The Ear and Hearing

Sound and sensations:
Physical attributes of sound: intensity, frequency, duration

The human auditory system converts **variations in air pressure** due to sound into **coded neural firings** in the auditory nerve.

Sound and Sensation

The physiological outputs (i.e. auditory nerve firings) can be measured and quantified.

But how do we define and measure the subjective sensations that arise?

Sensations are of the following types:

- Loudness
- Pitch
- Timbre

Study of Hearing

- Physiology

  Main stages:

  -Peripheral auditory system (ears) ← analysis
  -Auditory nervous system (brain) ← synthesis

- Psychoacoustics
  - Study of the relation between the magnitude of the stimulus (physical units) and the magnitude of the sensation experienced.

A study of hearing helps to build a *model for hearing*.

Why is this useful?

- Speech and audio signal compression
- Objective evaluation of audio quality
- Sound classification
- Hearing aid design

Some remarkable properties of human hearing:

1. Response to wide range of stimuli…

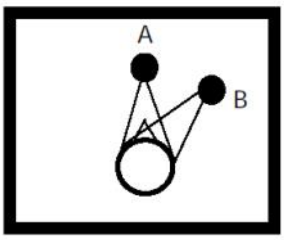
   … in (20 Hz, 20 kHz) over 120 dB amplitude range

2. Can distinguish closely spaced frequencies

3. Can identify pitch and timbre

4. …..with two ears…?

Binaural difference cues are used for source azimuth detection.



(Begault book)

Figure 2.1. A listener in an anechoic chamber, with a sound source oriented directly ahead on the median plane (A) and displaced to 60 degrees azimuth.

Sound pressure level (SPL):
Our ears respond to extremely small periodic variations in atmospheric pressure. The minimum pressure fluctuation to which the ear can respond is less than 10^{-9} of atmospheric pressure (=> ear drum vibration of 10^{-7} cm)

The "threshold of audibility" is frequency-dependent. At 1 kHz it corresponds to a rms sound pressure level of 2×10^{-5} N/m^2 or Intensity (α pressure^2) = 10^{-12} W/m^2.

Sound levels are typically ratios expressed in dB SPL by:

$L = 10 \log(I/I_0)$, $I_0 = 10^{-12}$ Watts/m^2

$= 20 \log(p/p_0)$, $p_0 = 2 \times 10^{-5}$ Newtons/m^2

Human hearing range

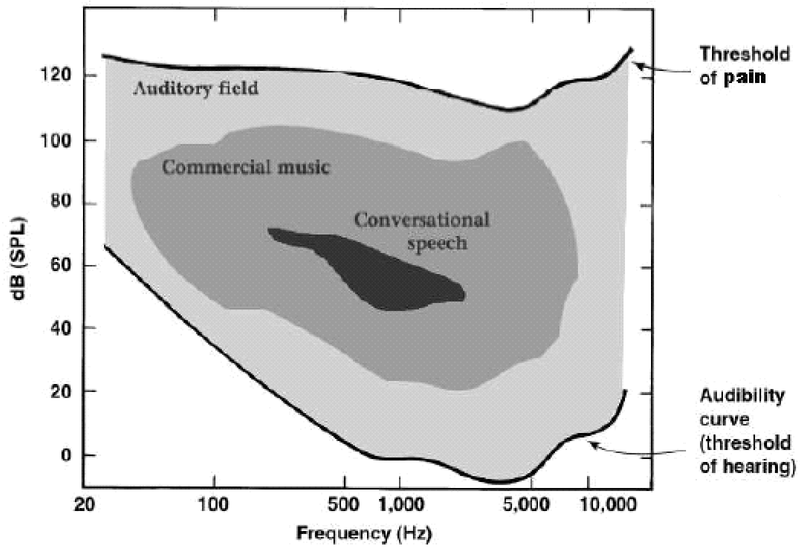
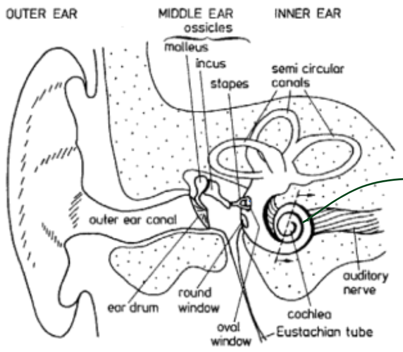


Fig. 1. The auditory field in the frequency-intensity plane. The sound pressure level is measured in dB with respect to the standard reference pressure level of 20 microPascals.

