Lecture 3: Perception

1. Ear Physiology
2. Auditory Psychophysics
3. Pitch Perception
4. Music Perception
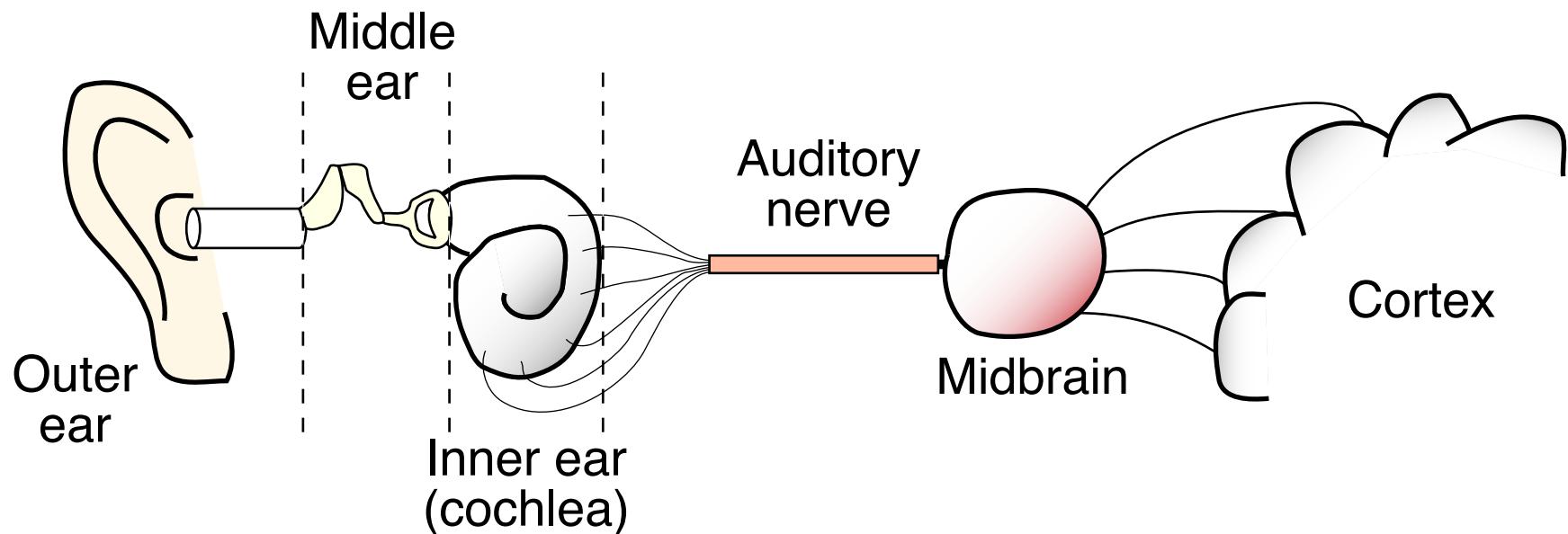
Dan Ellis

Dept. Electrical Engineering, Columbia University

dpwe@ee.columbia.edu   http://www.ee.columbia.edu/~dpwe/e4896/

1. Ear Physiology
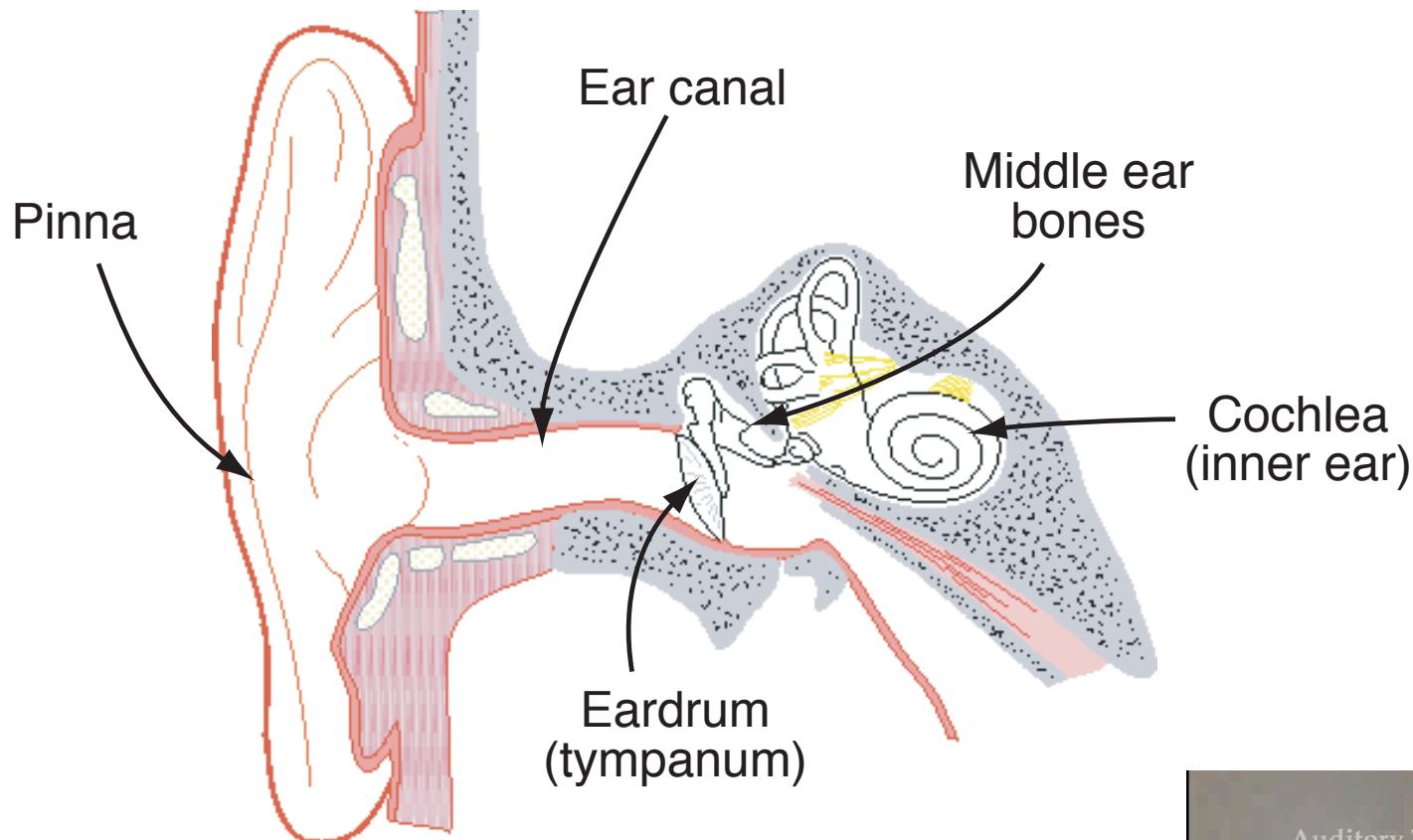
- **The ear is a very sensitive transducer of air pressure variations into nerve firings**
 - just above Brownian motion !?

