Harvard-MIT Division of Health Sciences and Technology
HST.723: Neural Coding and Perception of Sound
Instructor: Bertrand Delgutte

# Neural coding of pitch: Temporal and place representations

# Why study pitch?

- Important role in music (melody and harmony)
- Some role in speech (speaker identity, emotions, tonal languages)
- Present in animal vocalizations
- Differences in pitch are a major cue for segregating sound sources
- Thoroughly studied psychophysically

# Outline

- Rate and temporal neural codes
- Pitch based on resolved and unresolved harmonics
- Peripheral representations of pitch
  - Temporal representation in interspike intervals (autocorrelation model)
  - Harmonic templates and rate-place representation
  - Spatio-temporal representation
- Central processing of pitch
  - Are there pitch neurons?
  - Degradation in temporal representations of pitch
  - Neural tuning to temporal envelope

# Rate vs. Temporal Codes

- Temporal code conveys more information than rate code
- Information measured in *bits*, the number of binary choices that must be made to completely specify the code
- For 5 time bins:
  - Rate code: $\log_2(6)$=2.6 bits
  - Temporal code: 5 bits

Figure from: Rieke, F., D. Warland, de Ruyter van Steveninck R., and W. Bialek. "Spikes." In *Exploring the Neural Code*. Cambridge, MA: MIT Press, 1997.

# Classification of Neural Codes

| Neural Codes | Spike Counts | Spike Timing |
|---|---|---|
| **Single Neuron** | Rate code | Temporal code |
| **Neural Population (Map?)** | Rate-place or population-rate code | Temporal-place code (all information) |

