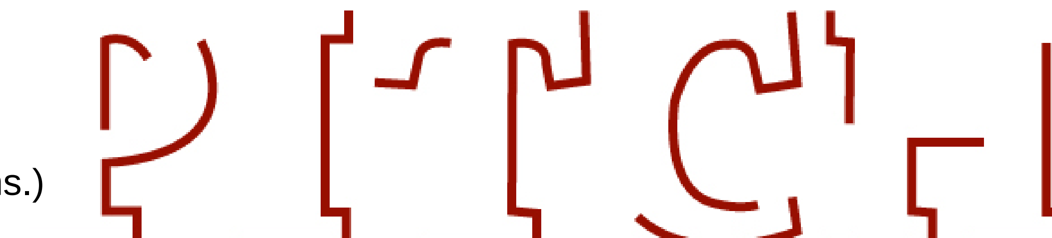
Harvard-MIT Division of Health Sciences and Technology
HST.725: Music Perception and Cognition
Prof. Peter Cariani

HST 725 Lecture 6
Temporal models for musical pitch

(Image removed due to copyright considerations.)

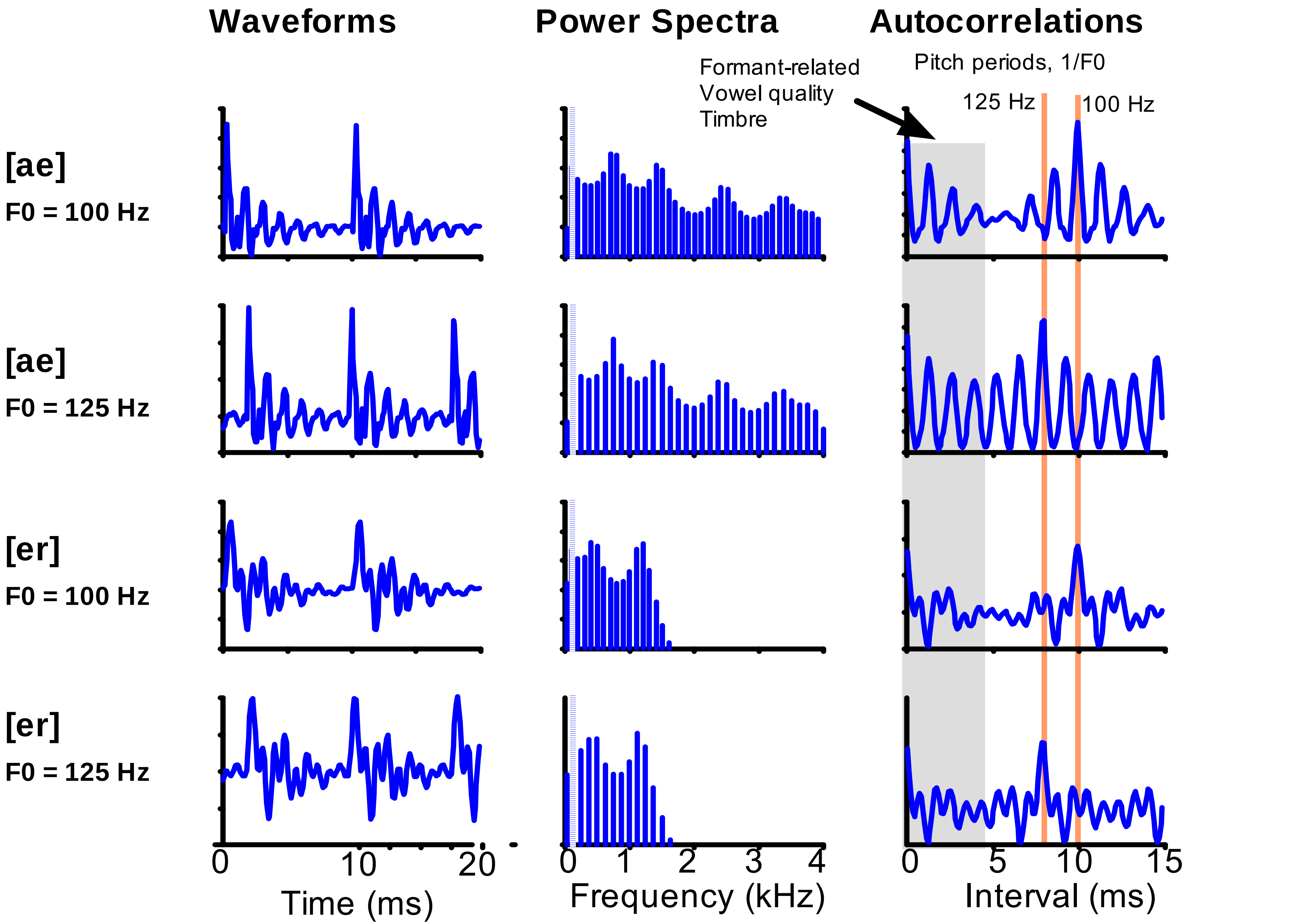


www.cariani.com

The search for the missing fundamental: theories & models of musical pitch

- **Temporal theories of pitch**
  - Residues: Beatings of unresolved harmonics (Schouten, 1940's)
  - Problems with residues and envelopes
  - Temporal autocorrelation models (Licklider, 1951)
  - Interspike interval models (Moore, 1980)
  - Correlogram demonstration (Slaney & Lyon, Apple demo video)
  - Population-interval models (Meddis & Hewitt, Cariani & Delgutte)
- **Interval-based representations of the spectrum**
  - Vowels
- **Competition between interspike interval patterns**
  - Pitch multiplicity
  - Masking
- **Problems & prospects**
  - The fate of temporal information in the auditory system
  - Rethinking the problem of pitch: neural timing nets
    - Pitch and auditory scene analysis
    - Formation and separation of auditory objects using temporal patterns

Synthetic vowels



Pitch : basic properties

- **Highly precise percepts**
  - **Musical half step: 6% change F0**
  - **Minimum JND's: 0.2%** at 1 kHz (20 usec time difference, comparable to ITD jnd)
- **Highly robust percepts**
  - **Robust quality** Salience is maintained at high stimulus intensities
  - **Level invariant** (pitch shifts < few % over 40 dB range)
  - **Phase invariant** (largely independent of phase spectrum, f < 2 kHz)
- **Strong perceptual equivalence classes**
  - **Octave similarities** are universally shared
  - **Musical tonality** (octaves, intervals, melodies) 30 Hz - 4 kHz
- **Perceptual organization ("scene analysis")**
  - **Fusion:** Common F0 is a powerful factor for grouping of frequency components
- **Two mechanisms? Temporal (interval-based) & place (rate-based)**
  - **Temporal:** predominates for periodicities < 4 kH (level-independent, tonal)
  - **Place:** predominates for frequencies > 4 kHz(level-dependent, atonal)

Periodic sounds produce distinct pitches

Many different sounds produce the same pitches

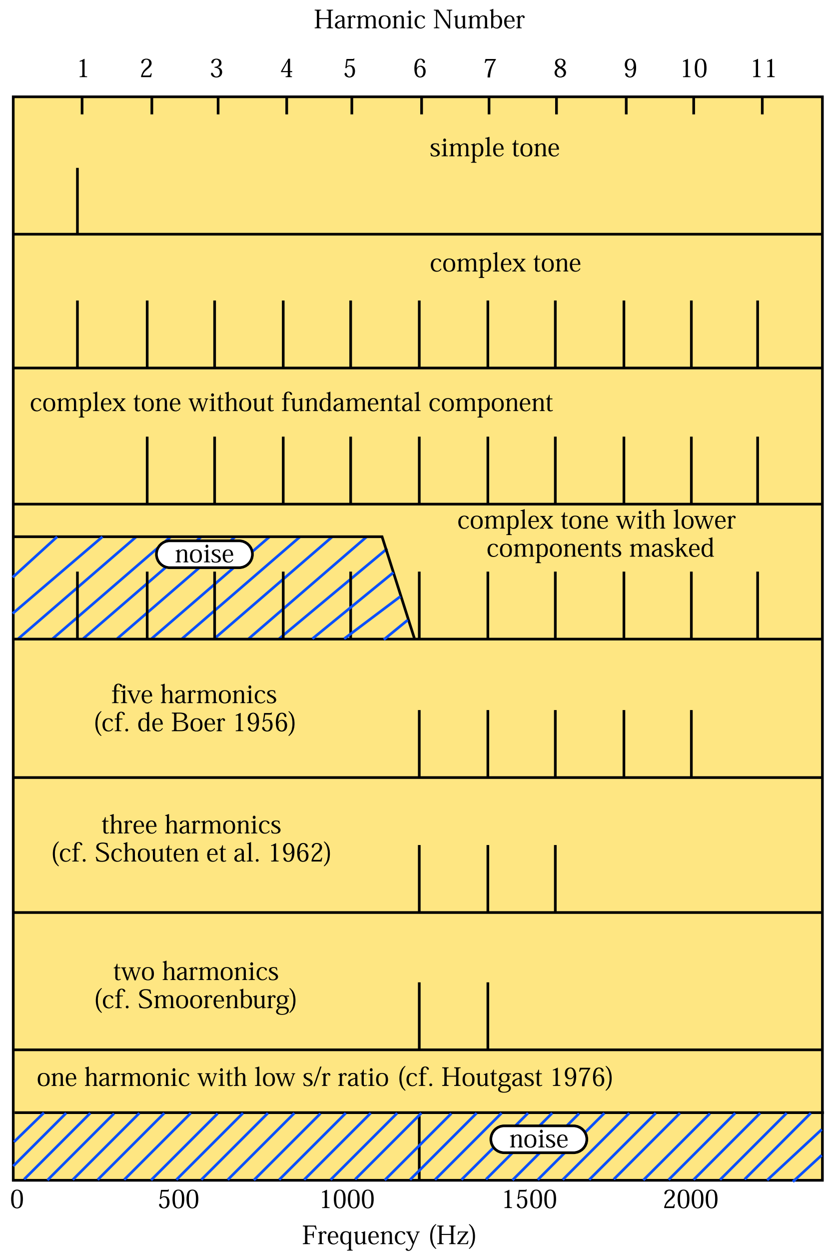
Strong
- Pure tones
- Harmonic complexes
- Iterated noise

Weaker
- High harmonics
- Narrowband noise

Very weak
- AM noise
- Repeated noise

Strong pitches

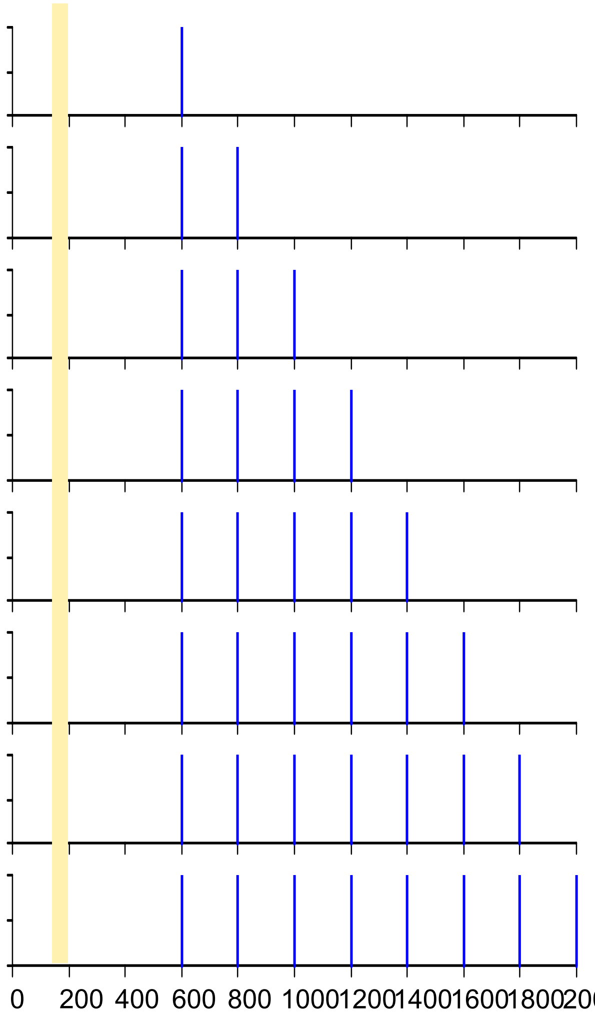
Weaker low pitches

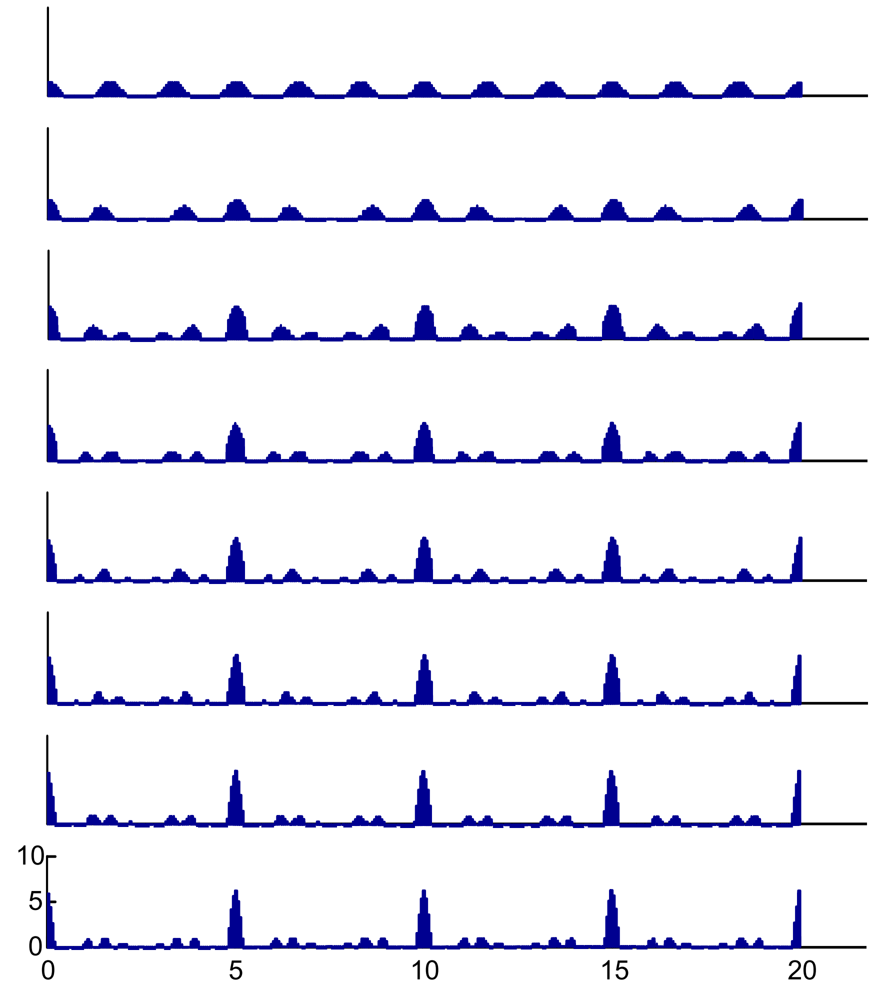


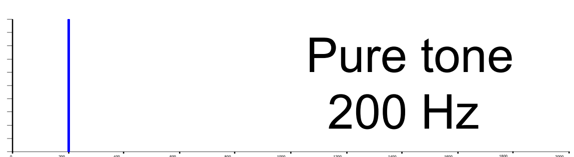
Schematic diagram representing eight signals with the same low pitch.

Pitch as a perceptual emergent

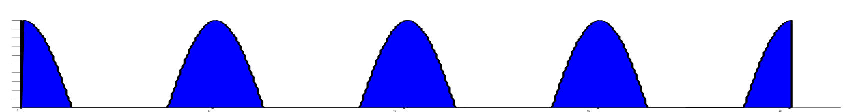
Missing F0 | **Line spectra** | **Autocorrelation (positive part)**