- **The cochlea is largely understood**
  - the brain is more difficult

The Ear

- Impedance matching & transduction
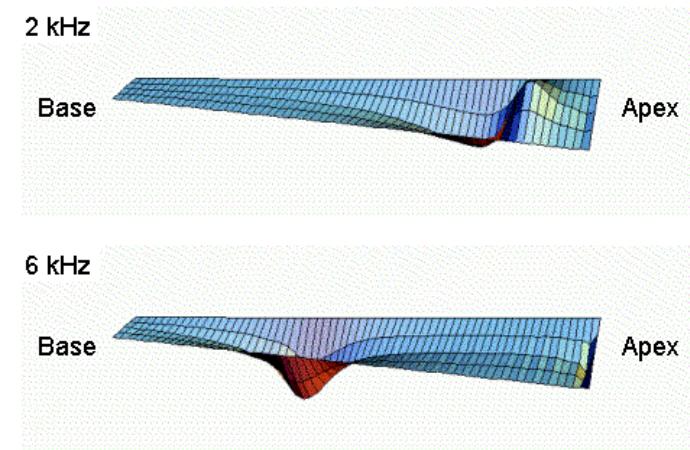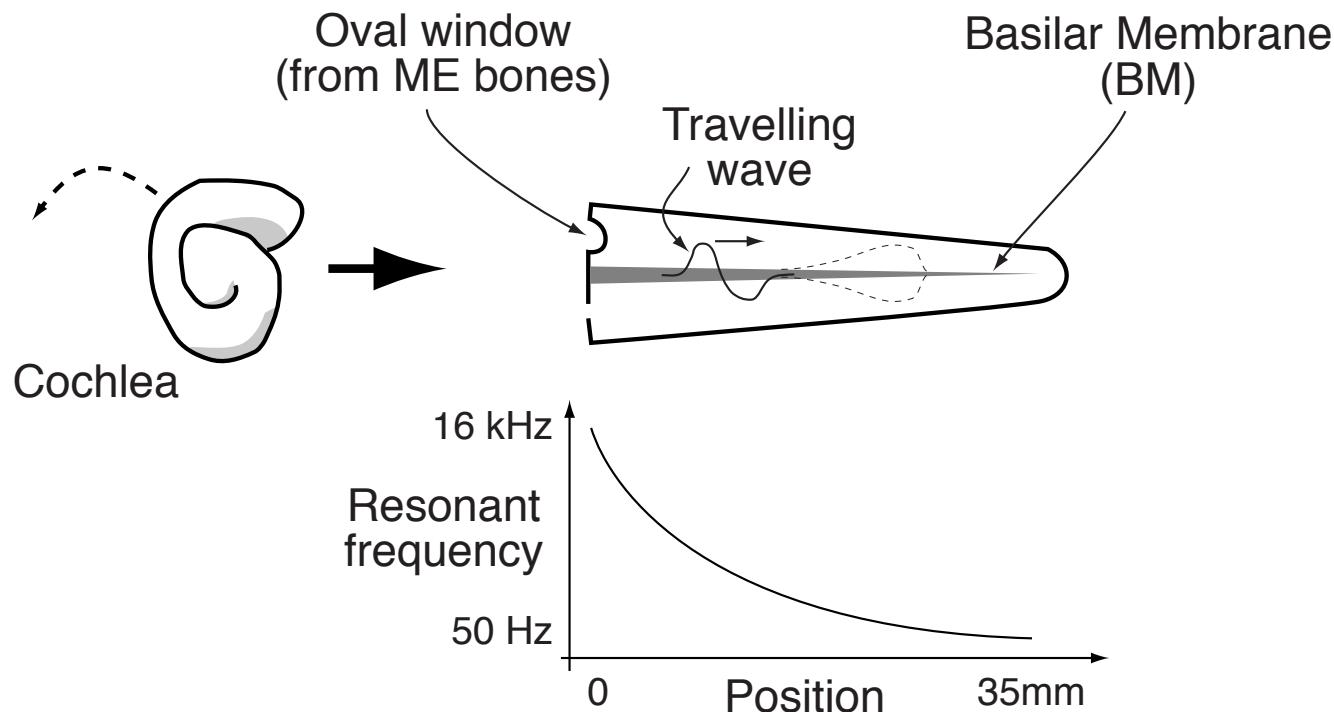
- pinna acts as horn
- eardrum + bones match impedance
- cochlea transduces to nerve firings

The Cochlea

- **Complex resonant structure**

Oval window
(from ME bones)

Travelling wave

Basilar Membrane
(BM)

Cochlea

2 kHz — Base — Apex

6 kHz — Base — Apex

16 kHz

Resonant
frequency

50 Hz

0   Position   35mm
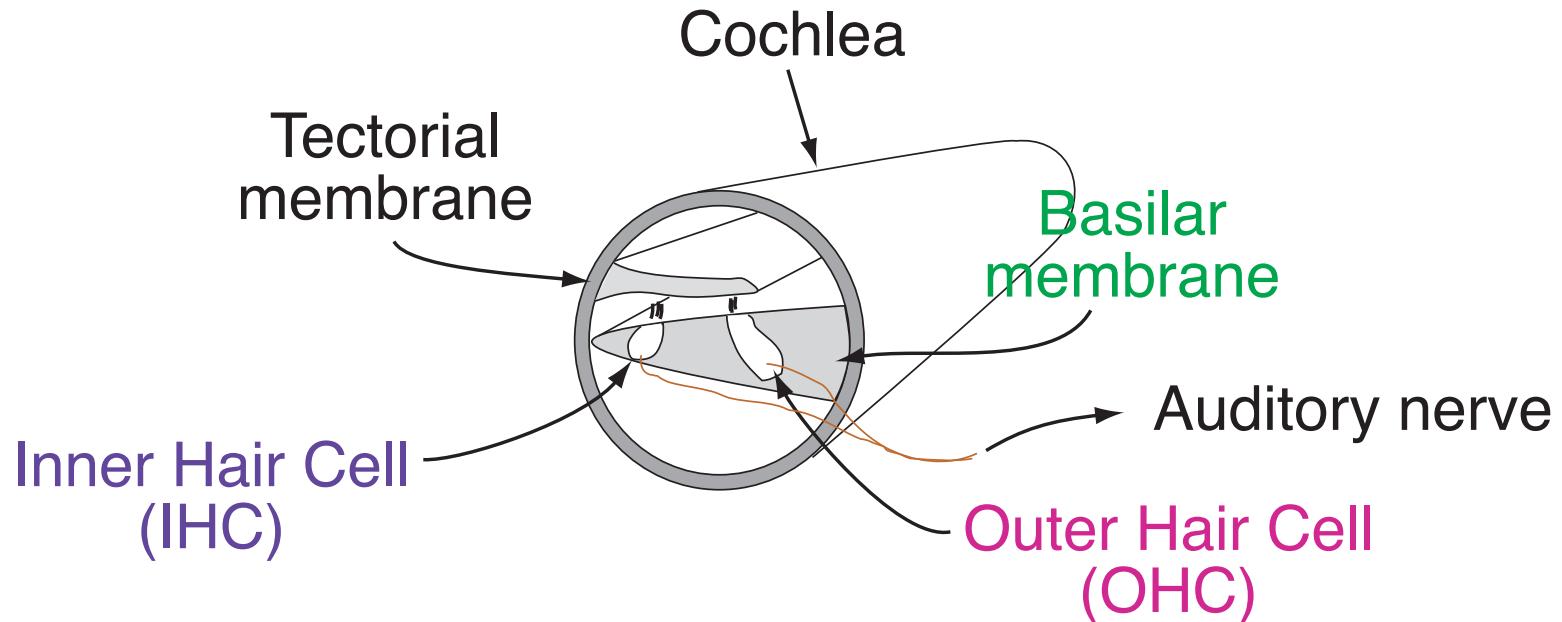
http://www.wadalab.mech.tohoku.ac.jp/FEM_BM-e.html

- active feedback to maintain near-ringing state
- efferent fibers?

Hair Cells
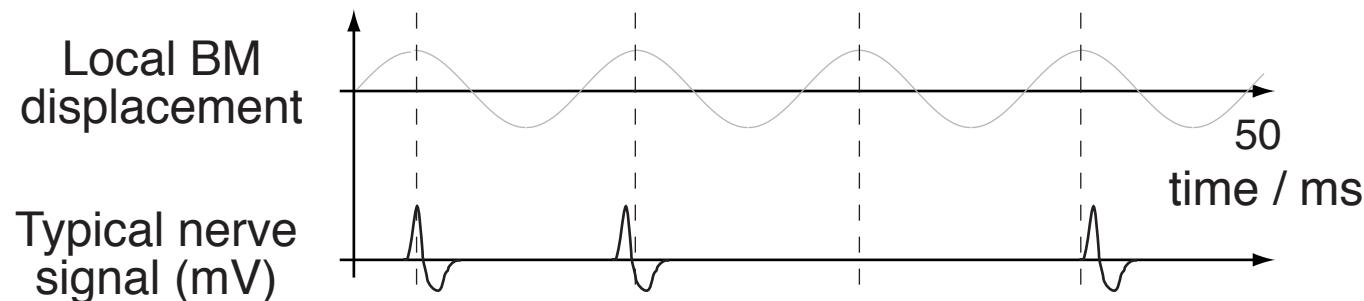
- Transduce mechanical motion to nerve firings

Cochlea

Tectorial membrane

Basilar membrane

Auditory nerve

Inner Hair Cell (IHC)

Outer Hair Cell (OHC)
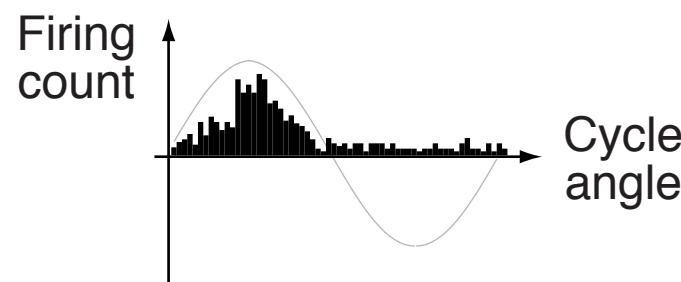
- ○ 3,000 IHCs driving 20,000 nerves
- ○ easily damaged