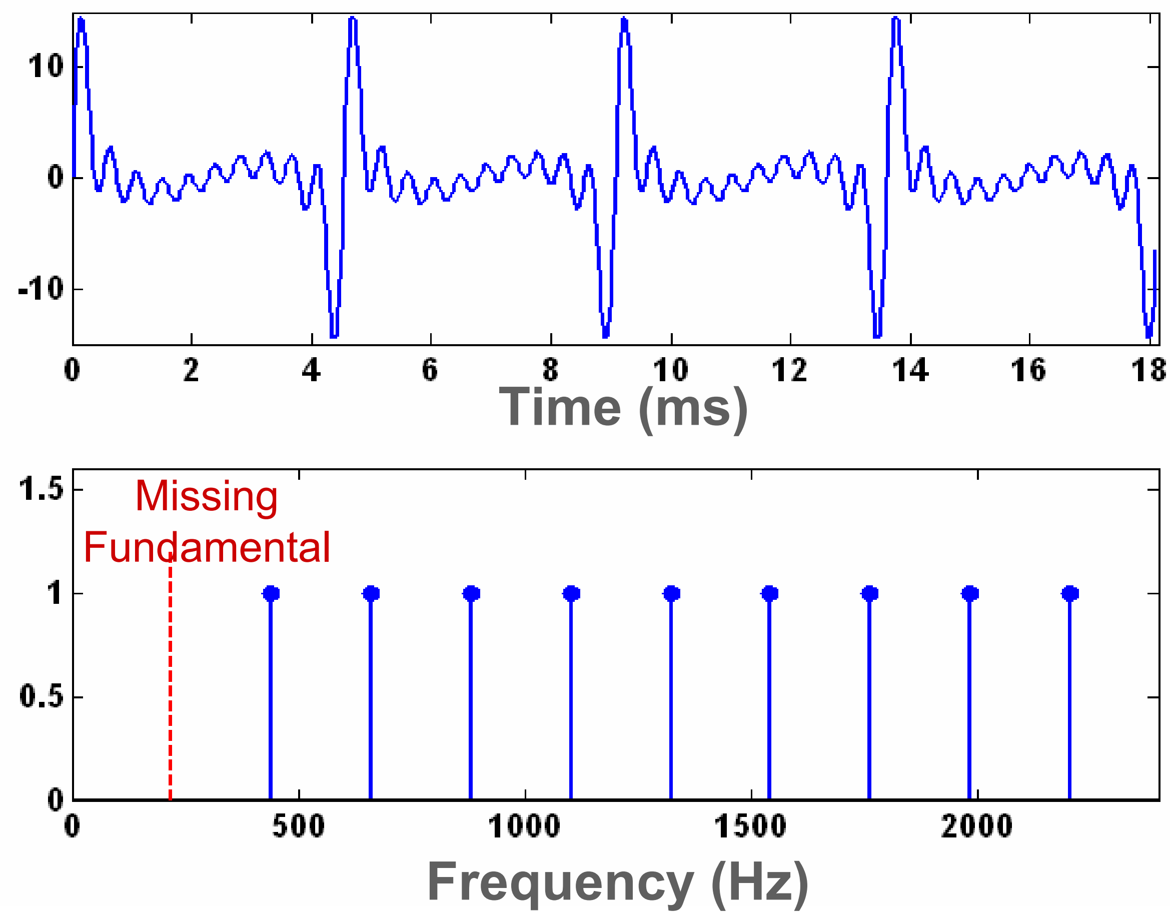
# Pitch of harmonic complex tones

Harmonic Complex Tone
F0=220 Hz

Pure Tone
220 Hz

Periodic Time Waveform

Harmonic Power Spectrum



# Periodicity and Harmonicity: Physiological Basis

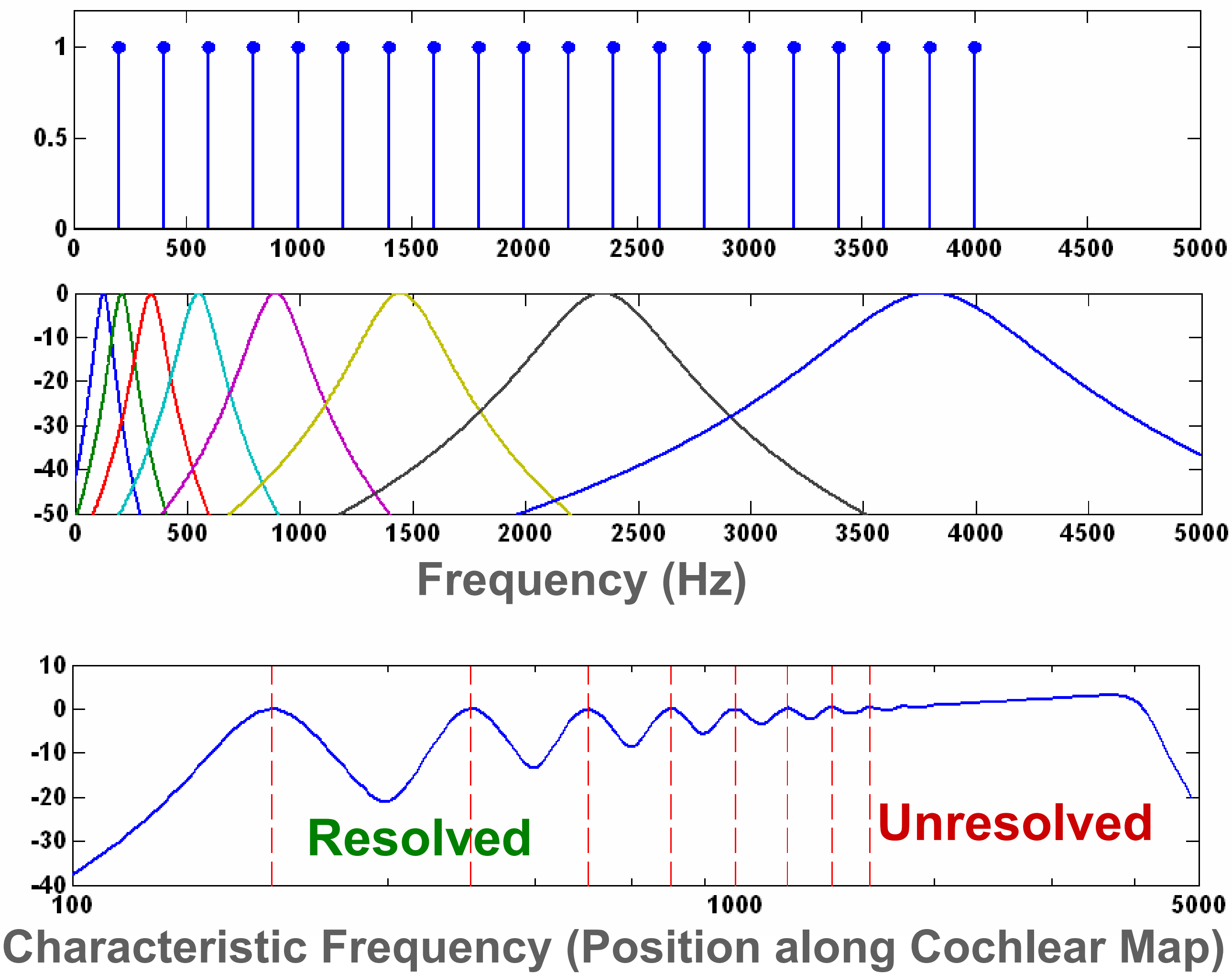
- Periodicity ⟷ Temporal representation
  - Based on neural phase locking to stimulus waveform
  - Limited by jitter in synaptic transmission (~5 kHz upper limit in cats)
- Harmonicity ⟷ Place representation
  - Based on mechanical frequency analysis and frequency-to-place mapping in the cochlea
  - Limited by cochlear frequency resolution (first 6-10 harmonics in humans)

# Harmonicity information is limited by frequency resolution of the cochlea

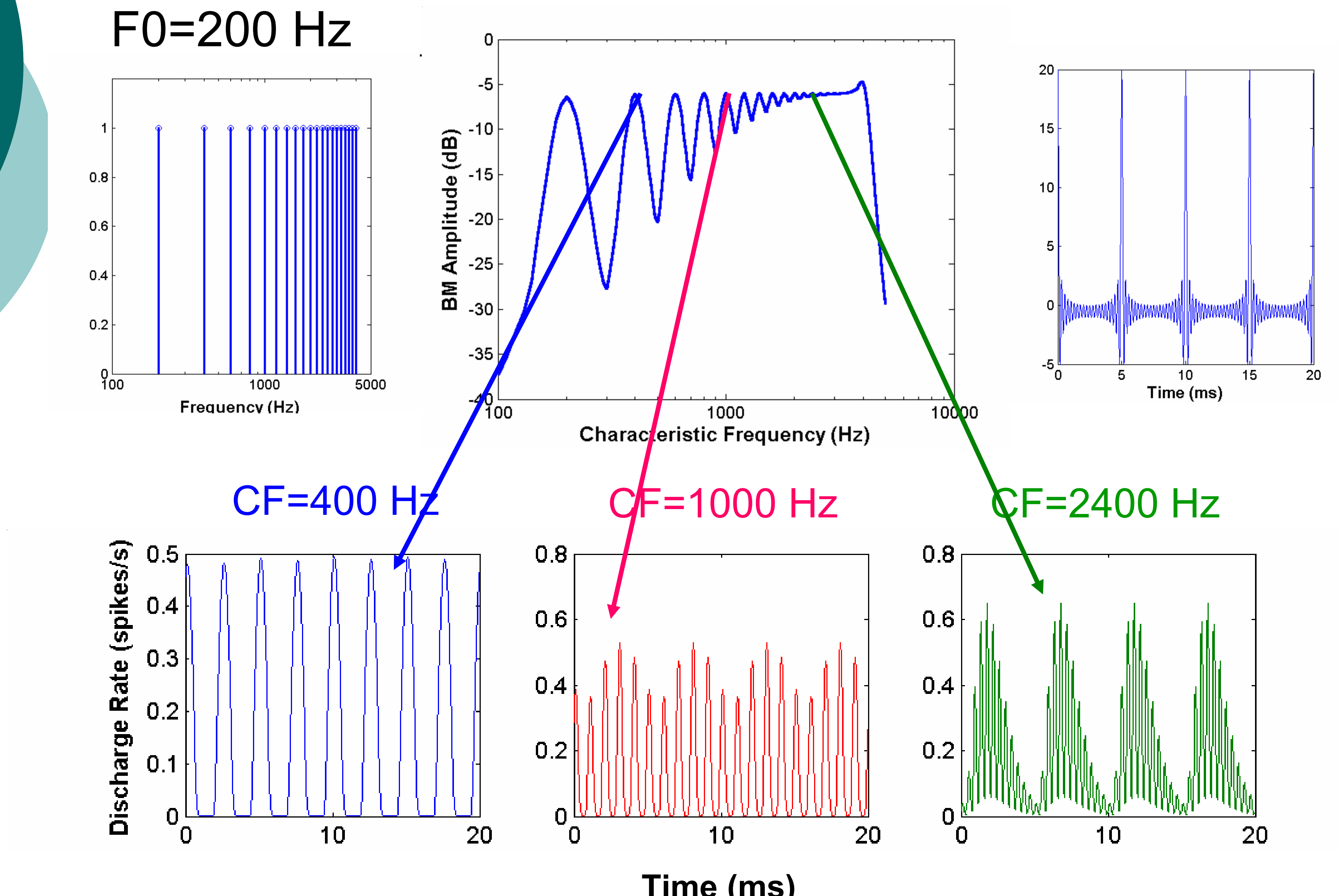
Stimulus Spectrum F0=200 Hz

Cochlear Filter Bank

Neural Activation Pattern



# Resolved and unresolved harmonics



# Resolved harmonics evoke stronger pitch than unresolved harmonics

Harmonic
complex tone,
12 components
in sine phase

Figure removed due to copyright considerations.

Please see: Bernstein, J. G., and A. J. Oxenham. "Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number?" *J Acoust Soc Am* 113, no. 6 (Jun 2003): 3323-34.

