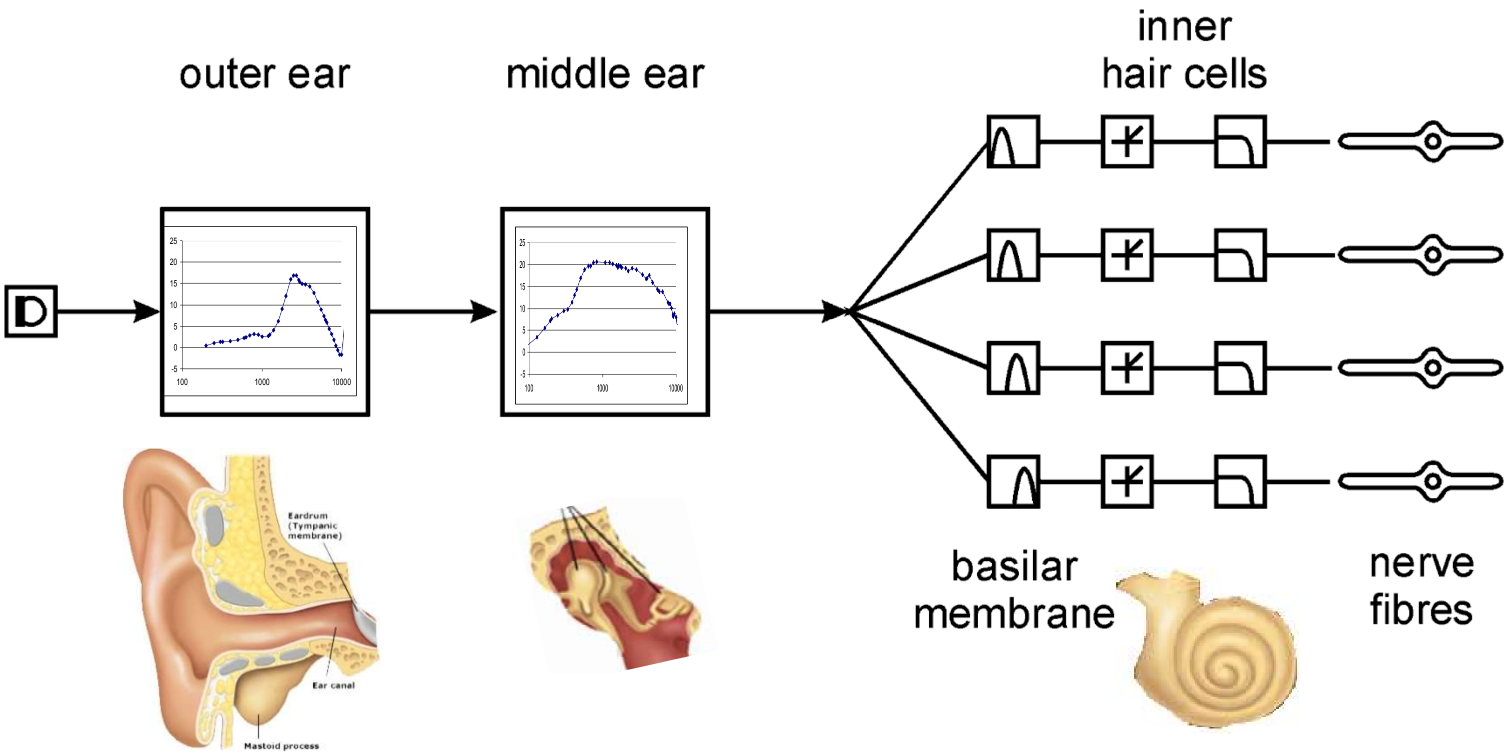
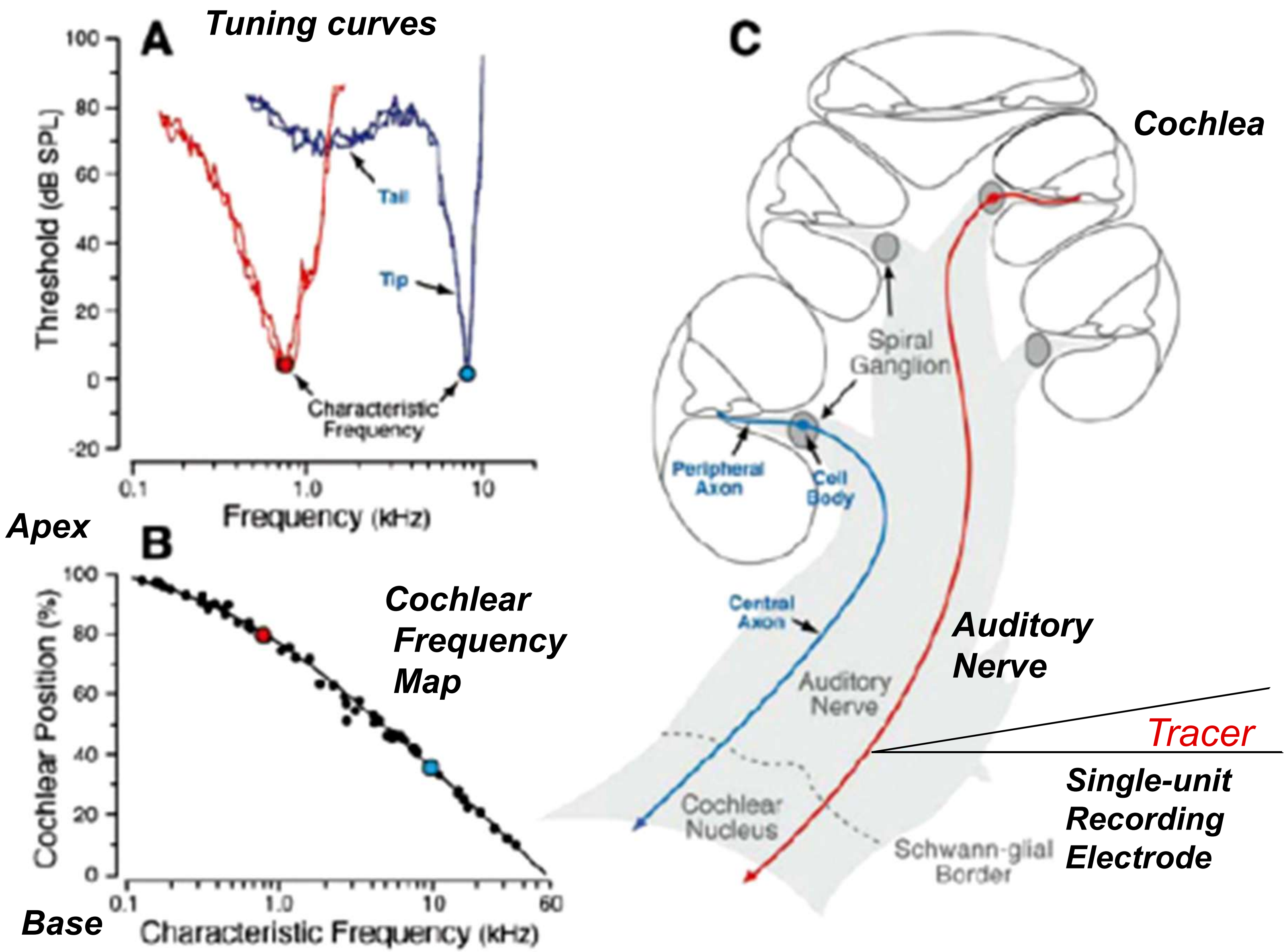
AUDL GS08/GAV1
Signals, systems, acoustics and the ear

Pitch & Binaural listening

Review

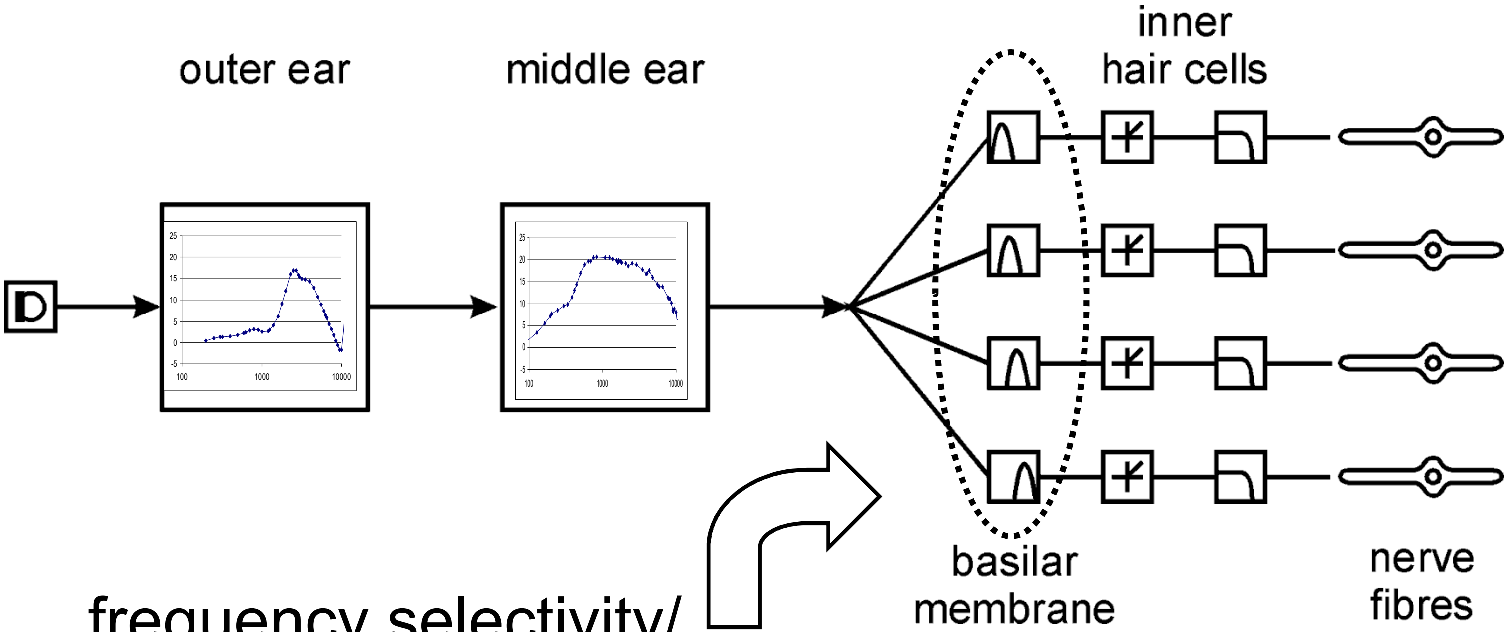

outer ear
middle ear
inner
hair cells
Eardrum
(Tympanic
membrane)
Ear canal
Mastoid process
basilar
membrane
nerve
fibres