Auditory Nerve

- IHC fires near maximum displacement
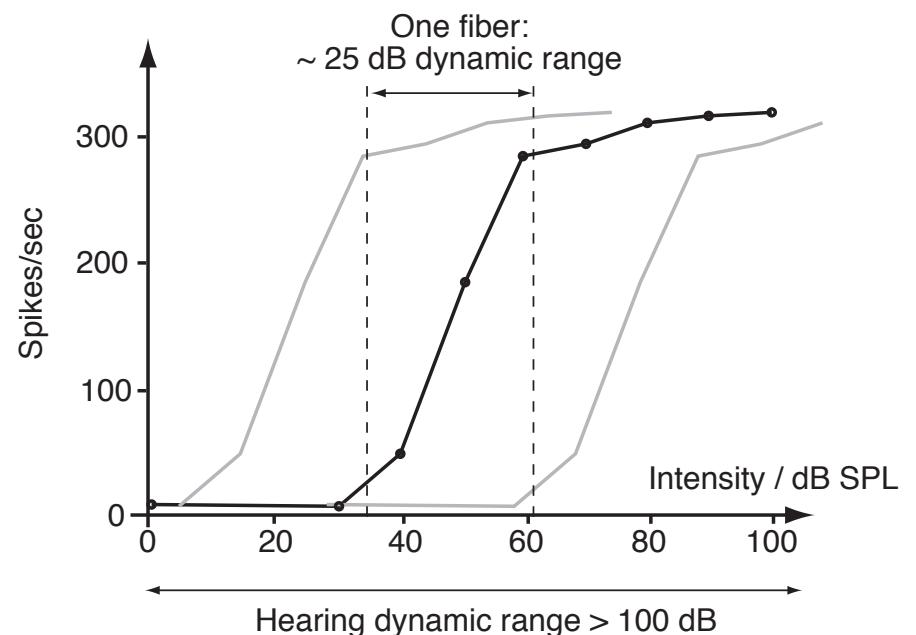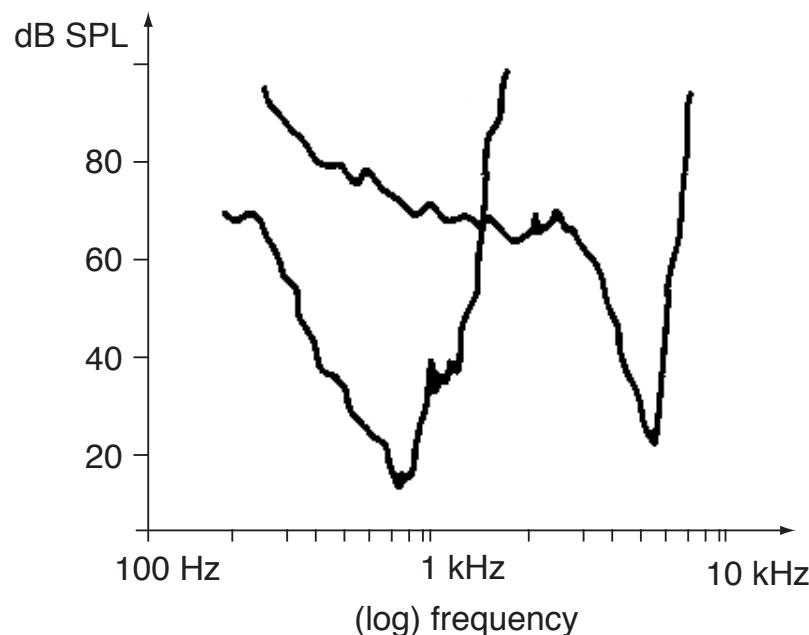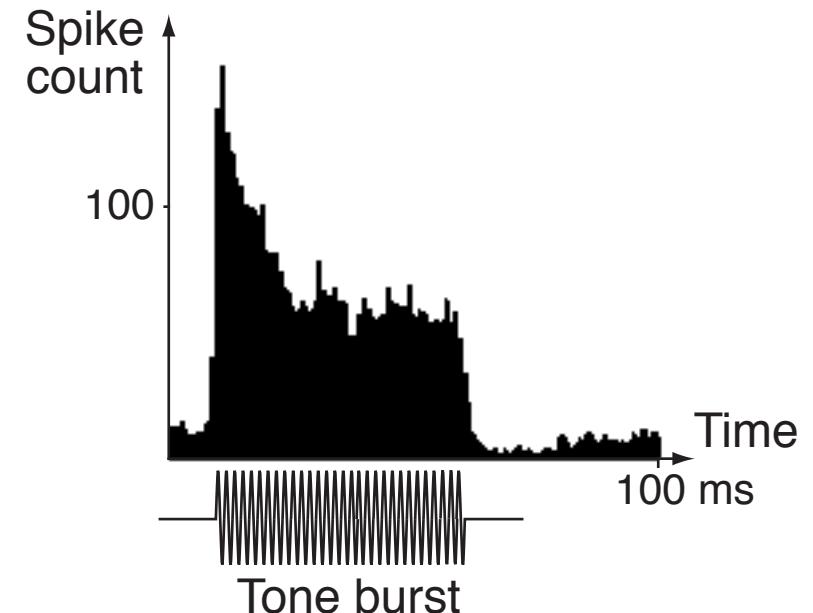
- cannot fire every cycle
- some "noise"
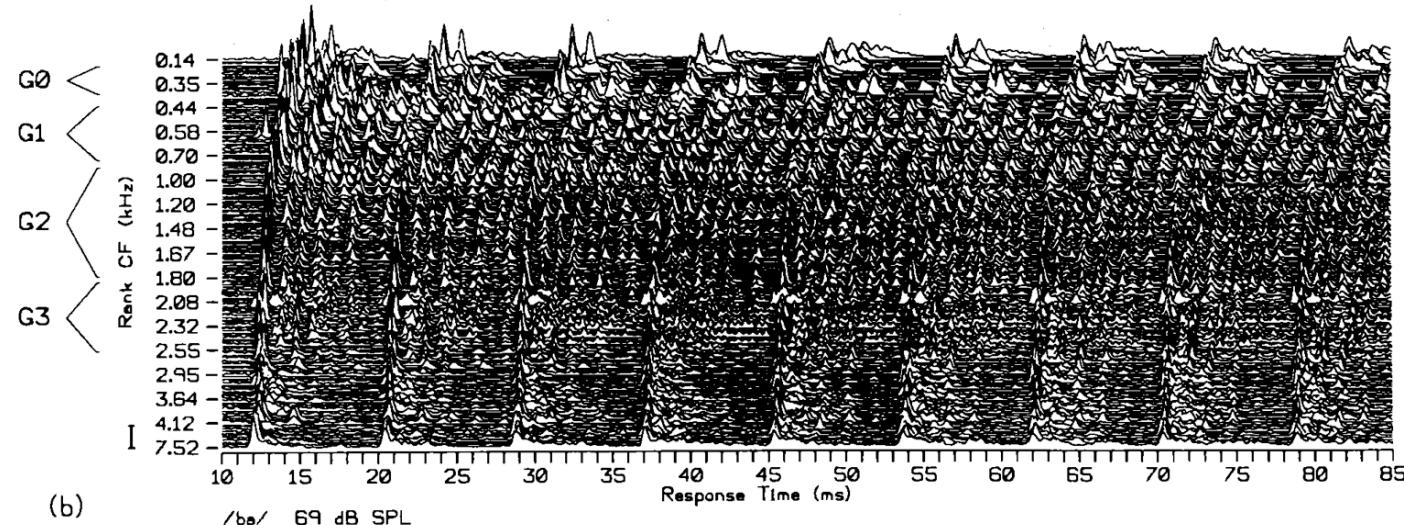
# Nerve Responses

- Onset enhancement

- Frequency selectivity

- Dynamic range

Spike count vs Time. Tone burst. 100 ms.

dB SPL vs (log) frequency. 100 Hz, 1 kHz, 10 kHz.

One fiber: ~ 25 dB dynamic range. Spikes/sec vs Intensity / dB SPL. Hearing dynamic range > 100 dB.
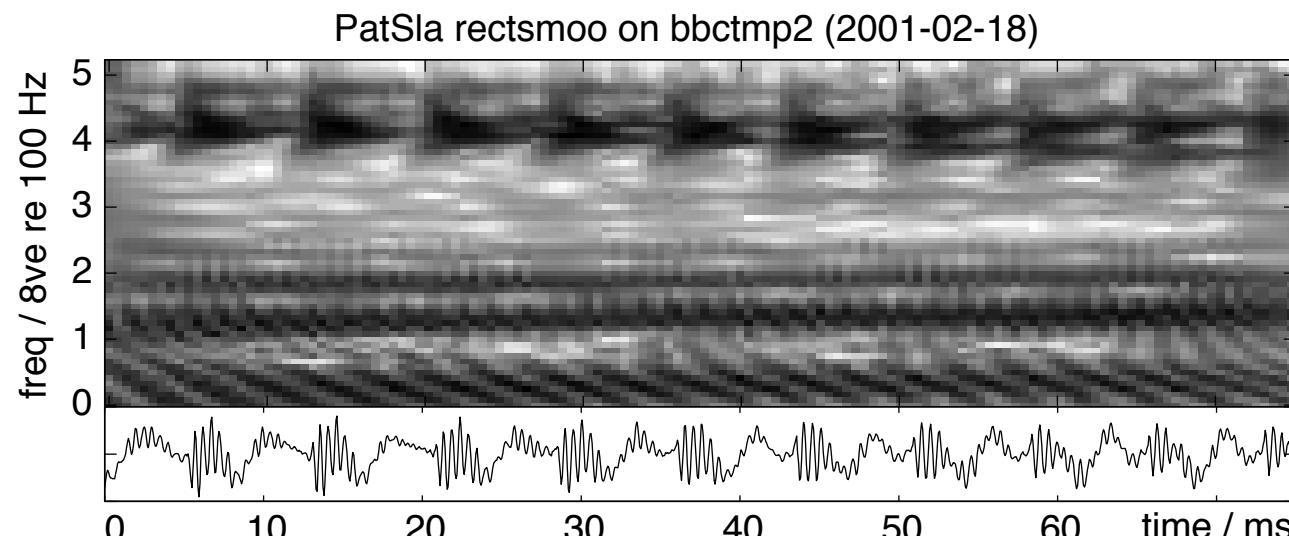
Auditory Nerve Ensemble