# Pitch based on resolved and unresolved harmonics

| | Available Neural Cues | Perceptual Salience | Dependent on component phases |
|---|---|---|---|
| Resolved Harmonics | Spatial and Temporal | Strong | No |
| Unresolved Harmonics | Temporal Only | Weak | Yes |

# A paradox

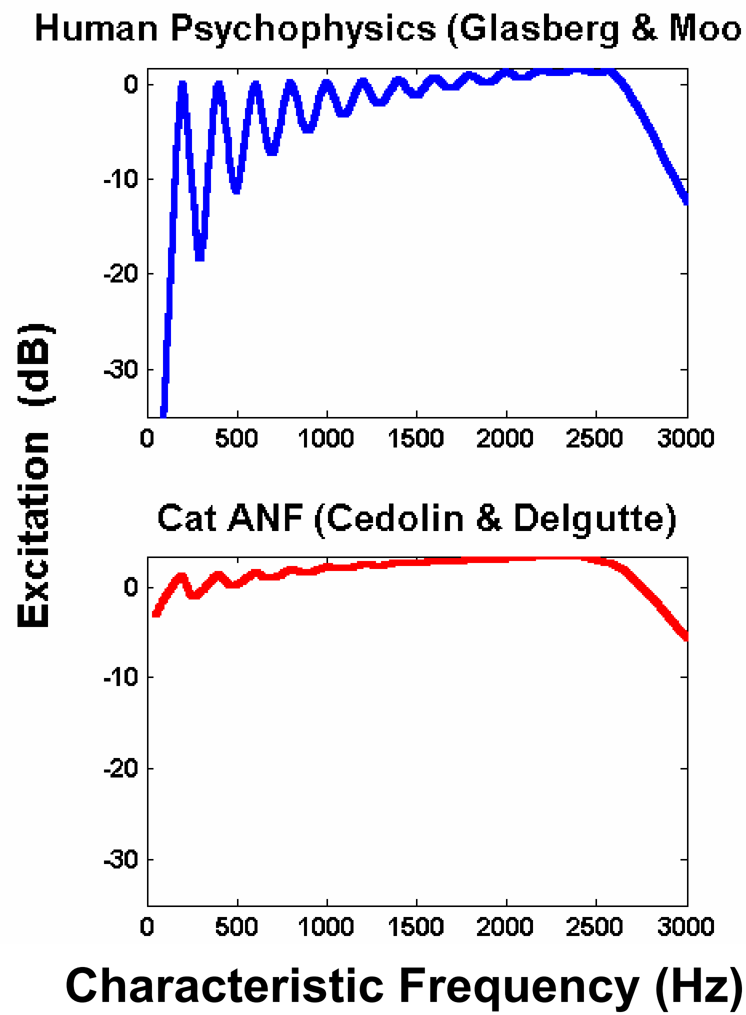
- For normal-hearing listeners outside of the laboratory, pitch perception is based entirely on resolved harmonics
- Studies of the neural coding of pitch have focused (almost) entirely on stimuli with unresolved harmonics

# No resolved harmonics in AN rate-place profiles for F0 in range of human voice

- A steady state vowel is a harmonic complex tone whose fundamental frequency (here 128 Hz) determines voice pitch. Even at low sound levels, the rate-place profile shows no obvious peaks at harmonics of the fundamental frequency.
- In general, data from the cat auditory nerve show few, if any, rate-place cues to pitch for harmonic complex tones with fundamental frequencies in the range of human voice (100-300 Hz).

# Auditory filters are more sharply tuned in humans than in experimental animals

F0=200 Hz



# Cat vocalizations have higher fundamental frequencies than human voice

Figures removed due to copyright reasons.

# Neurophysiological experiment

## Purpose

- Determine F0 range where individual harmonics of complex tones are resolved in rate responses of AN fibers
- Determine upper F0 limit of temporal representation of pitch

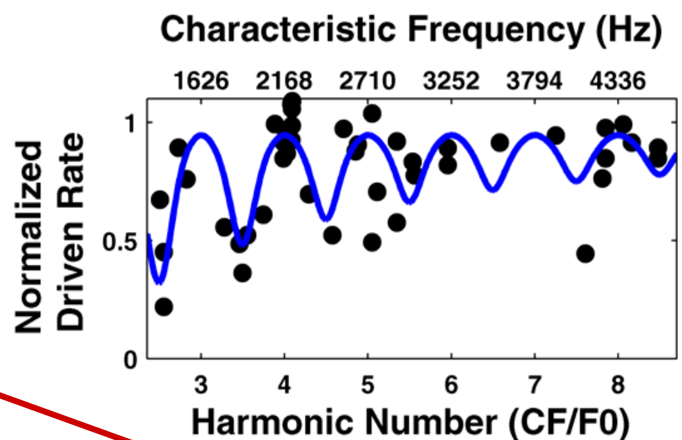
## Method

- Single-unit recordings from auditory nerve in anesthetized cats
- Use very wide range of F0s (110-3520 Hz)
- Missing-fundamental stimuli, equal-amplitude harmonics