Part I: Auditory frequency selectivity



Liberman (1982)

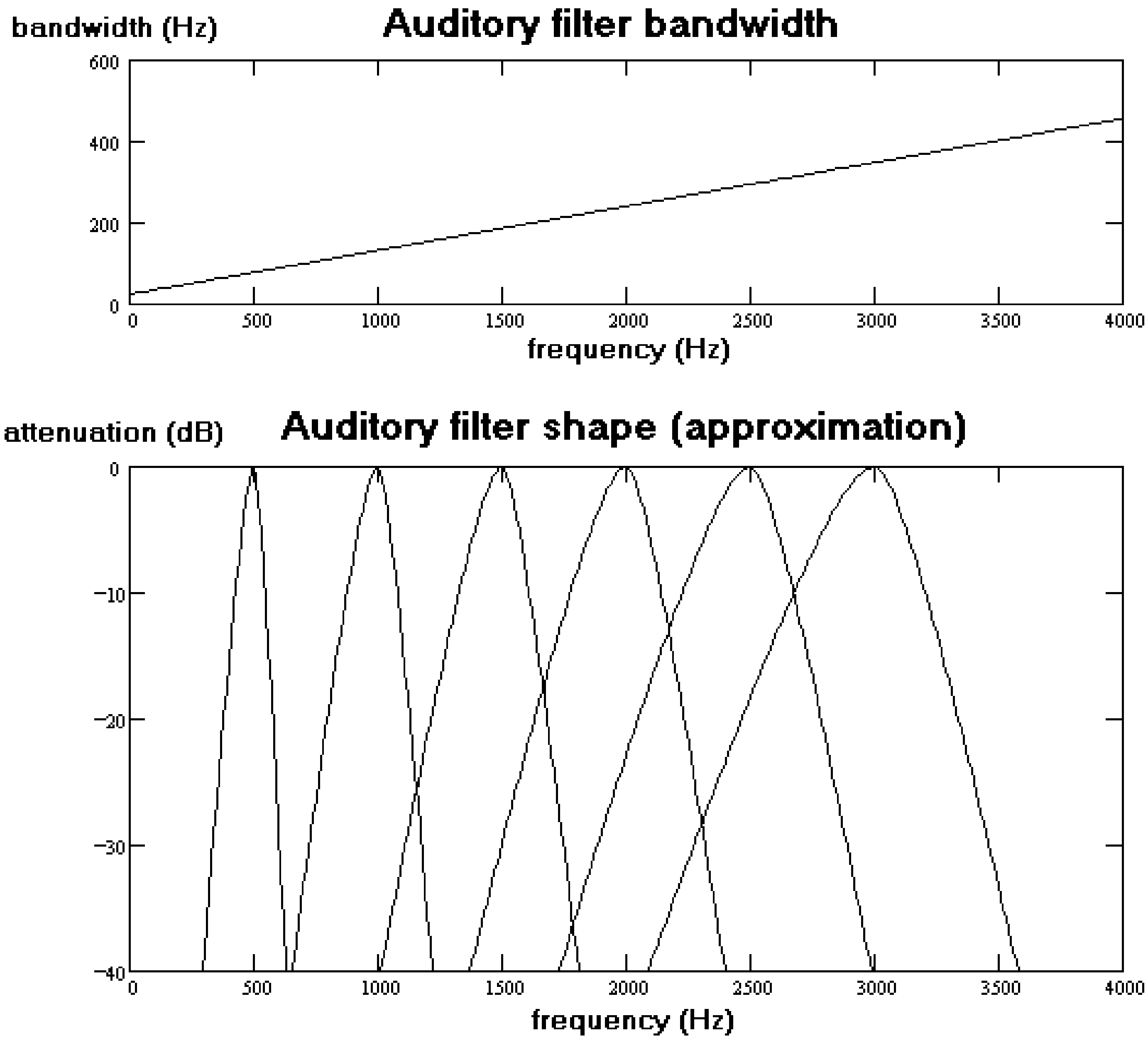
The auditory periphery as a signal processor



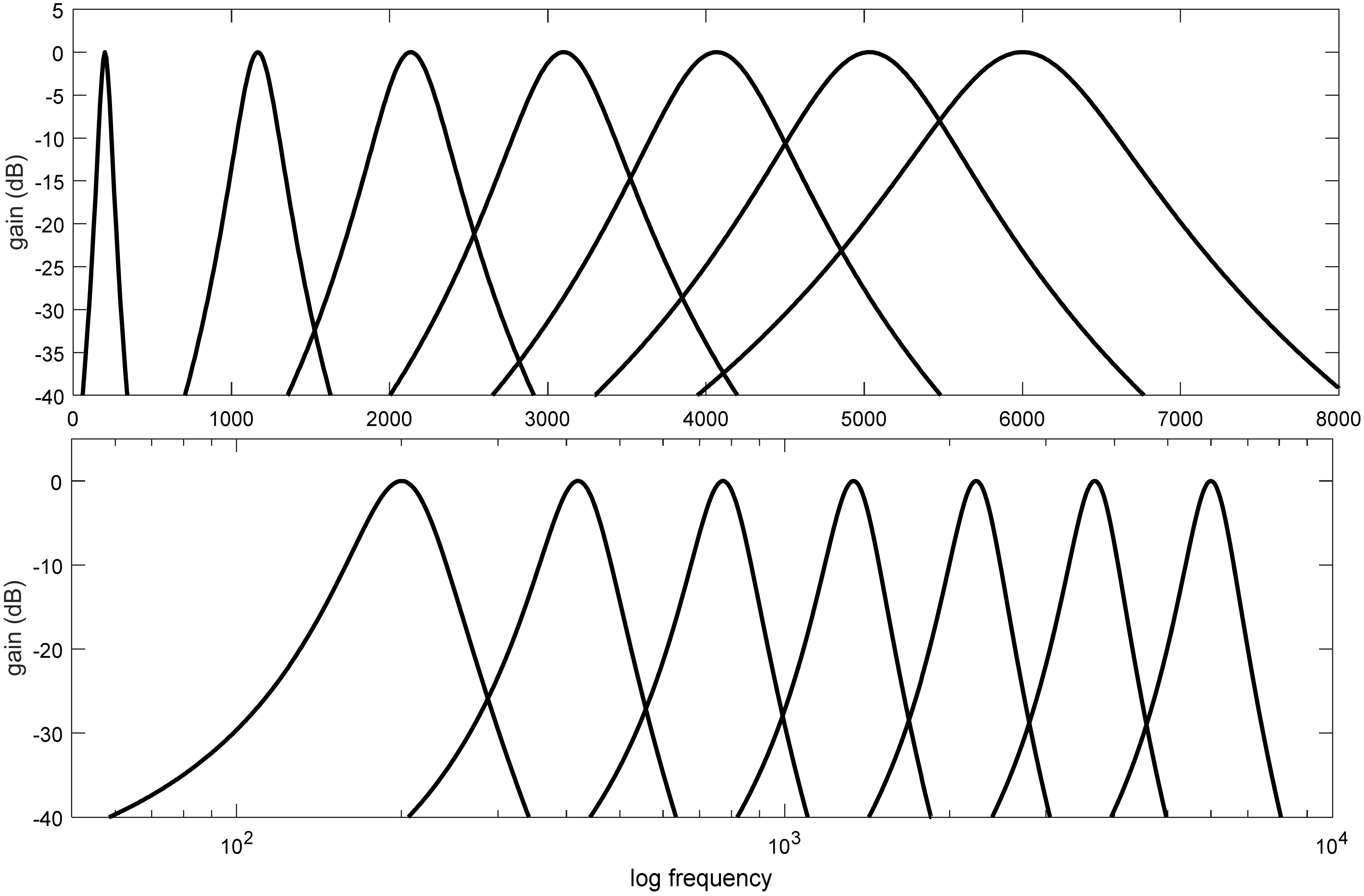
frequency selectivity/
frequency analysis

auditory filters & channels

Auditory filter band-widths vary with frequency (note linear frequency scale)