(From: Audio Signal Processing, Chapter 9, Springer book)

Ear anatomy and physiology



Screen clipping taken: 10/6/2010, 2:08 PM

"Sound"
Acoustic pressure signal
↓
Mechanical vibration on BM
↓
Neural activity in auditory nerve
↓
AUDITORY CORTEX
"Sensations"

Outer ear: Pinna

- Directs sound towards eardrum
- Ear canal is quarter-wave resonator, amplifying the 3-5 kHz range by 15 dB; resonance is broad
- Localises sound sources in medial plane (detecting elevation)

Middle ear: Ossicles (malleus, incus, stapes)

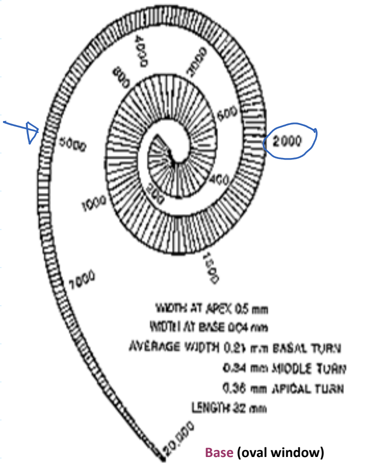
- Transmits eardrum vibrations to the oval window membrane
  => impedance matching

Inner ear: cochlea, semi-circular canals

Cochlea

- Contains endolymph fluid in chamber lined by basilar membrane <----- *ear's microphone*
- On basilar membrane is the organ of corti containing several rows of hair cells (inner + outer = 30,000).
- Each hair cell (with many cilia) connects to a nerve fiber
- Nerve fibers are bundled into auditory nerve

The BM varies gradually in tension and width along its length
=> frequency response varies along its length
Each location has a **characteristic center frequency** of vibration



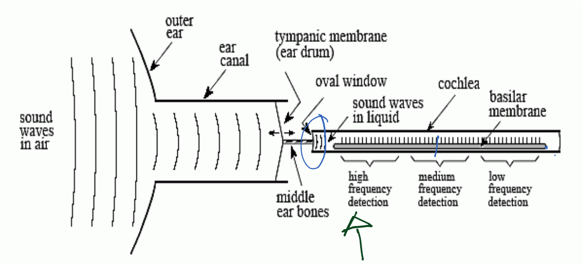


FIGURE 22-1
Functional diagram of the human ear. The outer ear collects sound waves from the environment and channels them to the tympanic membrane (ear drum), a thin sheet of tissue that vibrates in synchronization with the air waveform. The middle ear bones (hammer, anvil and stirrup) transmit these vibrations to the oval window, a flexible membrane in the fluid filled cochlea. Contained within the cochlea is the basilar membrane, the supporting structure for about 12,000 nerve cells that form the cochlear nerve. Due to the varying stiffness of the basilar membrane, each nerve cell only responses to a narrow range of audio frequencies, making the ear a frequency spectrum analyzer.

Screen clipping taken: 10/4/2010, 10:01 PM
From: Steven Smith, The Scientist and Engineer's Guide to DSP, E-book

- Stapes vibration at the oval window generates a traveling wave along the BM in the cochlear liquid.
- The traveling wave causes vibration of the BM. For a given frequency component of the traveling wave, the amplitude of vibration varies with the distance along the BM. High frequencies resonate near the base and low frequencies close to the apex.
- Vibration amplitude increases with increase in tone intensity

Historical perspective:

Helmholtz postulate (1863): subjective pitch is determined by a group of auditory nerve fibres related to place of maximal vibration of the BM.

This was based on the observation that listeners can "hear out " partials in a complex tone.