# Pitch is estimated by matching harmonic templates to rate response vs. CF

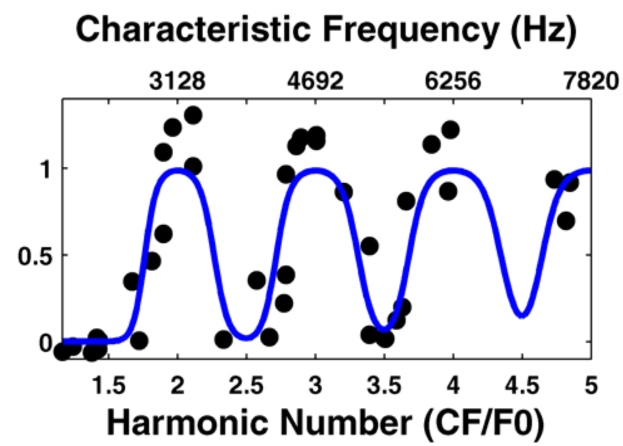
**Estimated F0: 554 Hz**

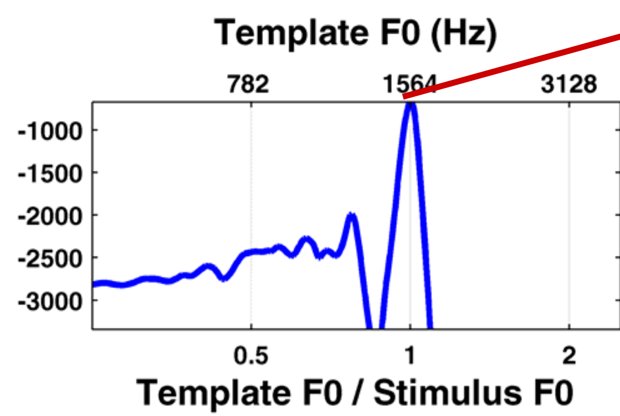
**Stimulus F0: 542 Hz**




**Estimated F0: 1566 Hz**

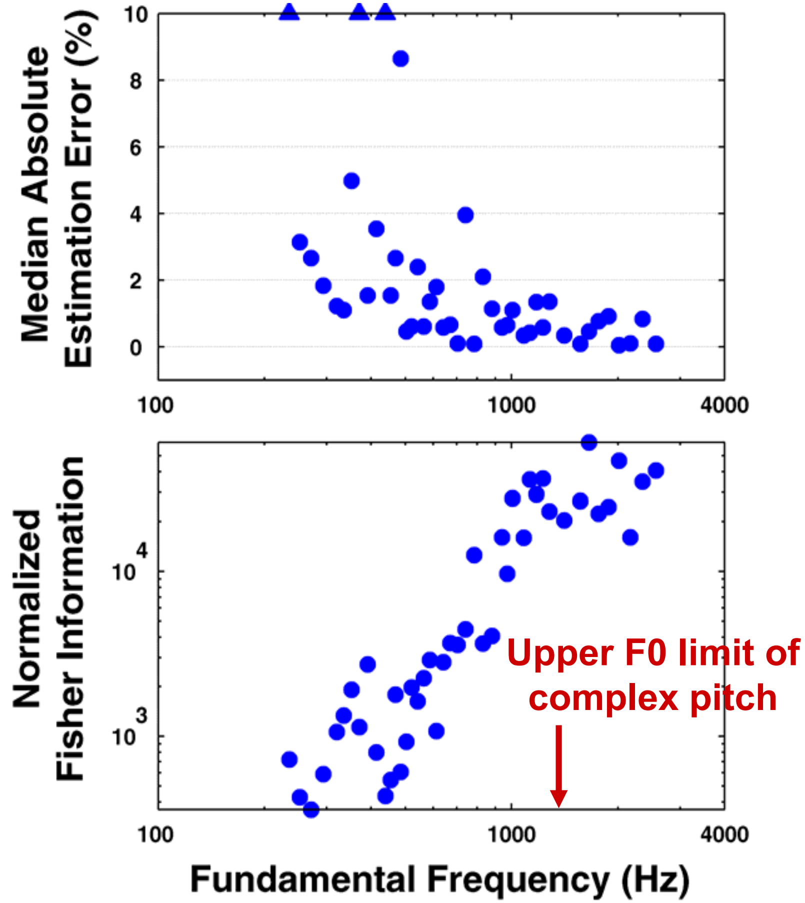
**Stimulus F0: 1564 Hz**




Please see: Cedolin, L., and B. Delgutte. "Pitch of Complex Tones: Rate-Place and Interspike-Interval Representations in the Auditory Nerve." *J Neurophysiol* (Mar 23, 2005).

# Rate-place estimates of pitch are accurate for F0 above 400 Hz




but strength of pitch representation increases with F0 beyond upper frequency limit of missing-fundamental pitch

Please see: Cedolin, L., and B. Delgutte. "Pitch of Complex Tones: Rate-Place and Interspike-Interval Representations in the Auditory Nerve." *J Neurophysiol* (Mar 23, 2005).

# Licklider's autocorrelation model

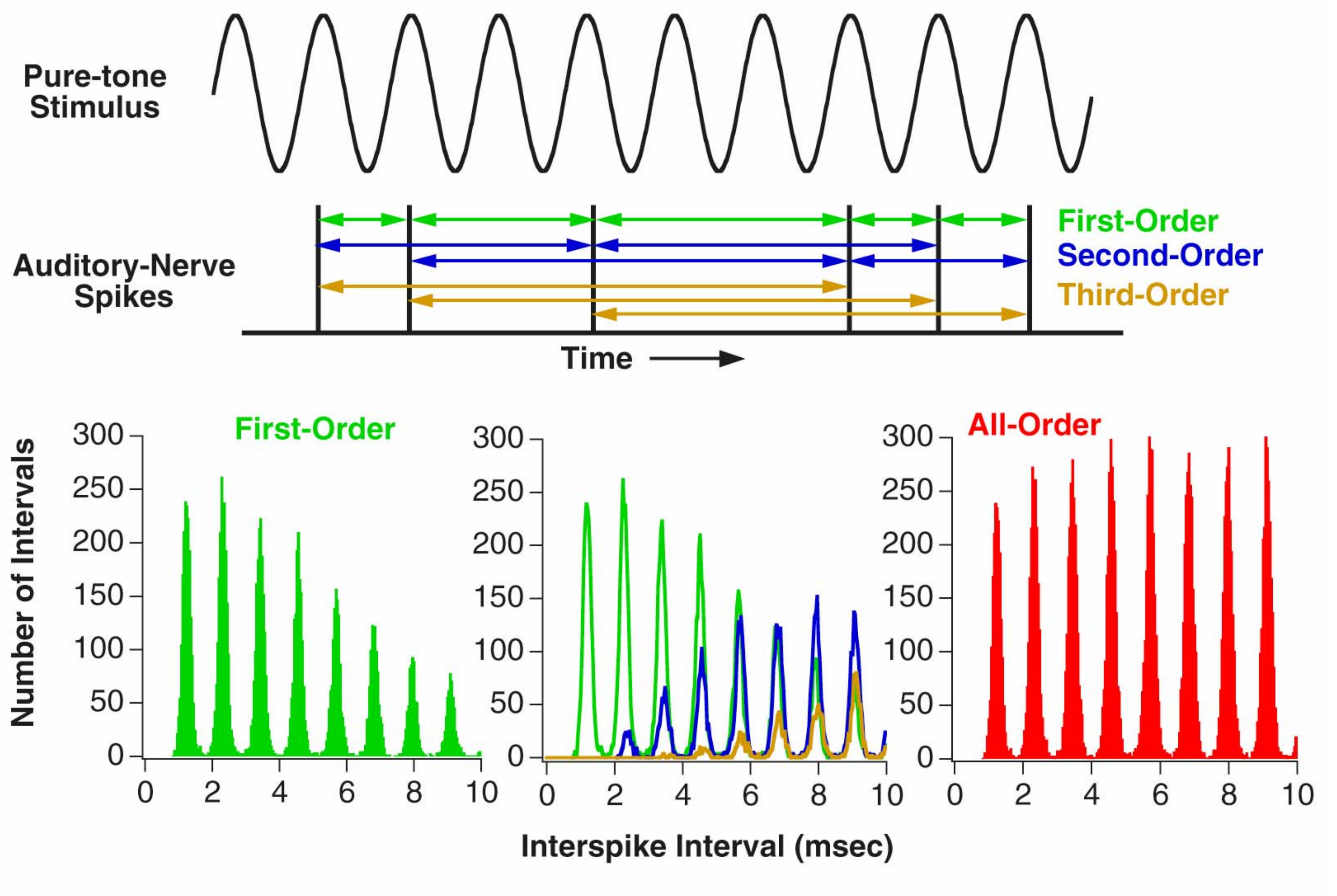
- *Autocorrelation function* :

$$r_t(\tau) = \frac{1}{T}\int_{-T}^{0} x(t)x(t-\tau)dt$$

  - Delay: chain of synapses or long, thin axon.
  - Multiplication: *coincidence detector* neurons.
- The autocorrelation function of a spike train is equivalent to its *all-order interspike interval*.
- Temporal code to place code transformation

Please see: Licklider, J. C. "A duplex theory of pitch perception." *Experientia* 7, no. 4 (Apr 15, 1951): 128-34.

# First-order and all-order interspike interval distributions



# Pitch is prominently represented in pooled interspike interval distribution

Single-formant vowel

Figures removed due to copyright reasons. Please see: Cariani, Peter A., and Bertrand Delgutte. "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch." *J Neurophysiology* 76, no. 3 (1996):1698-1734.