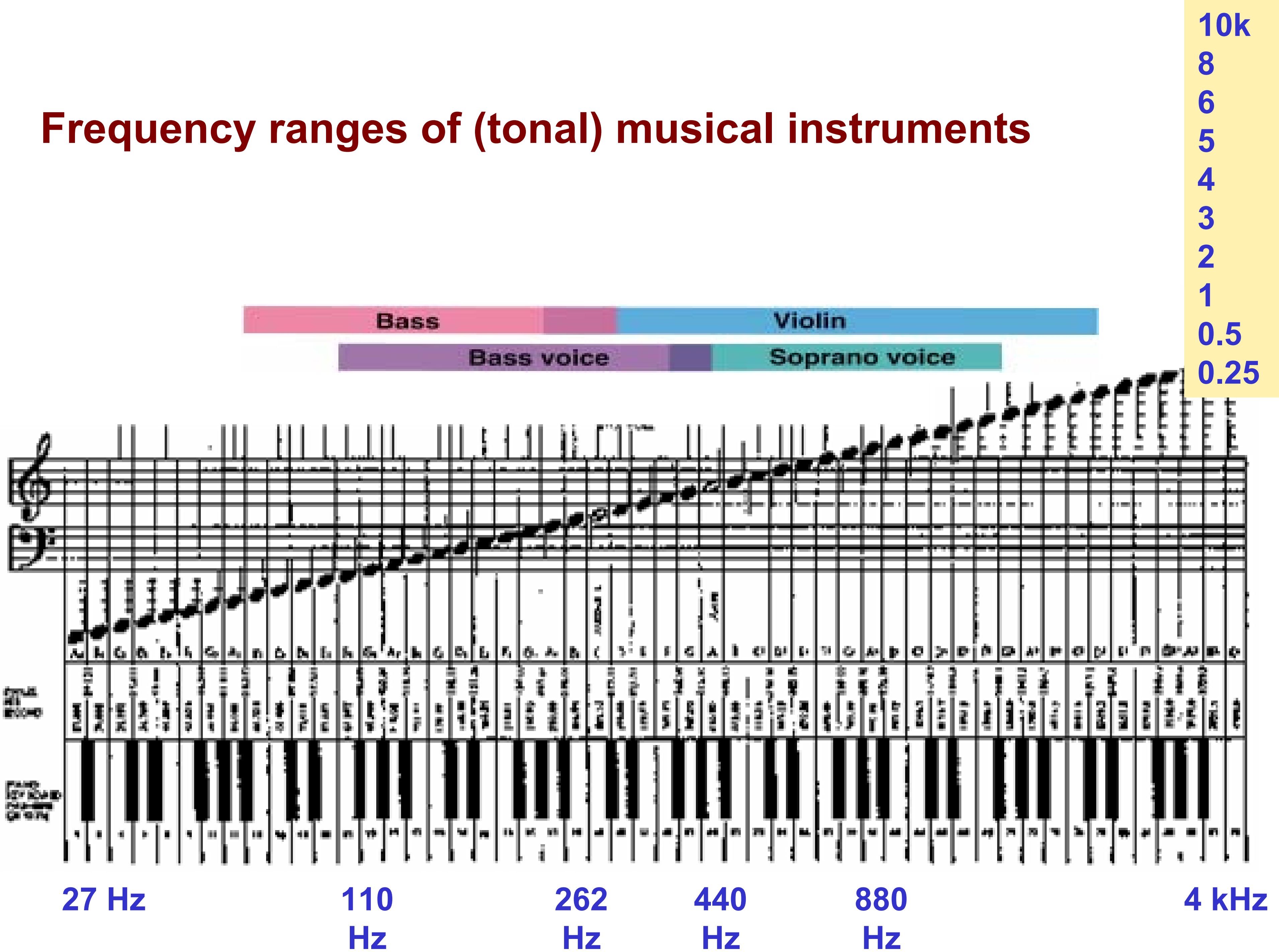




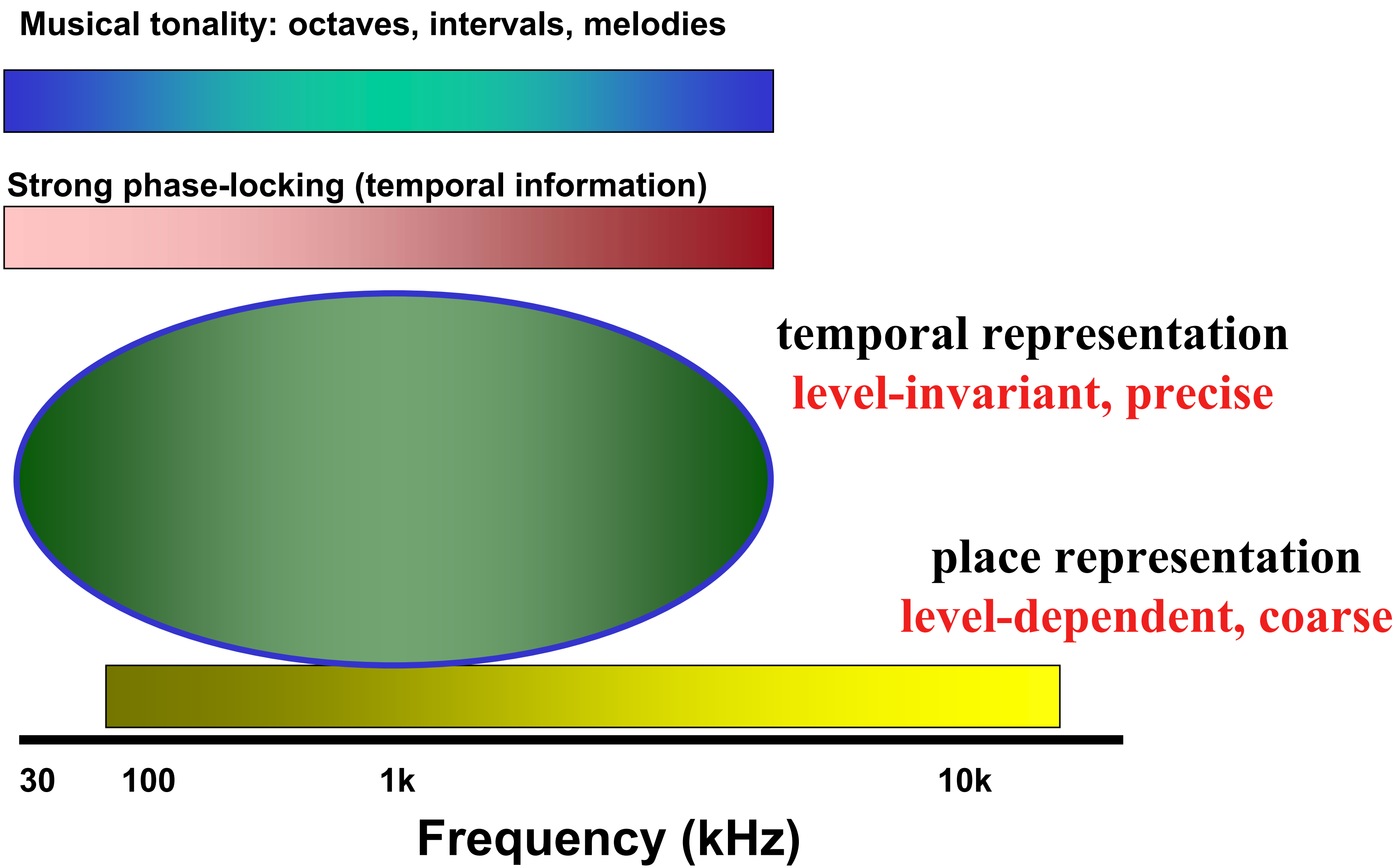
Pure tone
200 Hz



Frequency ranges of (tonal) musical instruments



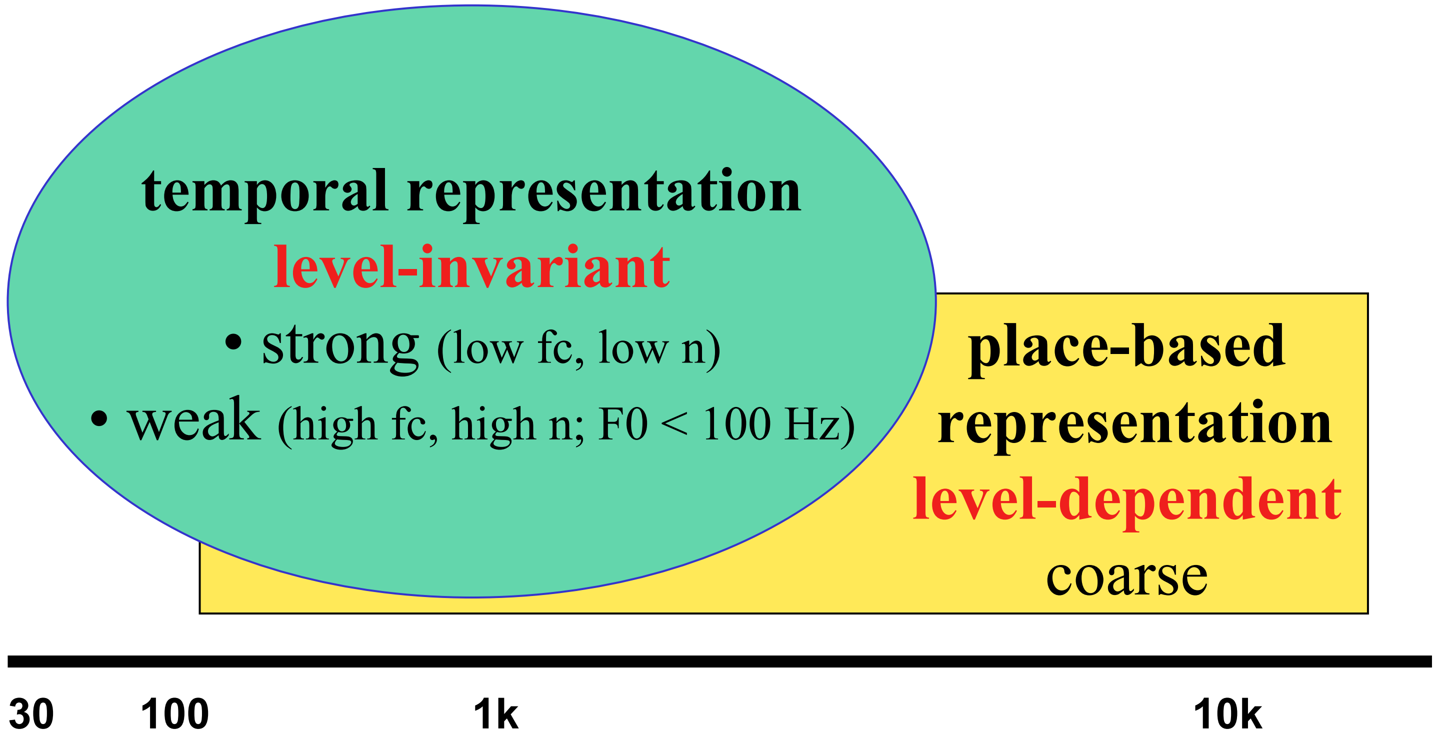
Duplex time-place representations
"Pitch is not simply frequency"

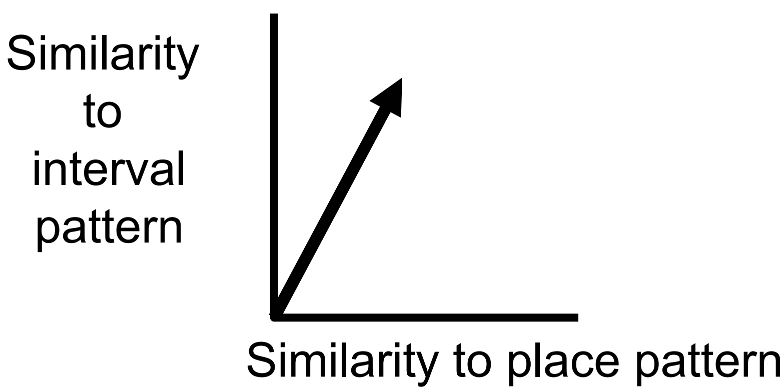


Pitch height and pitch chroma

Please see Figure 1, 2, and 7 in Roger N. Shepard. "Geometrical Approximations to the Structure of Musical Pitch." *Psychological Review* 89 no. 4 (1982): 305-322.

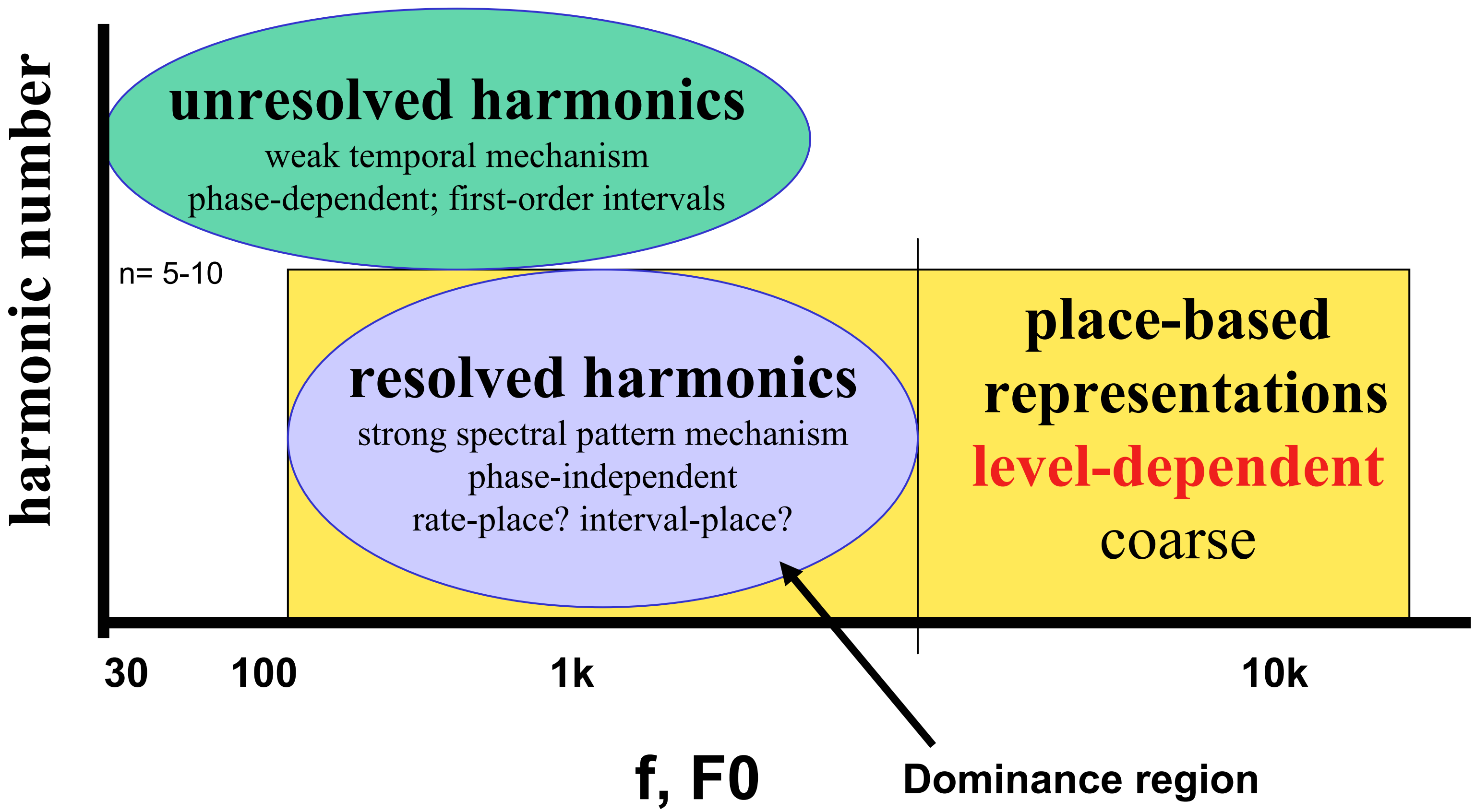
Duplex time-place representations

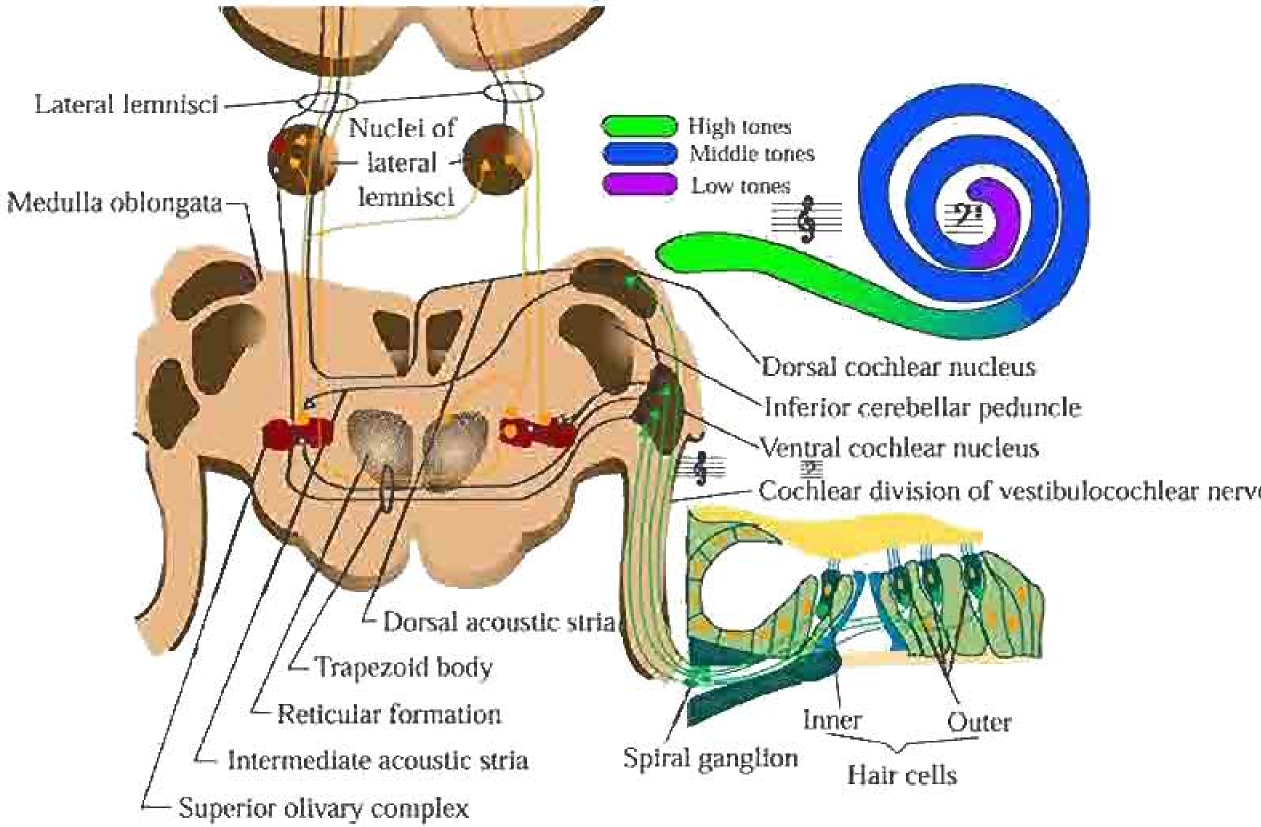




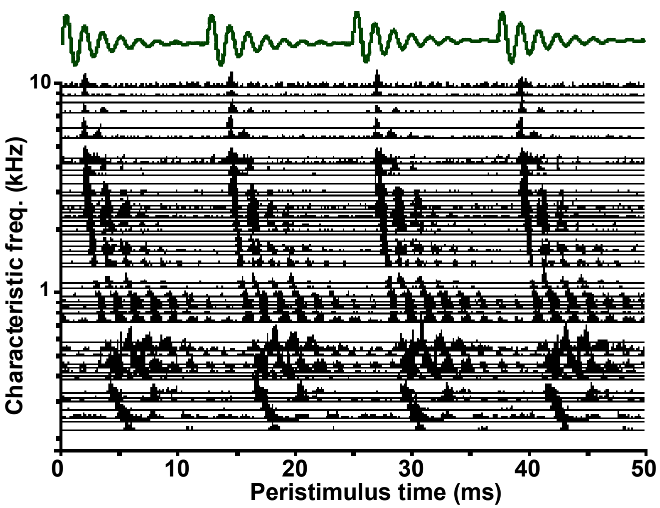
<u>cf. Terhardt's spectral and virtual pitch</u>

A "two-mechanism" perspective (popular with some psychophysicists)




Lateral lemnisci
Nuclei of lateral lemnisci
Medulla oblongata
High tones
Middle tones
Low tones
Dorsal cochlear nucleus
Inferior cerebellar peduncle
Ventral cochlear nucleus
Cochlear division of vestibulocochlear nerve
Dorsal acoustic stria
Trapezoid body
Reticular formation
Intermediate acoustic stria
Superior olivary complex
Spiral ganglion
Inner
Outer
Hair cells

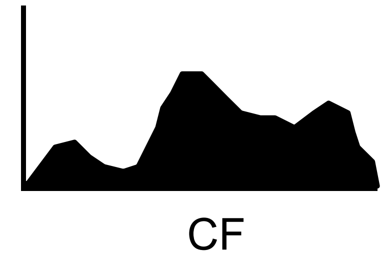
Some possible auditory representations



Local

Rate-place →

Central spectrum



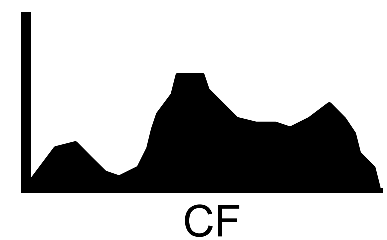
Masking phenomena
Loudness

Synchrony-place

Phase-place

Interval-place

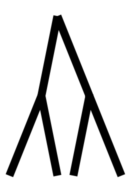
Central spectrum



Pure tone pitch JNDs: Goldstein

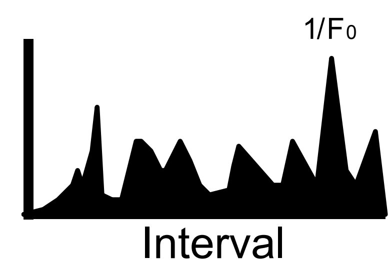
Stages of integration

All-at-once



Population-interval →

Population interval



Complex tone pitch

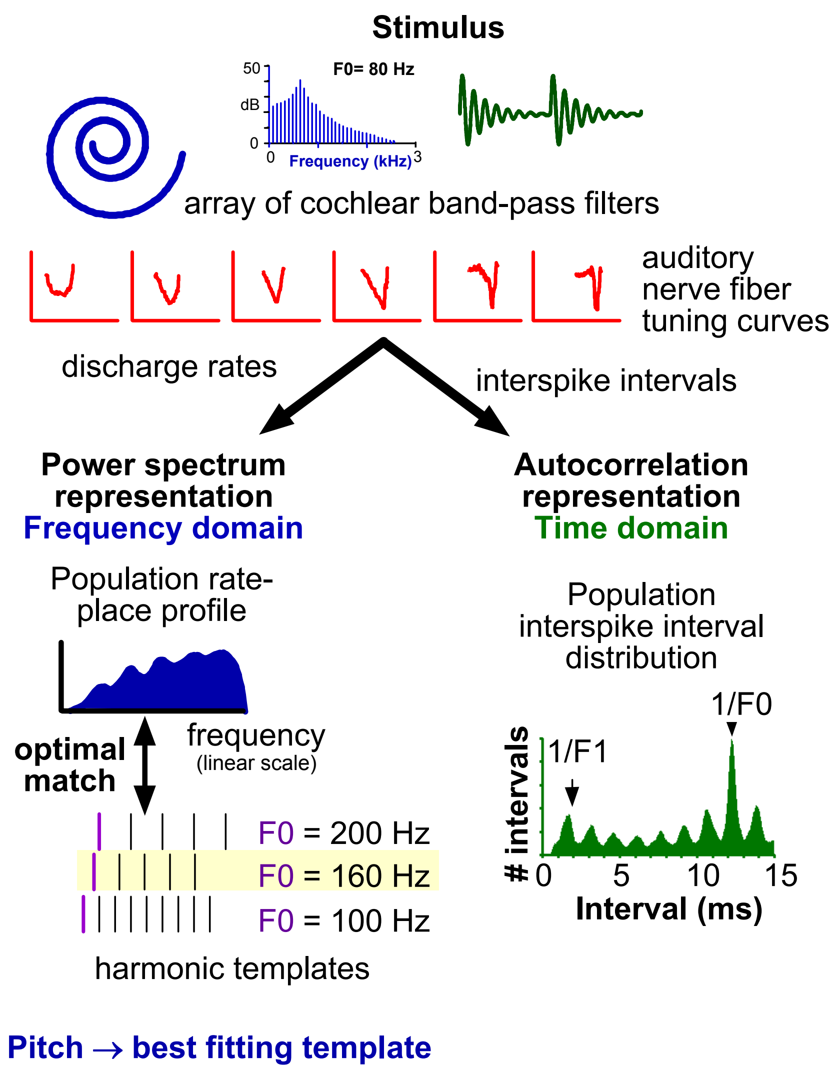
Global

General theories of pitch

1. Distortion theories
– reintroduce F0 as a cochlear distortion component (Helmholtz)
–sound delivery equipment can reintroduce F0 through distortion
–however, masking F0 region does not mask the low pitch (Licklider)
–low pitch thresholds and growth of salience with level not consistent with distortion processes (Plomp, Small)
–binaurally-created pitches exist

2. Spectral pattern theories
–Operate in frequency domain
–Recognize harmonic relations on resolved components

3. Temporal pattern theories
–Operate in time domain
–Analyze interspike interval dists.



Basic problems to be solved

- "Hyperacuity problem"
- Account for the precision of pitch discriminations given the relatively coarse tunings of auditory neurons (at all levels), especially lower-frequency ones (BFs < 2 kHz)

- "Dynamic range problem"
- Account for the ability of listeners to discriminate small fractional changes ($\Delta I/I$) in intensity over a large dynamic range, and especially at high SPLs, where the vast majority of firing rates are saturated.

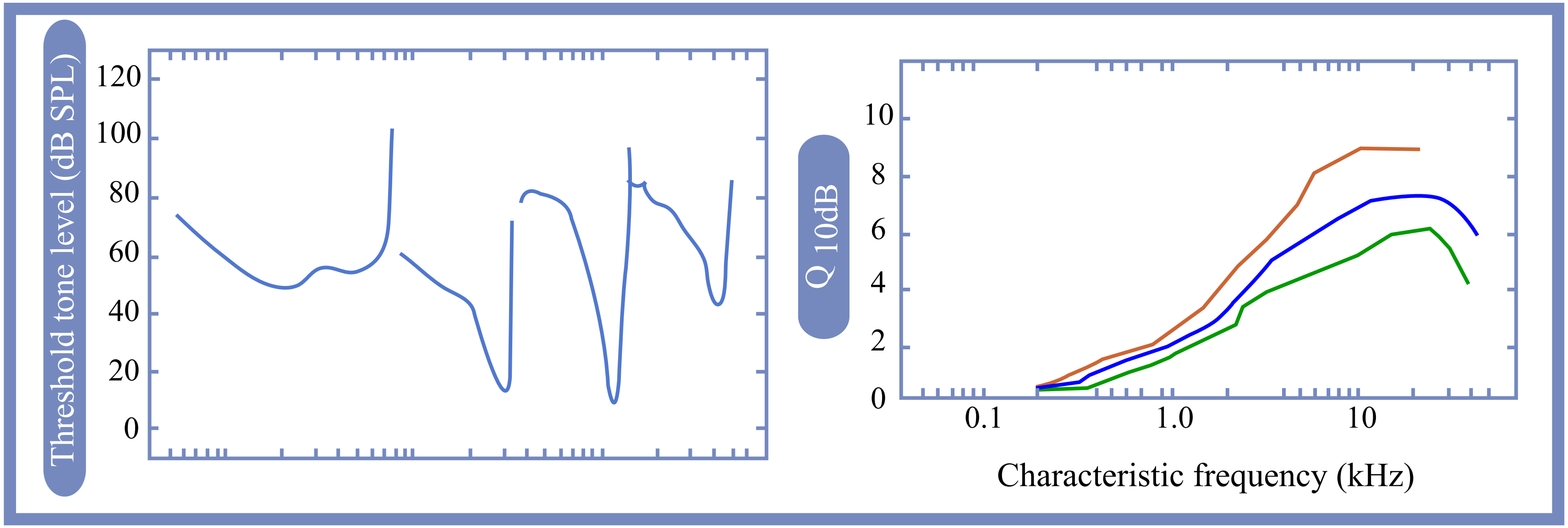
- "Level-invariance problem"
- Account for the invariance (and precision) of auditory percepts over large dynamic ranges given the profound changes in neural response patterns that occur over those ranges (rate saturation, rate non-montonicities).