Historical perspective

Georg von Bekesy (1899-1972)

Georg von Bekesy, born in Budapest, qualified in Chemistry. But he worked with the Hungary Telephone and Post Office Laboratory on telecommunications problems leading him, in 1928, to the discovery of the mechanical characteristics of the cochlea by observing the traveling waves on the basilar membrane produced by sound. The location of the maximum amplitude depended on the frequency.

Nobel Prize in Physiology and Medicine (1961)

From: B. Golstein, Sensation and Perception, Chapter 11

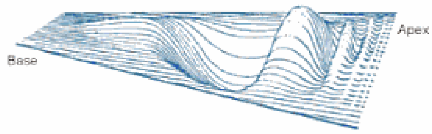
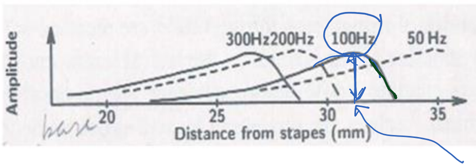


Figure 11.21 A perspective view showing the traveling wave motion of the basilar membrane. This picture shows what the membrane looks like when the vibration is "frozen" with the wave about two thirds of the way down the membrane. *(From Tonndorf, 1960.)*

Screen clipping taken: 10/6/2010, 4:47 AM

T.W. amplitude envelopes for pure tones



From Bekesy optical observations in human cadaver ears using very intense sounds

<- vibration pattern...dB vs distance on linear scale

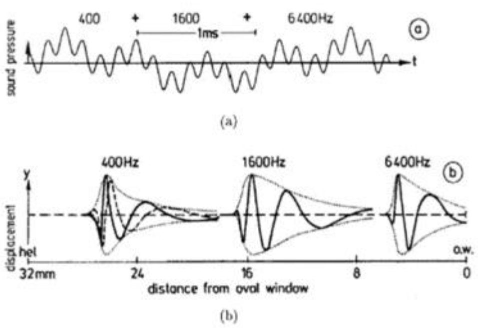
Frequency selectivity of the BM => a Fourier analysis of limited resolution

Distance from apex of maximum is roughly proportional to log(frequency)
(1 octave ~ 3.6 mm)

We thus obtain a transformation of frequency -> **place**

Each "place" shows frequency selectivity.

Frequency-to-place transformation

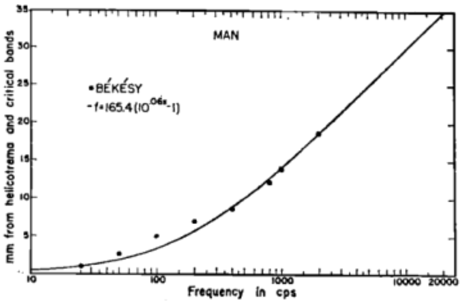


Screen clipping taken: 10/6/2010, 9:26 AM

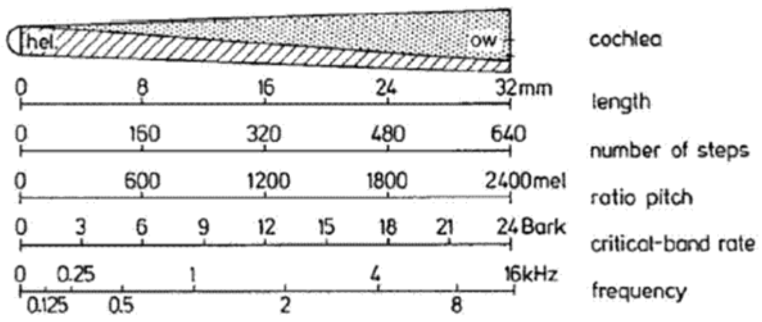
From Lanciani, Auditory Perception and the MPEG Audio Standard, Georgia Tech., 1995

Functional fit of measured position of maximum amplitude on BM to frequency...

From Greenwood, JASA, vol. 33,
no. 10, 1961



← log scale



Response to clicks: alternate positive and negative sound pulses of
100 μs at rate of one every 5 ms

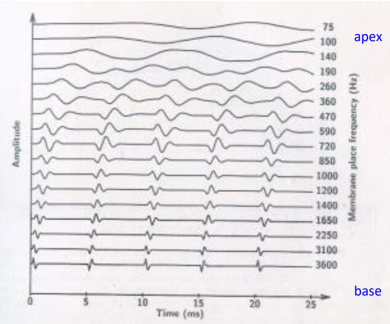


Figure 4.4 Computer simulation for BM displacement as a function of time and CF (noted along the ordinate); each trace is the response for a different point (separated by 0.5 mm) along the basilar membrane to alternating positive and negative sound pulses of 100 μs duration at a rate of 200 pulses/s. (After Flanagan [8].)