- Waveform (above) and autocorrelation function (top right) of "single formant vowel" show periodicity at 160 Hz fundamental frequency
- All-order interspike interval histograms of most AN fibers arranged in order of CF show peak at period of single-formant vowel
- Pooled interval distributions, formed by summing interval histograms from all fibers, shows prominent representation of stimulus period

# Pitch representation is more robust in all-order interspike interval distributions than in first-order distributions

Single-formant vowel
F0=80Hz

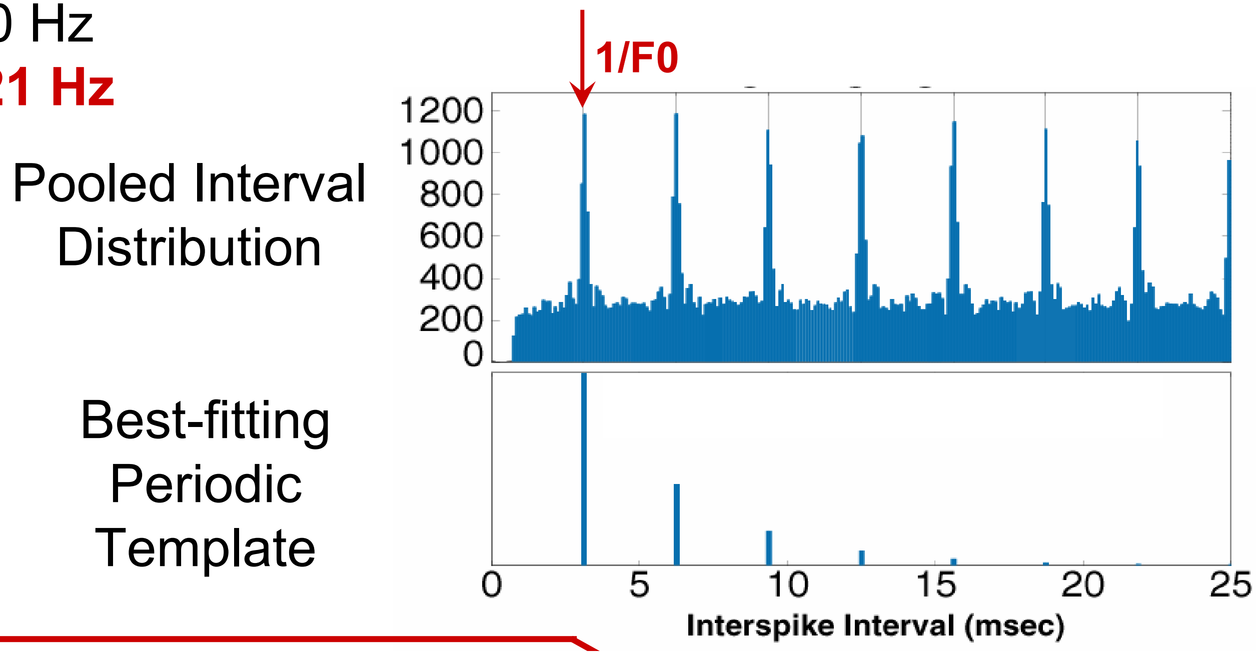
Figures removed due to copyright reasons. Please see: Cariani, Peter A., and Bertrand Delgutte. "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch." *J Neurophysiology* 76, no. 3 (1996):1698-1734.

- Pooled interval distribution constructed from all-order interspike intervals (left) gives more stable representation of F0 across stimulus levels than pooled distribution based on first-order intervals

# Pitch is estimated by matching periodic template to pooled interspike interval distribution

Stimulus F0: 320 Hz

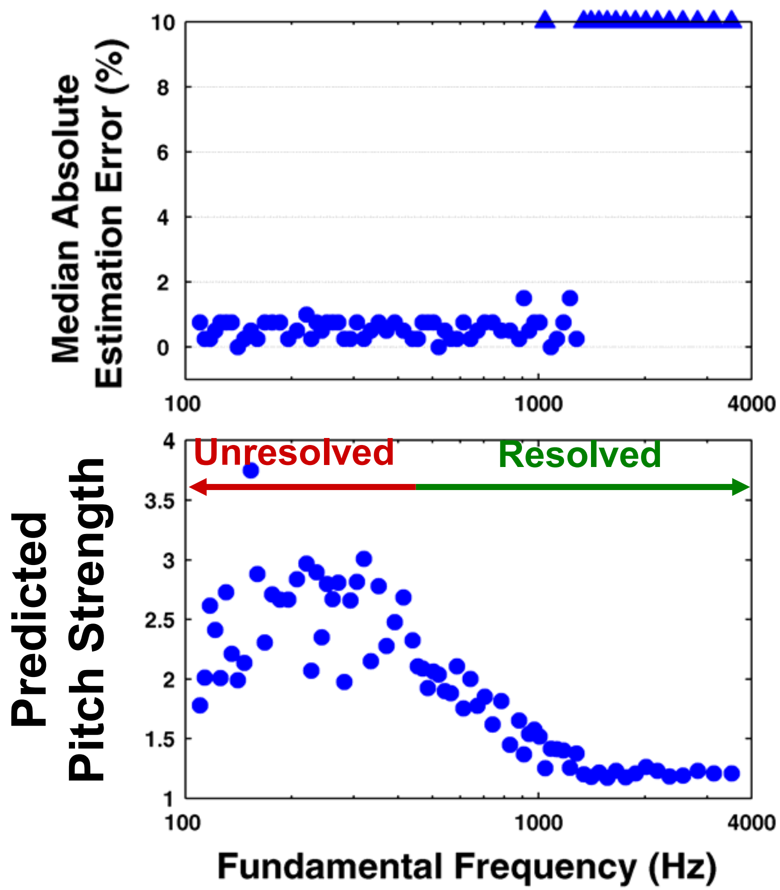
**Estimated F0: 321 Hz**




Please see: Cedolin, L., and B. Delgutte. "Pitch of Complex Tones: Rate-Place and Interspike-Interval Representations in the Auditory Nerve." *J Neurophysiol* (Mar 23, 2005).

# Interval-based estimated of pitch are highly accurate for F0 below 1300 Hz

Upper F0 limit matches limit of missing-fundamental pitch in human




but pitch salience of unresolved harmonics is overestimated

Please see: Cedolin, L., and B. Delgutte. "Pitch of Complex Tones: Rate-Place and Interspike-Interval Representations in the Auditory Nerve." *J Neurophysiol* (Mar 23, 2005).

# Autocorrelation model gives similar predictions for tones with resolved and unresolved harmonics

Figures removed due to copyright reasons. Please see: Cariani, Peter A., and Bertrand Delgutte. "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch." *J Neurophysiology* 76, 3 (1996): 1698-1734.

# Interspike-interval (autocorrelation) model of pitch

- <u>Strengths:</u>
  - Works with both resolved and unresolved harmonics
  - Accounts for many pitch phenomena: Pitch equivalence, phase and level invariance, pitch shift of inharmonic tones, dominance region
  - Account for upper frequency limit of musical pitch (~5 kHz)
- <u>Weaknesses:</u>
  - Fails to predict greater salience of pitch produced by resolved harmonics
  - Underestimates pitch salience of pure tones
  - Requires long neural delay lines (33 ms for 30 Hz lowest pitch)

# Neither rate-place, nor interspike-interval representation of pitch is satisfactory

| | Psychophysically-desirable properties | | | |
|---|---|---|---|---|
| | Level robust | Depends on harmonic resolvability | Predicts upper F0 limit | Phase invariant |
| Rate-place | **No** | **Yes** | **No** | **Yes** |
| Interspike Intervals | **Yes** | **No** | **Yes** | **Yes** |

# Spatio-temporal representation of pitch

- Seeks to combine advantages and overcome limitations of rate-place and interspike-interval representations
- Based on idea by Shamma (1985)
- Tested using peripheral auditory models and recording from AN fibers

# Spatio-temporal representation of pure-tone frequency in AN model response

Figures removed due to copyright reasons.