- ## Ensemble of nerves provide full information
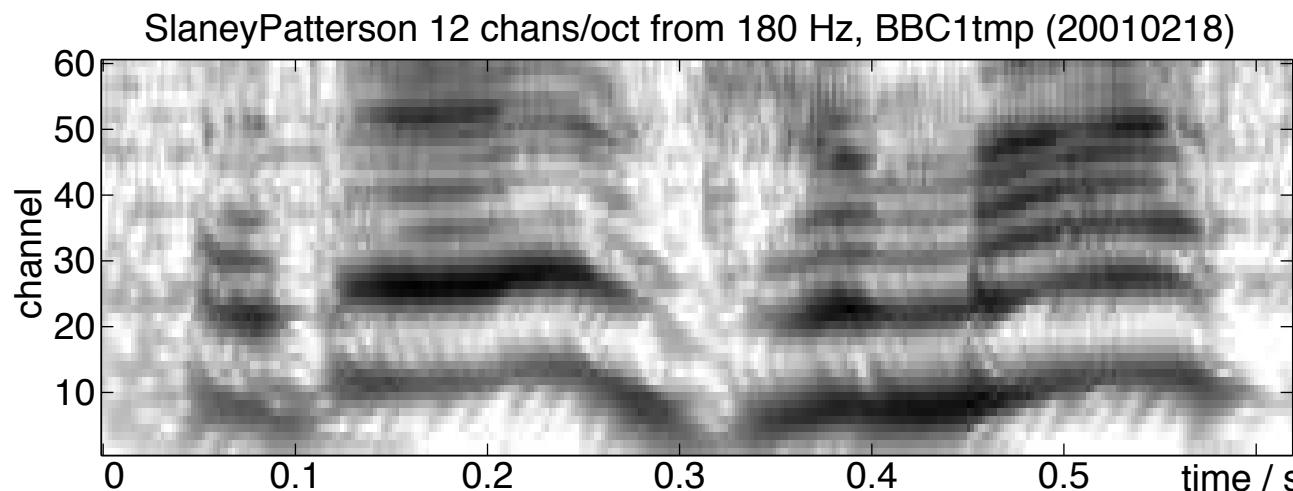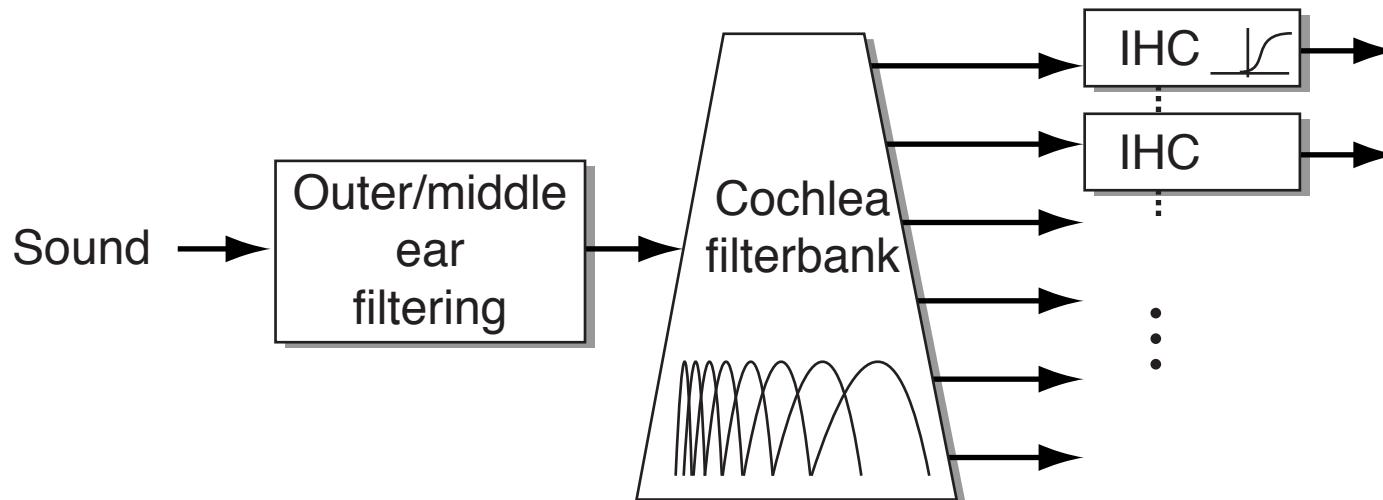


Secker-Walker & Searle '90

- ○ similar to constant-Q log-intensity spectrogram

PatSla rectsmoo on bbctmp2 (2001-02-18)

Auditory Models

- ## Filterbank + nonlinearity
  - ○ varying (but broad) bandwidth

SlaneyPatterson 12 chans/oct from 180 Hz, BBC1tmp (20010218)

2. Auditory Psychophysics

- Extensive study of relationship between physical (Φ) and psychological (Ψ) values
  - perception is not "direct"!

- Common across all perceptual modalities
  - proprioceptive force, body positioning
  - vision
  - hearing

- Φ - Ψ distinction is important!

Loudness perception

- Perception of physical parameter
  - just noticeable difference - jnd
  - magnitude scaling

- Weber's law:

$$\Delta I \propto I \quad \Rightarrow \quad \log(L) \propto \log(I)$$
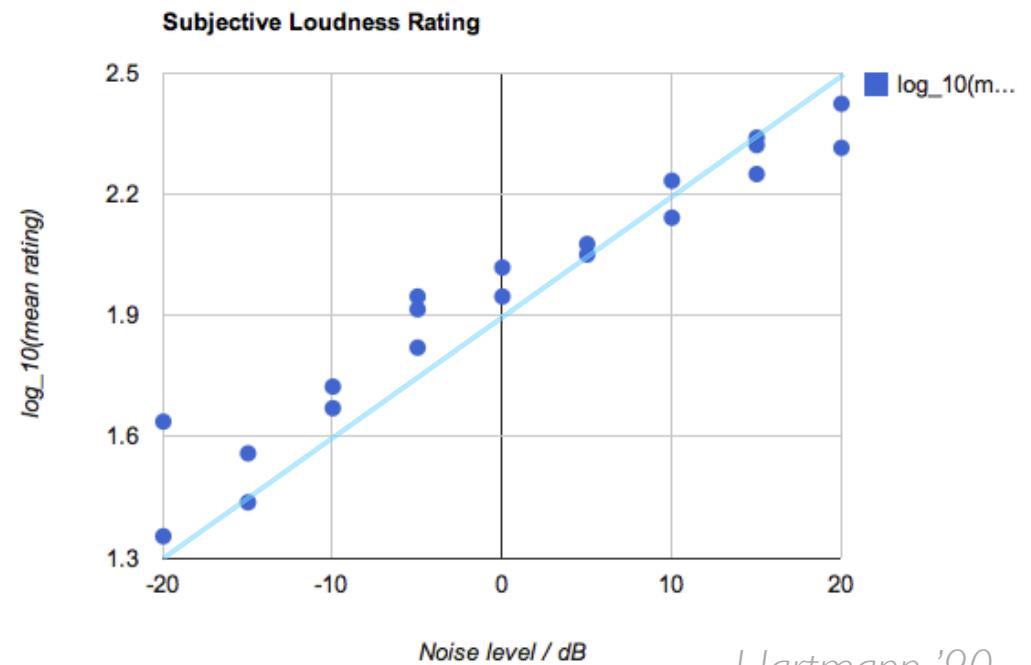
- Loudness