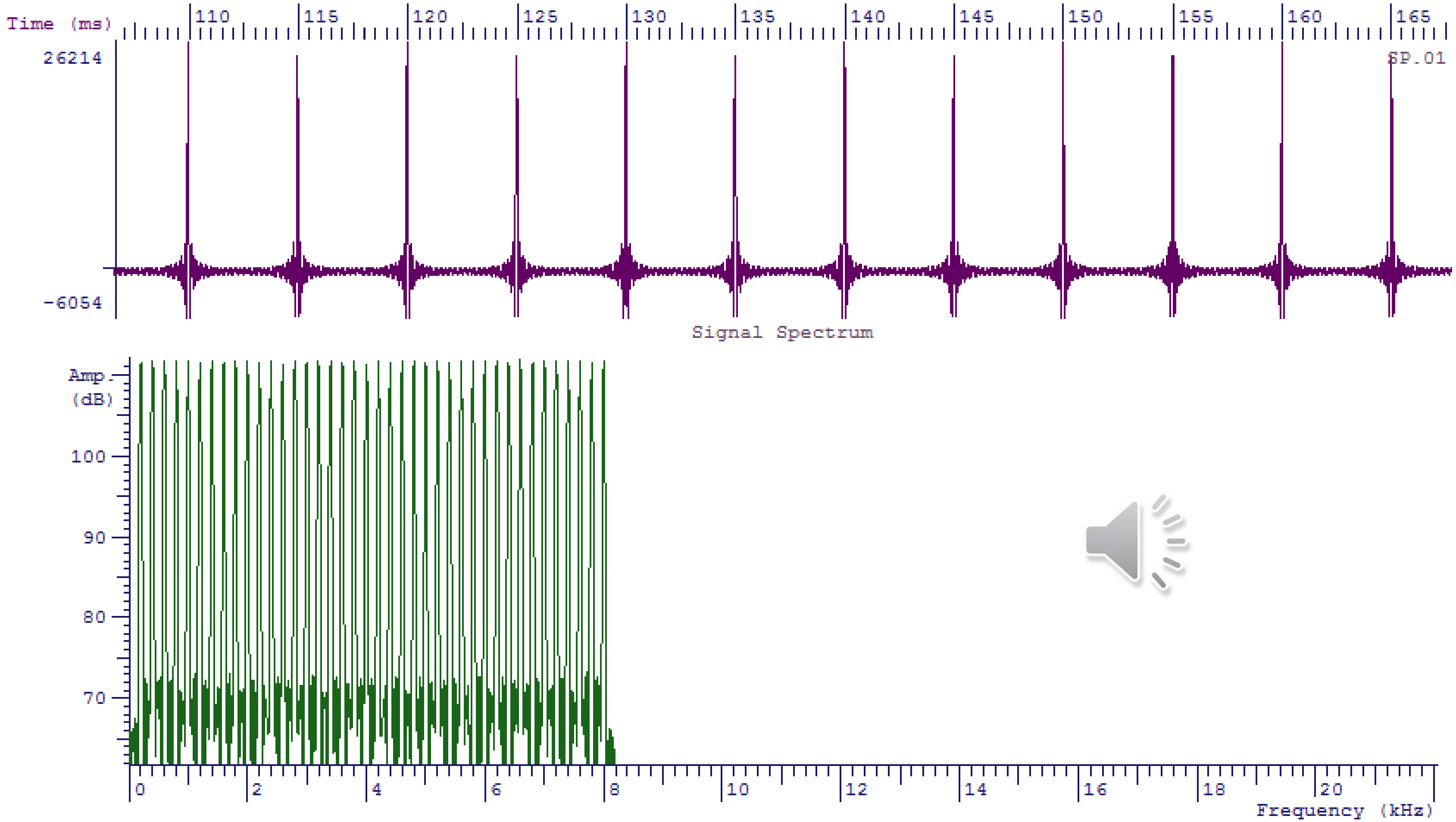


Auditory filters on linear & log frequency scales

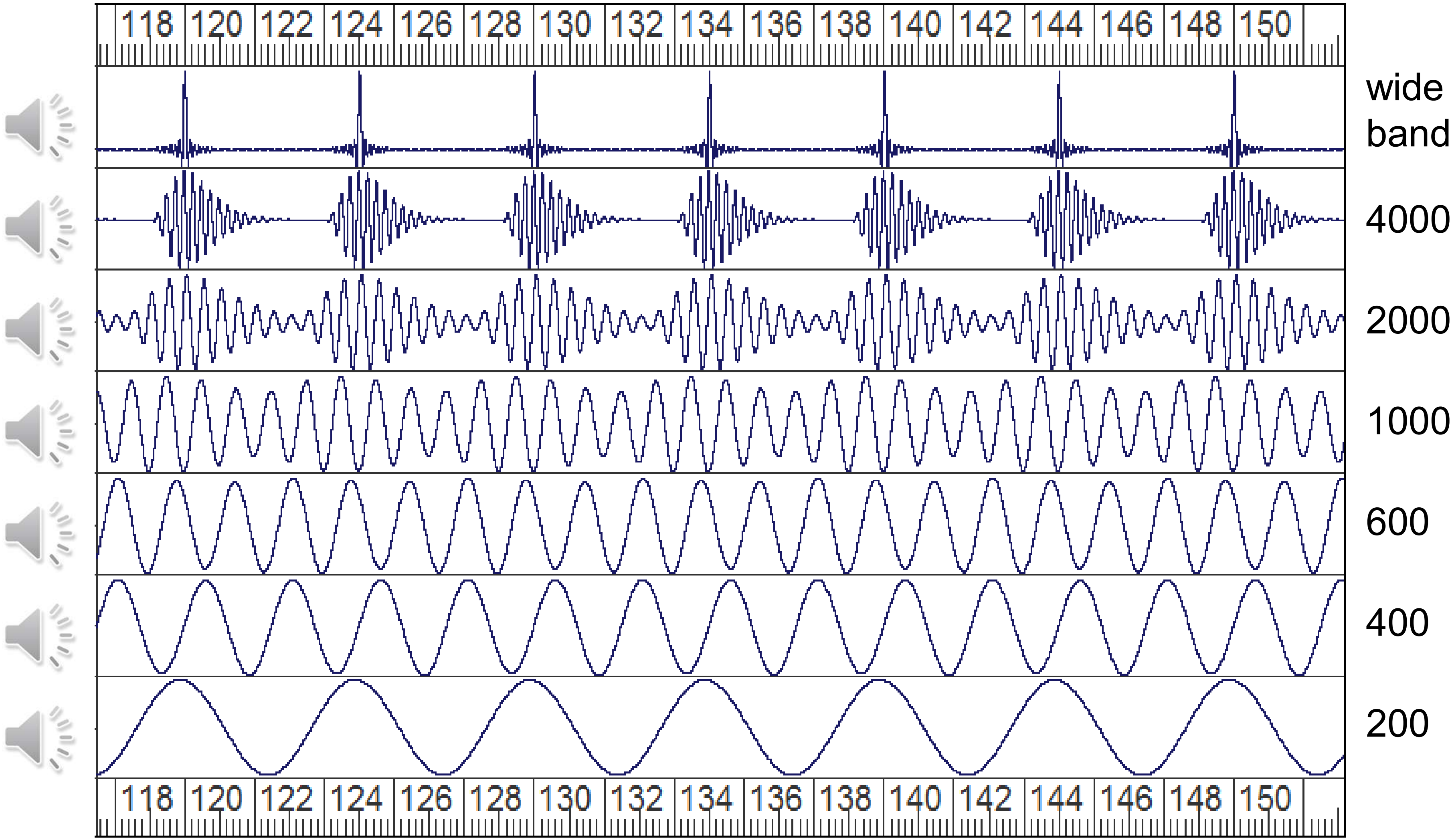


A complex periodic wave
(40 equal-amplitude harmonics of 200 Hz)
through an auditory-like filter bank

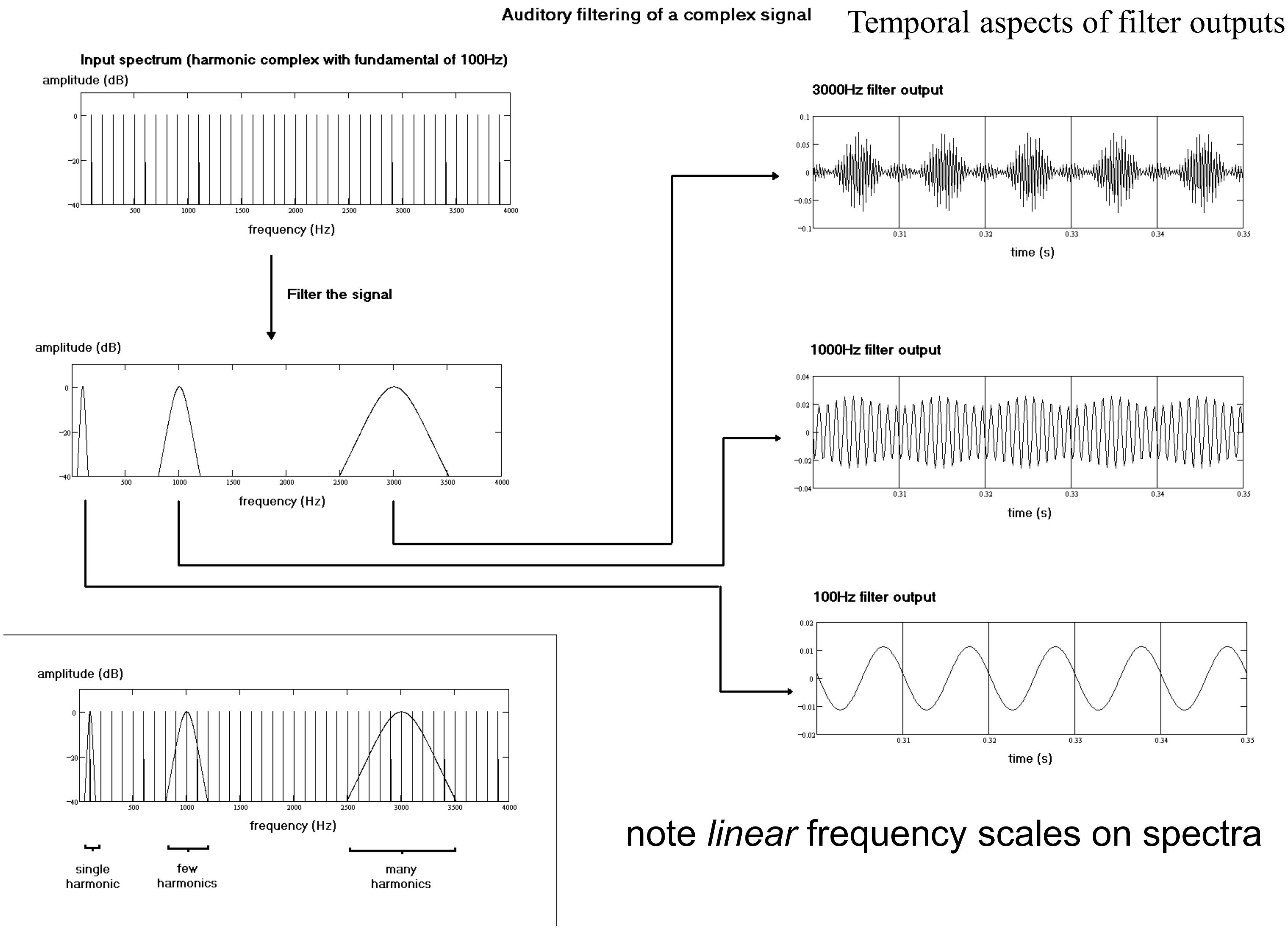
The complex periodic wave (20 equal-amplitude harmonics of ? Hz)