Zhang et al. (2001) model
Human cochlear bandwidths

# Spatio-temporal representation of resolved harmonics in AN model response

Figures removed due
to copyright reasons.

Zhang et al. (2001) model
Human cochlear bandwidths

# Spatio-temporal representation of pitch

- Principle:
  - Spatial variation of phase of basilar membrane motion gives cues to locations of resolved harmonics
  - These phase cues can be extracted by mechanism sensitive to relative timing of spikes in AN fibers innervating neighboring cochlear locations (e.g. lateral inhibition, coincidence detection, etc…)
- Possible advantages over interspike interval representation:
  - Depends on harmonic resolvability
  - Requires no delay lines

# Scaling invariance in cochlear mechanics (Zweig, 1976)

Figures removed due
to copyright reasons.

Larsen et al., 2005

# Spatio-temporal representation of pure-tone frequency in AN fiber response

Figures removed due to copyright reasons.

# Spatio-temporal representation of resolved harmonics in AN fiber response

Figures removed due
to copyright reasons.

# Spatio-temporal representation is robust with stimulus level

Figures removed due to copyright reasons.

# Spatio-temporal representation predicts upper F0 limit of pitch of complex tones

Figures removed due
to copyright reasons.

Zhang et al. (2001) model
Human cochlear bandwidths

# Neural representations of pitch

| | Psychophysically-desirable properties | | | |
|---|---|---|---|---|
| | Level robust | Depends on harmonic resolvability | Predicts upper F0 limit | Phase invariant |
| Rate-place | **No** | **Yes** | **No** | **Yes** |
| Interspike Intervals | **Yes** | **No** | **Yes** | **Yes** |
| Spatio-temporal | **Partly** | **Yes** | **Yes** | **Yes?** |

# What might a "pitch neuron" look like?

- Tuned to fundamental frequency of harmonic complex tones
- Pitch equivalence: Similar tuning to stimuli that have the same pitch, regardless of spectral content
- Stronger response to stimuli that produce stronger pitch
- Multiple tuning to harmonically-related stimuli (If T is a period, so are 2T, 3T, etc)
- Basic properties do not depend on whether mechanism is temporal or spectral

# Neurons with multi-peaked tuning curves in monkey primary auditory cortex



- Awake marmoset preparation
- Peak frequencies tend to be harmonically-related
- Other neurons (not shown) show single peak TC, but with inhibition at harmonically related frequencies

# Pitch neuron in auditory cortex of awake monkey?

- B. Neuron responds to 3-component harmonic complex tones (A) whose F0 matches pure-tone BF (182 Hz, see C)
- E. Neuron does not respond to individual harmonics (except fundamental)

Figures removed due to copyright reasons.

Bendor & Wang, unpublished

# Pitch neurons are located in restricted area near anterolateral border of primary auditory cortex (A1)

Figures removed due to copyright reasons.

Bendor & Wang, unpublished

# CN unit types differ in their phase locking to pure tones

Figures removed due to copyright reasons.
Please see: Rhode, W. S., and P. H. Smith. "Encoding timing and intensity in the ventral cochlear nucleus of the cat." *J Neurophysiol* 56, no. 2 (Aug 1986): 261-86.

- Good phase lockers: Primarylike, some Onset
- Poor phase lockers: Choppers, many DCN neurons (not shown)

# Phase locking to pure tones degrades with each successive synapse in auditory pathway

Figures removed due to copyright reasons.

1: Auditory nerve
2: Cochlear nucleus
3: Superior olive
5: Medial geniculate

$$F_{\max} \,\alpha\, \frac{1}{\sqrt{N}}$$

Rouiller et al. (1979)

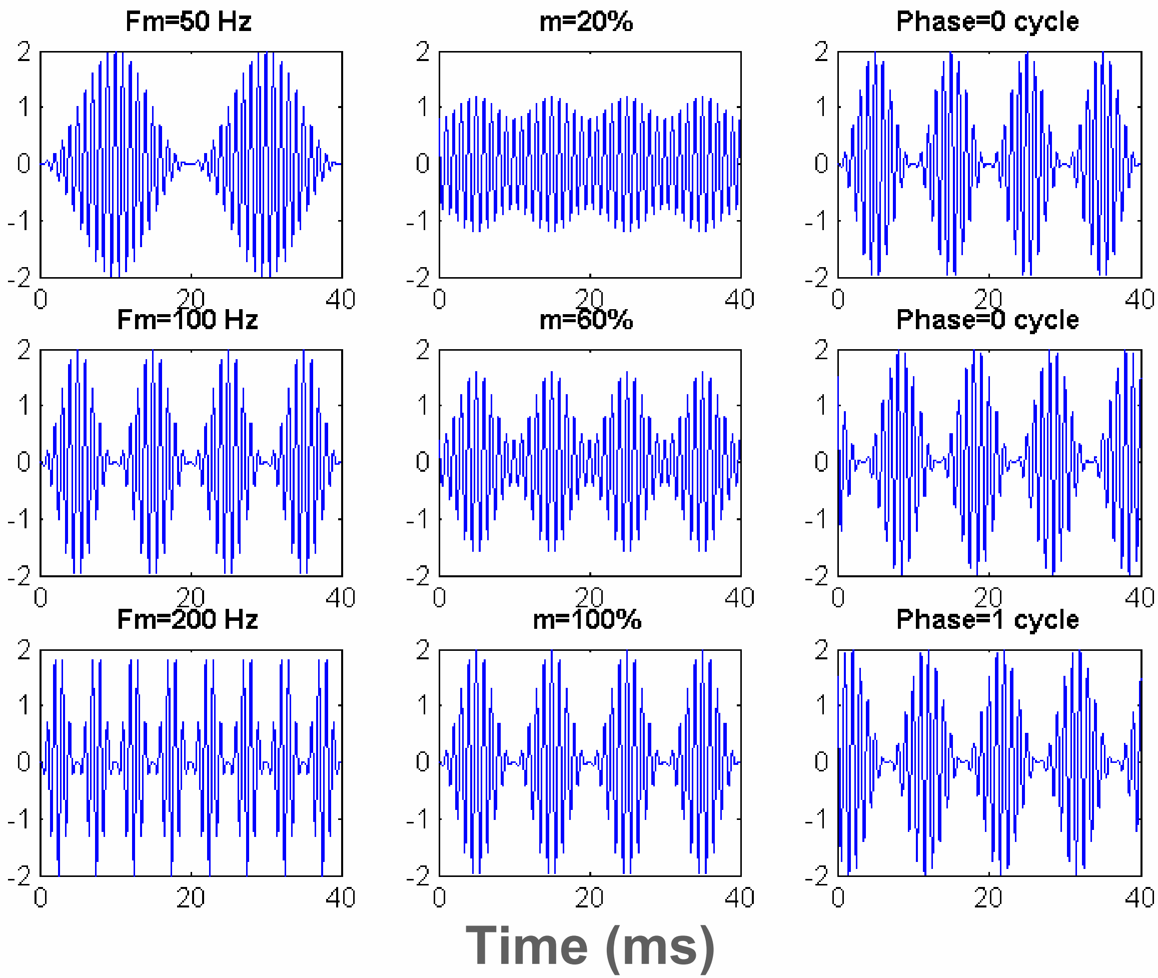
# Pitch is poorly represented in pooled interval distributions of IC neurons

Figures removed due
to copyright reasons.