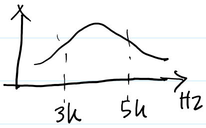
One trace every 0.5 mm along BM

Computational model of hearing: "Auditory excitation" patterns

Motivation? To describe a sound by its auditory nerve level representation.

~ BM "excitation pattern" → neural transduction

Outer-middle ear freq resp ->



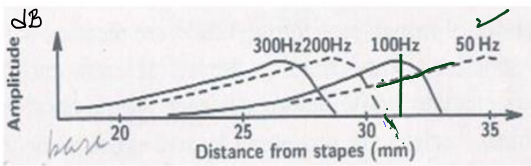
Next:

- The BM frequency selectivity can be modeled by a bank of bandpass filters
- Each point on the BM acts like a tuned filter with a specific center frequency

Can we derive these filter responses from the observed BM vibration pattern?

Screen clipping taken: 10/6/2010, 4:47 AM

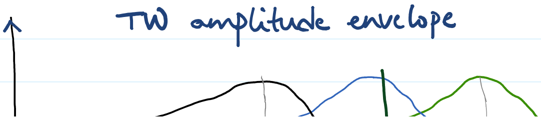
T.W. amplitude envelopes for pure tones



Frequency selectivity of the BM => a Fourier analysis of limited resolution

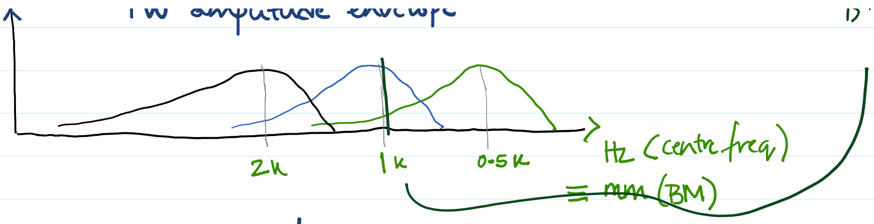
Can we derive a BPF shape at any fixed location on BM? Yes, we can, via the known BM vibration pattern for each frequency component.

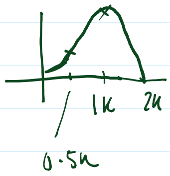
TW amplitude envelope

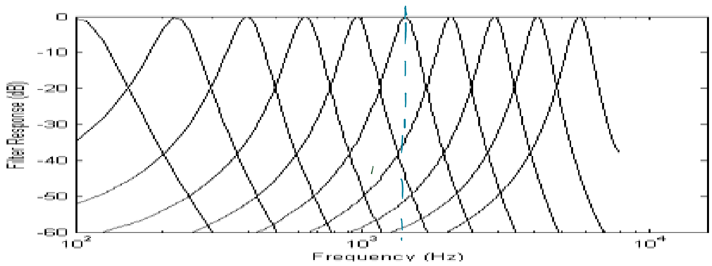


"BPF"

TW amplitude envelope



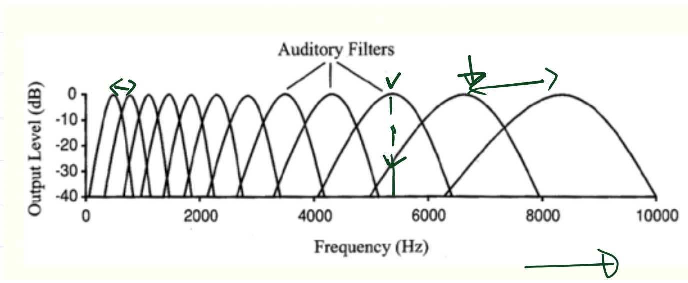




BPF shape at given CF (BM location)

← CF ≡ BM dist (mm)