Auditory filtering of a harmonic complex

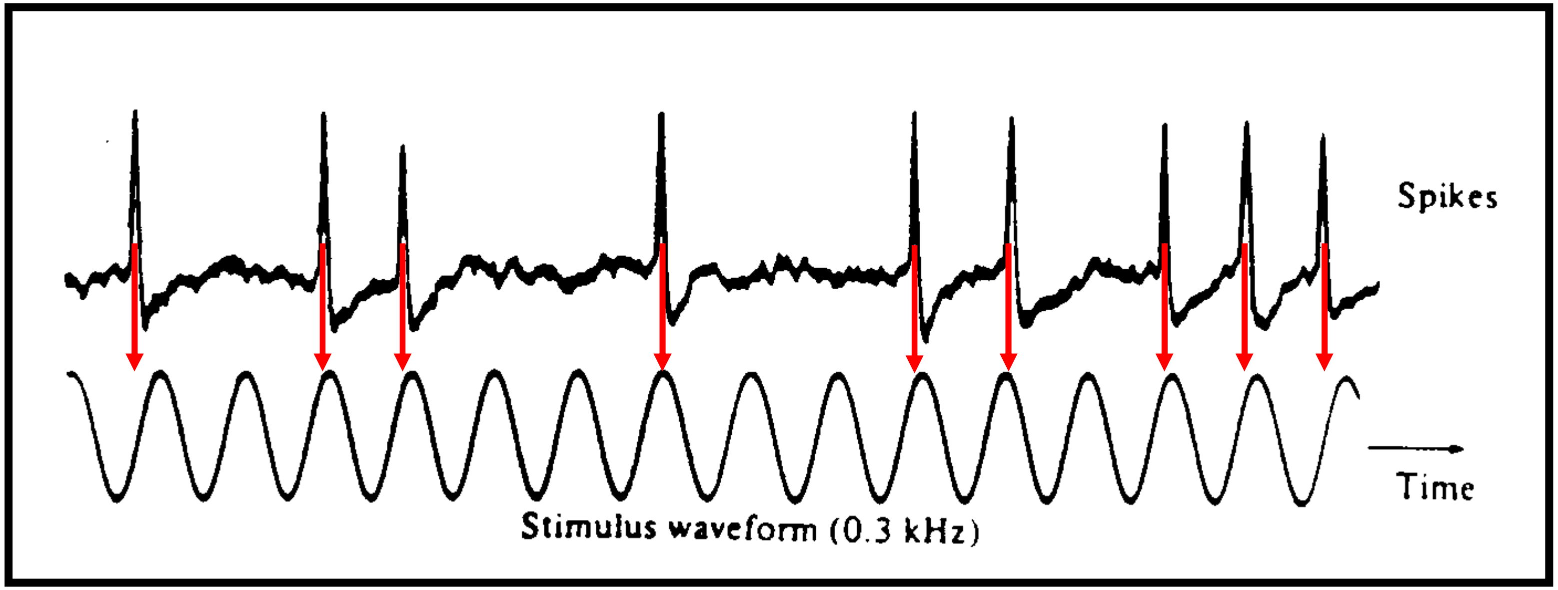


Auditory filter bandwidth varies across frequency



note *linear* frequency scales on spectra

Part II:
The auditory nerve
phase locks to low-frequency tones



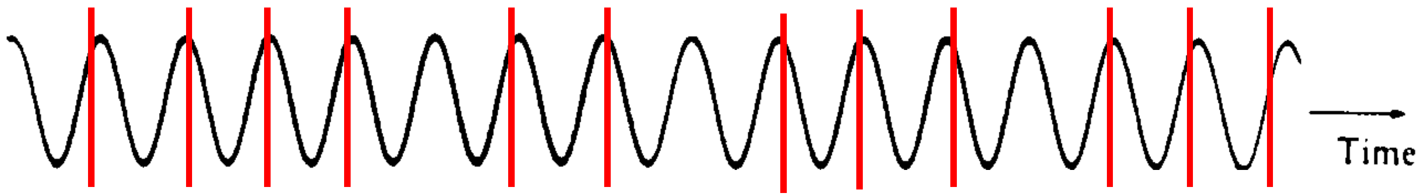
Not the same as firing *rate*!

Evans (1975)

Repeat:
Not the same as firing *rate*!