$$L \propto \quad I^{0.3}$$

$$\Rightarrow \log(L) = \quad 0.3\log(I)$$

$$\Rightarrow \log_{10}(L) = \quad 0.03\,\mathrm{dB}(I)$$

$$\Rightarrow \mathrm{dB}(I) = \quad 33.3\log_{10}(L)$$

Subjective Loudness Rating

■ log_10(m…

y-axis: log_10(mean rating), values 1.3, 1.6, 1.9, 2.2, 2.5

x-axis: Noise level / dB, values -20, -10, 0, 10, 20

Hartmann '90 el
tracks 19+20
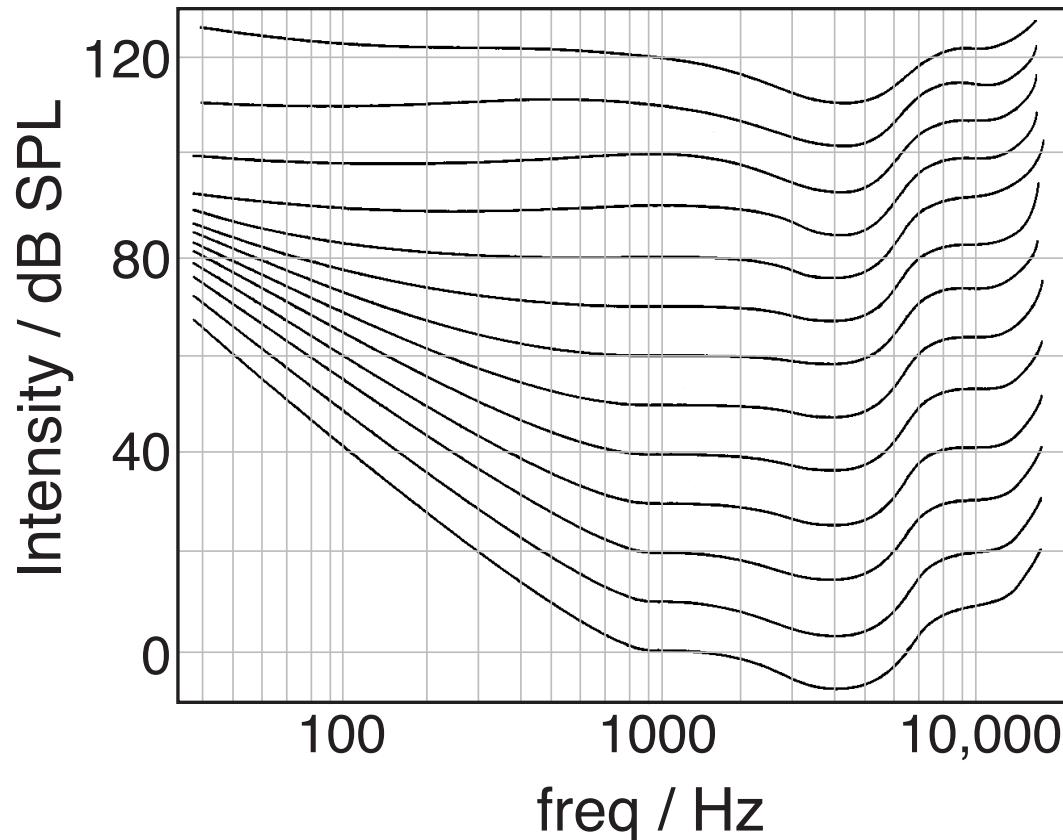
Equal Loudness
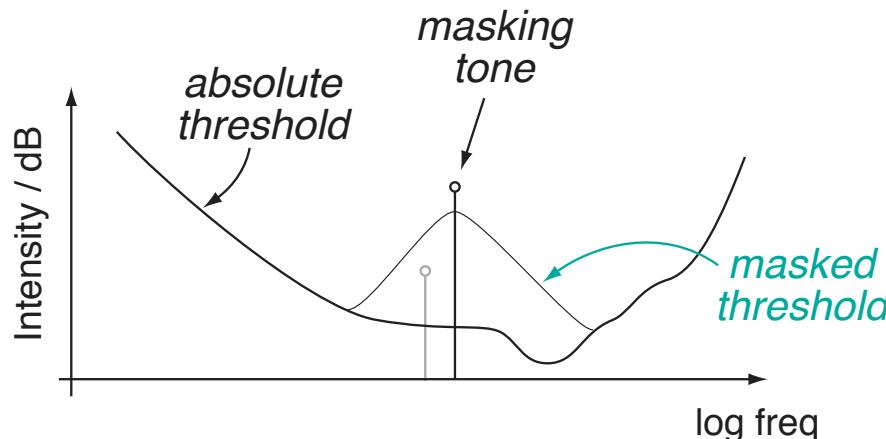
- Fletcher-Munson curves (1937)
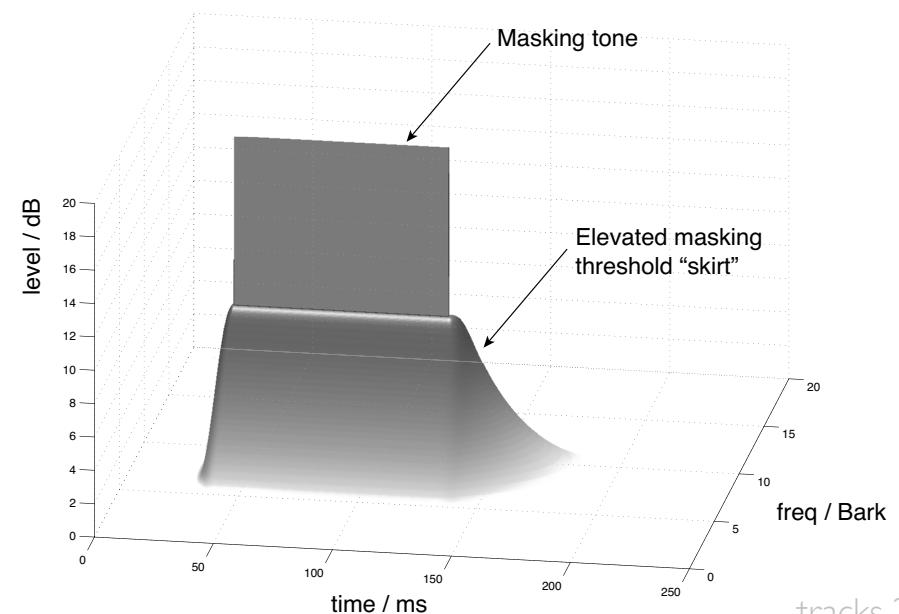
- ○ match intensity to specific 1 kHz tone

- ○ loudness growth

Masking

- Limited dynamic range in cochlea
  - effect within frequency "critical bands"

*absolute threshold* — *masking tone* — *masked threshold*

Intensity / dB — log freq

  - basis of MPEG Audio
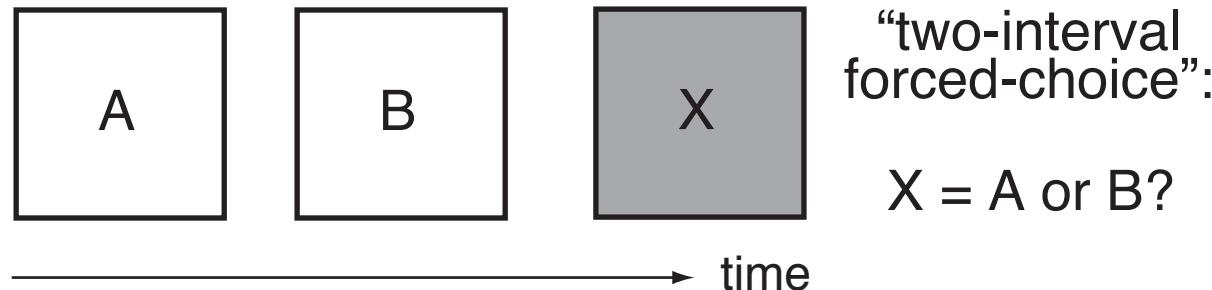
- Forward/backward temporal effects

Masking tone

Elevated masking threshold "skirt"

level / dB — time / ms — freq / Bark

Limits of Hearing

- Test what listeners can discriminate

"two-interval forced-choice":

X = A or B?

time
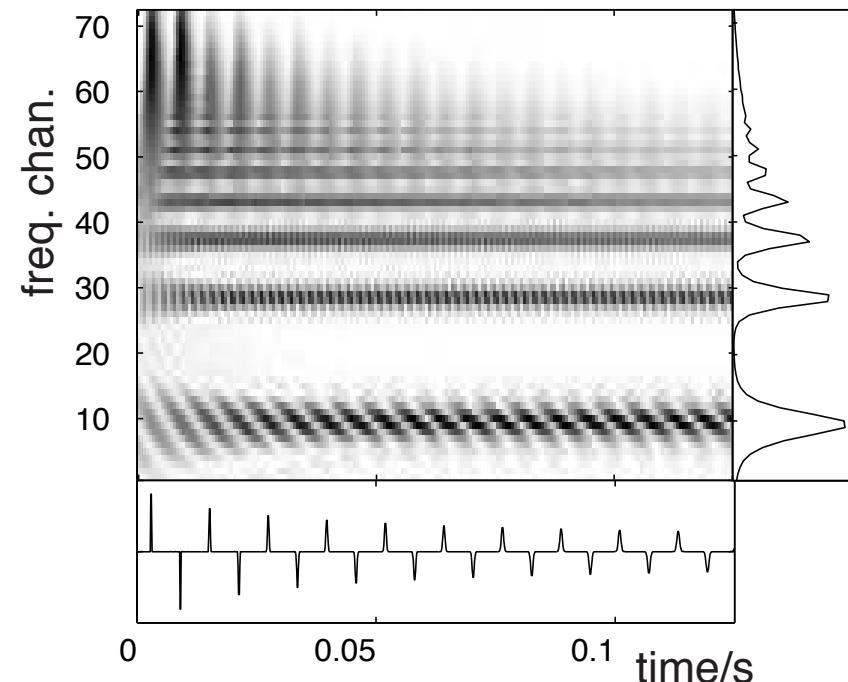
- Roughly...
  - timing: 2 ms difference, 20 ms ordering
  - tuning: ~1%
  - spectral profile: single components ~ 2 dB

  - phase?
  - tones vs. noise...