McKinney (2001)

# Sinusoidally amplitude-modulated tones: Modulation frequency, depth and phase

$$s(t) = \left(1 + m\ \cos(2\pi f_m t + \phi_m)\right)\cos 2\pi f_c t$$



# Modulation Transfer Function (MTF)

- A sinusoidally amplitude-modulated (SAM) tone is defined by:

$$s(t) = A\left(1 + m\ \cos(2\pi f_m t + \phi_m)\right)\cos 2\pi f_c t$$

  - $f_m$ and $f_c$ are the modulation and carrier frequencies, respectively
  - $m$ is the *modulation depth*, a number between 0 and 1
  - $\phi_m$ is the *modulation phase*
- If an AM tone is used as input to a system, the *modulation gain* is the ratio of the modulation depth in the system's output to that in the input
- A linear system's sensitivity to modulation can be characterized by its *modulation transfer function* (MTF), which expresses the modulation gain as a function of modulation frequency $f_m$

# MTFs of AN fibers are lowpass
# MTFs of some CN neurons are bandpass



- A: AN fibers have *lowpass* modulation transfer functions (MTFs) for AM tones at the CF.
- B-H: MTFs of CN neurons can be either *lowpass* or *bandpass*. MTF shape correlates with unit type based on tone-burst response pattern.
- Best modulation frequencies (BMFs) are 50-500 Hz

# Map of best modulation frequencies orthogonal to frequency map in IC?

- Rate MTFs of many IC neurons are tuned to narrow range of modulation frequencies: The temporal code in AN & CN is transformed into a rate code in IC.
- Some *multi* units in IC have BMFs up to 1000 Hz.
- High BMFs are represented in the caudal and lateral parts of iso-frequency bands.
- Multi units could represent inputs to the IC rather than IC neurons.
- Role in pitch processing?

Figures removed due to copyright reasons. Please see: Schreiner, C. E., and Langner G. "Periodicity coding in the inferior colliculus of the cat. II. Topographical organization." *J Neurophysiol* 60, no. 6 (Dec 1988): 1823-40.

AND

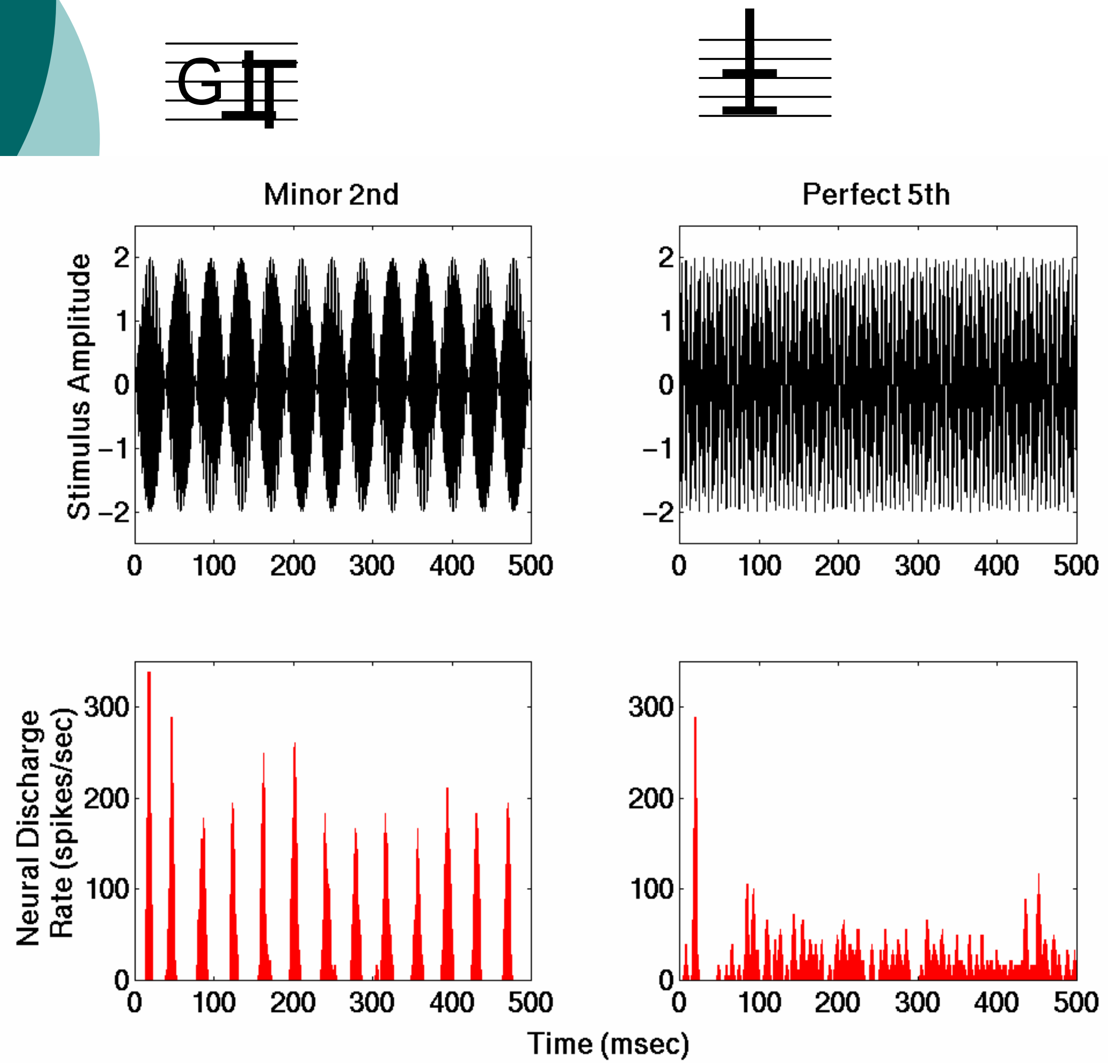
Langner, G., and C. E. Schreiner. "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms." *J Neurophysiol* 60, no. 6 (Dec 1988): 1799-822.