Linear scale

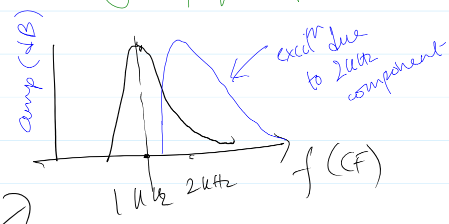


<- BPF at CF= 2 kHz

Let us consider BM Excitation Patterns for

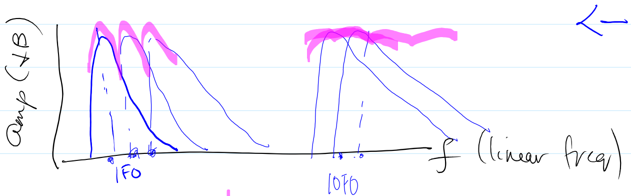
1. A single tone
2. A vowel

Excit^n pattern for single freq component: (1 kHz)

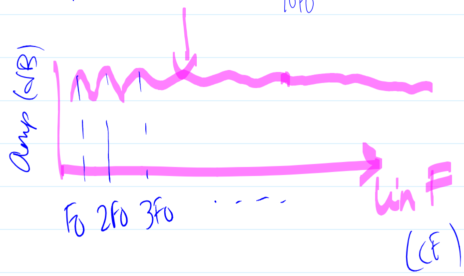


reflected version of BM vibration pattern

Excitn pattern due to a "vowel"



← harmonics at lower freqs are well resolved but no so at higher freqs.



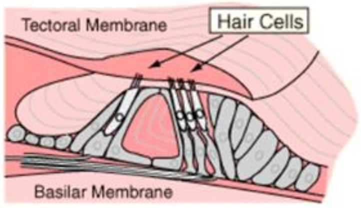
Supporting psychoacoustic expt.

"Hearing out" one tone in a tone complex

→ is easier for lower frequency tones

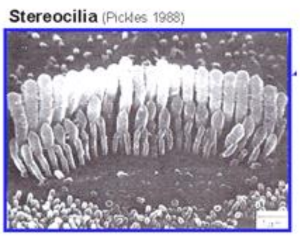
We saw that the cochlea is tuned to frequency as a function of distance along the BM.

The BM is lined with several rows of hair cells. Hair cells are 'active' participants in the mechanoelectric transduction process. Outer hair cells change shape under movement and contribute to active feedback.



HyperPhysics Sound and Hearing, Georgia State University (http://hyperphysics.phy-astr.gsu.edu/hbase/sound/soucon.html#soucon)

Screen clipping taken: 10/10/2010, 9:18 AM



Screen clipping taken: 10/6/2010, 2:23 PM

30,000 sensory hair cells in several rows

From: B. Golstein, Sensation and Perception, Chapter 11

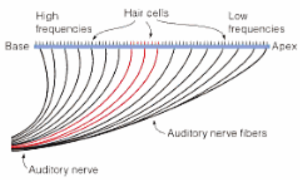


Figure 11.20 Hair cells all along the cochlea send signals to nerve fibers that combine to form the auditory nerve. According to place theory, low frequencies cause maximum activity at the apex end of the cochlea, and high frequencies cause maximum activity at the base. Activation of the hair cells and auditory nerve fibers indicated in red would signal that the stimulus is in the middle of the frequency range for hearing.

Screen clipping taken: 10/6/2010, 2:24 PM

Sensory hair cells are activated when their stereocilia bend in particular direction

=> increase in the *firing rate* of the many auditory neurons connected to the hair cell

Neuronal firing rate increases with increasing vibration amplitude of corresp. BM location

Each nerve fiber follows a ***tuning curve***: the sound intensity needed to lift its firing rate (out of spontaneous rate) as a function of tone frequency