3. Pitch Perception
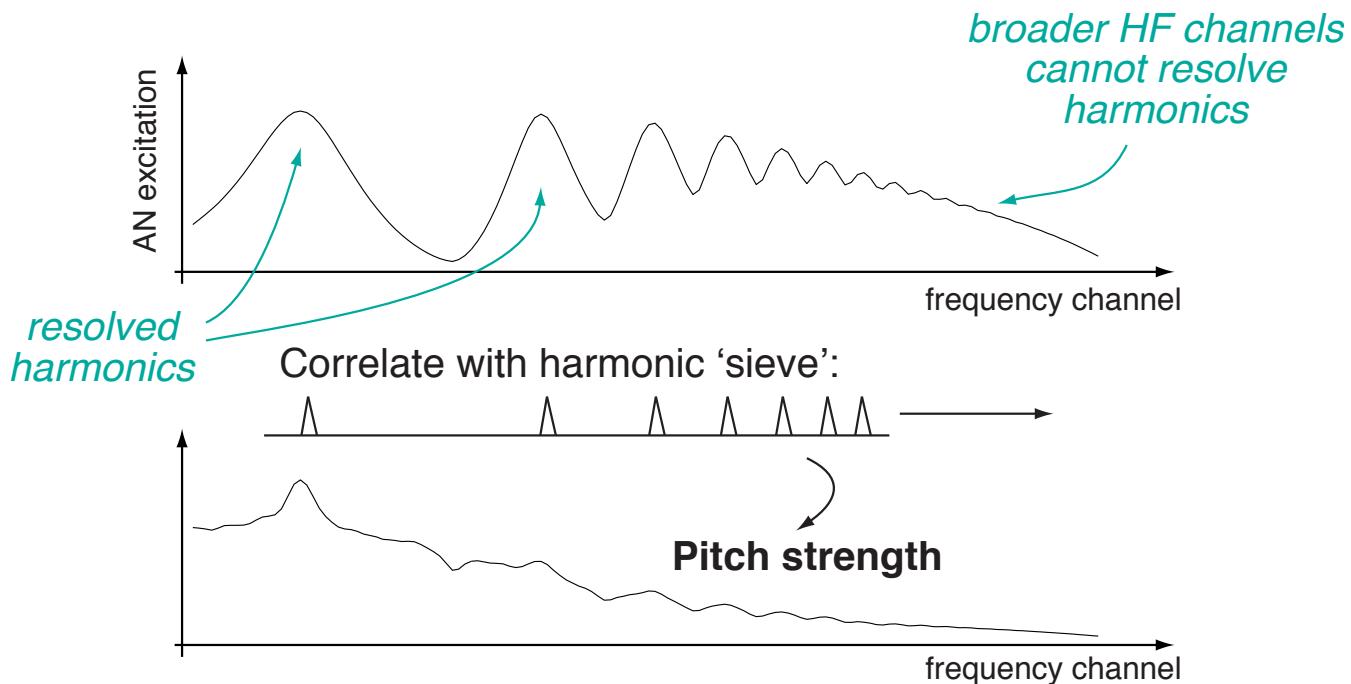
- **Complex (non-sinusoidal) tones give a single, fused percept**
  - despite harmonics resolved by cochlea

  - percept is of a single pitch
  - .. but pitch does NOT rely on the fundamental

"Place" models of pitch
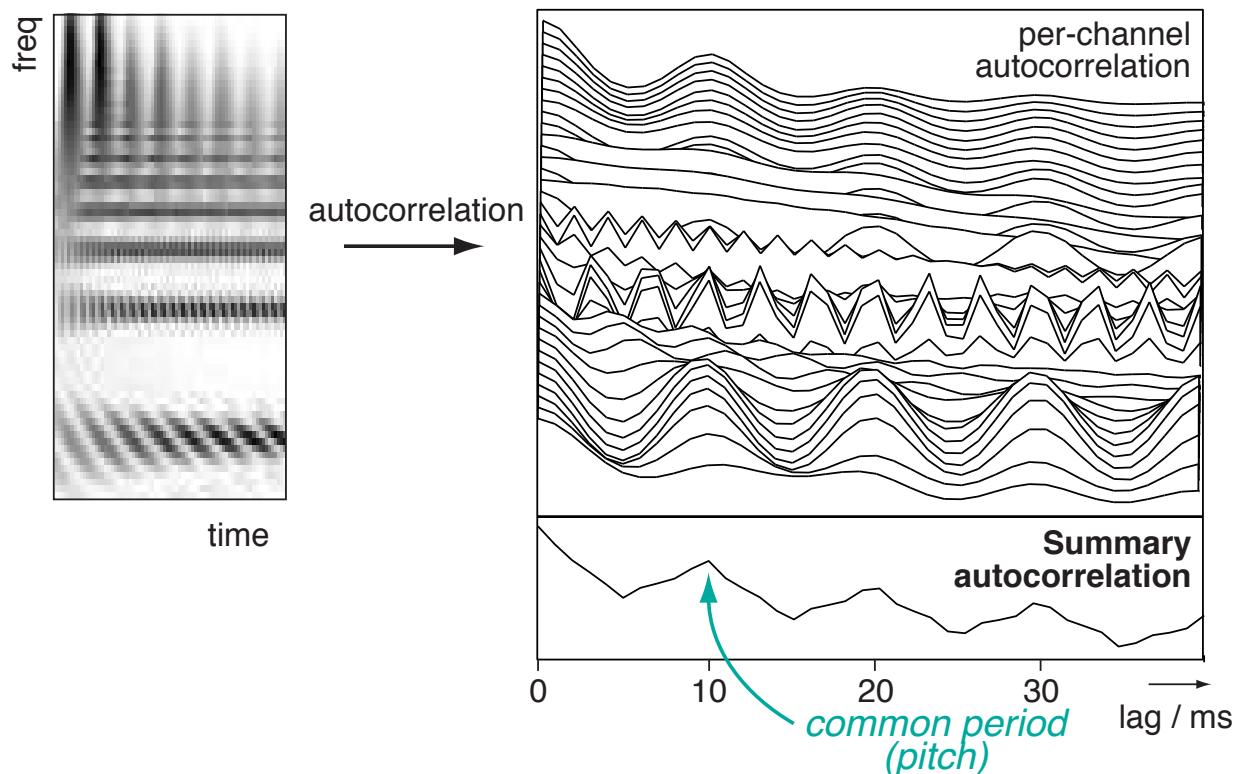
- Hypothesis:
Pitch results from activation pattern

AN excitation

broader HF channels cannot resolve harmonics

resolved harmonics

frequency channel

Correlate with harmonic 'sieve':