- Pitch equivalence
- Account for the ability to precisely match pitches of pure and complex tones (pitch equivalence, metamery) given differences in spectra and under conditions where stimulus levels are roved 20 dB or more

- Multiple pitches
- Account for the abilityof the auditory system to simultaneously represent multiple pitches ($\Delta F0$ > 10% apart)

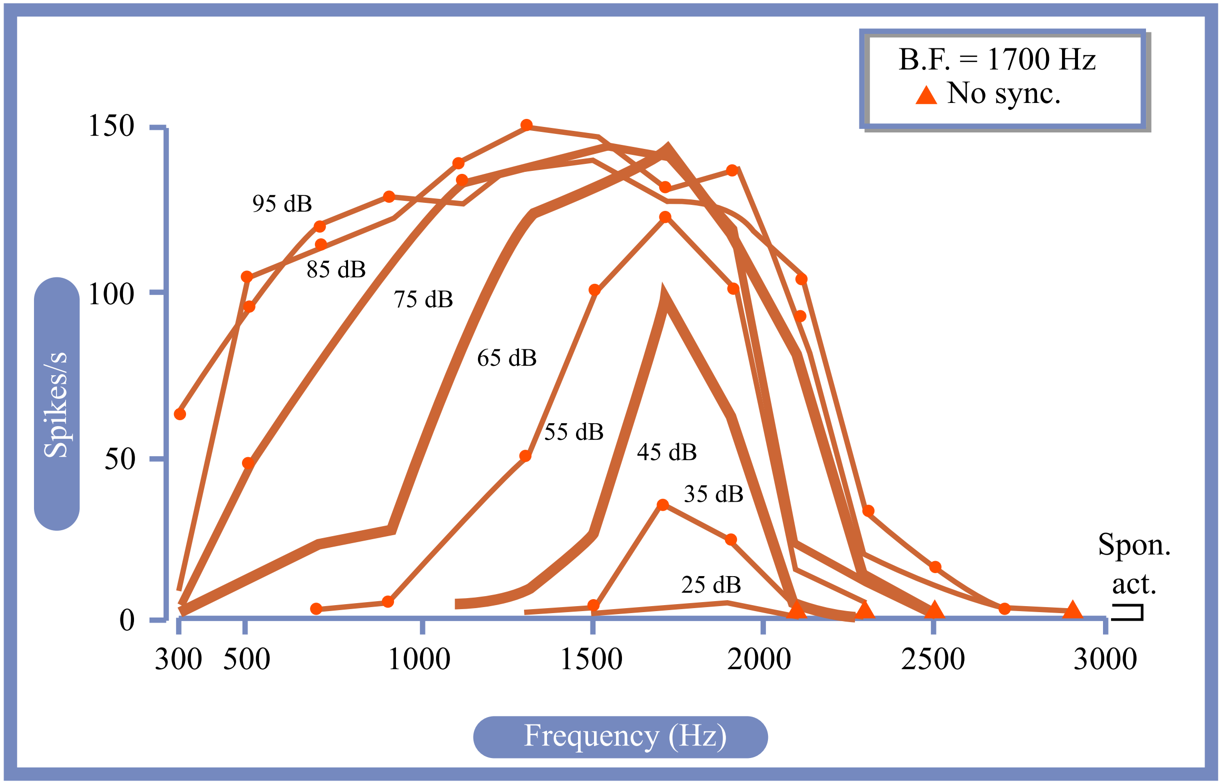
- Relative nature of pitch & transpositional invariance
- Account for the ability to precisely match pitches an octave apart (and/or to recognize patterns of pitch sequences) in the absence of an ability to identify absolute frequencies/periodicities. Account for ability to recognize transposed melodies as similar.

Neural tuning as a function of CF



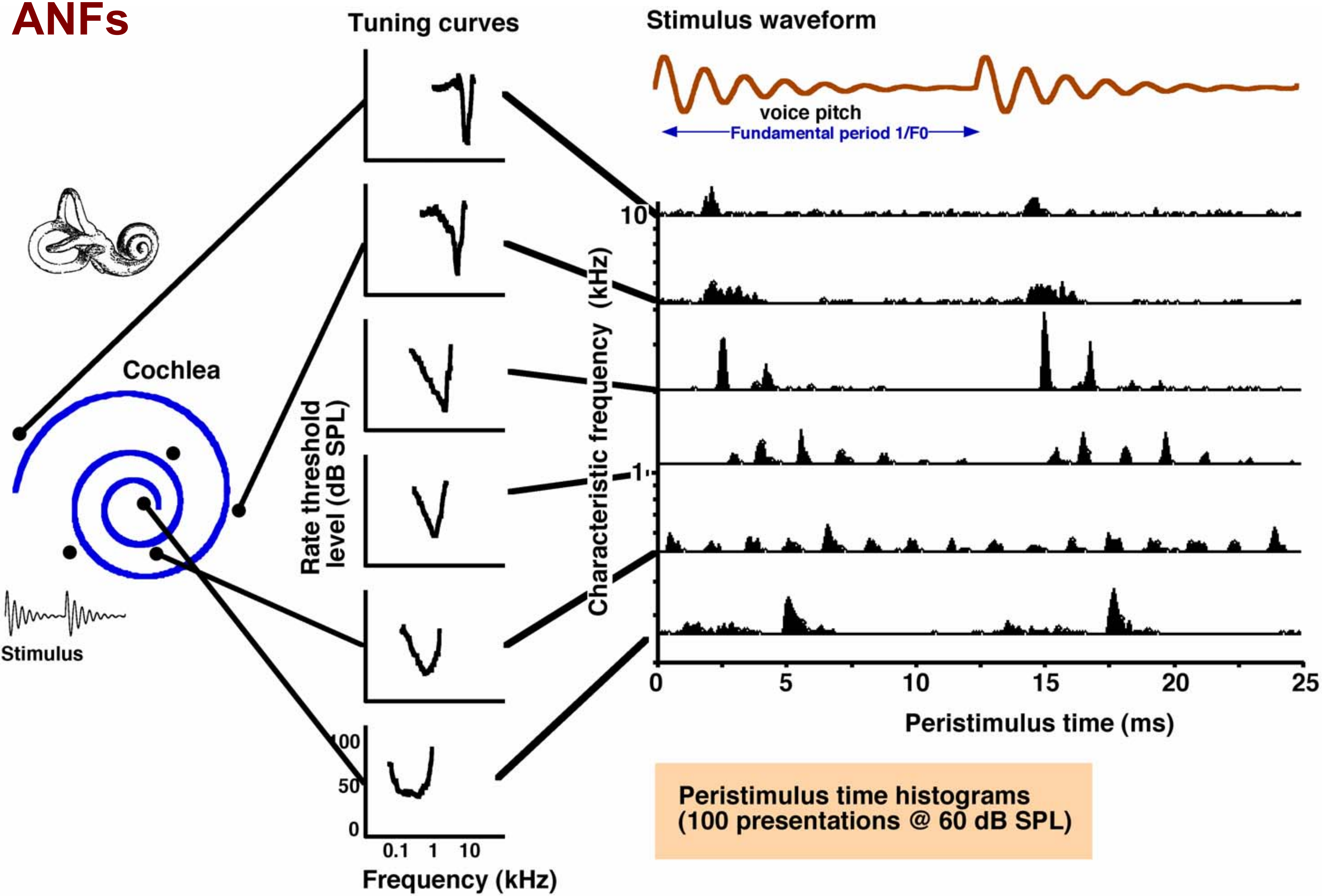
Broad tuning and rate saturation at moderate levels in low-CF auditory nerve fibers confounds rate-based resolution of harmonics.

Low SR auditory nerve fiber



Alan Palmer

ANFs



Peristimulus time histograms
(100 presentations @ 60 dB SPL)

Temporal pattern theories

- Schouten's temporal theory (1940's) depended on interactions between unresolved (high) harmonics. It was displaced by discovery of dominance region and binaural combination pitches in the 1960's. The idea persists, however in the form of spectral mechanisms for resolved harmonics and temporal ones for unresolved harmonics.

Σ **First-order intervals (renewal density)**

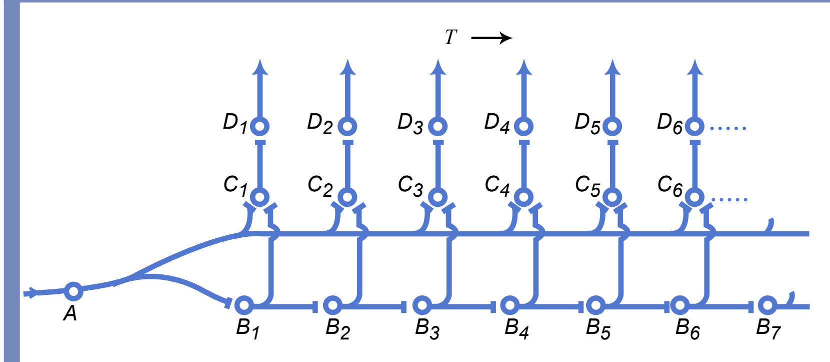
Please see van Noorden, L. Two channel pitch perception. In Music, Mind and Brain. Edited by M. Clynes. New York: Plenum.

Σ **All-order intervals (temporal autocorrelation)**

Please see Moore, BCJ. An introduction to the psychology of hearing. 5th ed. San Diego: Academic Press. 2003.

Licklider (1951)

Basic schema of neuronal autocorrelator. A is the input neuron, B_1, B_2, B_3, ... is a delay chain.



Please see Figure 1 in Meddis, R., and M. J. Hewitt. Virtual pitch and phase sensitivity of a computer model of the Auditory periphery. I. Pitch identification. J. Acoust. Soc. Am. 89 no. 6 (1991): 2866-2882.

Interval-based theories of pitch

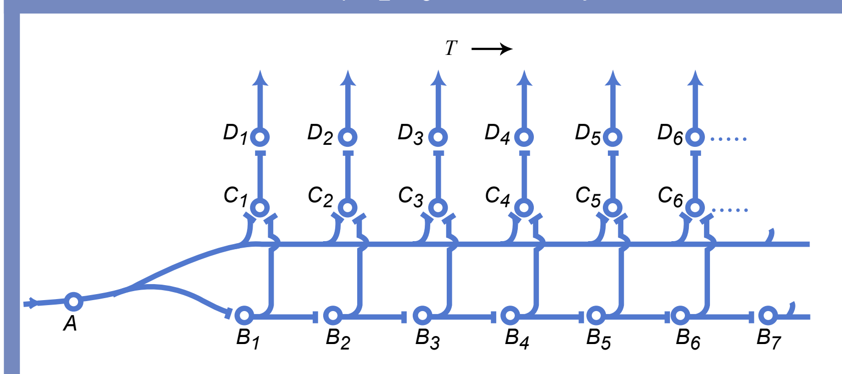
First-order intervals
(renewal density)

Please see van Noorden, L. Two channel pitch perception. In Music, Mind and Brain. Edited by M. Clynes. New York: Plenum.

Please see Moore, BCJ. An introduction to the psychology of hearing. 5th ed. San Diego: Academic Press. 2003.

All-order intervals
(temporal autocorrelation)

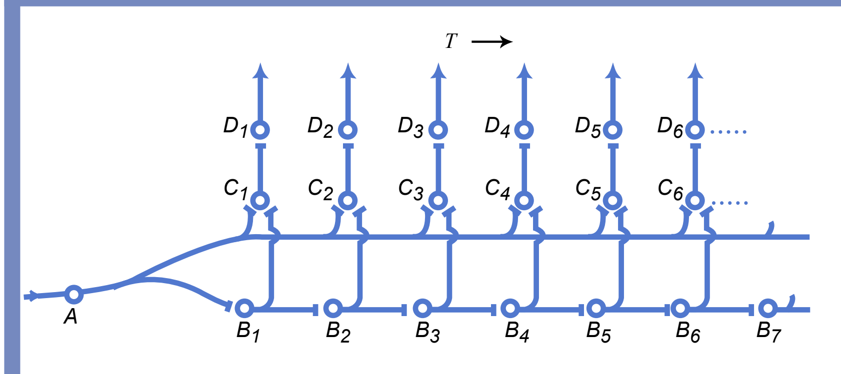
Basic schema of neuronal autocorrelator. A is the input neuron, B_1, B_2, B_3, ... is a delay chain.



Licklider (1951)

Please see Figure 1 in Meddis, R., and M. J. Hewitt. Virtual pitch and phase sensitivity of a computer model of the Auditory periphery. I. Pitch identification. J. Acoust. Soc. Am. 89 no. 6 (1991): 2866-2882.

Licklider's (1951) duplex model of pitch perception



Basic schema of neuronal autocorrelator. A is the input neuron, B_1, B_2, B_3, ... is a delay chain.

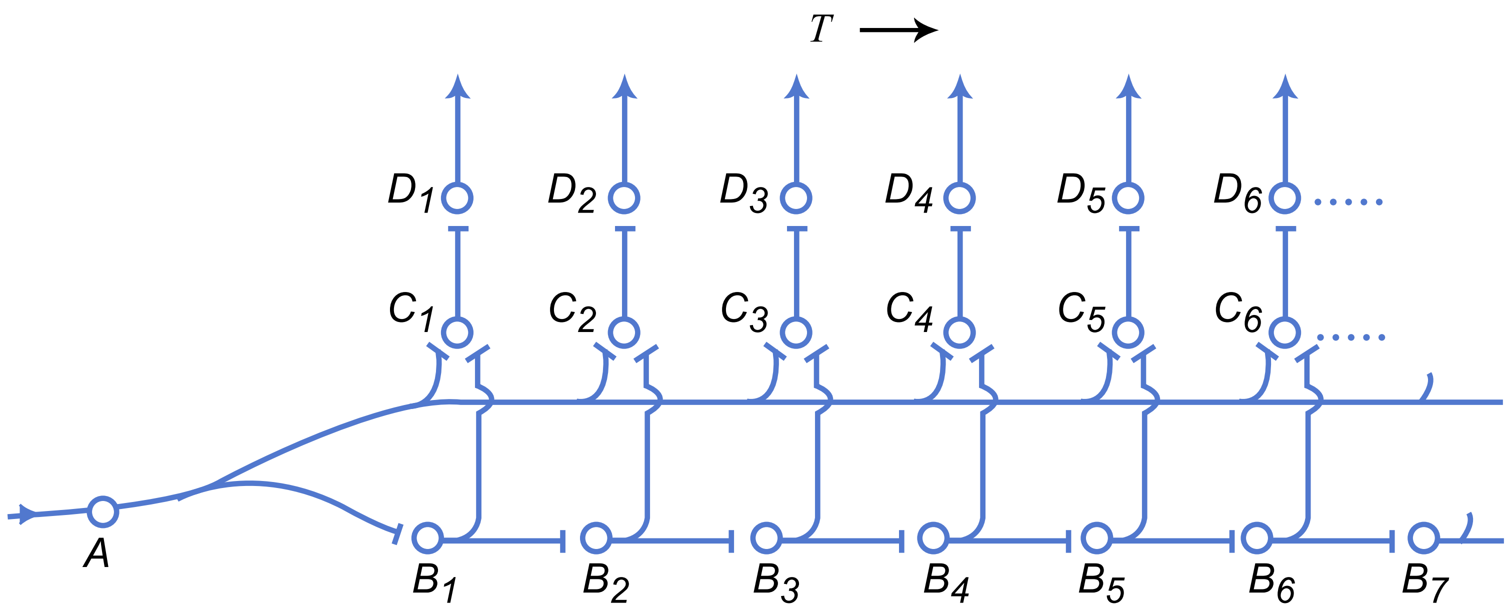
Licklider's binaural triplex model

(Image removed due to copyright considerations.)

J.C.R. Licklider "Three AuditoryTheories" In Psychology: A Study of a Science, Vol. 1, S. Koch, ed., McGraw-Hill, 1959, 41-144.

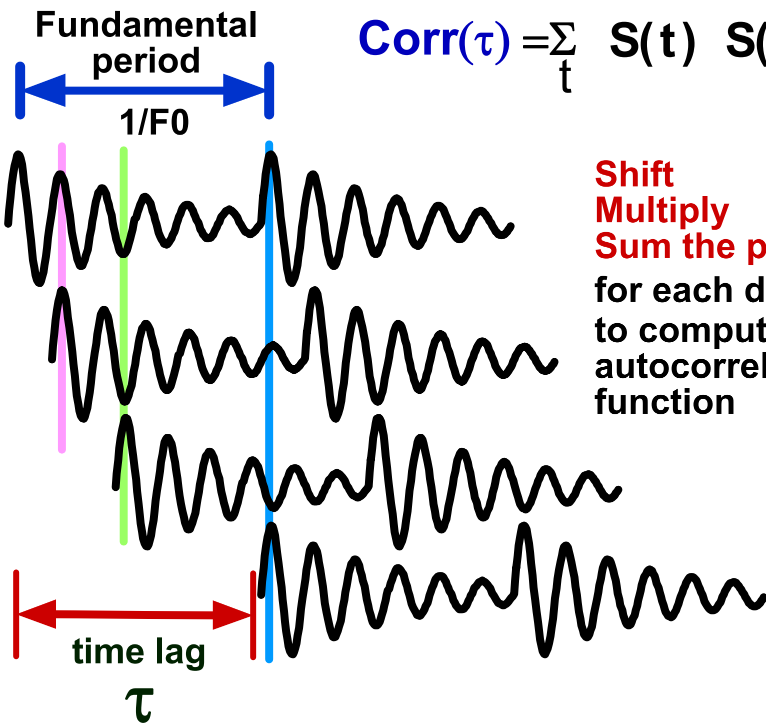
Delay lines

Basic schema of neuronal autocorrelator. A is the input neuron, B_1, B_2, B_3, ... is a delay chain.



Autocorrelation and interspike intervals

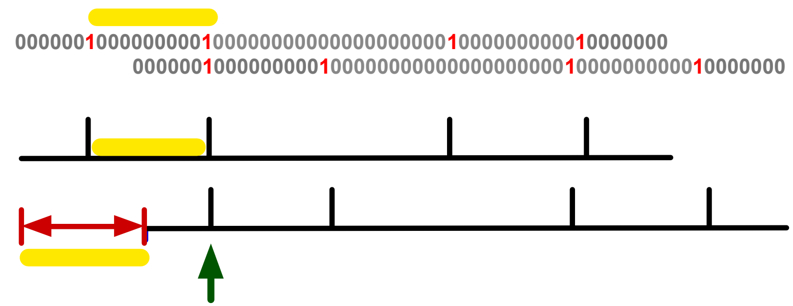
Autocorrelation functions



$$Corr(\tau) = \sum_t S(t)\ S(t - \tau)$$

Shift
Multiply
Sum the products
for each delay τ to compute autocorrelation function

Autocorrelations of spike trains = Histograms of all-order intervals



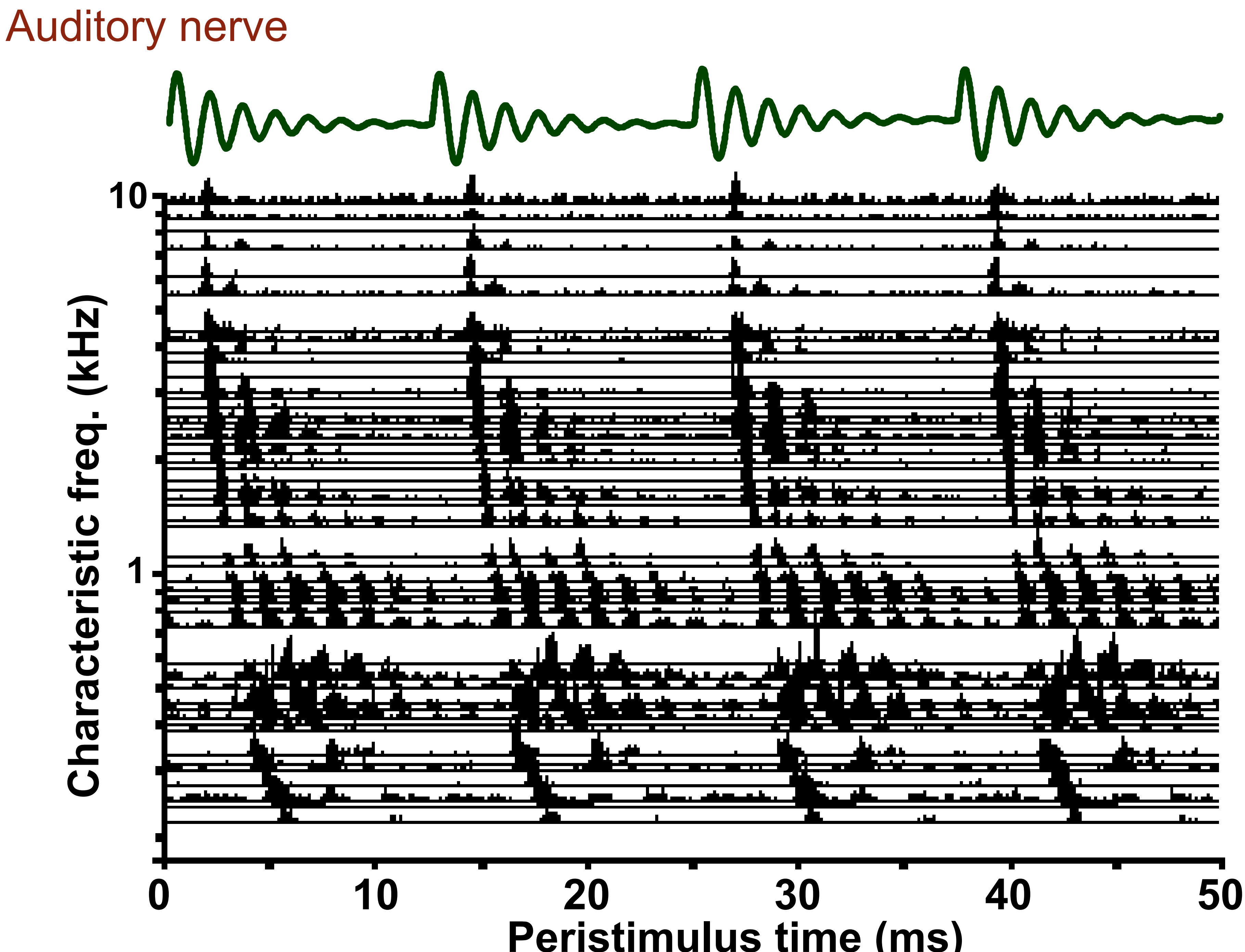
Please see Figure 6.16A-D in Lyon, Richard, and Shihab Shamma. "Auditory Representations of Timbre and Pitch." In Auditory Computation. Edited by R. R. Fay. New York: Springer Verlag. 1996.

Correlograms: interval-place displays (Slaney & Lyon)

Please see Slaney, Malcolm, and Richard F. Lyon. "On the Importance of Time -A Temporal Representation of Sound." In Visual Representations of Speech Signals. Edited by M. Crawford. New York: John Wiley. 1993.

Correlograms

Please see Figure 6.17 in Slaney, Malcolm, and Richard F. Lyon. "On the Importance of Time - A Temporal Representation of Sound." In Visual Representations of Speech Signals. Edited by M. Crawford. New York: John Wiley. 1993.


