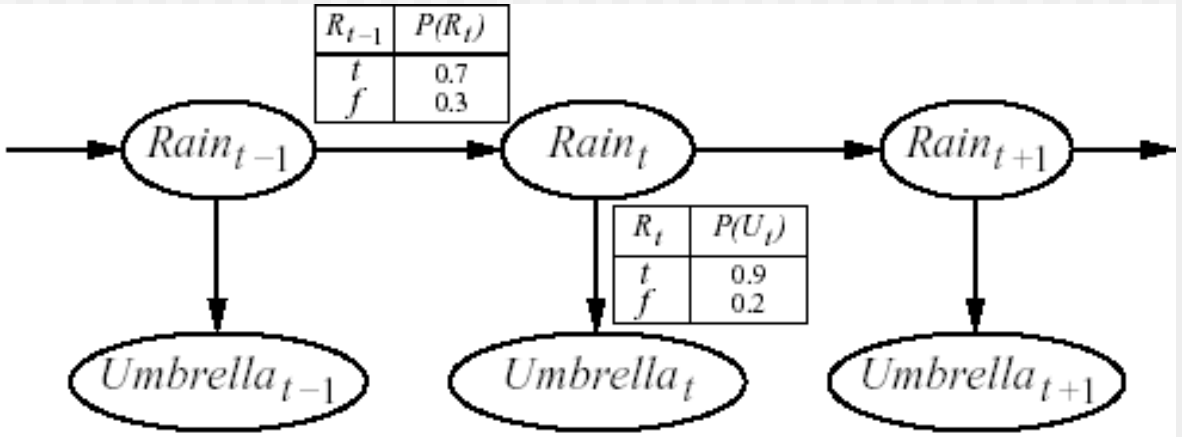


Hidden states : the (TRUE) states of a system that may be described by a Markov process (e.g., the weather).

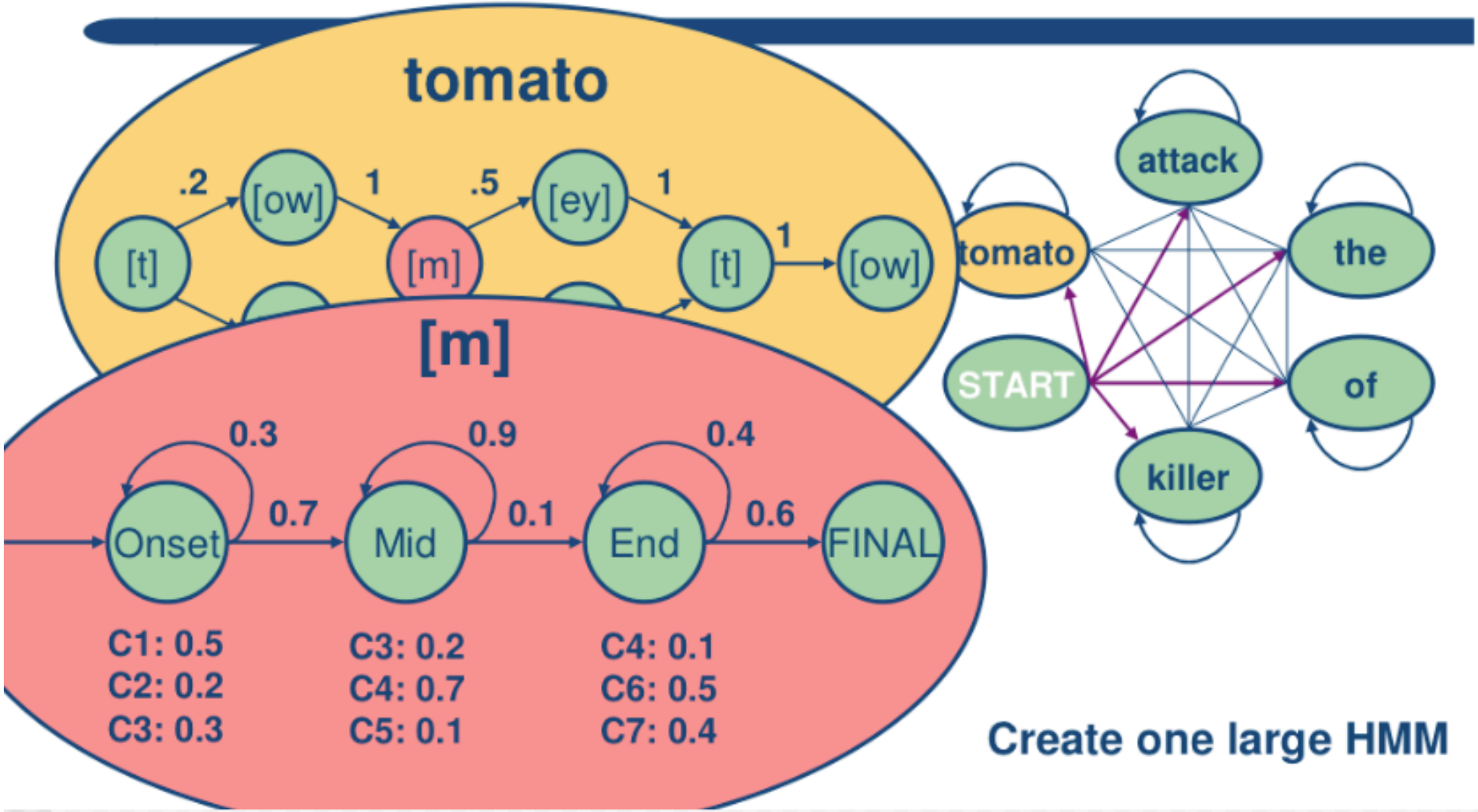
Observable states : the states of the process that are `visible' (e.g., seaweed dampness).