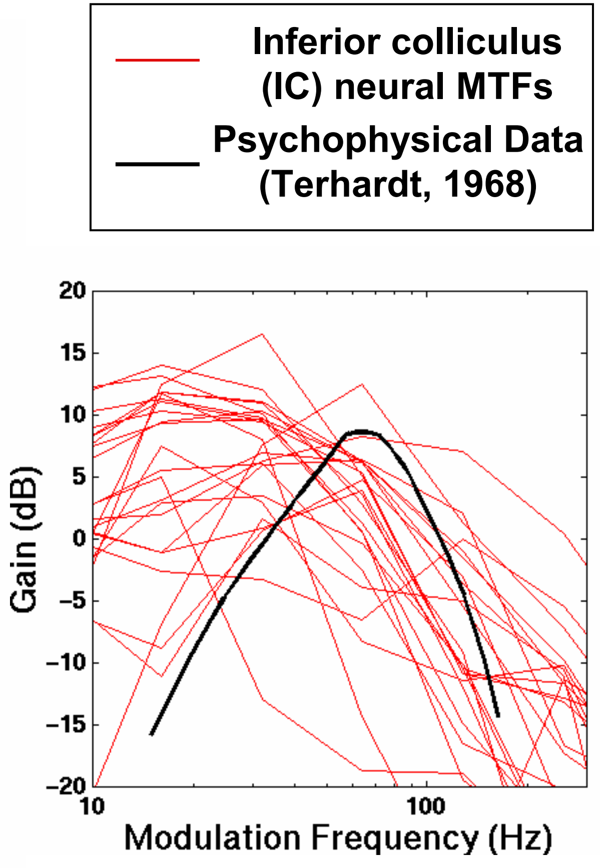
# Most best modulation frequencies in IC are below pitch range of human voice

Figures removed due to copyright reasons.
Please see: Krishna, B. S., and M. N. Semple. "Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus." *J Neurophysiol* 84, no. 1 (Jul 2000): 255-73.

- IC BMFs: 1-150 Hz
- Upper frequency cutoffs of phase locking to AM in IC: 10-300 Hz
- Range in IC appears too low to support general pitch mechanism

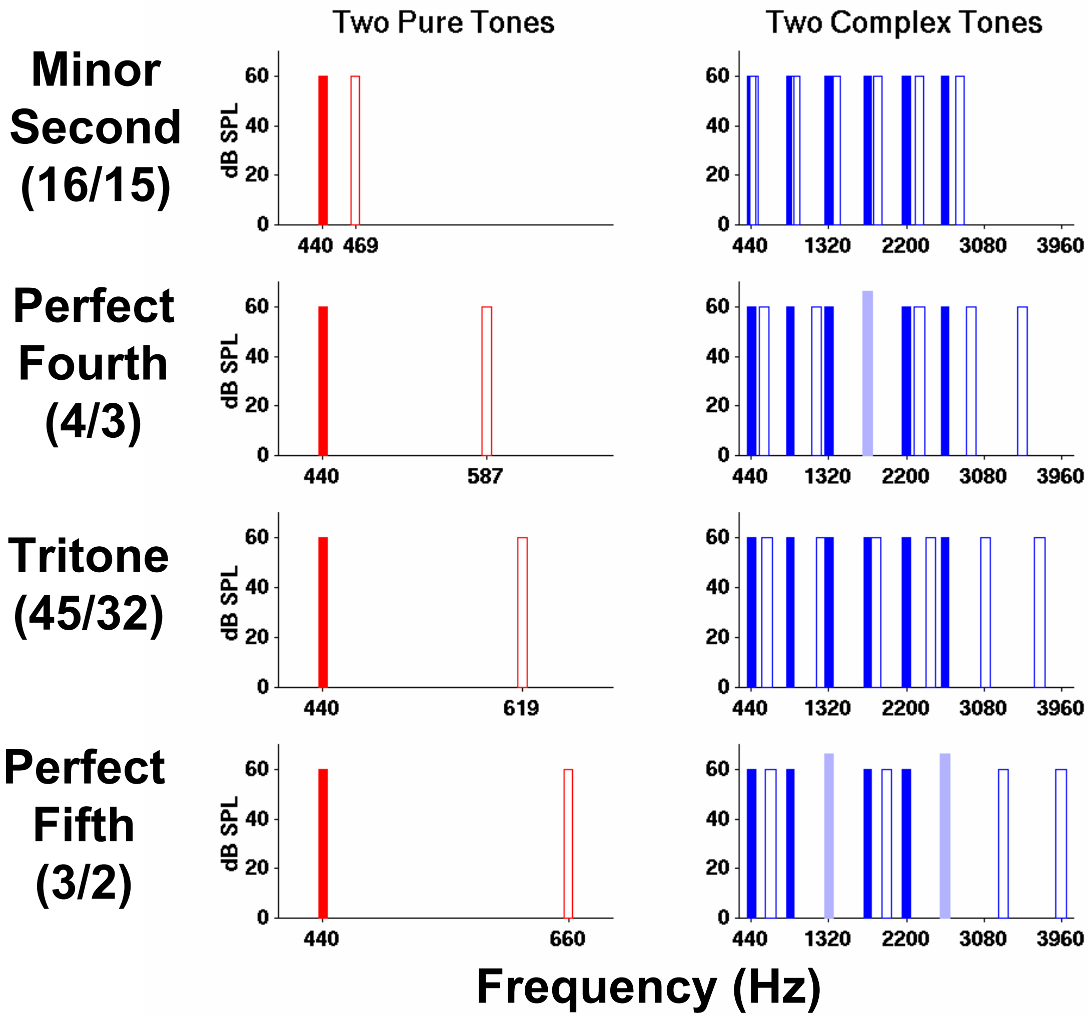
# Temporal envelope modulations in roughness range are reflected in IC responses



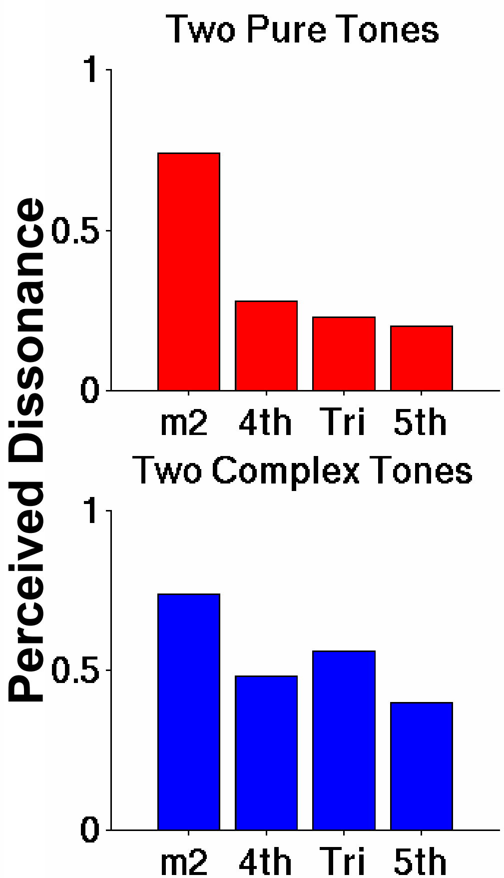




McKinney & Delgutte (2001)

# Consonant and Dissonant Stimuli




**Psychoacoustic Data**




(Adapted from Terhardt, 1984)

# Dissonance is reflected in average rates and rate fluctuations of IC neurons

Figures removed due
to copyright reasons.

# Population neural responses in IC correlate with psychophysical dissonance ratings

Figures removed due to copyright reasons.

McKinney 2001