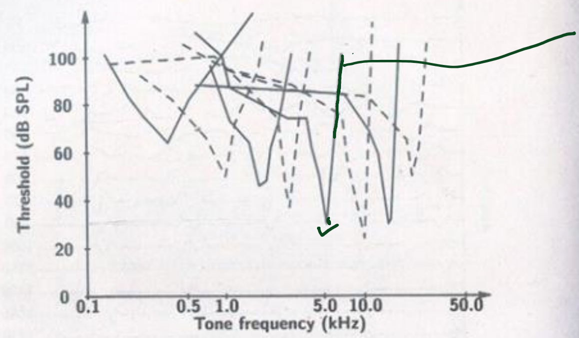
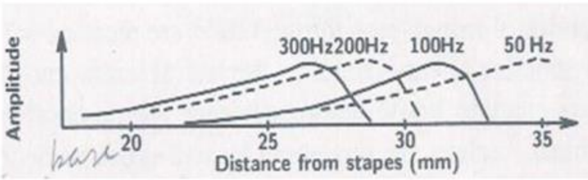


Figure 4.5 Representative tuning curves of cat auditory fibers, for eight different frequency regions. For each fiber, the frequency threshold is a function of stimulating frequency. The curves have been arbitrarily positioned on the ordinate for clarity. (After Evans [33].)

Screen clipping taken: 10/6/2010, 2:25 PM

..obtained from single nerve fibers in the auditory bundle using electrodes

T.W. amplitude envelopes for pure tones

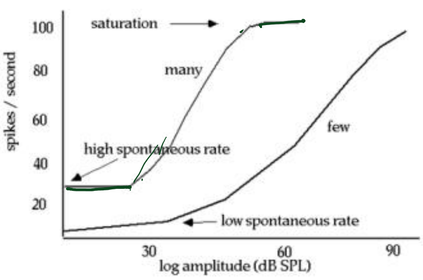


Frequency selectivity of the BM => a Fourier analysis of limited resolution

We note that the neural tuning curves resemble inverted forms of the BPFs with approx constant Q at frequencies above 500 Hz. Only they are more sharply defined than BM responses due to nonlinear amplification via the OHCs' active feedback.

Intensity coding by the ear...

Neural transduction



Each auditory nerve fiber can typically signal only a 20-40 dB range after which it attains saturation

Two types of fibers:

- Low threshold, low dyn range
- High threshold, wide dyn range

=> a max range of 60 dB in which the integrated firing rate increases with sound intensity

So, how to explain the 120 dB dyn range of hearing?

Firing rate of a single fiber of CF = 1700 Hz as a function of stimulus tone frequency and intensity level (inverted tuning curve)

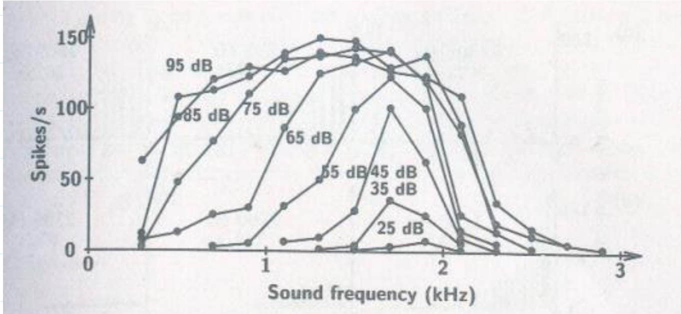


Figure 4.6 Firing rates as a function of the frequency of a stimulus tone for an auditory nerve fiber in a squirrel monkey, plotted for eight different sound intensities (25–95 dB) (iso-intensity contours). The fiber's CF is 1700 Hz. (After Rose *et al.* [43].)

Screen clipping taken: 10/11/2010, 7:42 PM

=> indicates that for all tone frequencies in the neighborhood of 1700 Hz, we have that as stimulus intensity increases, firing rate of 1700 Hz fiber increases until it reaches saturation rate.

Intensity coding by neurons

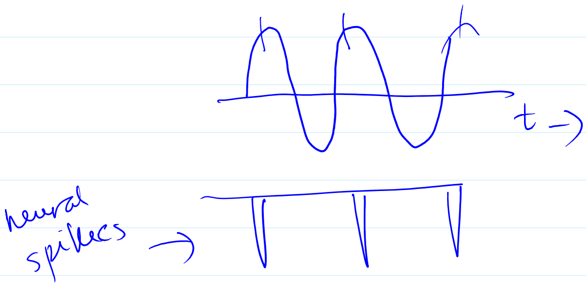
As stimulus tone intensity increases, the stimulus enters the excitatory areas of the other fibers which respond to that frequency only at higher intensities ("recruitment of adjacent neurons"). Thus the intensity increment is coded by the increased overall firing rates among more fibers over a wider frequency range.