Simple HMM

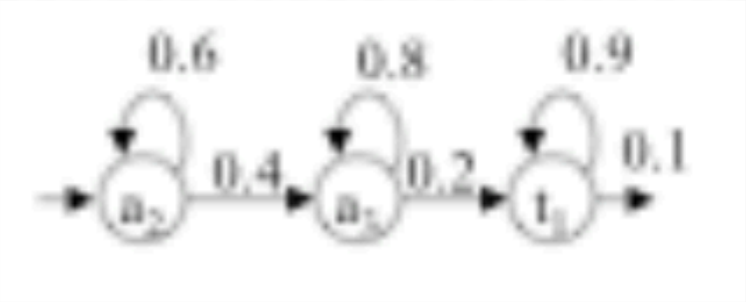
- Security guard resides in underground facility (with no way to see if it is raining)
- Wants to determine the probability of rain given whether the director brings an umbrella
- $P(Rain_0 = t) = 0.50$

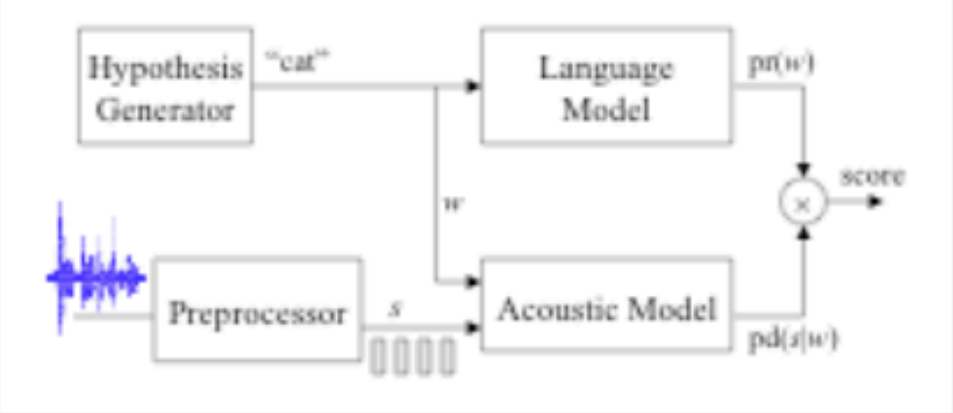


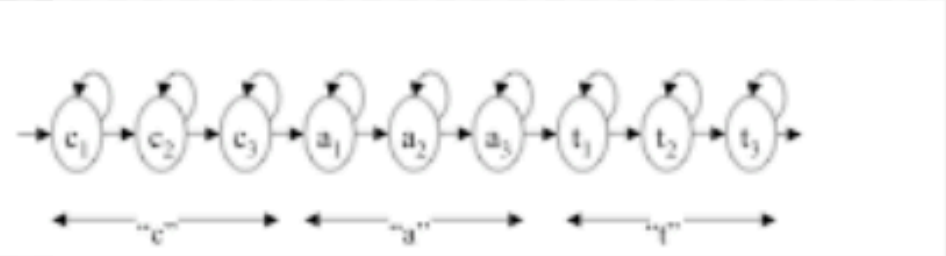
Combining Models

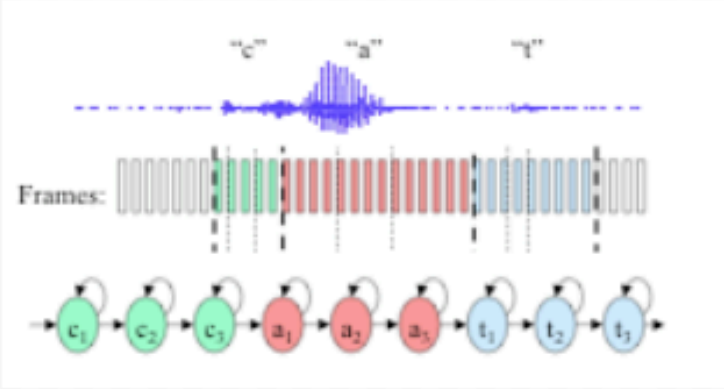


HMM for speech recognition









Signal, Time and Brain

- Shorter time period: Time Warping
  - FFT
- Longer time period: Representation of states and transition between them (HMM)
  - STFT, Wavelets

Persistence of Vision

- According to the theory of persistence of vision, the perceptual processes of the brain or the retina of the human eye retains an image for a brief moment
  - A visual form of memory known as iconic memory has been described as the cause of this phenomenon
  - Persistence of vision is said to account for the illusion of motion which results when a series of film images are displayed in quick succession, rather than the perception of the individual frames in the series

Persistence of Vision

- Although psychologists and physiologists have rejected the relevance of this theory
- .....
- ...
- Some scientists nowadays consider the entire theory a myth.

Process of temporal integration, perceptual identity (Pöppel)

- It has been demonstrated by a number of psychophysical and neuropsychological experiments that the human brain povides a temporal platform of just a few seconds
- Some neurophysiological observations support the existence of such a temporal platform with a duration of 2 to 3 seconds

- Ernst Pöppel
- http://www.imp-muenchen.de/Ernst_Poeppel.357.0.html

Process of temporal integration, perceptual identity (Pöppel)

- The reproduction of visual or auditory stimuli of different duration can be done veridically and with small variance up to approx. 2 to 3 seconds
- (of course with some interindividual variance), but not beyond

- Ernst Pöppel
- http://www.imp-muenchen.de/Ernst_Poeppel.357.0.html

- Sound and Light
- The Ear
  - Cochlea
- Auditory Pathway
- Speech Spectrogram
  - Vocal Cords
  - Formant Frequencies
- Time Warping
- Hidden Markov Models
- Signal, Time and Brain
  - Process of temporal integration, perceptual identity (Pöppel)