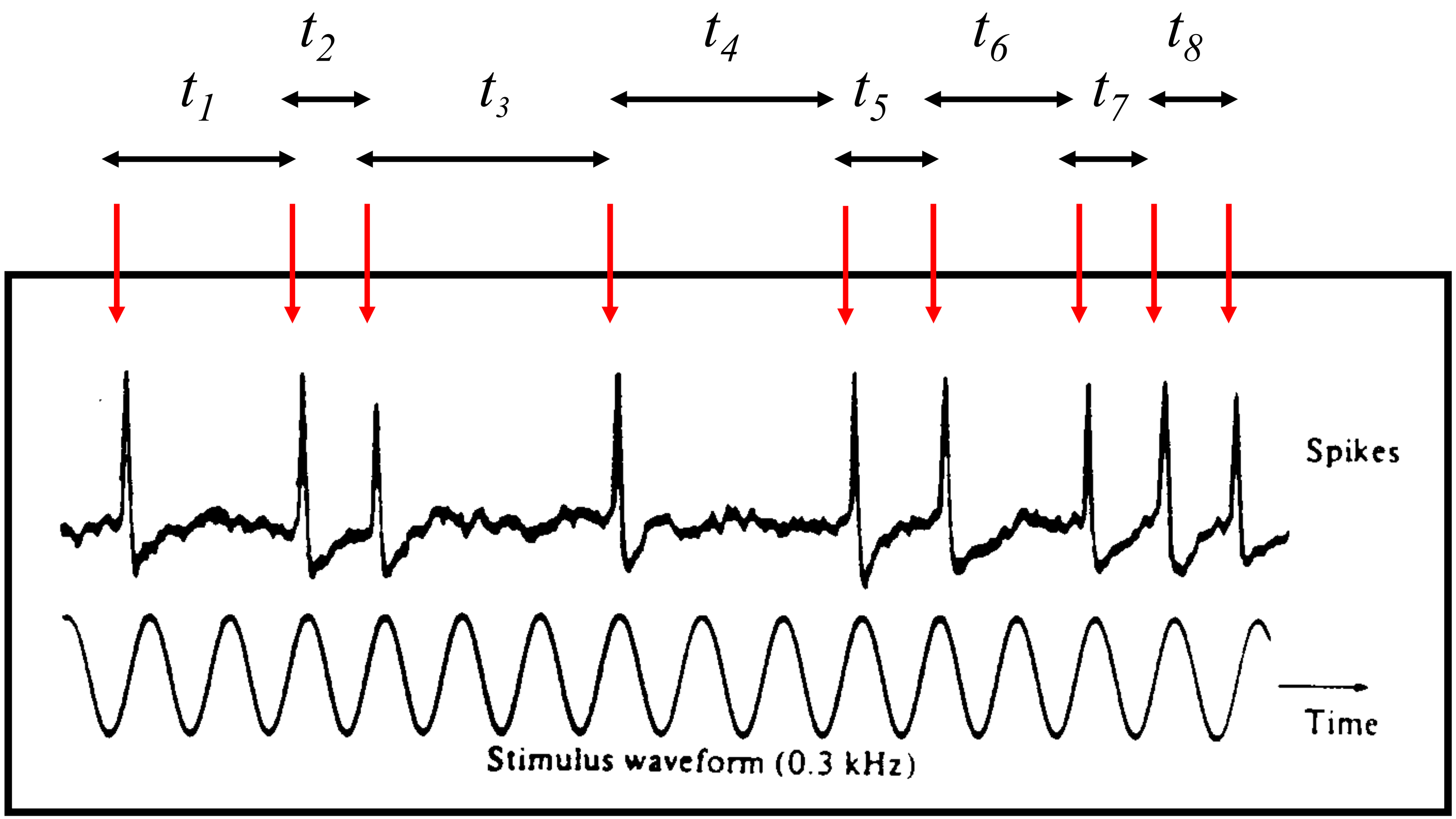
as previously: 9 spikes

faster: 12 spikes

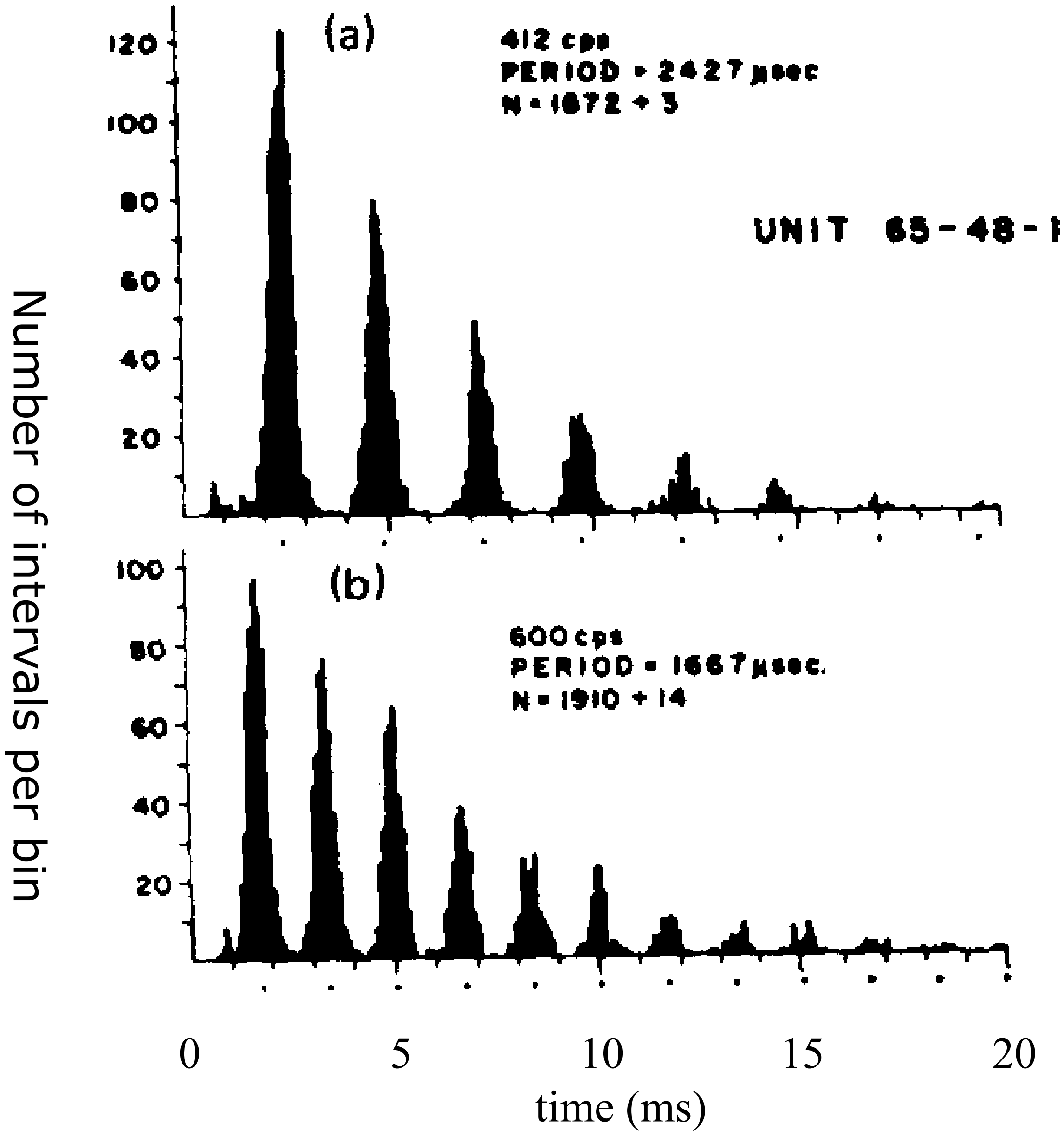


slower: 4 spikes

Constructing an *interval histogram*



Interval histograms for a single AN fibre at two different frequencies



Phase locking is limited to lower frequencies

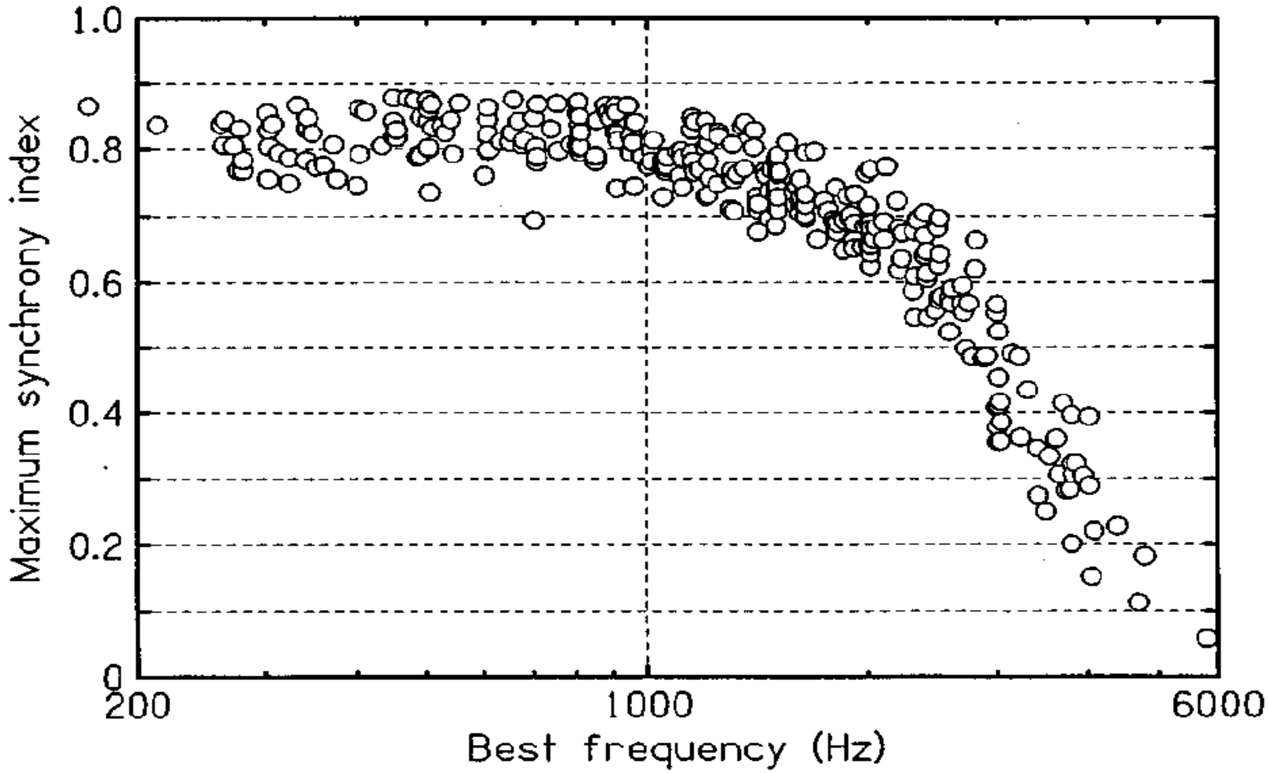
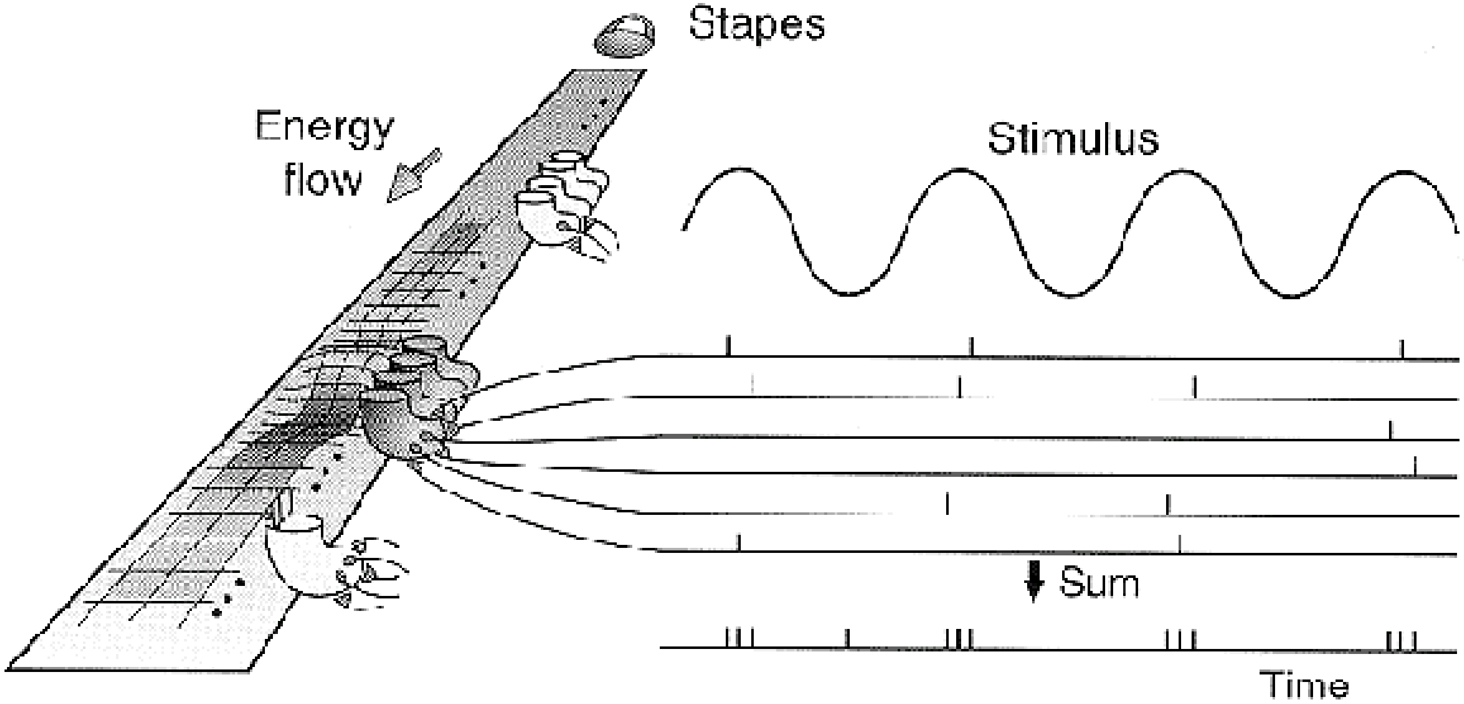


FIG. 10. Synchrony coefficient for 315 neurons as measured by D. H. Johnson who provided data for this plot (courtesy AIP Press).

Neural stimulation to a low frequency tone



Sound energy propagates to the characteristic place of the tone where it causes deflection of the cochlear partition. Neural spikes, when they occur, are synchronized to the peaks of the local deflections. The sum of these neural spikes tends to mimic the wave shape of the local deflections.