Auditory nerve
10
1
Characteristic freq. (kHz)
0
10
20
30
40
50
Peristimulus time (ms)

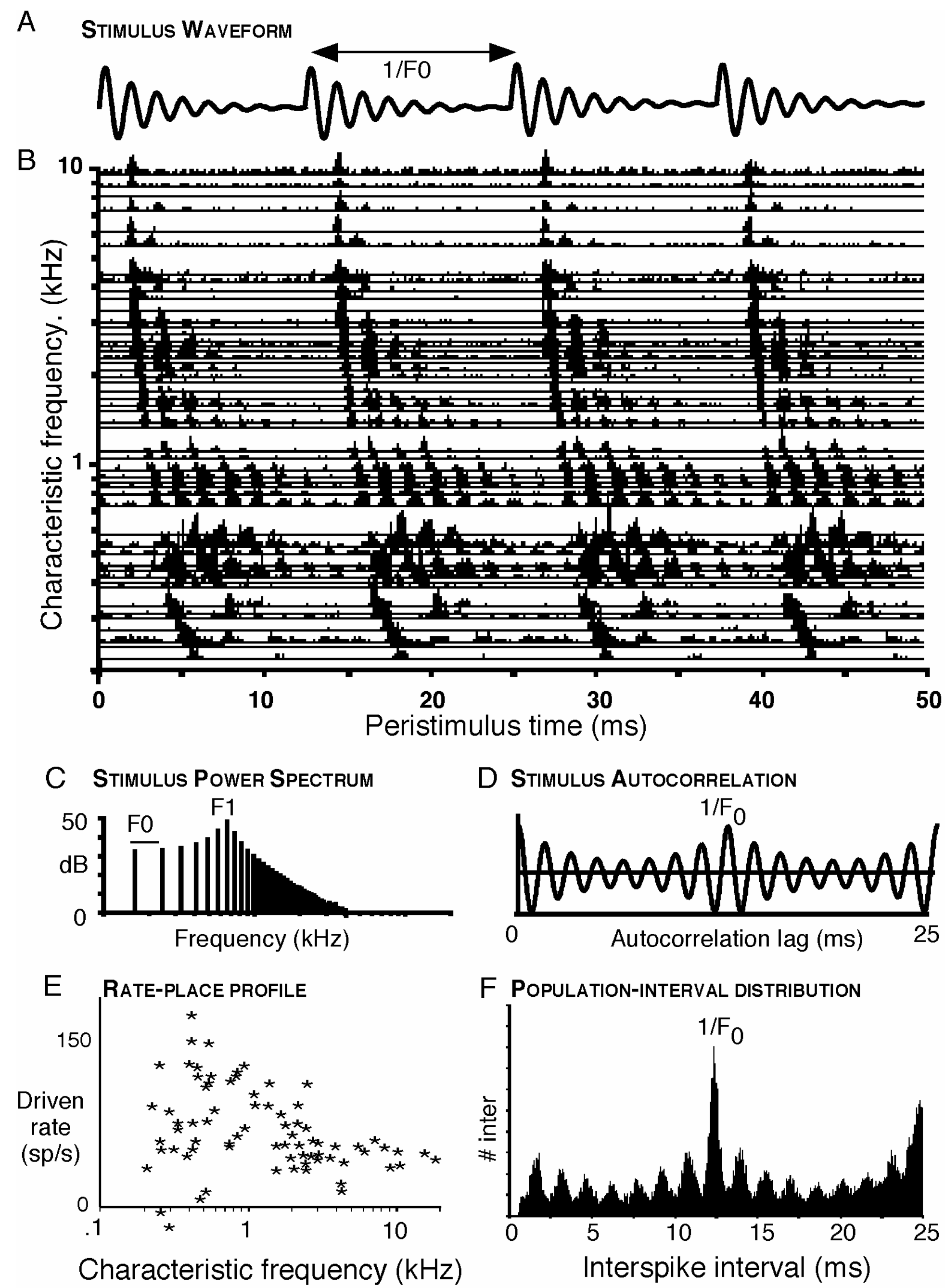
Temporal coding in the auditory nerve

Work with Bertrand Delgutte
Cariani & Delgutte (1996ab)

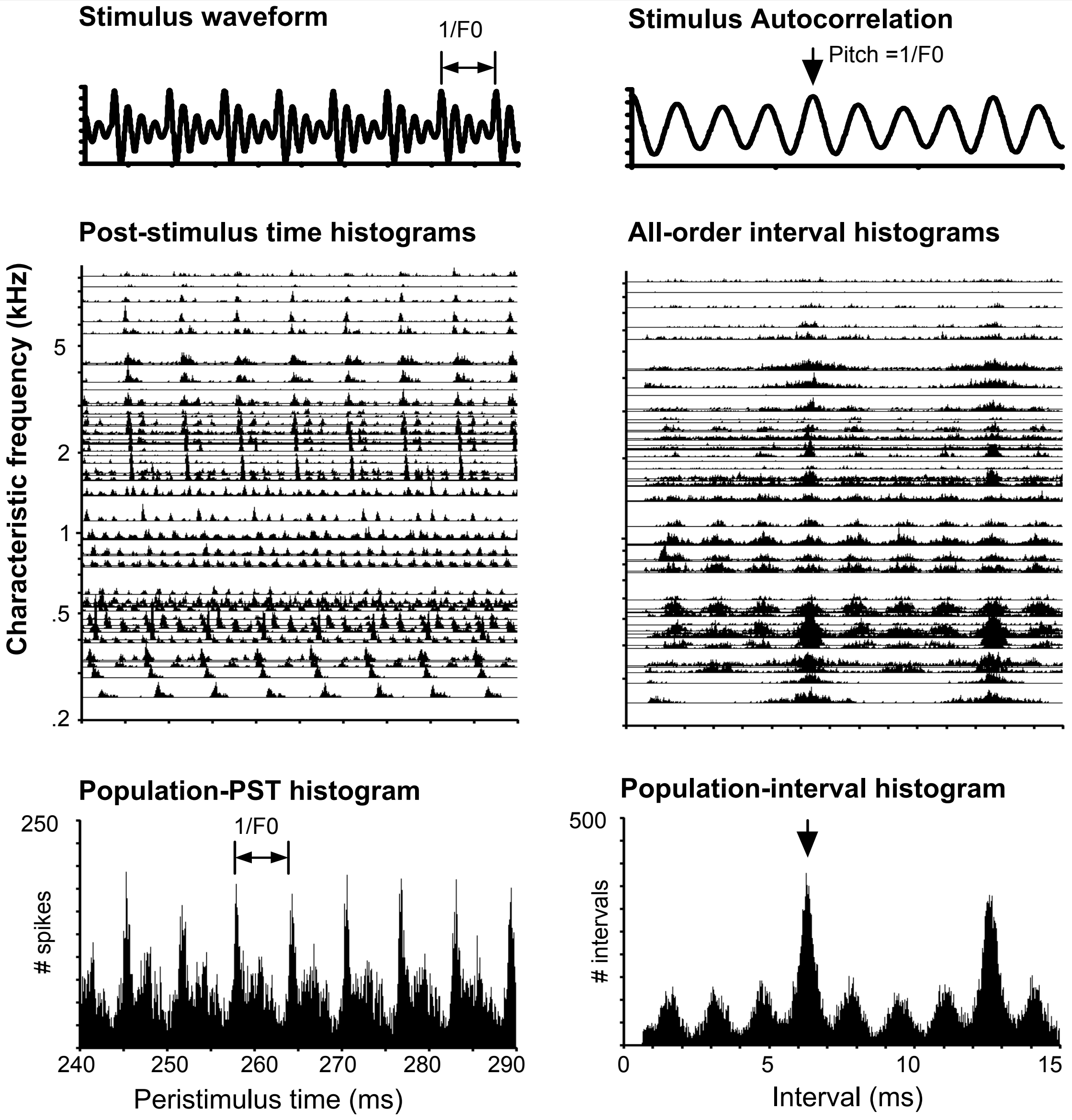
Dial-anesthetized cats.
100 presentations/fiber
60 dB SPL

Population-interval distributions are compiled by summing together intervals from all auditory nerve fibers.

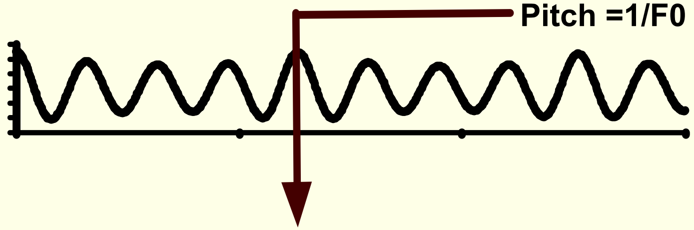
The most common intervals present in the auditory nerve are invariably related to the pitches heard at the fundamentals of harmonic complexes.



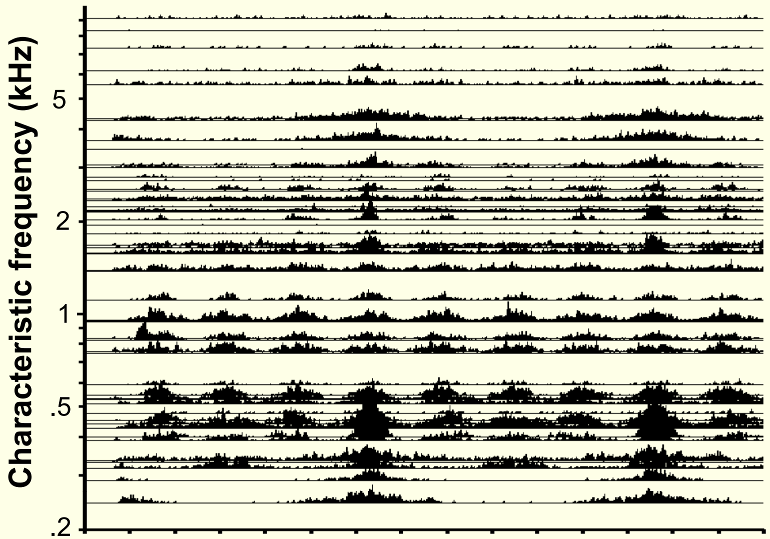
The population-interval distribution of the auditory nerve



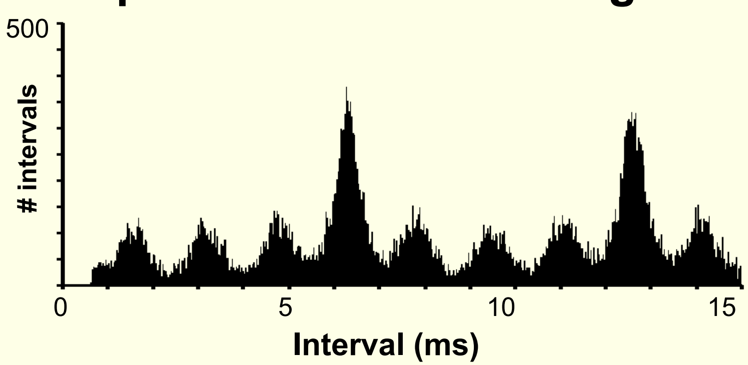
Stimulus Autocorrelation



All-order interval histograms

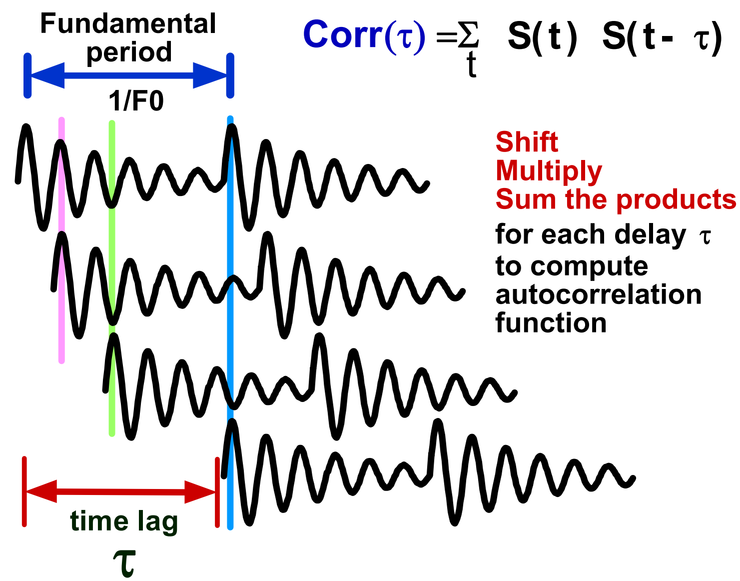


Population-interval histogram

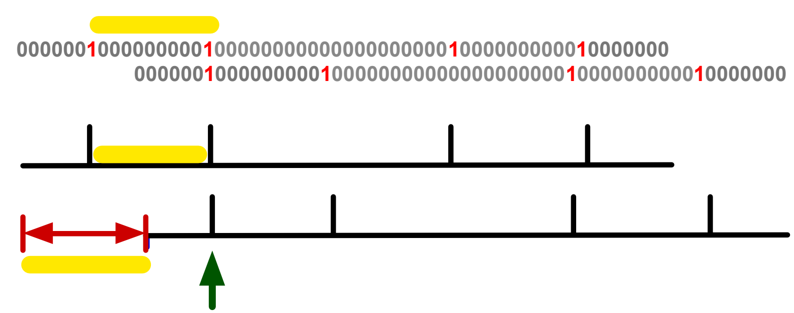


Autocorrelation functions

$$\mathrm{Corr}(\tau) = \sum_t S(t)\ S(t-\tau)$$

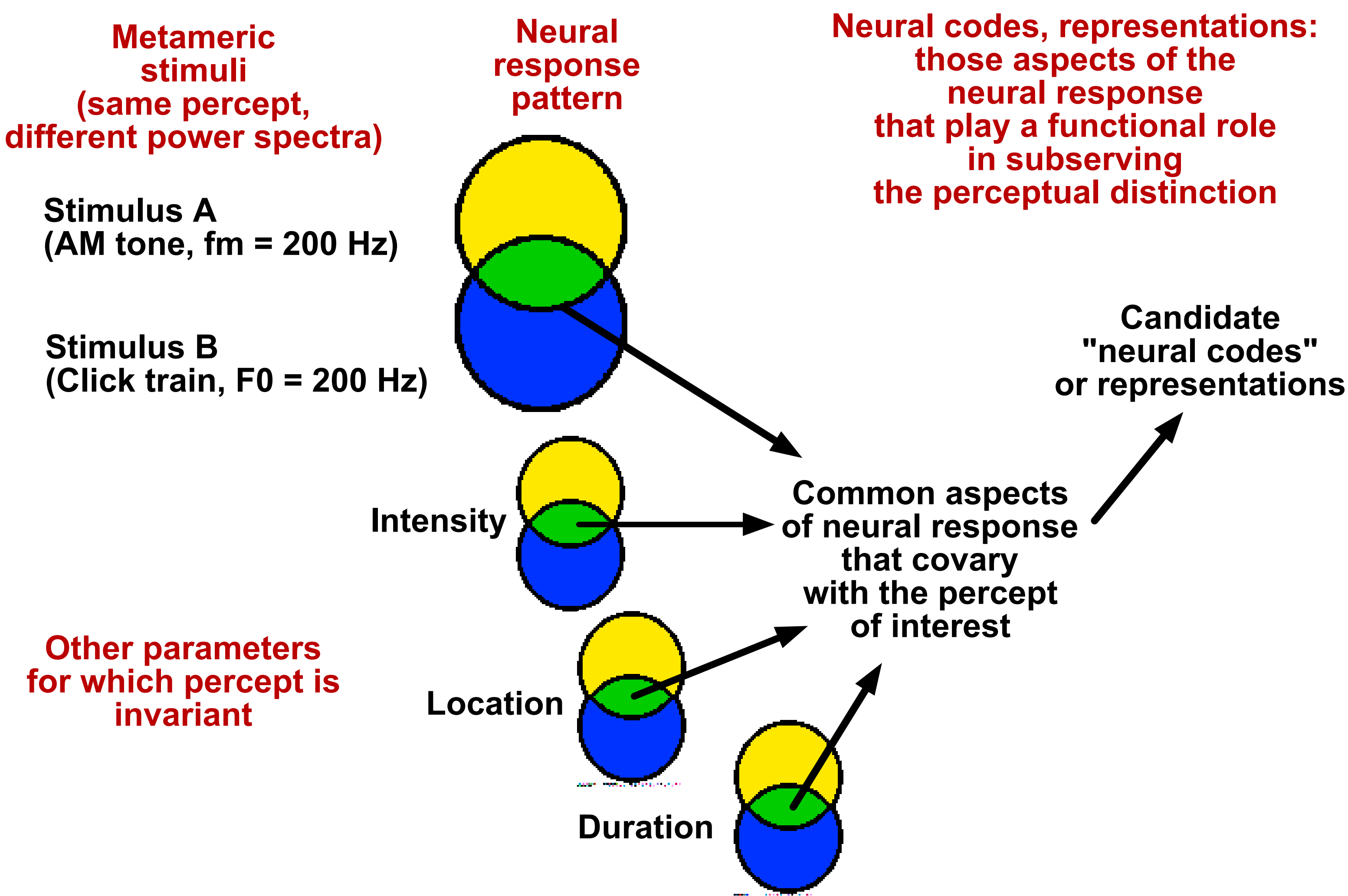


Autocorrelations of spike trains = Histograms of all-order intervals



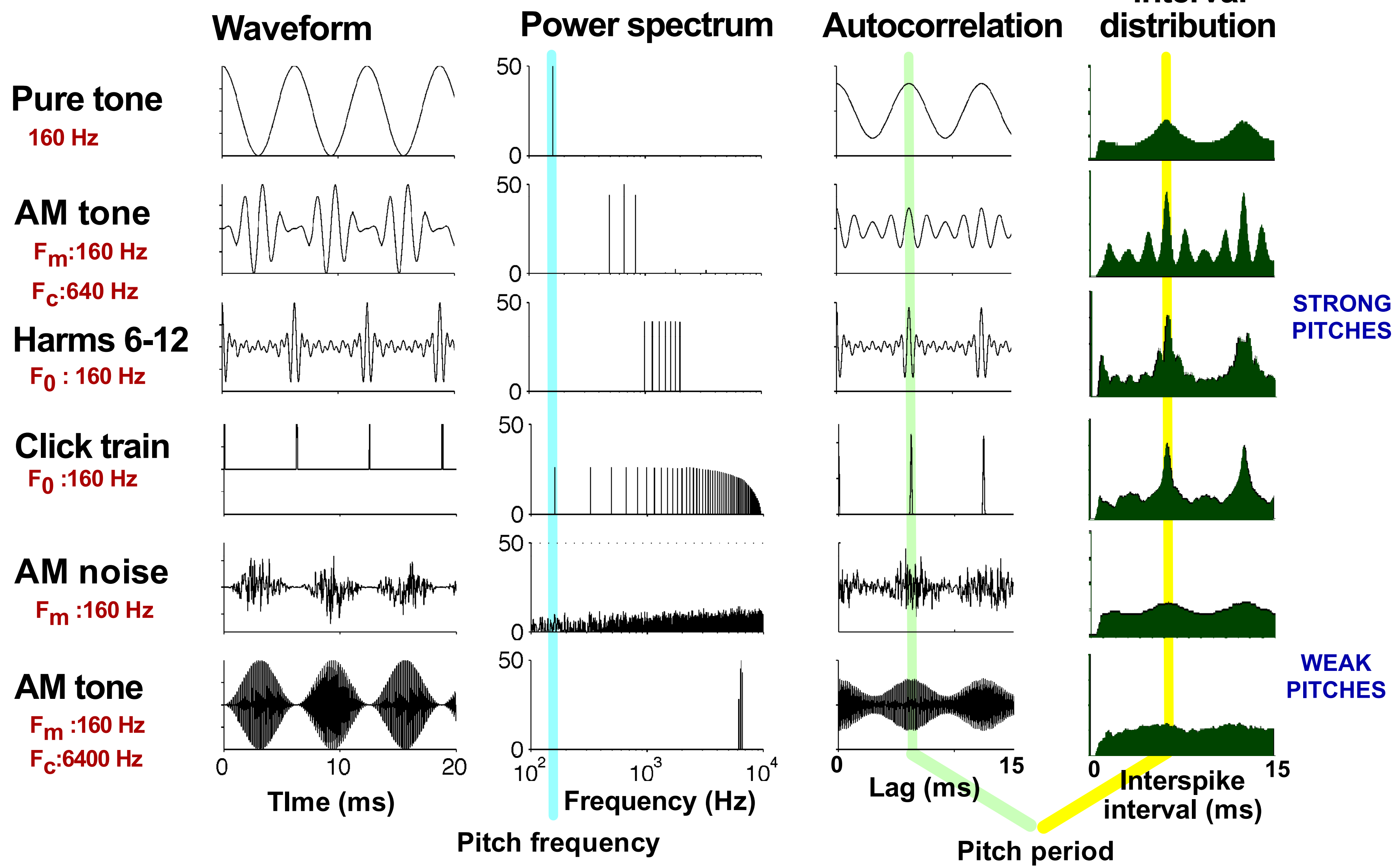
Percept-driven search for neural codes:

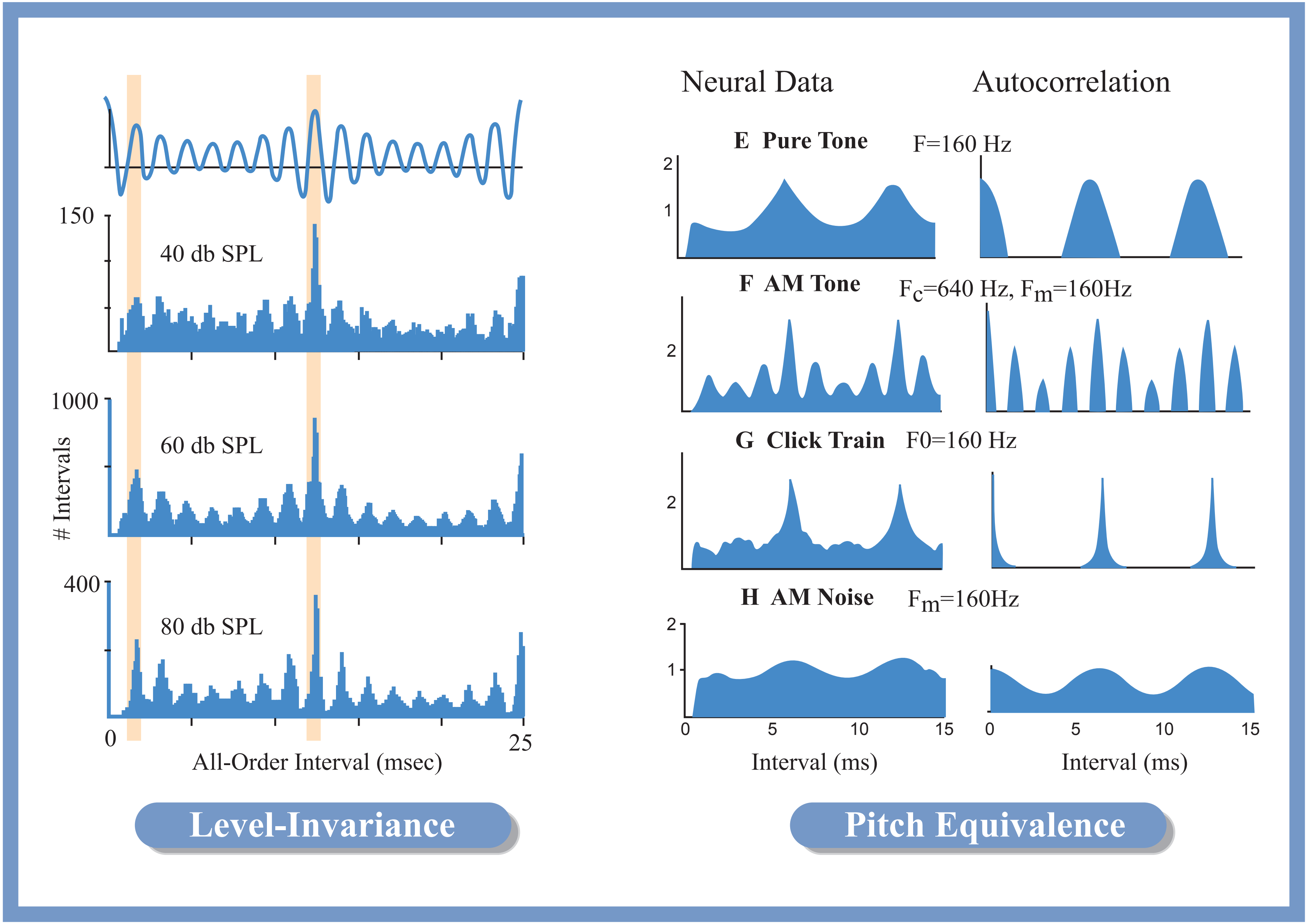
1. Use stimuli that produce equivalent percepts
2. Look for commonalities in neural response
3. Eliminate those aspects that are not invariant