However, the precise mechanism of neural coding of intensity is still unresolved

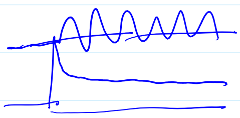
Temporal coding by neurons (*alternate mechanism for coding frequency*)

"Phase locking": At a given tone frequency, the auditory neurons *prefer* to fire in a given phase of the cycle.

Figure:



Interval between successive firings cannot be below 1 ms => at low frequencies (< 1 kHz) of BM vibration, a neuron's spikes can be time-synchronised (phase-locked) to a tonal sound waveform. Phase-locking is completely absent above 5 kHz.



Adaptation: The changing sensitivity in response to a continued stimulus. Neurons fire at high rates at the onset of the stimulus but then adapt slowly to half the rate within 15-20 ms later. Serves to emphasize sudden spectral transitions.

Lecture-Oct11-b
Monday, October 11, 2010 6:30 PM

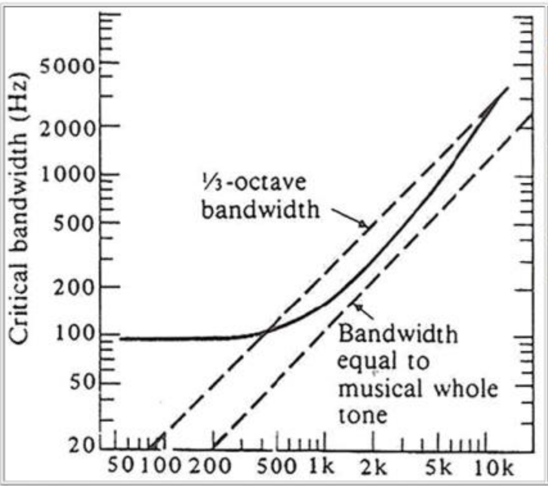
Cochlea: Frequency analysis

The cochlea can be modeled as a bank of overlapping bandpass filters.
The auditory filter bandwidth is known as "critical bandwidth".

Critical bandwidth has been measured in many distinct ways

- Bekesy-type observations of BM vibration (or neural tuning curves)
- Psychoacoustic experiments on loudness, pitch and masking whose interpretation points naturally to critical bandwidth (to be discussed later)

Critical bandwidth as a function of filter center frequency (from Rossing, 1982)
The audible range has typically been modeled by 24 critical-bandwidth filters



Center frequency (Hz)	Critical bandwidth (Hz)
100	90
200	90
500	110 ISO
1,000	280
2,000	700
5,000	1,200

Screen clipping taken: 10/11/2010, 6:36 PM

The critical bandwidth corresponds to a fixed distance (1.5 mm) on the basilar membrane (= approx 1200 hair cells)

Next : BM excit^n Pattern → auditory nerve activity

(Spatial distribution)

So far, we have seen how we can in principle derive the BM excitation pattern by a filterbank model for the cochlea. The filter parameters are based on the observed excitation patterns which are invariant in the log freq scale.

To complete the computational auditory model, we need to fill in details about the nonlinear conversion from BM vibration pattern to neural spike train. We cannot rely any further completely on physiological observation/measurements. So we resort to psychoacoustic studies….

Psychoacoustics

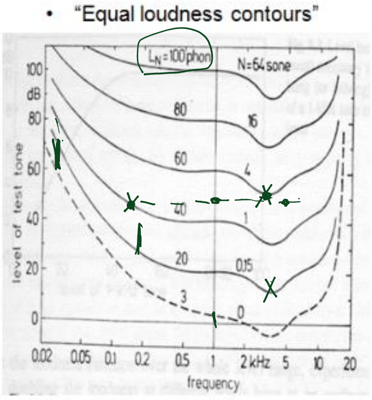
- Using listening experiments to understand the relationship between perceived sensation and the physical features of a stimulus

- Sensations:
  loudness (includes masking), pitch, roughness, fusion

Loudness perception

Absolute thresholds and equal loudness contours

• "Equal loudness contours"



Screen clipping taken: 10/13/2010, 9:52 AM

<= Measured via listening experiments

- Tones of equal SPL but different frequencies do not sound equally loud ->
  Loudness depends on sound pressure level and frequency!