Auditory Temporal Coding (phase locking)

Neural responses are synchronised to the input signal

- No synchrony at rates above ~ 1.5 kHz
- Important for the encoding of pitch
- Synchrony is to input ***after*** the band-pass filtering of the basilar membrane

Pitch

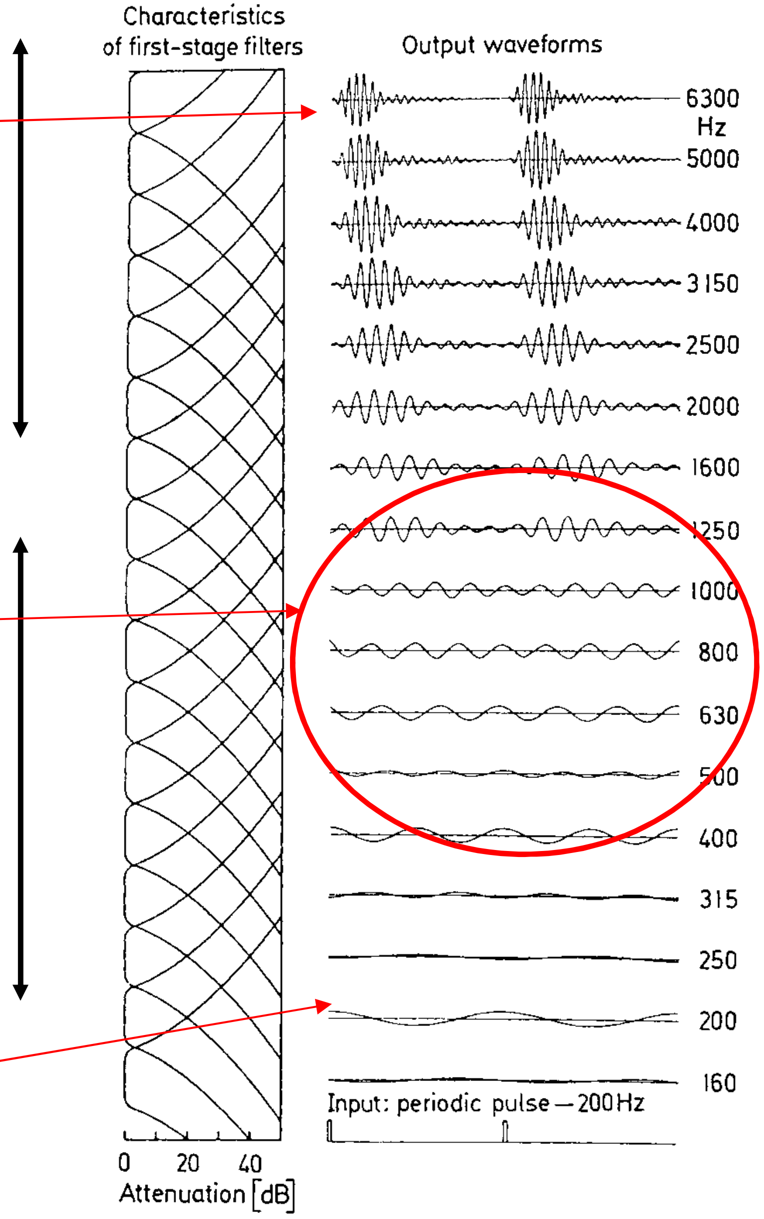
- In speech
  - Linguistic pitch variation conveys ...
    - intonation, which indicates lexical stress and aspects of syntax, etc
    - lexical information in tone languages
  - Paralinguistic (or indexical) information
    - age and sex
    - emotional affect
- In music, melody
- In environmental sounds?

Where are cues to pitch?

Output of high frequency filters shows periodicity corresponding to (F0)

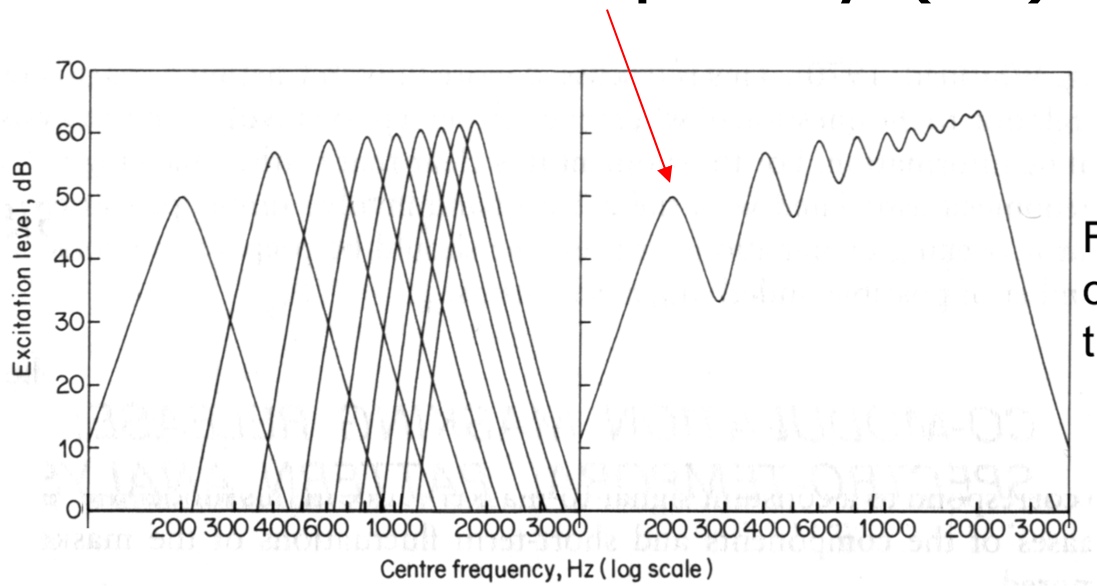
First few harmonics are also resolved

Harmonic at fundamental frequency (F0)



Classical Place account of pitch

- Pitch of a complex sound determined by position of peak in excitation pattern due to basilar membrane response to fundamental frequency (F0) component



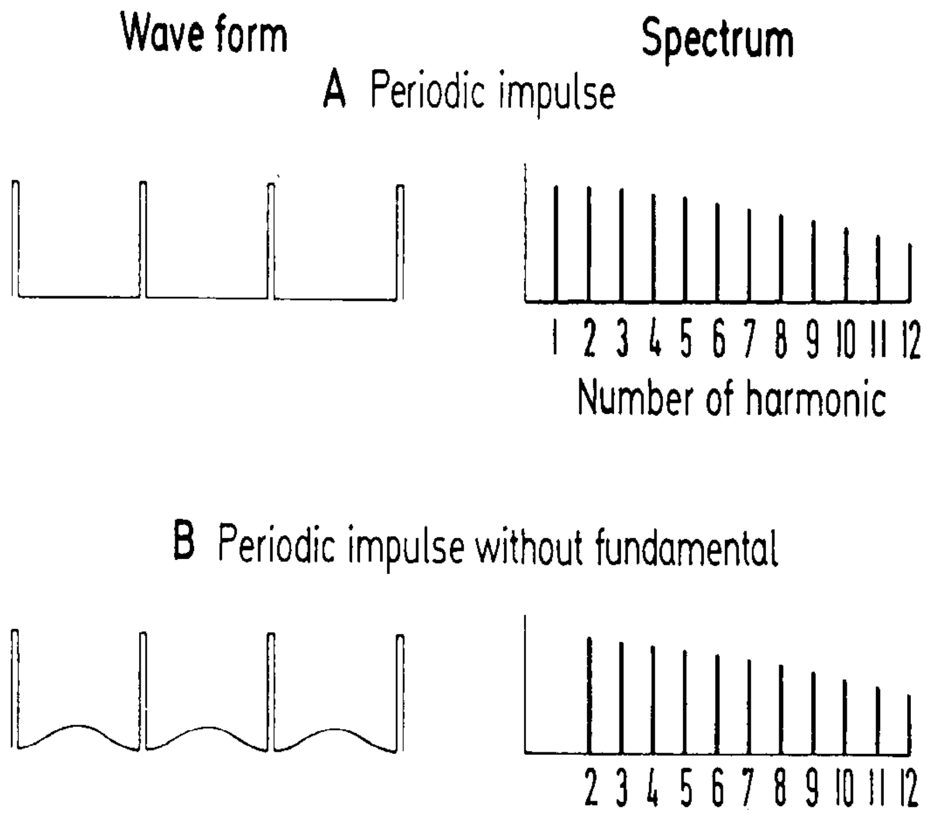
First 9 harmonics of 200 Hz pulse train

Excitation patterns of each component individually

Overall excitation pattern (harmonics all present together)

The missing fundamental

Schouten (1938, 1940) used a train of narrow pulses (many harmonics).



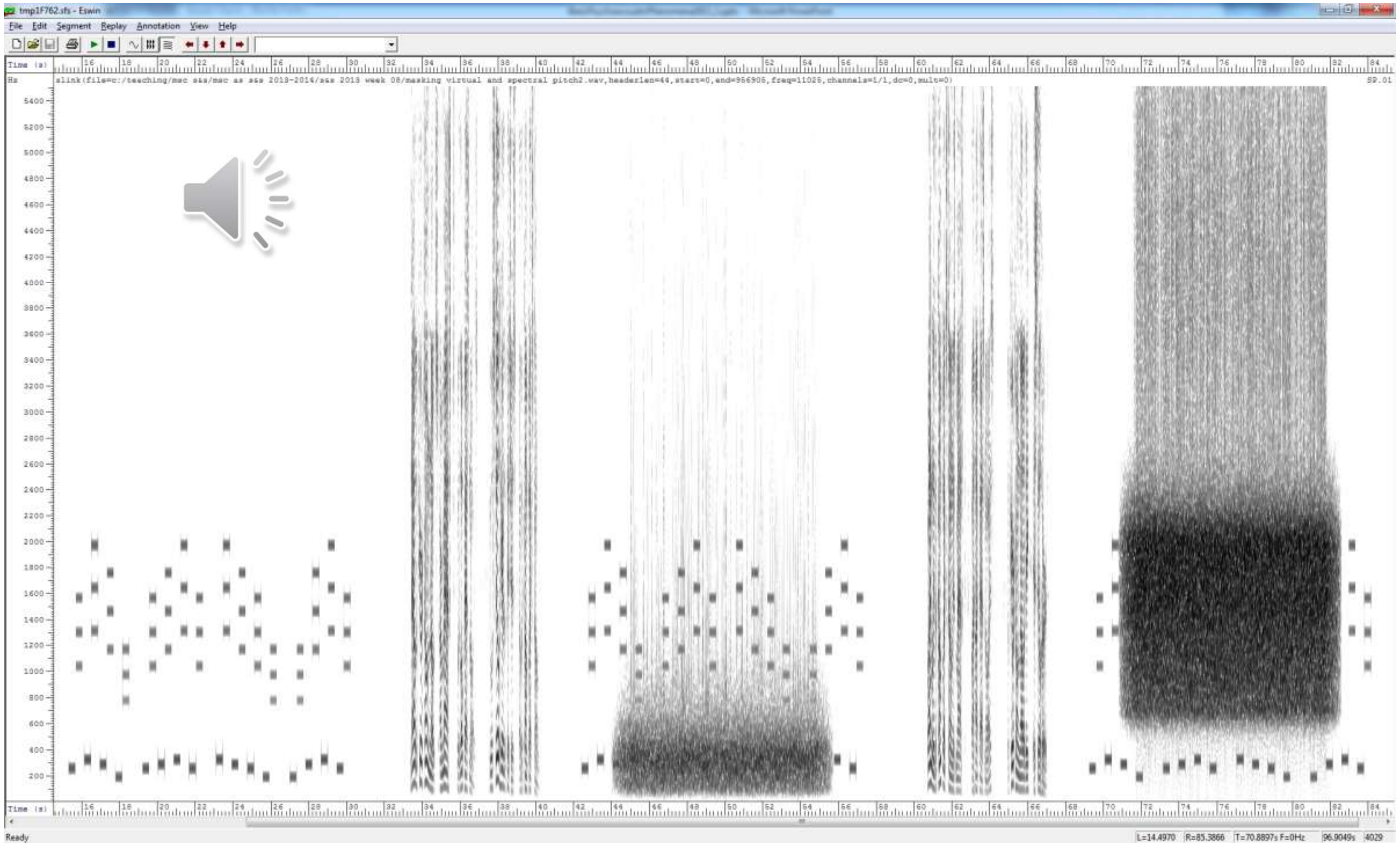
Place theory says: The pitch is due to the lowest frequency component, at the fundamental frequency (F0).