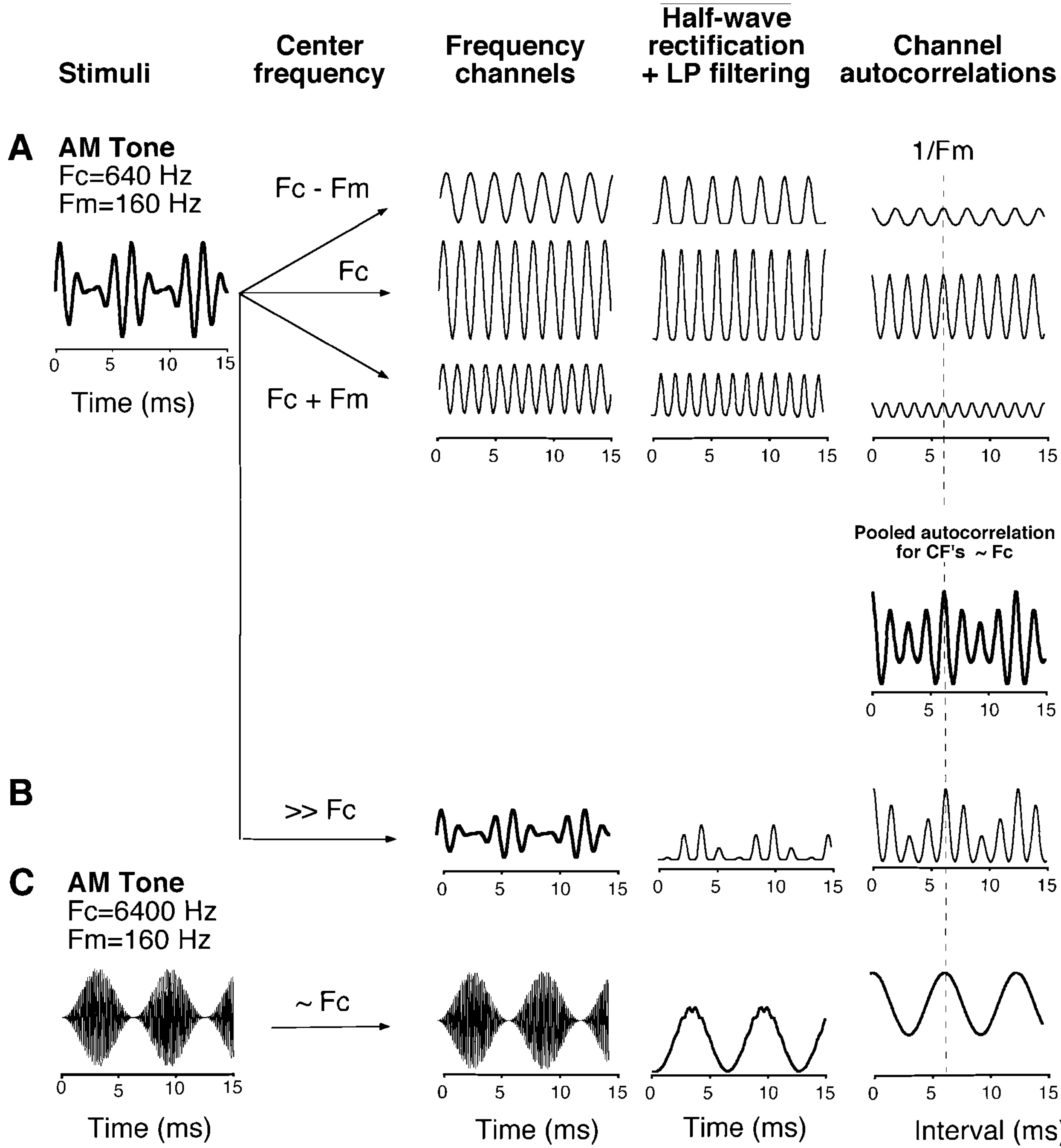
Pitch equivalence

Six stimuli that produce a low pitch at 160 Hz



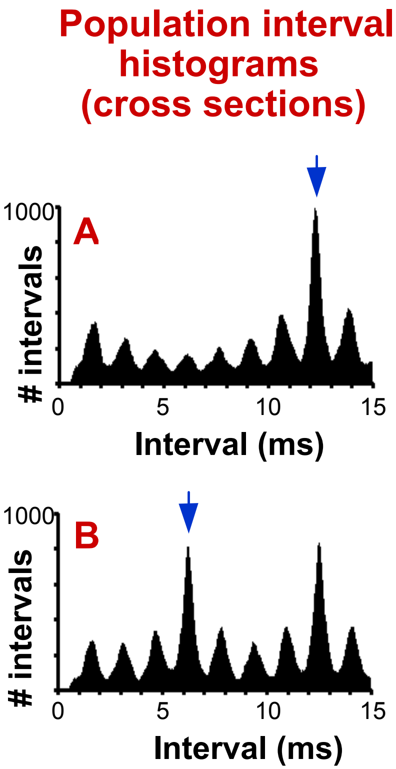


150
40 db SPL
1000
60 db SPL
Intervals
400
80 db SPL
0
25
All-Order Interval (msec)
Level-Invariance
Neural Data
Autocorrelation
E Pure Tone
F=160 Hz
F AM Tone
Fc=640 Hz, Fm=160Hz
G Click Train
F0=160 Hz
H AM Noise
Fm=160Hz
Interval (ms)
Interval (ms)
Pitch Equivalence



Stimuli
Center frequency
Frequency channels
Half-wave rectification + LP filtering
Channel autocorrelations
A
AM Tone
Fc=640 Hz
Fm=160 Hz
Time (ms)
Fc - Fm
Fc
Fc + Fm
1/Fm
Pooled autocorrelation for CF's ~ Fc
B
>> Fc
C
AM Tone
Fc=6400 Hz
Fm=160 Hz
~ Fc
Time (ms)
Time (ms)
Time (ms)
Interval (ms)

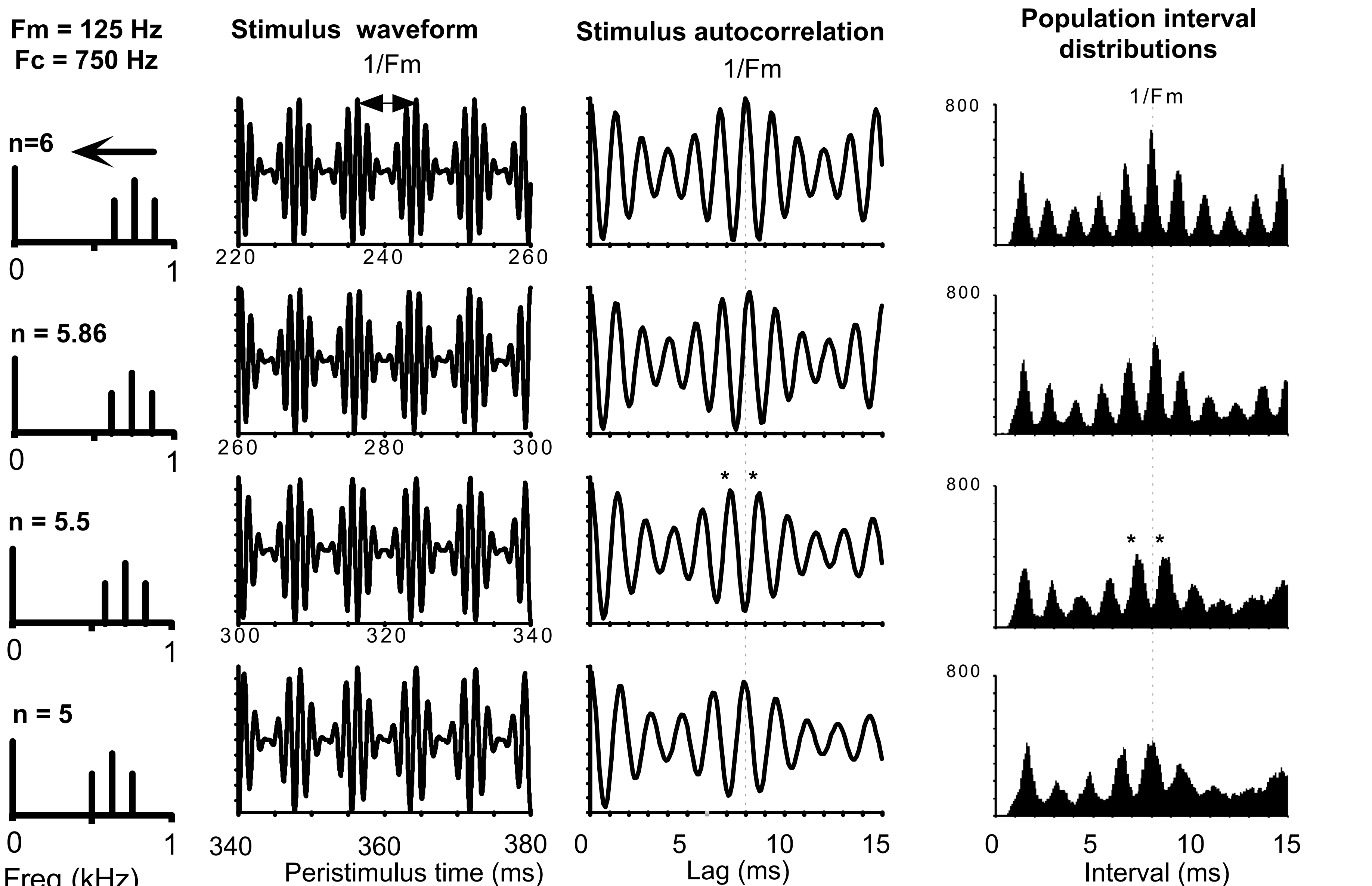
The running population-interval distribution



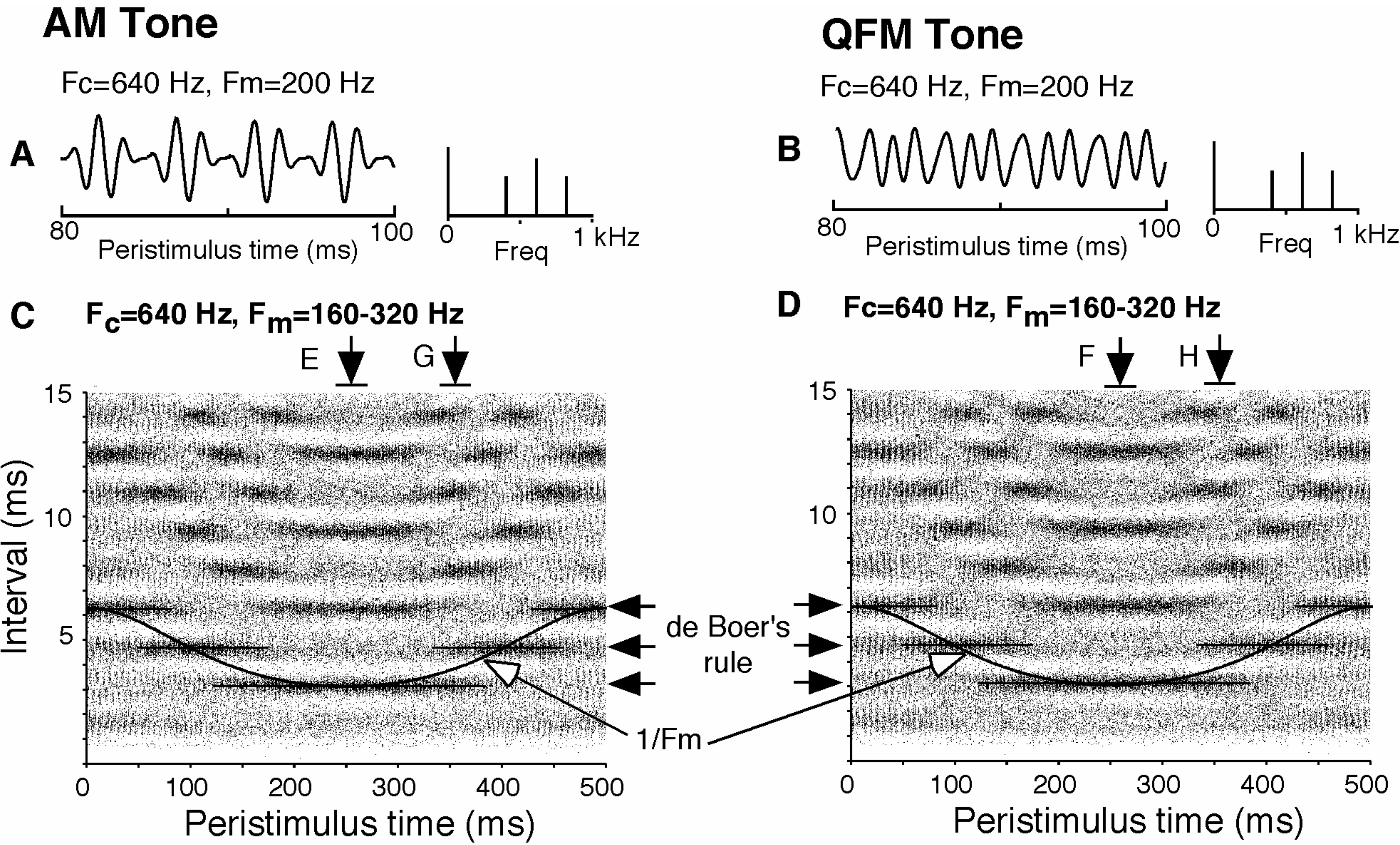
Pitch height and pitch chroma

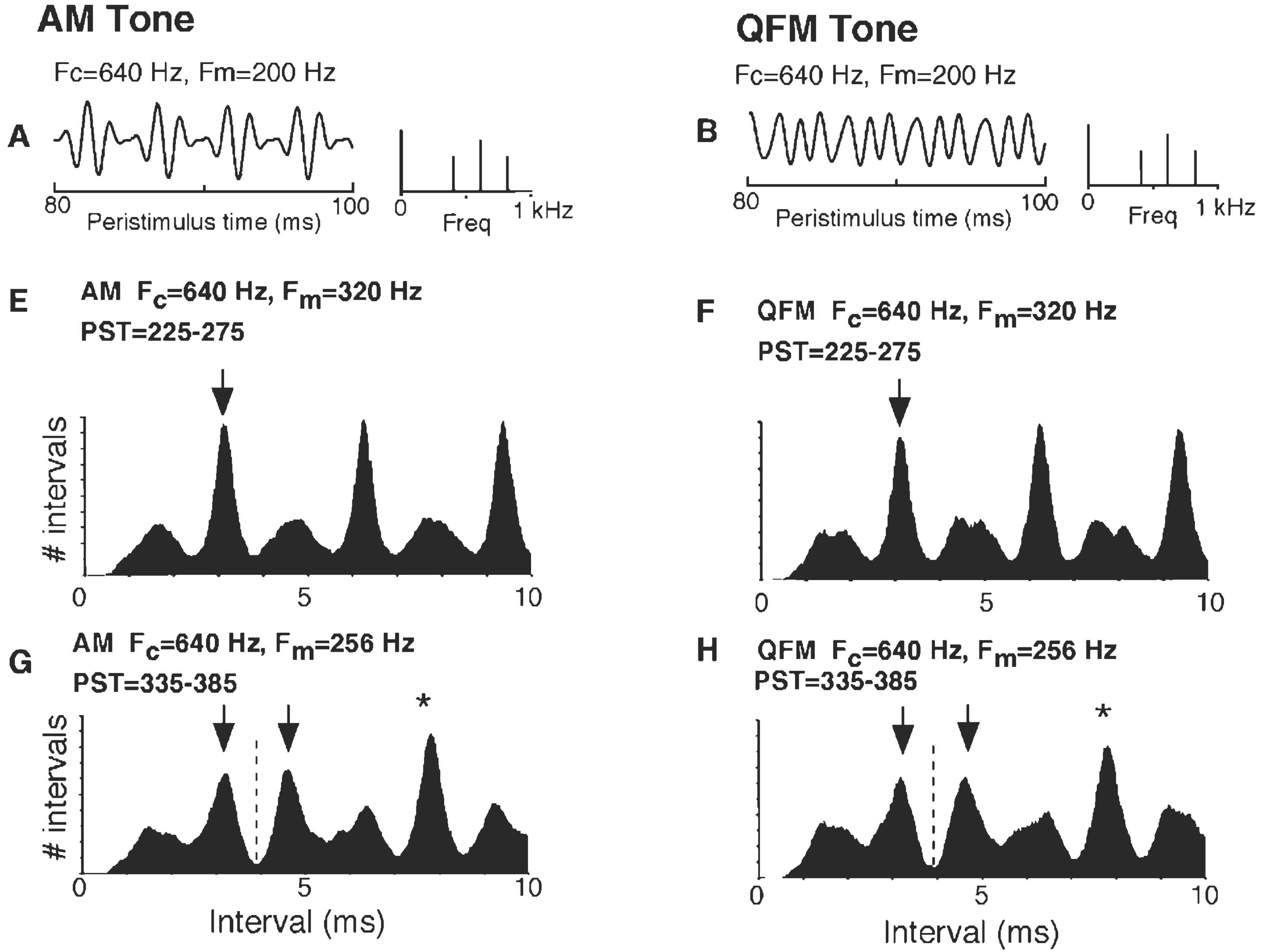
Please see Figure 1, 2, and 7 in Roger N. Shepard. "Geometrical approximations to the structure of musical pitch." *Psychological Review* 89 no. 4 (1982): 305-322.

Pitch shift of inharmonic complex tones



Pitch shift of inharmonic complex tones
Phase-invariant nature of all-order interval code

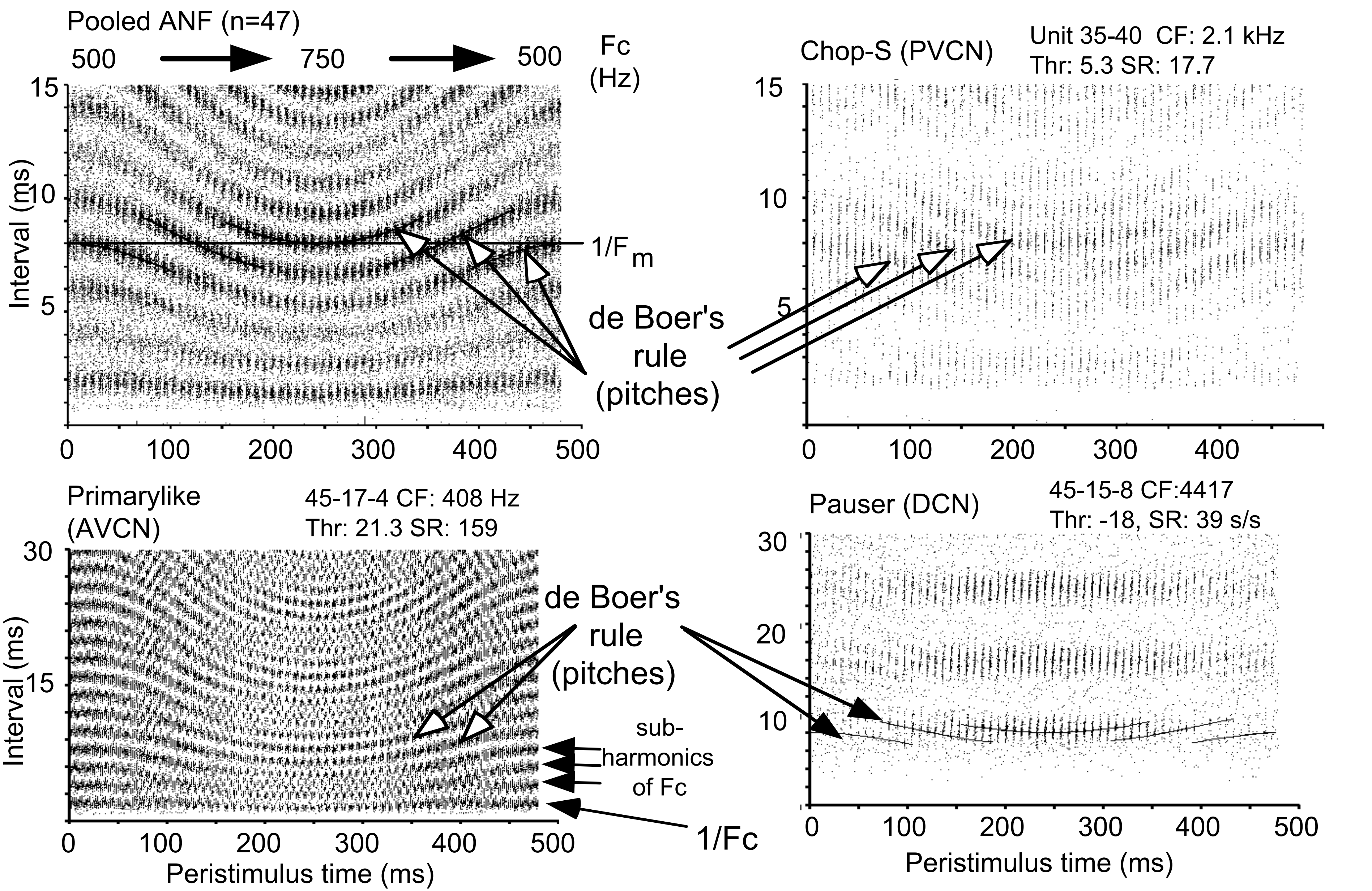




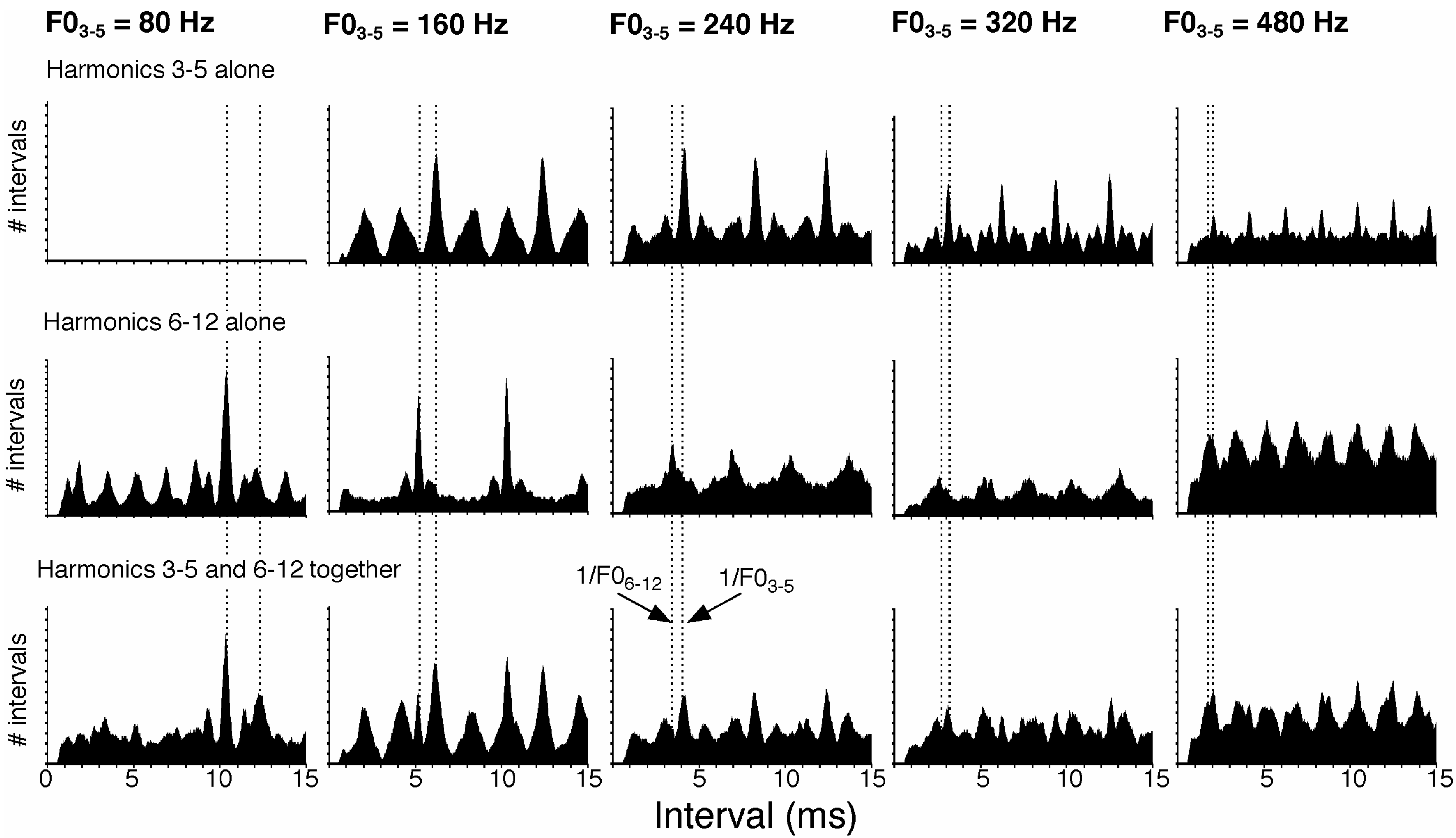
AM Tone
Fc=640 Hz, Fm=200 Hz
A
80
100
Peristimulus time (ms)
0
Freq
1 kHz
QFM Tone
Fc=640 Hz, Fm=200 Hz
B
80
100
Peristimulus time (ms)
0
Freq
1 kHz
E
AM Fc=640 Hz, Fm=320 Hz
PST=225-275
# intervals
0
5
10
F
QFM Fc=640 Hz, Fm=320 Hz
PST=225-275
0
5
10
G
AM Fc=640 Hz, Fm=256 Hz
PST=335-385
*
intervals
0
5
10
Interval (ms)
H
QFM Fc=640 Hz, Fm=256 Hz
PST=335-385
*
0
5
10
Interval (ms)