  The bandpass effect is attributed to the transfer function of the outer-middle ear and to a drop in the number of hair cells towards the extremes of the BM.

- Growth of loudness is different for different frequencies.

- For a sound to be perceived as twice as loud (*unit*: **sone**), its intensity must be increased by a factor of 10 at 1 kHz frequency region.

  Loudness sensation comes from total neural activity which is a nonlinear function of stimulus intensity (see intensity coding by nerve fiber).

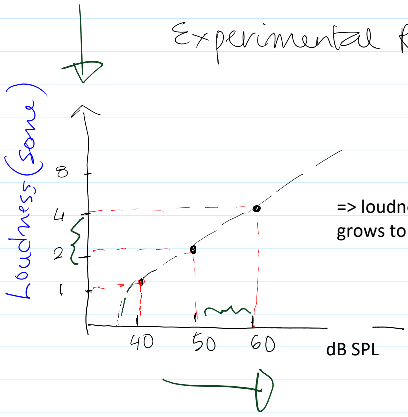
Note the phon unit of loudness. It indicates that loudness in phon increases linearly with dB intensity (like nerve cell response)

Experimental Results for the 1 kHz tone.

Linear variation on log-log axes

$L = kI^p$

$\log L = \log k + p \log I$



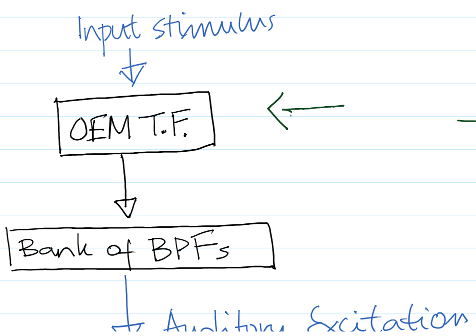
=> loudness doubles when intensity grows to 10 times

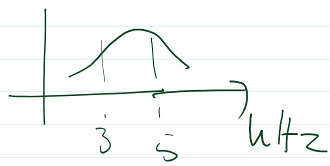
Steven's law
$L\text{ (sone)} = k \times I^p$

p=0.3
k depends on frequency

Computational model of hearing

Acoustic pressure signal
↓ *Frequency selectivity*
Mechanical vibration pattern on basilar membrane
| *Compressive nonlinearity*





Mechanical vibration pattern on basilar membrane

↓ *Compressive nonlinearity*

Spatial distribution of neural activity in auditory nerve

Bank of BPFs

↓ Auditory Excitation Pattern

Non-linear conversion to Loudness

↓ Spatial distribution of loudness

Integrate

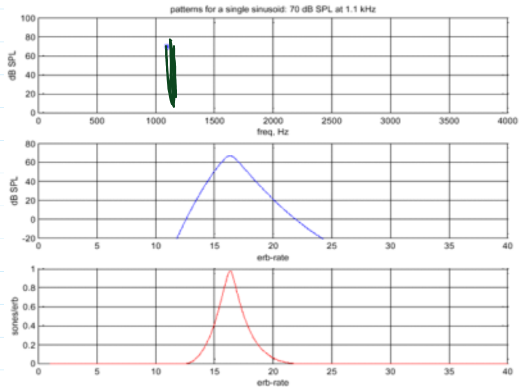
↓

Total Loudness

Critical band

Outputs of the computational model for a tone stimulus

(1100 Hz)



nerve activity

spatial dist

Screen clipping taken: 10/13/2010, 8:43 PM

Extraction of *features* from the temporal and spatial pattern of neural activity such as:

pitch, loudness, timbre, other cues

- loudness ←→ *total neural activity caused by the sound*
- timbre ←→ *spatial distribution pattern of neural activity*
- pitch ←→ *position of local peaks in the spatial distribution*

So, how does our model explain well-known psychoacoustic observations.....