Now remove the component at F0.

Place theory says: The pitch should change

The pitch of complexes with a missing fundamental

frequency →



time→

CoolEdit - masking virtual and spectral pitch

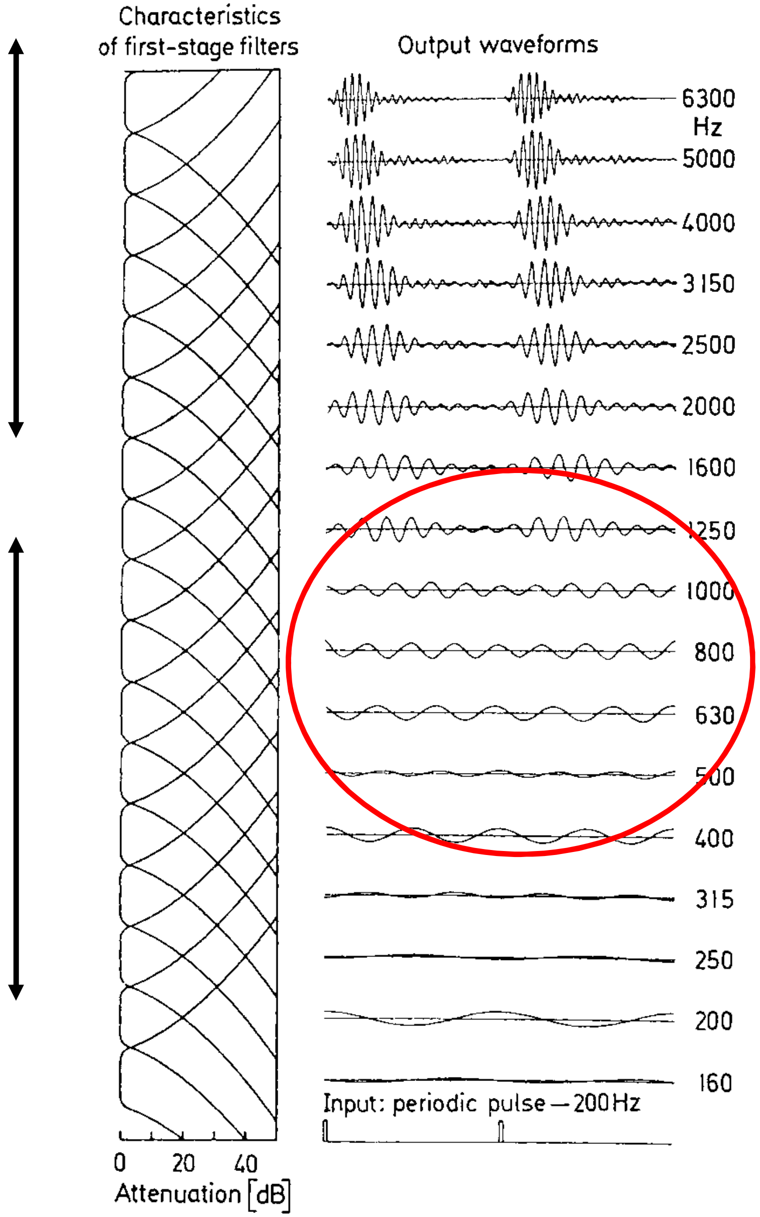
Pitch cues

- The harmonic at F0 is NOT a primary cue
  - Pitch doesn’t change when it is absent
- Periodicity at $1/F_0$ from unresolved harmonics is NOT a primary cue either
  - Provides only a weak pitch percept

Where are cues to pitch?

Primary pitch cues are from resolved harmonics – whose frequencies are individually encoded in the time pattern of nerve firing

A pattern recognition process estimates the period of a sound from the relationship between the periods of its resolved harmonic components



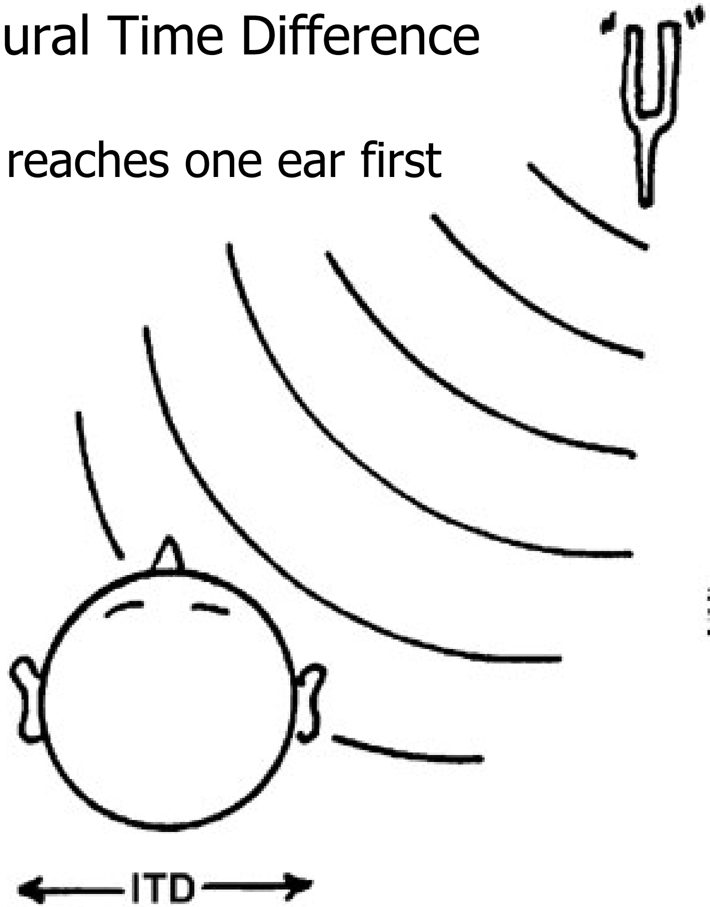
Binaural Hearing

- What does the auditory system gain from having input from two ears rather than one?
  - **Localisation in azimuth**
    - determine the direction from which a sound is coming
  - **Masking release**
    - under certain circumstances the auditory system can detect sounds at lower levels when using two ears instead of only one
    - contributes to 'cocktail party effect'

Duplex theory of sound localization

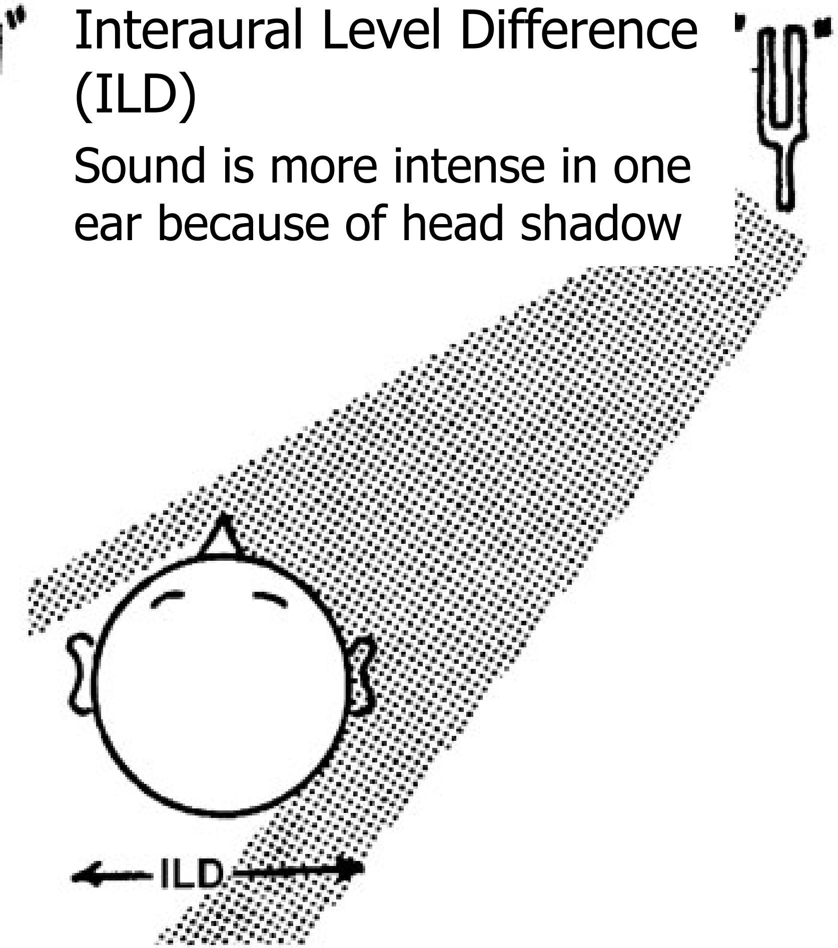
Interaural Time Difference (ITD)

Sound reaches one ear first



Interaural Level Difference (ILD)

Sound is more intense in one ear because of head shadow



low frequencies

high frequencies

Sound Localisation

We're very good at judging the location of broadband sounds

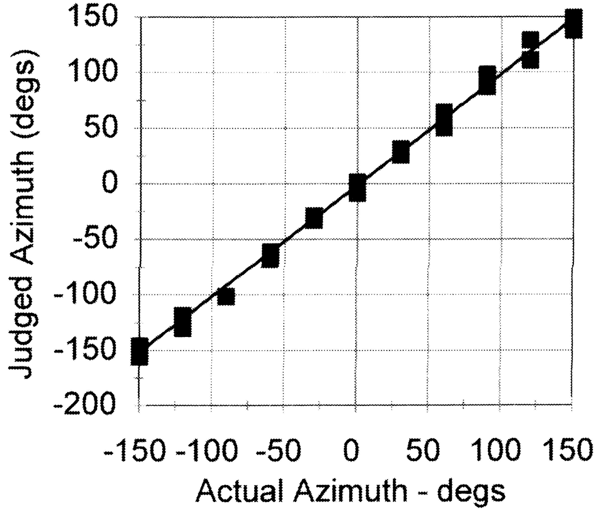


FIGURE 12.5 The judged location in the horizontal or azimuthal direction of a broadband noise source presented at different locations. The diagonal straight line represents perfect judgments (data from one listener, see Figure 12.9). Adapted with permission from Wightman and Kistler (1989b).