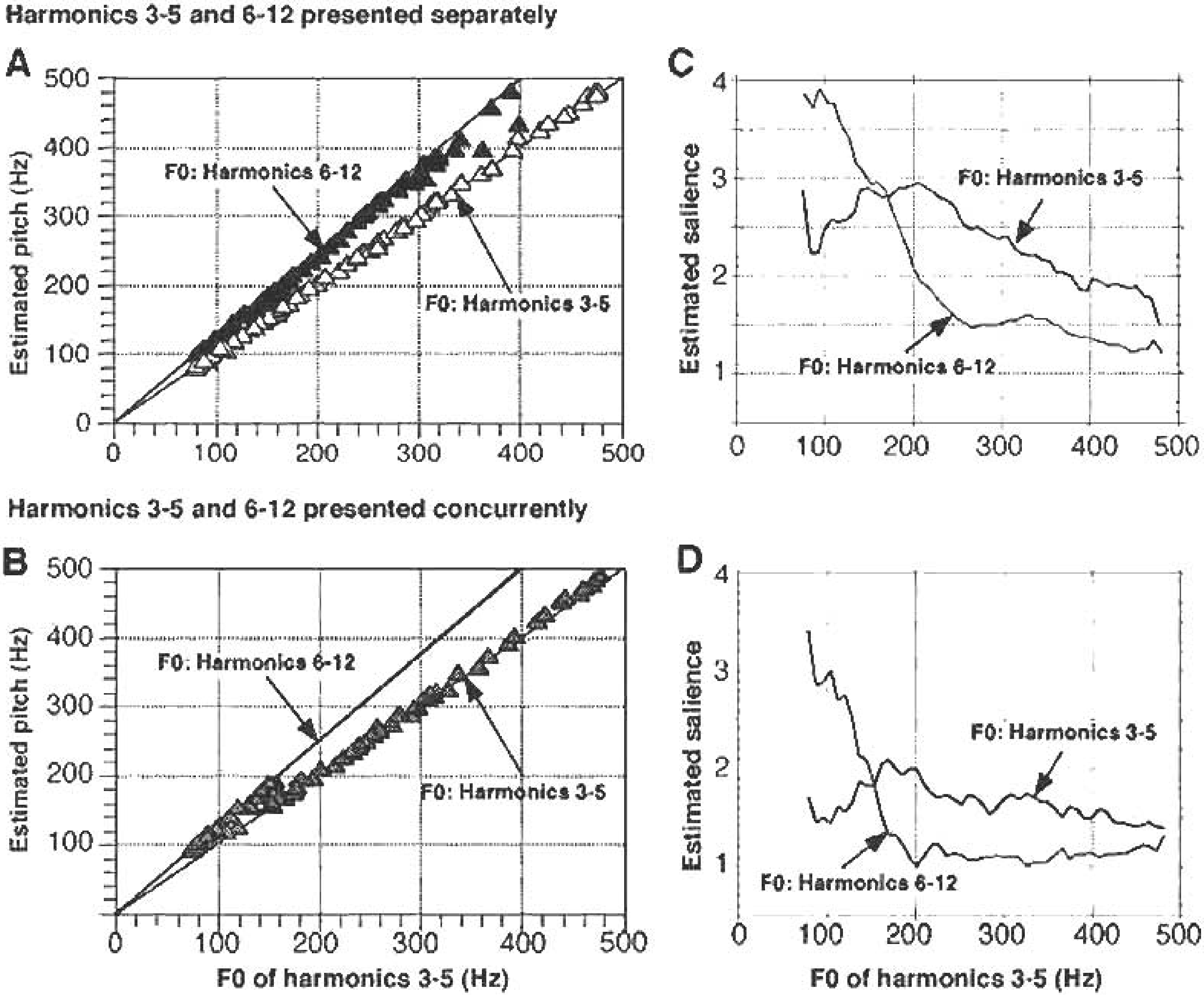
Cochlear nucleus IV: Pitch shift

Variable-Fc AM tone F_m = 125 Hz F_c = 500-750 Hz Pitch ~ de Boer's rule



Dominance region for pitch (harmonics 3-5 or partials 500-1500 Hz)

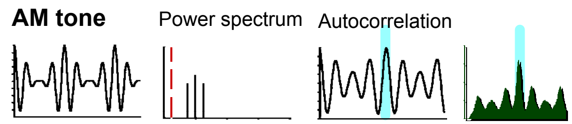



Harmonics 3-5 and 6-12 presented separately
A
C
F0: Harmonics 6-12
F0: Harmonics 3-5
Estimated pitch (Hz)
Estimated salience
Harmonics 3-5 and 6-12 presented concurrently
B
D
F0 of harmonics 3-5 (Hz)
0
100
200
300
400
500
1
2
3
4

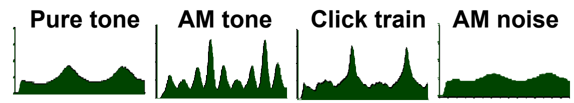
Summary

Population-interval representation of pitch at the level of the auditory nerve

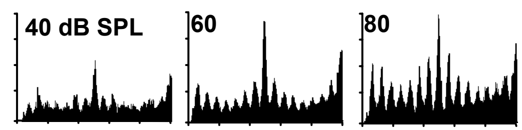
Pitch of the "missing fundamental"



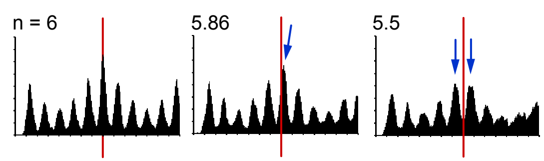
Pitch Equivalence



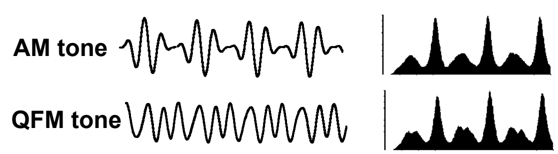
Level invariance



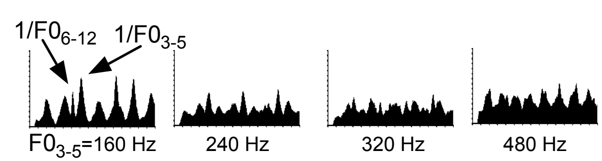
Pitch shift of inharmonic AM tones



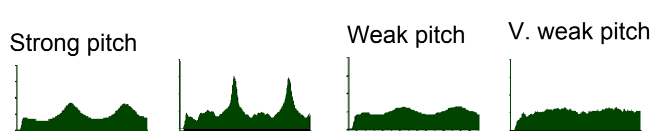
Phase invariance



Dominance region



Pitch salience



Temporal theories - pros & cons

Make use of spike-timing properties of elements in early processing (to midbrain at least)

Interval-information is precise & robust & level-insensitive

No strong neurally-grounded theory of how this information is used

Unified model: account for pitches of perceptually-resolved & unresolved harmonics in an elegant way (dominant periodicity)

Explain well existence region for F0 (albeit with limits on max interval durations)

Do not explain low pitches of unresolved harmonics

Interval analyzers require precise delays & short coincidence windows

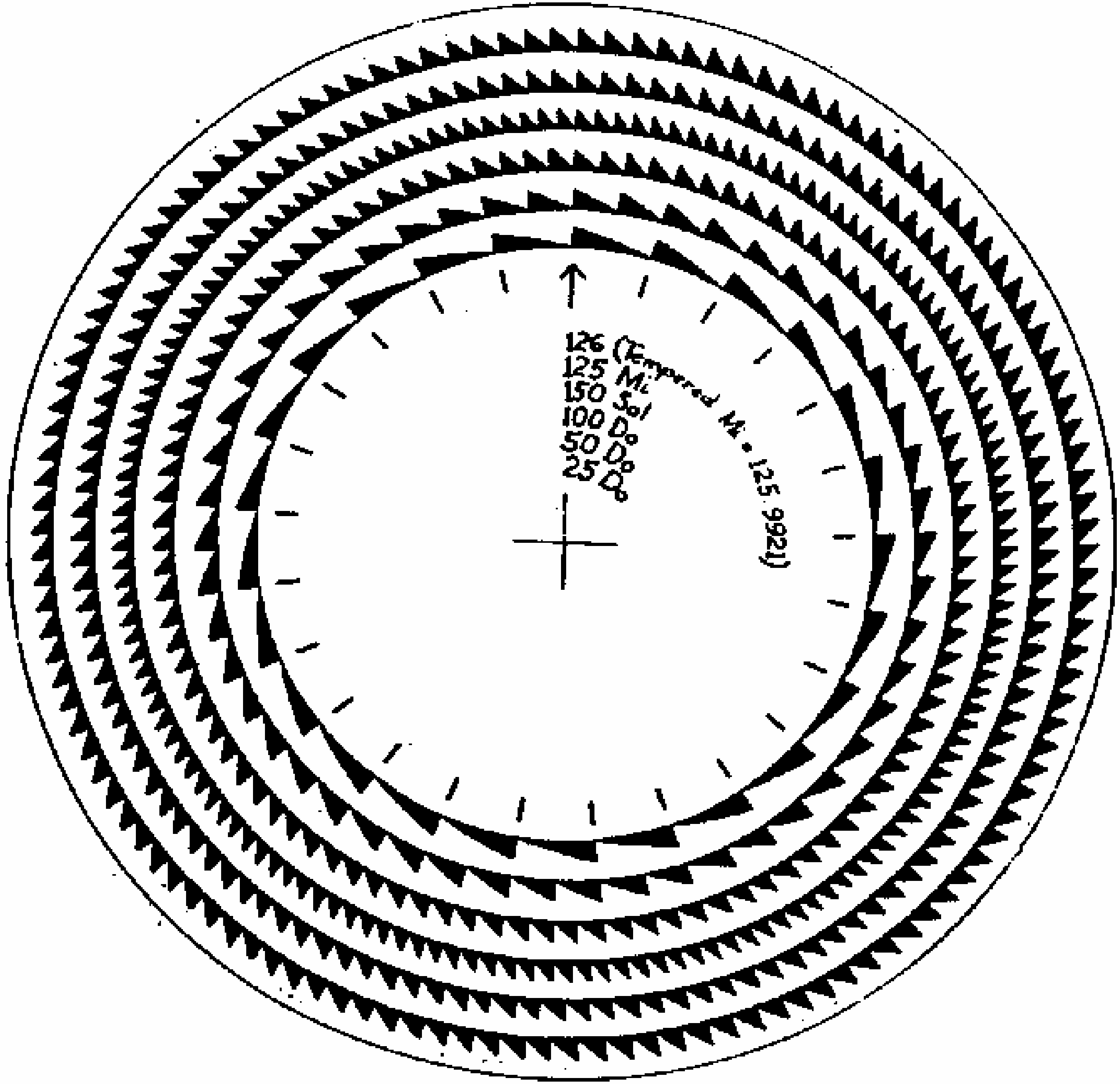
Little or no neural evidence for required analyzers

Pitch classes and perceptual similarity

Associationist theory:
Build up harmonic associations from repeated exposure to harmonic complex tones

Innate mechanisms:
Harmonic similarity relations are direct consequences of the inherent structure of interval codes and the neurocomputational mechanisms used for their analysis

A Harmony wheel of Boomsliter and Creel (1962)



Temporal coding and music: a few references

Boomsliter P, Creel W (1962) The long pattern hypothesis in harmony and hearing. J Music Theory 5:2-31.

Cariani P (2002) Temporal codes, timing nets, and music perception. J New Music Res 30:107-136.

McKinney MF, Delgutte B (1999) A possible neurophysiological basis of the octave enlargement effect. J Acoust Soc Am 106:2679-2692.

McKinney MF, Tramo MJ, Delgutte B (2001) Neural correlates of musical dissonance in the inferior colliculus. In: Physiological and Psychophysical Bases of Auditory Function (Houtsma AJM, ed), pp 71-77. Maastricht: Shaker Publishing.

Ohgushi K (1978) On the role of spatial and temporal cues in the perception of the pitch of complex tones. J Acoust Soc Am 64:764-771.

Ohgushi K (1983) The origin of tonality and a possible explanation for the octave enlargement phenomenon. J Acoust Soc Am 73:1694-1700.

Patterson RD (1986) Spiral detection of periodicity and the spiral form of musical scales. Psychology of Music 14:44-61.

Rose JE (1980) Neural correlates of some psychoacoustical experiences. In: Neural Mechanisms of Behavior (McFadden D, ed), pp 1-33. New York: Springer Verlag.

Tramo MJ, Cariani PA, Delgutte B, Braida LD (2001) Neurobiological foundations for the theory of harmony in western tonal music. Ann N Y Acad Sci 930:92-116.

INTERVAL DISTRIBUTIONS AND OCTAVE SIMILARITY

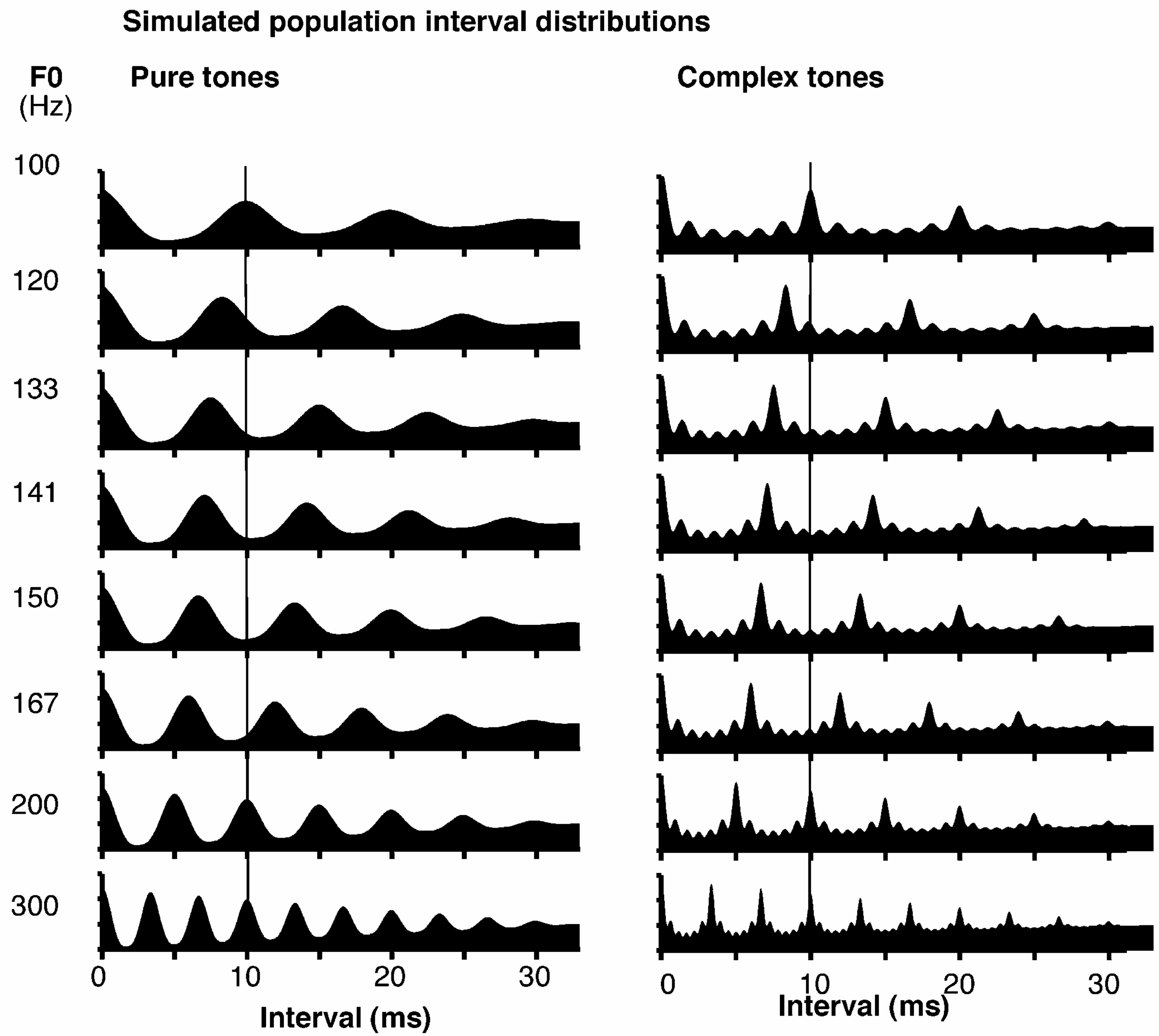


Figure 4. Similarities between population-interval representations associated with different fundamental frequencies. Simulated population-interval distributions for pure tones (left) and complex tones (right) consisting of harmonics 1-6.

Octave similarity

Please see Cariani, Peter. *Journal of New Music Research* 30, no. 2 (2001): 107-135.

Pitch height and pitch chroma

Please see Figure 1, 2, and 7 in Roger N. Shepard. "Geometrical Approximations to the Structure of Musical Pitch." *Psychological Review* 89, no. 4 (1982): 305-322.

Please see Cariani, Peter. *Journal of New Music Research* 30, no. 2 (2001): 107-135.

Musical tonal relations

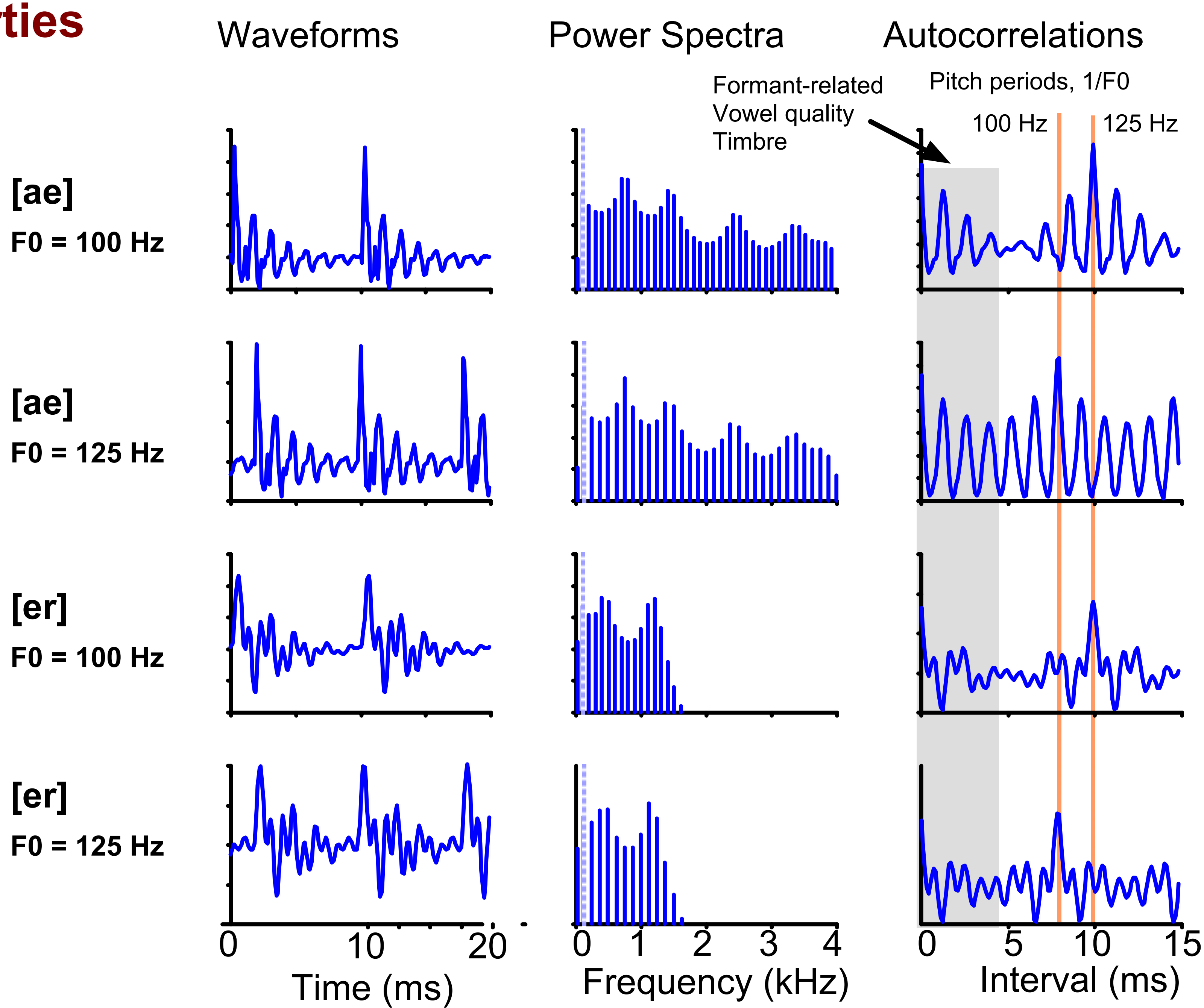
Please see Shepard, Roger N. "Geometrical Approximations to the Structure of Musical Pitch." *Psychological Review* 89, no. 4 (1982): 305-322.