Pitch strength

frequency channel

Duifhuis et al. '82

- support: low harmonics are important
- but: pitch of noisy signals

"Time" models of pitch

- Autocorrelation
neatly unifies pitch phenomena

- but: high-frequency modulation evokes weak pitch

Competing Cues
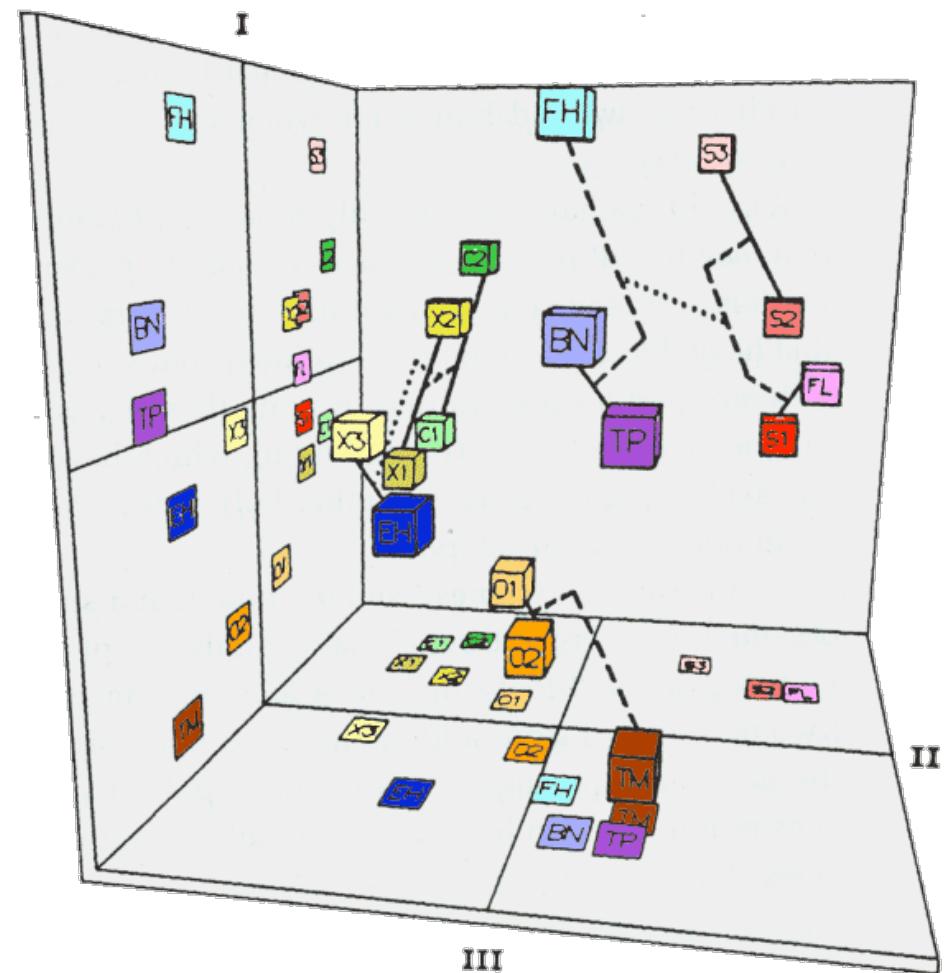
- Perhaps brains use both place & time cues
  - common perceptual strategy:
   opportunistic combination of information

- e.g. Probabilistic combination

$$\arg\max_{\theta} Pr(\theta|x) \propto \arg\max_{\theta} \frac{Pr(\theta|x_1)Pr(\theta|x_2)}{Pr(\theta)}$$

- if x_1, x_2 are independent given θ
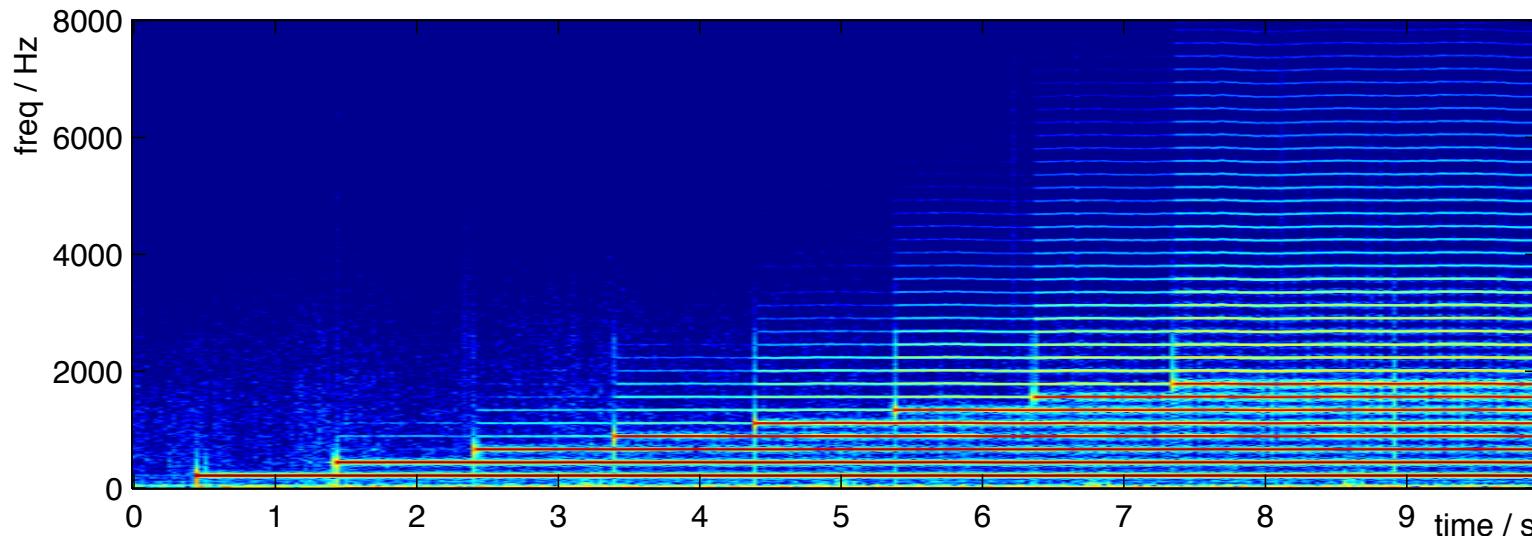
4. Music Perception

- Hearing music involves
  - instruments
  - notes
  - rhythm

- Can study with subjective experiments

Grey '75

Scene Analysis

- **Detect separate events**
  - common onset
  - common harmonicity
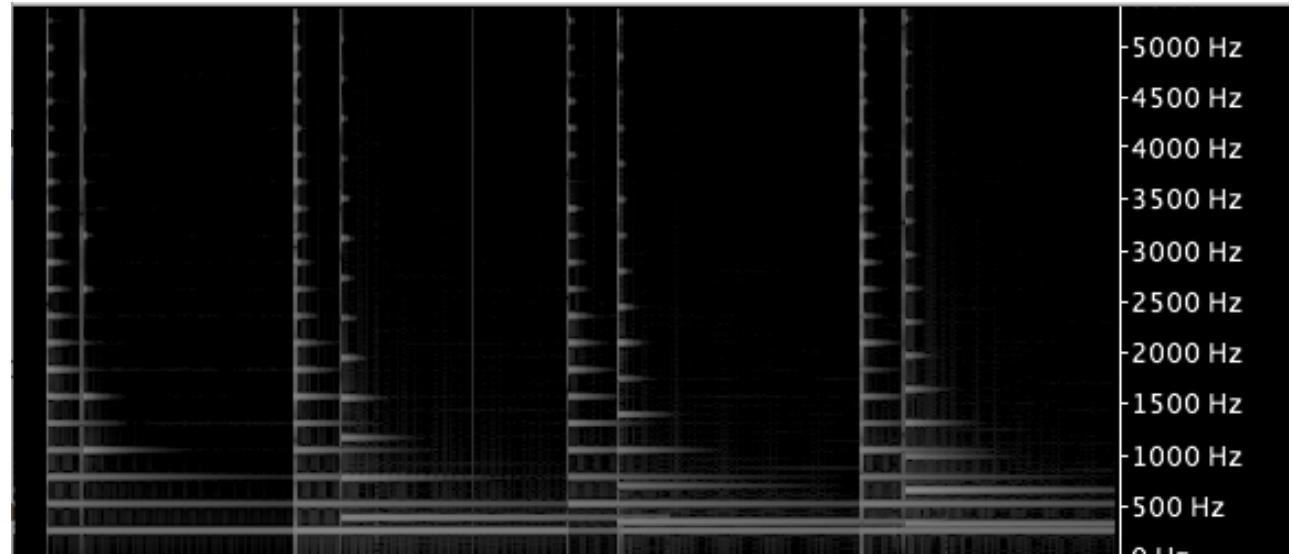
Pierce '83