But for pure tones localisation errors increase for frequencies around 2 kHz

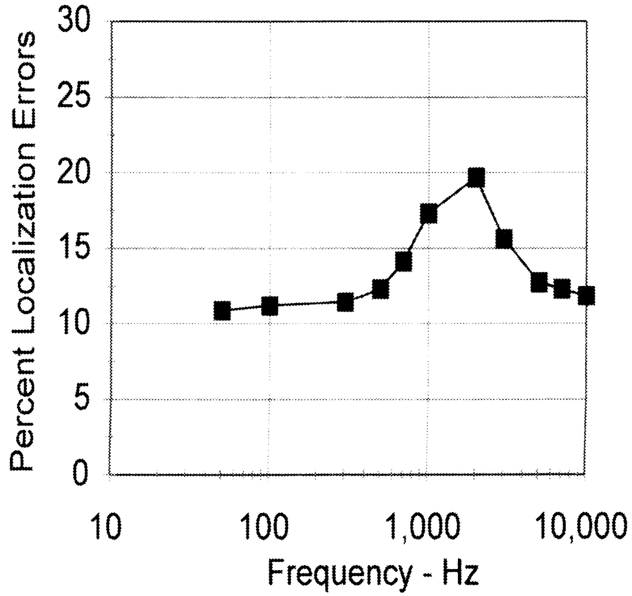


FIGURE 12.4 Errors (in terms of percentage of judgments made) in judging the location of a sinusoidal sound source shown as a function of frequency. Adapted with permission from Stevens and Newman (1936).

Sound Localisation

- ILDs dominate for high frequencies
  - Head shadow affects high frequencies much more than low

- ITDs dominate for low frequencies
  - ITDs become ambiguous when frequency becomes too high (in relation to time for sound to travel through head)

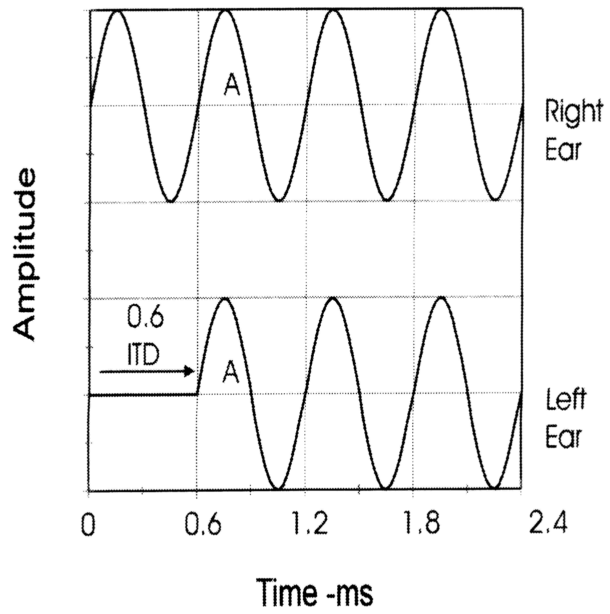


FIGURE 12.8 A 1666-Hz tone presented to the right side of a listener so that it reaches the right ear 0.6 msec before it reaches the left ear. (**Top**) Sinusoid at right ear; (**Bottom**) sinusoid at left ear. After the first peak (A), the waveforms arriving at each ear are in phase, which would indicate that the sound is in front rather than to the right.

Localisation for elevation is not binaural

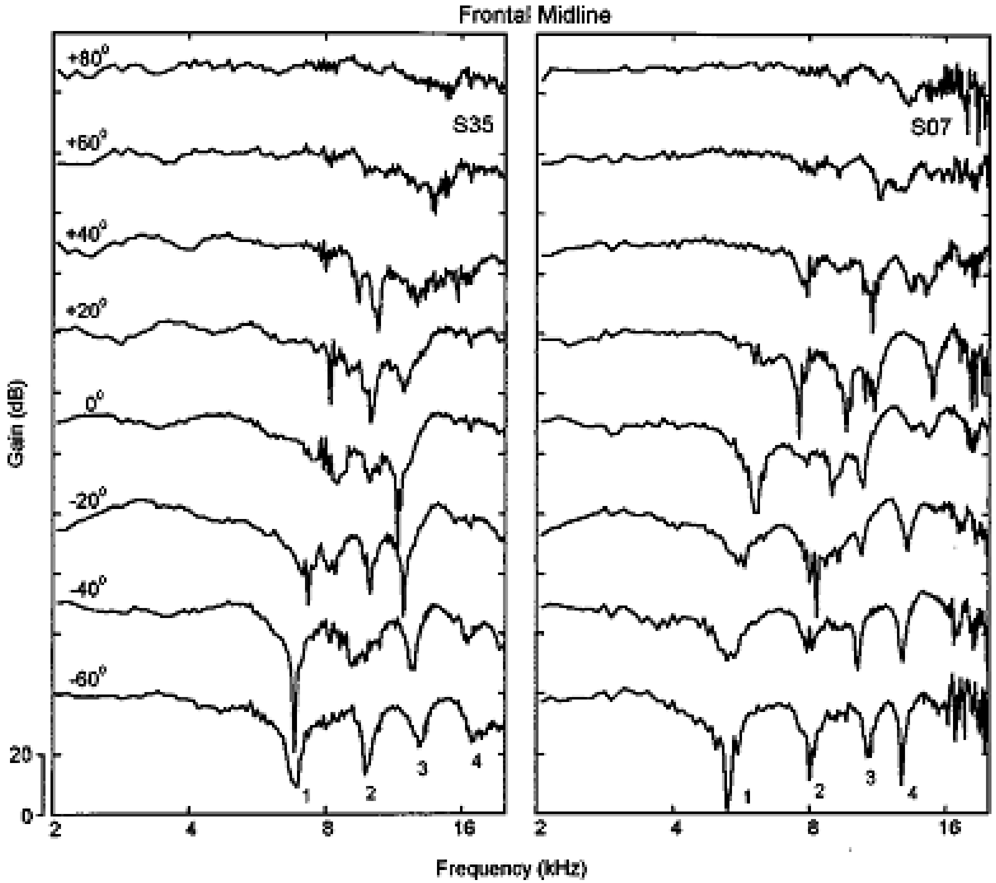


FIG. 1. Directional transfer functions (DTFs) for directions on the frontal midline at the elevations indicated relative to the horizontal plane. Left and right panels show DTFs measured from the right ears of subjects S35 and S07, respectively. Numbers 1 through 4 indicate spectral notches discussed in the text.

Because there are no or few interaural differences

Changes in spectrum appear to be crucial, in particular *spectral notches*

Middlebrooks 1999 JASA: 106 (3)

Masking Release -

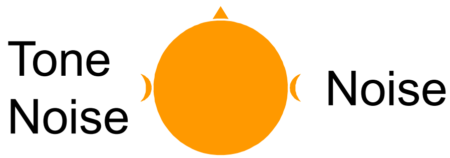
Binaural Masking Level Difference (BMLD)

- 3 stimulus configurations
  - **Monotic**: one ear only
  - **Diotic**: identical in each ear
  - **Dichotic**: different in each ear
- No differences in thresholds for a masked signal (a signal in noise) between *monotic* and *diotic* presentations…
- …but a masked signal can be easier to detect with *dichotic* presentation

Binaural Masking Level Difference (BMLD)

- Measure the threshold of a tone in noise presented to one ear only

- Present the identical noise to the other ear as well (signal still only in one ear)



- The threshold will now be lower
  - Even though all we've done is add more noise!

Binaural Masking Level Difference (BMLD)

- Difference in threshold for tone presented out of phase between ears vs in phase.

  (Same noise in both ears)

- The amount of release from masking varies with the frequency of the tone.

- BMLD relies on ITDs

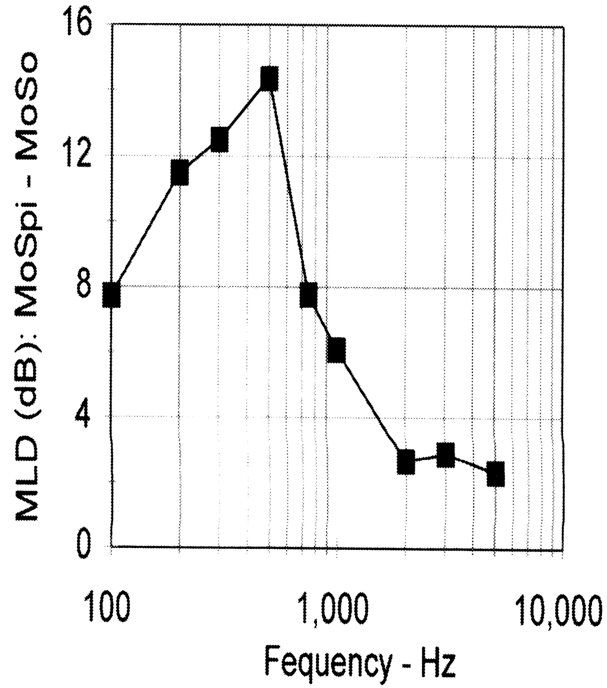


FIGURE 12.11 Difference in masked thresholds (in dB) between M_0S_π and M_0S_0 conditions (MLD) plotted as a function of signal frequency. Adapted with permission from Webster (1951).

Summary

- Temporal coding of frequency of acoustic components (constrained by the degree of place coding)
  - Crucial to encoding of pitch
- Two ears are better than one!