Please see Cariani, Peter. *Journal of New Music Research* 30, no. 2 (2001): 107-135.

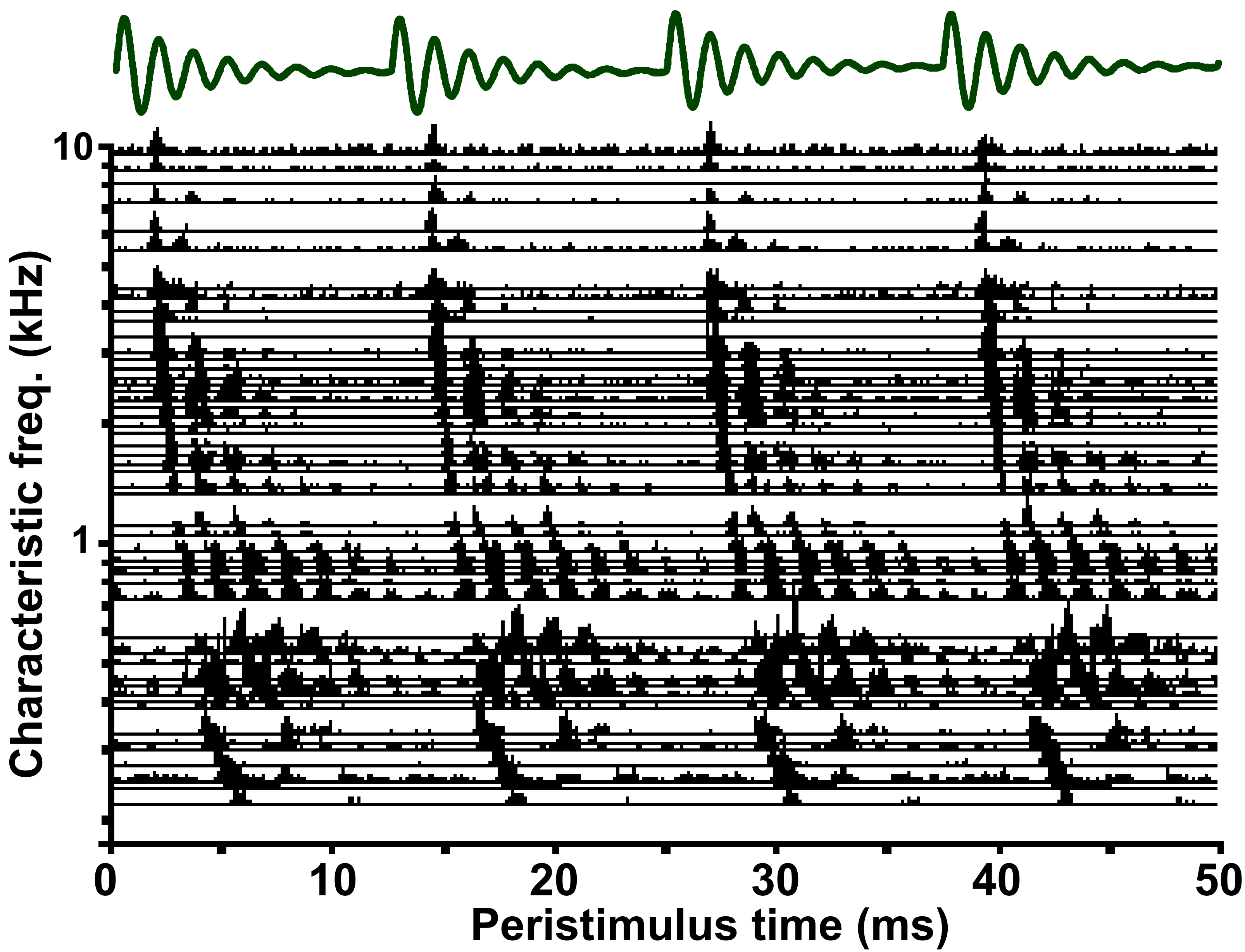
Note-chord relations
(Harms 1-12)

Please see Cariani, Peter. *Journal of New Music Research* 30, no. 2 (2001): 107-135.

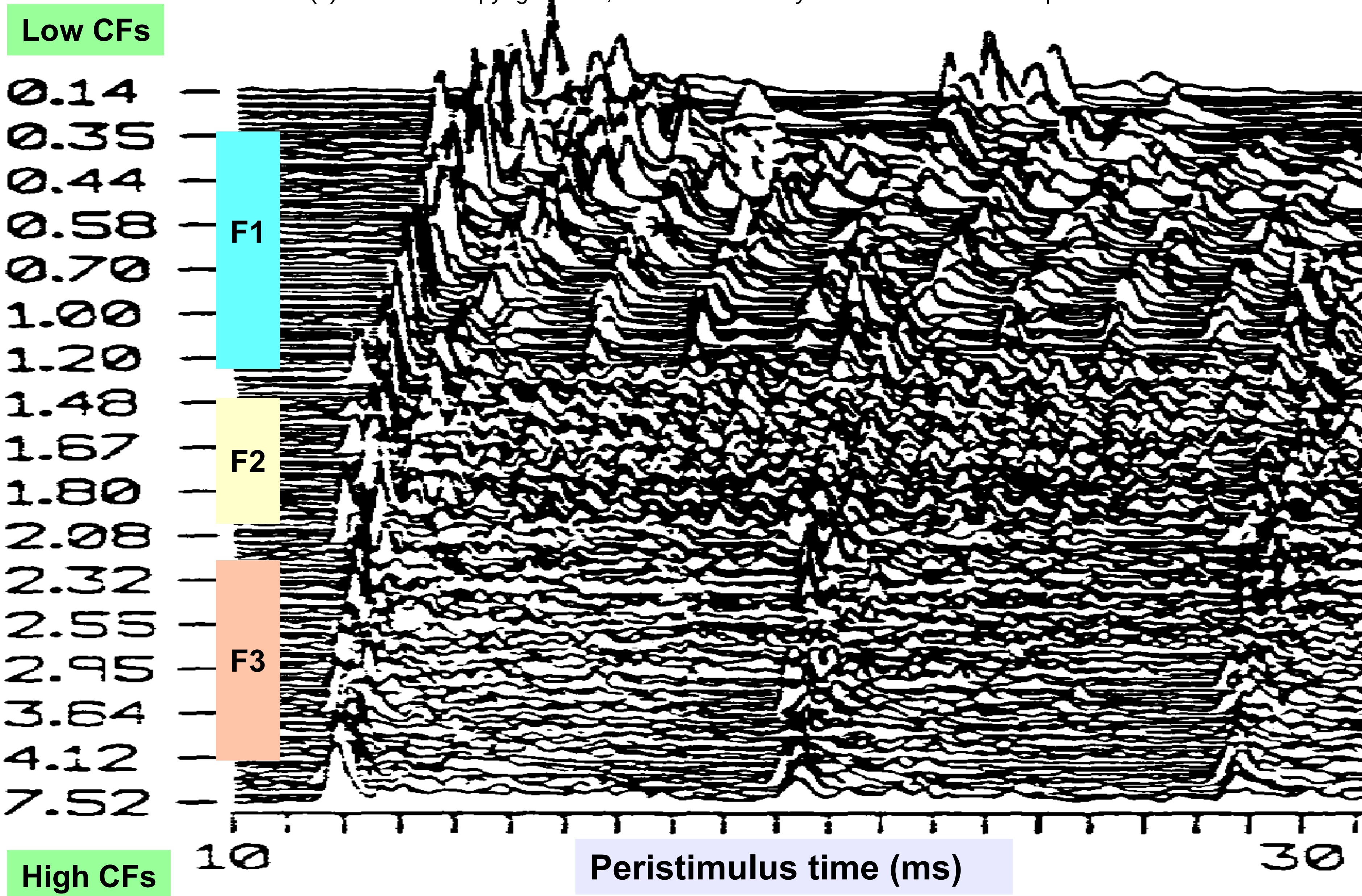
Vowel properties



Phase-locking of discharges in the auditory nerve

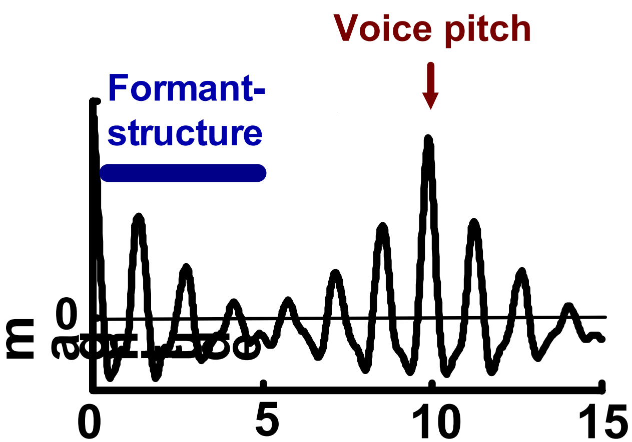


Time domain analysis of auditory-nerve fiber firing rates.
Hugh Secker-Walker & Campbell Searle, J. Acoust. Soc. 88(3), 1990
Neural responses to /da/ @ 69 dB SPL from Miller and Sachs (1983)

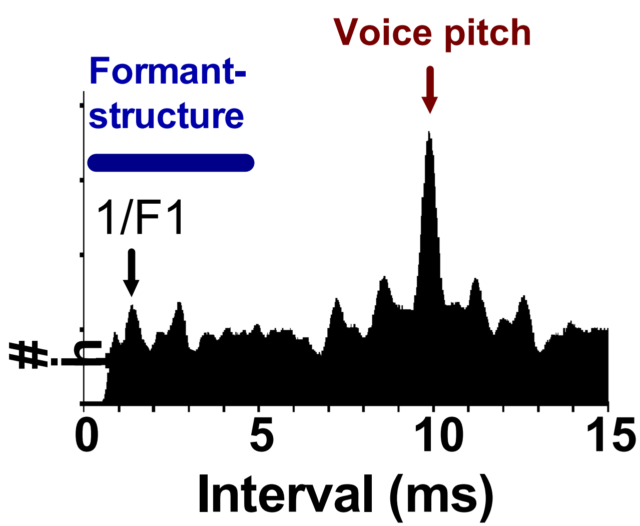


Population-interval coding of timbre (vowel formant structure)

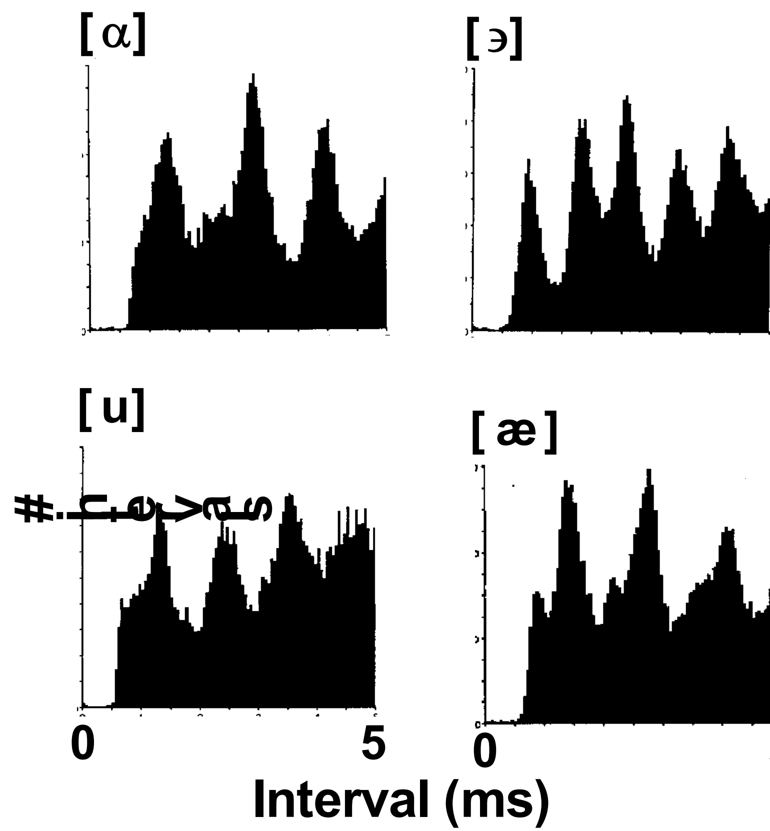
Signal autocorrelation [ae]



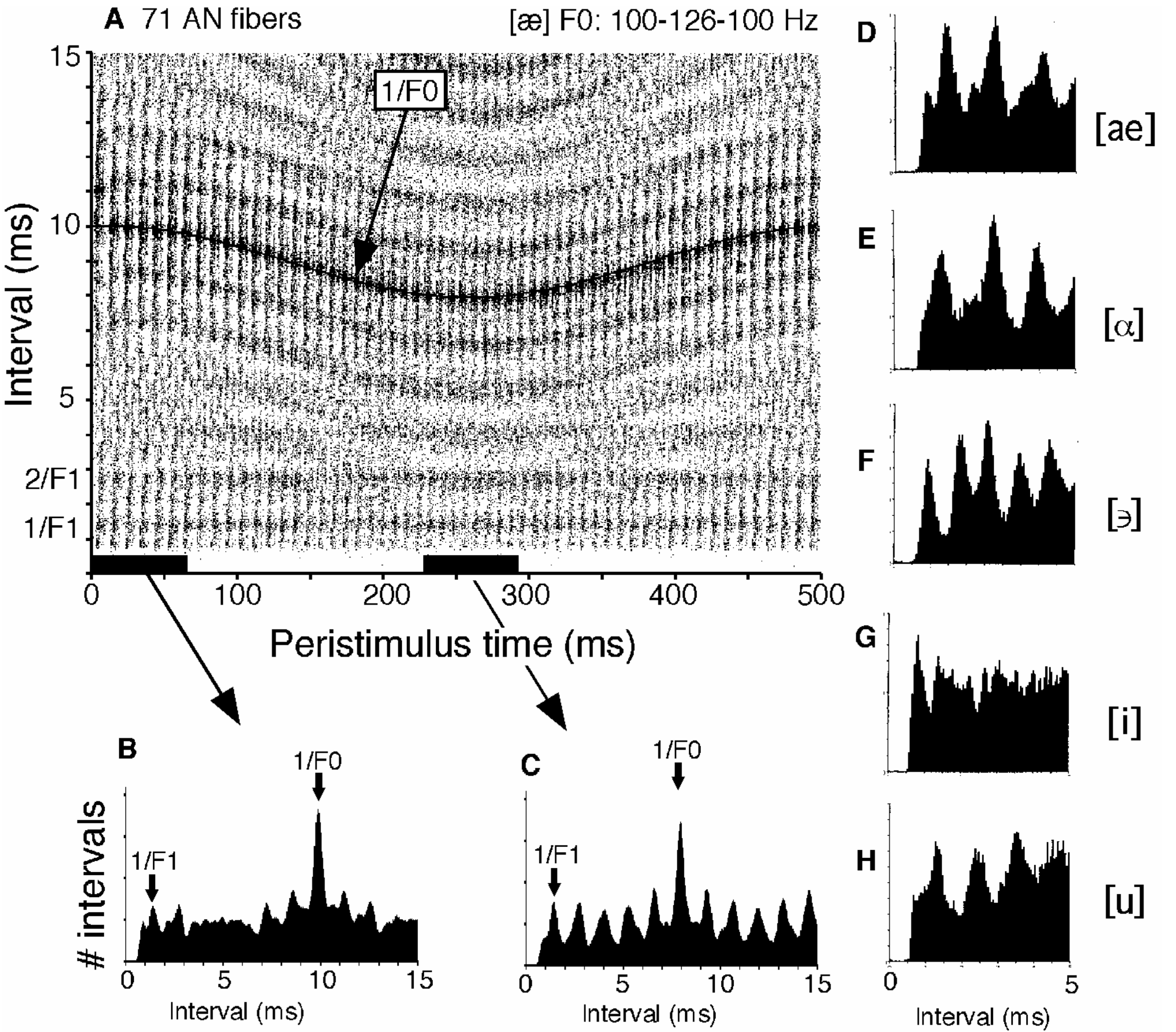
Population interval histogram



Population-wide distributions of short intervals for 4 vowels



Coding of vowel quality (timbre)



Population-interval representations thus far

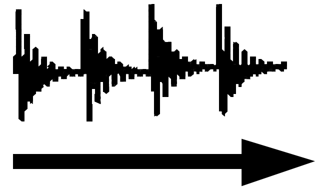
- **Periodicity:** capable of accounting for a wide range of pure and complex tone pitches that are heard; salience model less developed.
- **Spectrum:** capable of robustly representing the stimulus spectrum below 4 kHz (our neural data, Palmer neural data Meddis simulations)
- **Sound separation:** multiple interval bands apparent for $\Delta F0 > 10\%$

Auditory nerve model

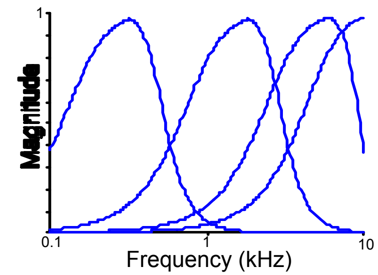
Basic features:

Gammatone filters
Adaptive gain control
Mapping of output amplitude to firing rates
Spontaneous activity