  - instruments & timbre
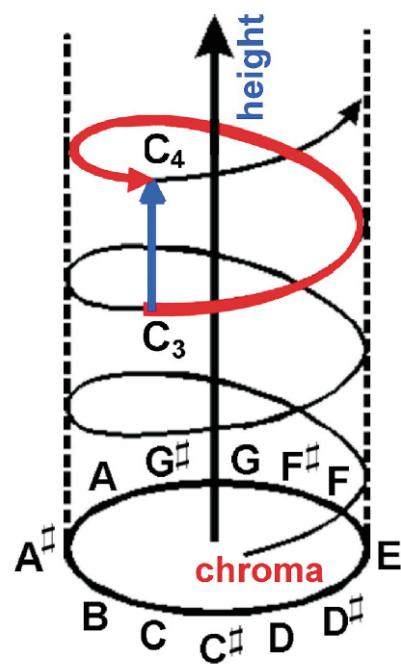
Consonance

- Musical intervals relate to harmonic proximity

- Pitch Helix
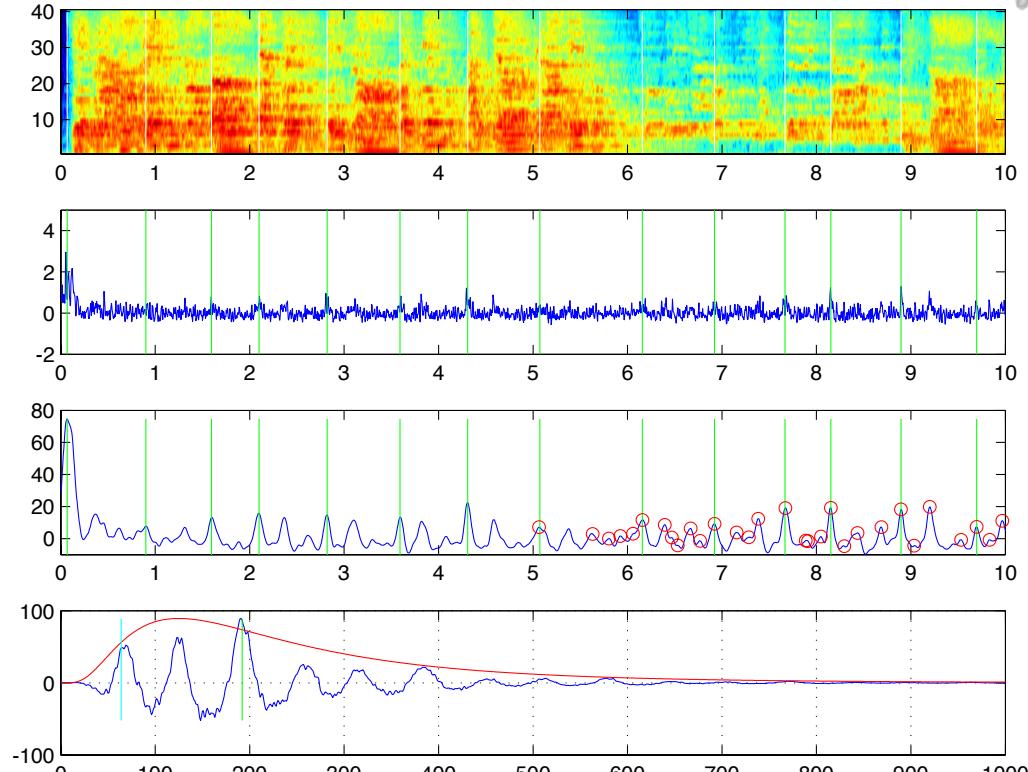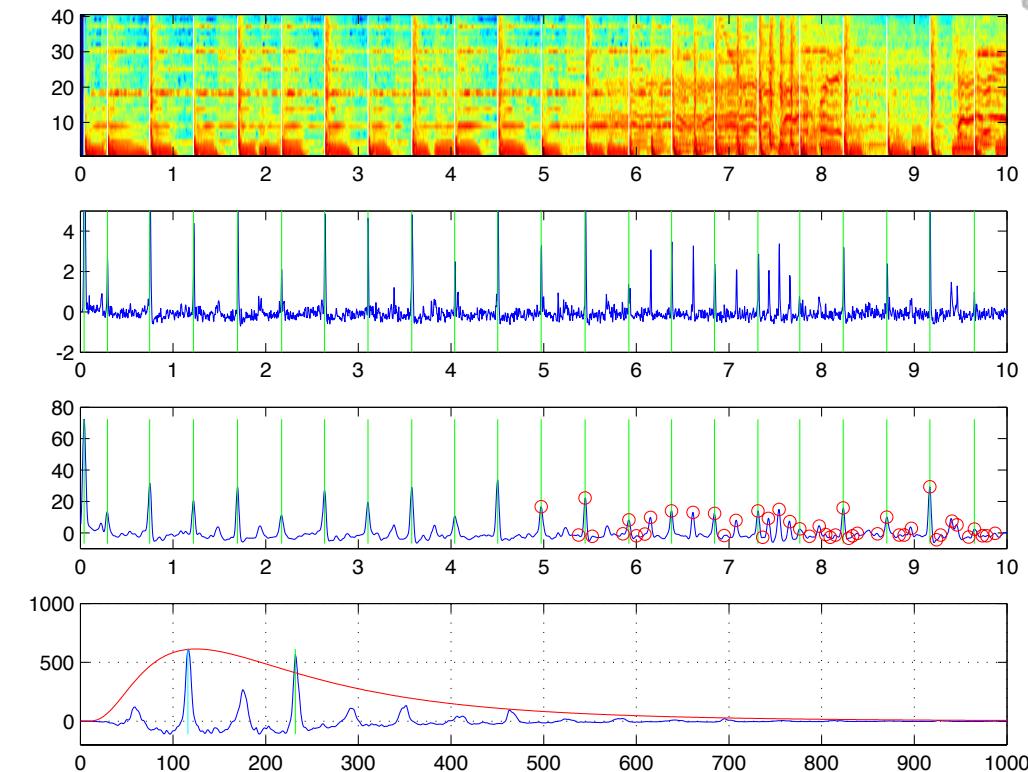
Warren et al. 2003
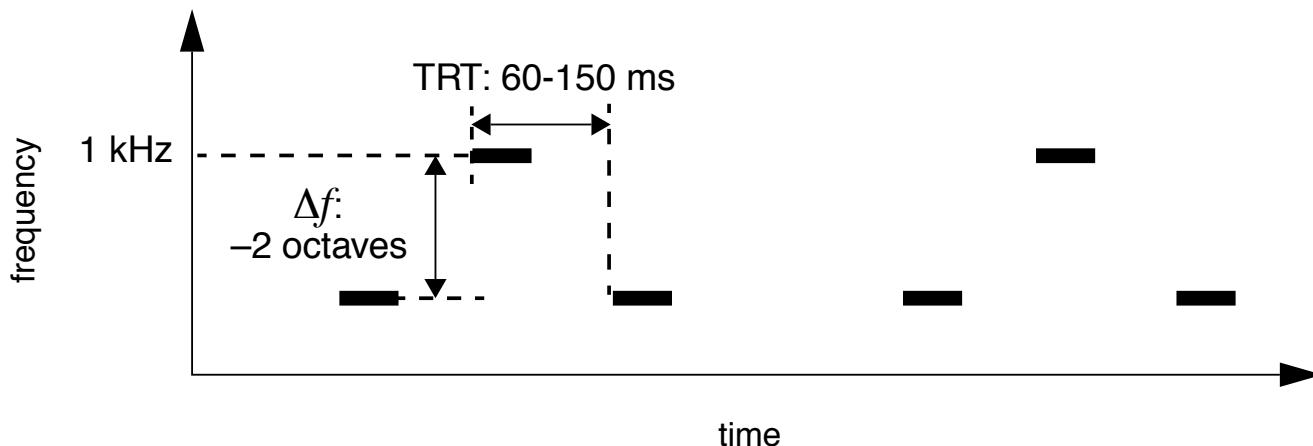
Rhythm

- **Sensitive to periodicity**
  - speech? breathing? brain?

- **Onsets + autocorrelation?**
  - variations in tapping
  - 4/4 vs 3/4

Sequences

- **Perceptual effects of sequences**
  - e.g. streaming

- **Music is built of sequences**
  - different sensitivities

track 36

Summary

- Ear converts air pressure to nerve firings
  - spectral analysis

- Brain does a lot with scarce information
  - dealing with the real world

- Music is a complex signal
  - multiple sources
  - harmonic structures
  - temporal patterns

References

- H. Duifhuis, L. F. Willems, & R. J. Sluyter, "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," *J. Acoust. Soc. Am.,* 71(6):1568-1580, 1982.

- H. Fletcher & W. A. Munson, "Relation between loudness and masking," *J. Acoust. Soc. Am.,* 9:1-10, 1937.

- B. Gold, N. Morgan, & D. Ellis, *Speech and Audio Signal Processing (2nd ed.)*, Wiley, 2011.

- J. M. Grey, *An Exploration of Musical Timbre,* Ph.D. Dissertation, Stanford CCRMA, No. STAN-M-2, 1975.

- W. M. Hartmann, "Auditory demonstrations on compact disk for large N," *J. Acoust. Soc. Am.,* 93(1):1-16, 1993.

- R. Meddis & M. J. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," *J. Acoust. Soc. Am.,* 89(6):2866-2882, 1991.

- B. C. J. Moore, *An Introduction to the Psychology of Hearing (4th ed.)*, Academic Press, 1997.

- J. R. Pierce, *The Science of Musical Sound*, Scientific American Press, 1983.

- H. E. Secker-Walker & C. L. Searle, "Time-domain analysis of auditory-nerve-fiber firing rates," *J. Acoust. Soc. Am.,* 88(3):1427- 1436, 1990.

- J. D. Warren, S. Uppenkamp, R. D. Patterson, & T. D. Griffiths, "Separating pitch chroma and pitch height in the human brain," *PNAS* 100(17):10038-42, 2003.