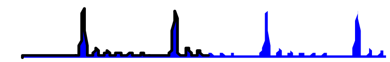
INPUT SIGNAL



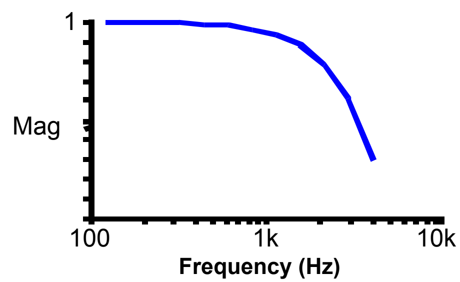
FILTER BANK



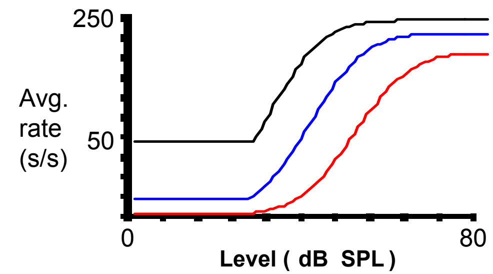
RECTIFICATION



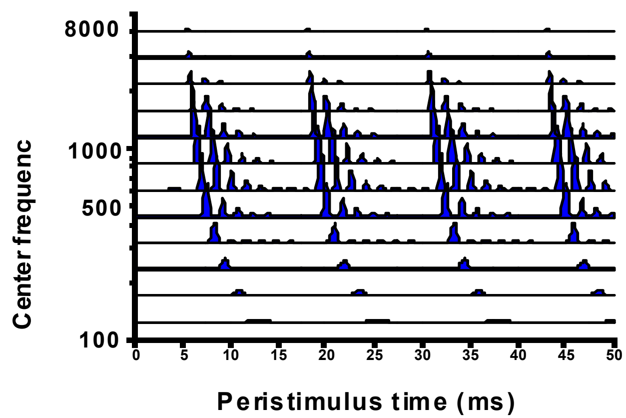
LOW PASS FILTER



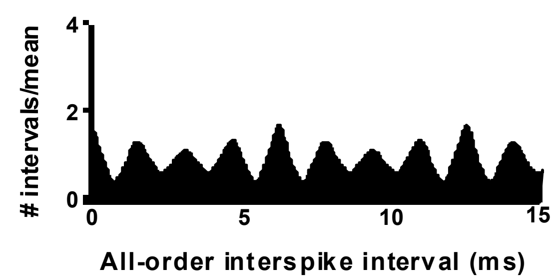
COMPRESSION



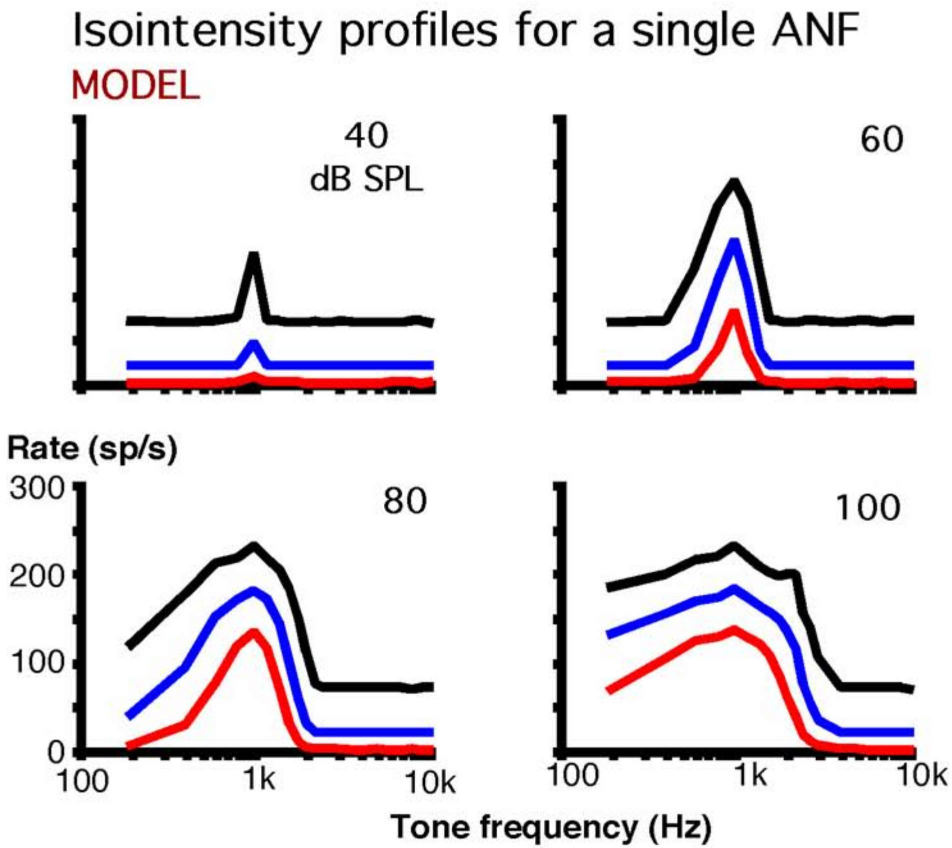
PST NEUROGRAM



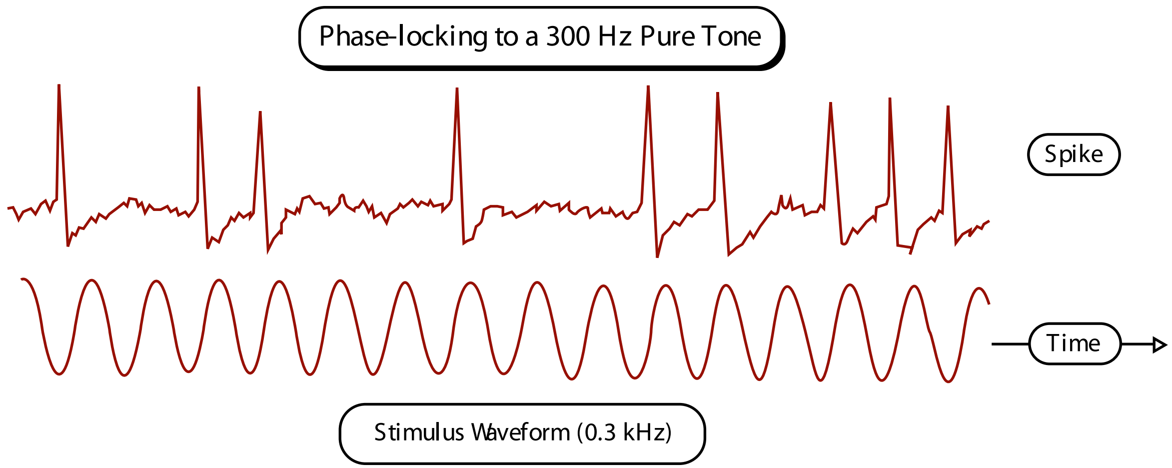
POPULATION-INTERVAL DiSTRIBUTION

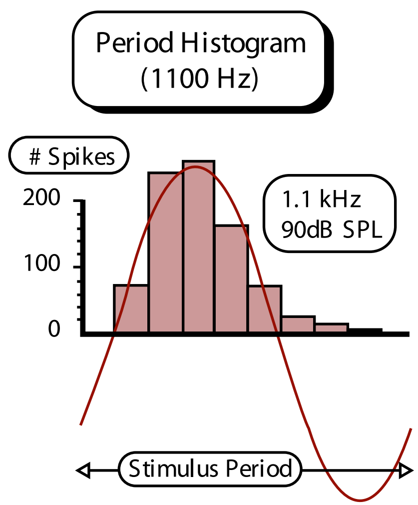


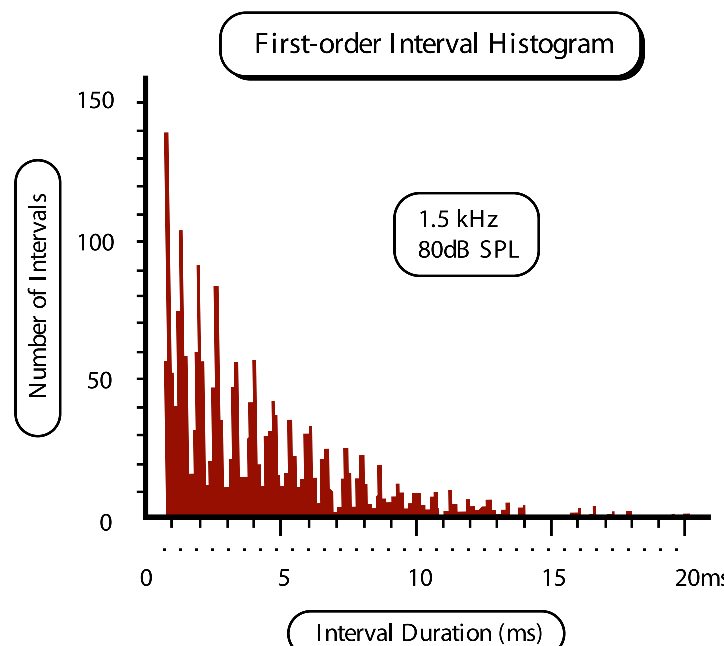
Please see Figure 60 and 61 in Keidel, D., Wolf., S. Kallert, M. Korth. *The physiological basis of hearing: a review*. *Edited by Larry Humes*. New York: Thieme-Stratton, Stuttgart, George ThiemeVerlag, 1983. ISBN: 0865770727.



Please see Hind, J. E., D. J. Anderson, J. F. Brugge, and J. E. Rose. "Coding of Information Pertaining to Paired Low-frequency Tones in Single Auditory Nerve Fibers of the Squirrel Monkey." *J Neurophysiol* 30 (July, 1967): 794 -816.


Phase-locking to a 300 Hz Pure Tone
Spike
Time
Stimulus Waveform (0.3 kHz)

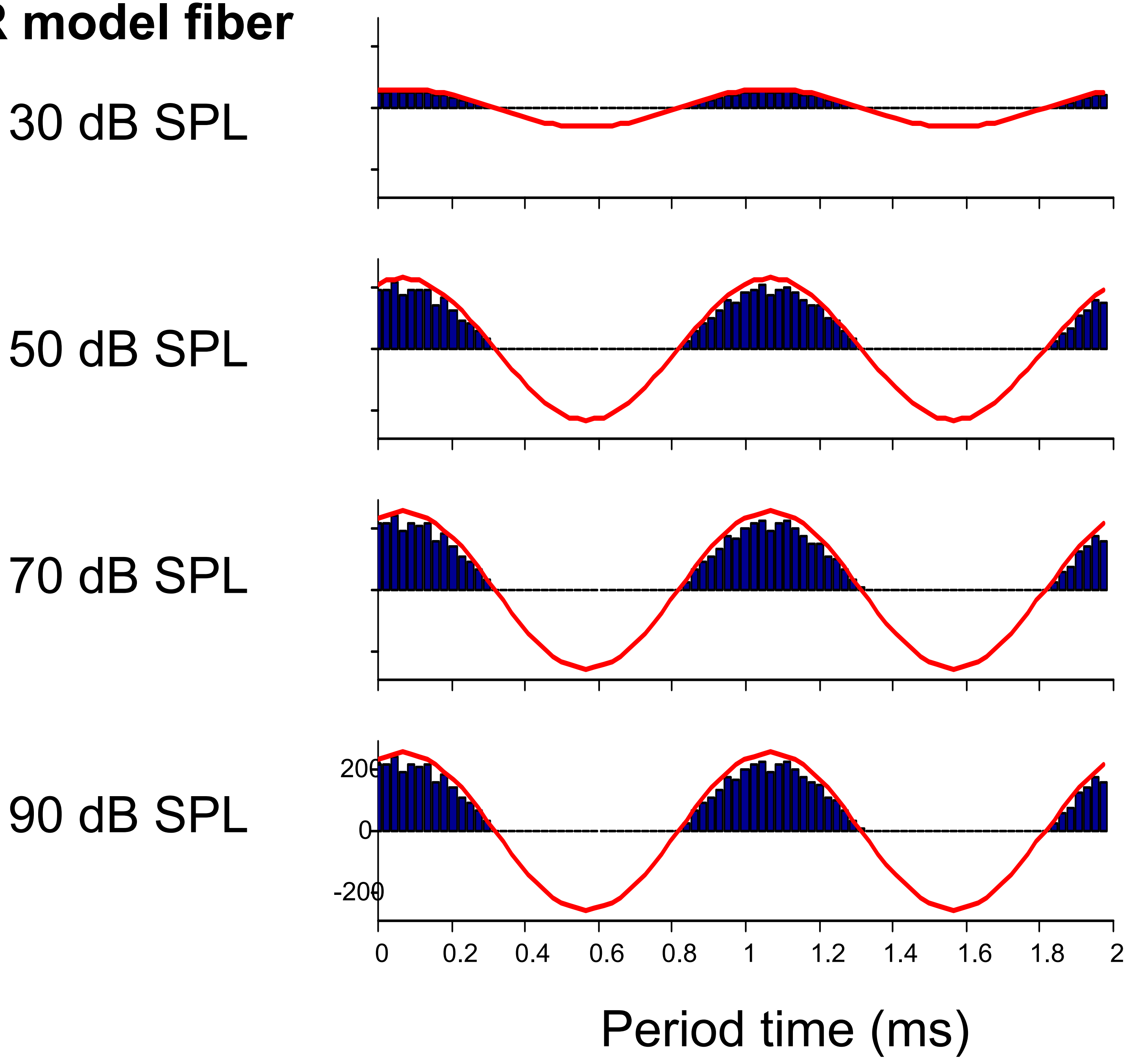

Period Histogram
(1100 Hz)
Spikes
200
100
0
1.1 kHz
90dB SPL
Stimulus Period


First-order Interval Histogram
150
100
50
0
Number of Intervals
1.5 kHz
80dB SPL
0
5
10
15
20ms
Interval Duration (ms)

Temporal discharge patterns 40-90 dB SPL

Please see Rose, Hind, Anderson, Brugge J. Neurophysiology 34 (1971): 685-99. Reprinted in D. Keidel, Wolf., S. Kallert, and M. Korth. *The Physiological Basis of Hearing: A Review.* Edited by Larry Humes. New York: Thieme-Stratton, Stuttgart, George ThiemeVerlag, 1983. ISBN: 0865770727.

Medium-SR model fiber



Temporal discharge patterns to pure tone dyad: 600 & 800 Hz

Please see Rose, Hind, Anderson, Brugge J. Neurophysiology 34 (1971): 685-99. Reprinted in D. Keidel, Wolf., S. Kallert, and M. Korth. *The Physiological Basis of Hearing: A Review*. Edited by Larry Humes. New York: Thieme-Stratton, Stuttgart, George ThiemeVerlag, 1983. ISBN: 0865770727.

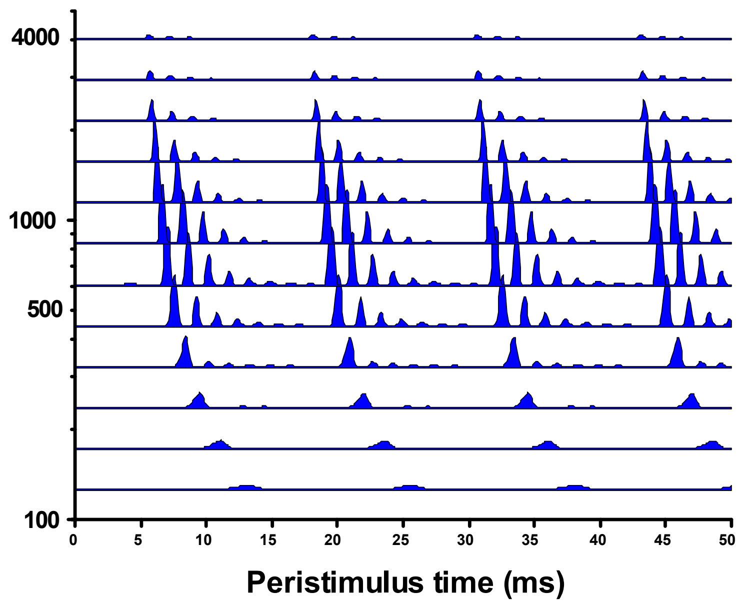
Physiological vs. simulated activity

Simulated vs. observed auditory nerve activity

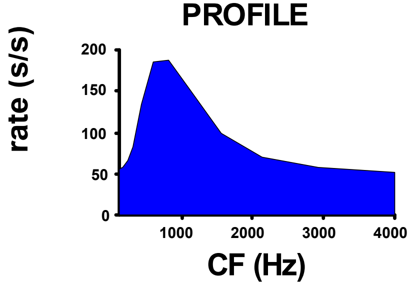
We are mainly interested here in replicating basic phase-locked time-frequency patterns.

SIMULATED RESPONSES

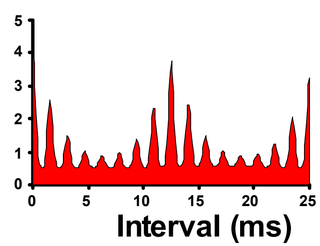
PST NEUROGRAM

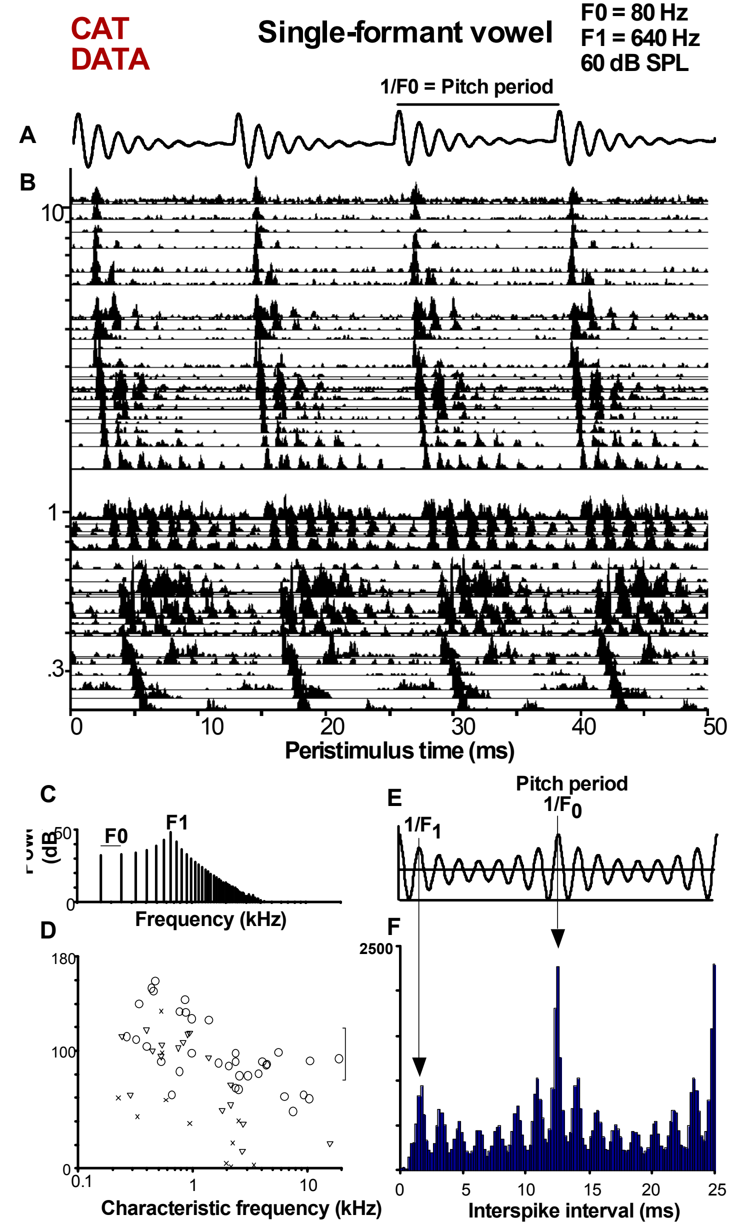


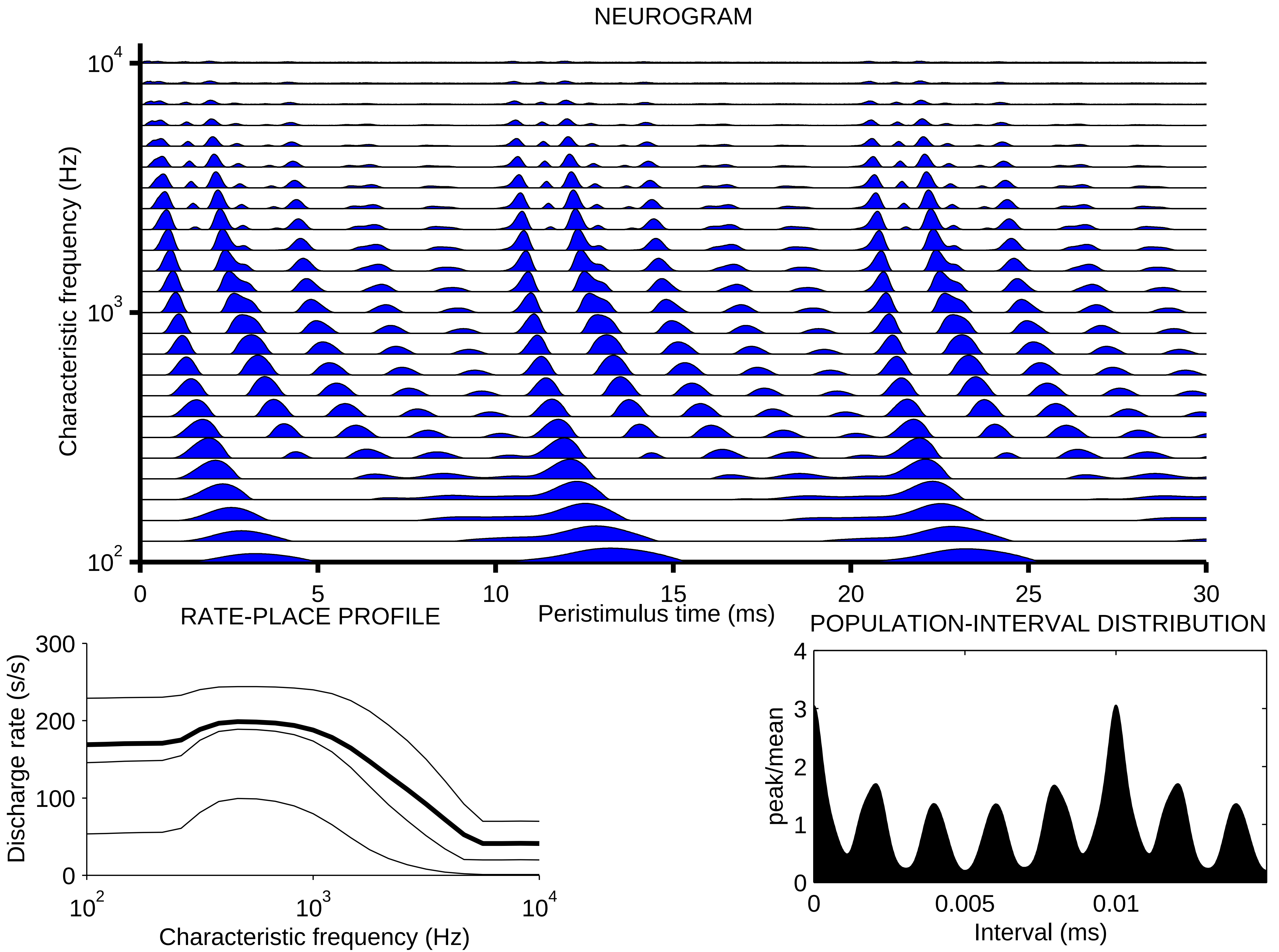
RATE-PLACE PROFILE



POPULATION-INTERVAL DISTRIBUTION







NEUROGRAM
Characteristic frequency (Hz)
10^4
10^3
10^2
0
5
10
15
20
25
30
Peristimulus time (ms)
RATE-PLACE PROFILE
Discharge rate (s/s)
300
200
100
0
10^2
10^3
10^4
Characteristic frequency (Hz)
POPULATION-INTERVAL DISTRIBUTION
peak/mean
4
3
2
1
0
0
0.005
0.01
Interval (ms)

Competition between sounds
Competition between respective interval patterns

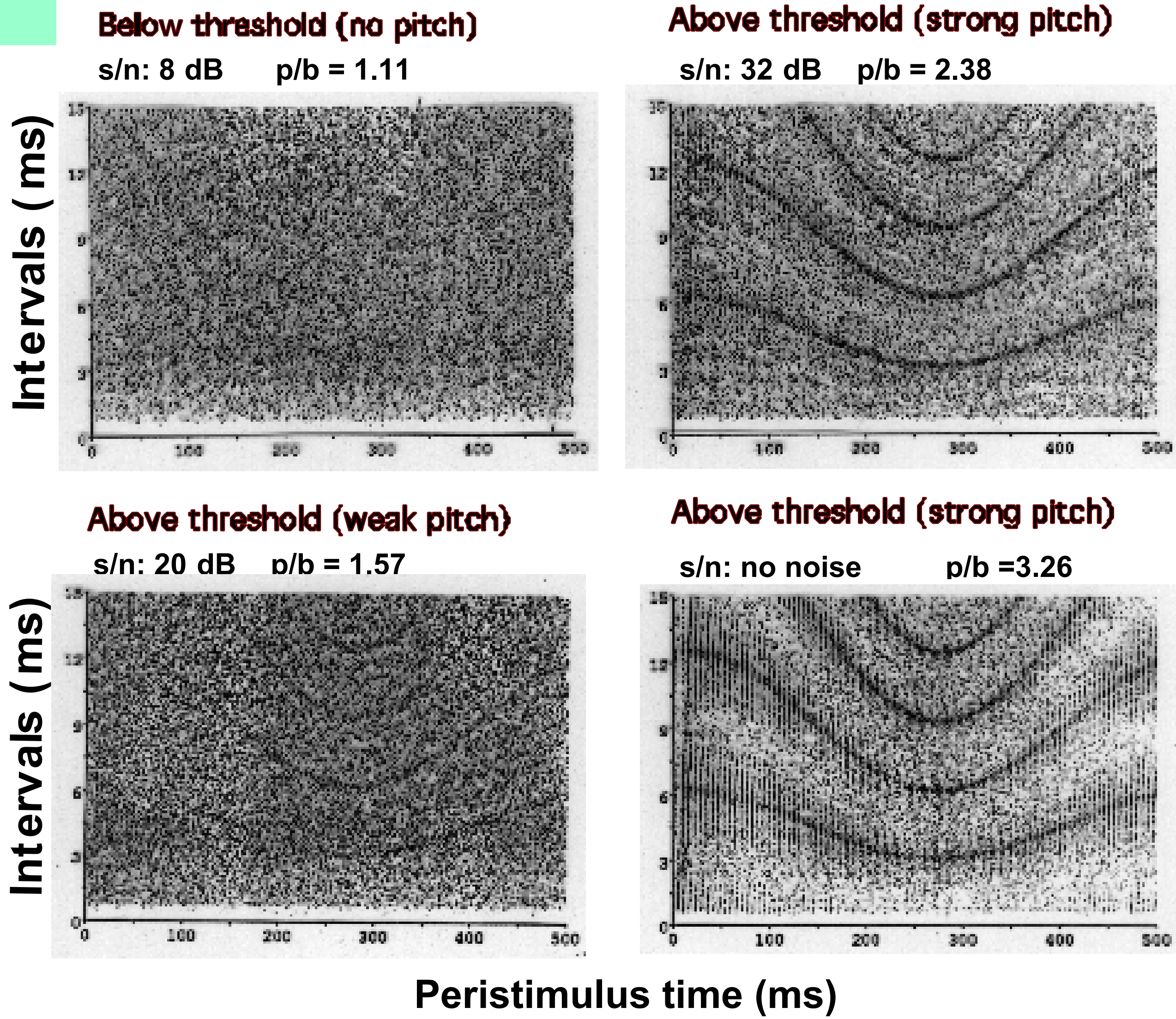
Towards interval-based accounts of interacting sounds:

- mutual masking
- pitch multiplicity
- roots of dyads and triads (chords)

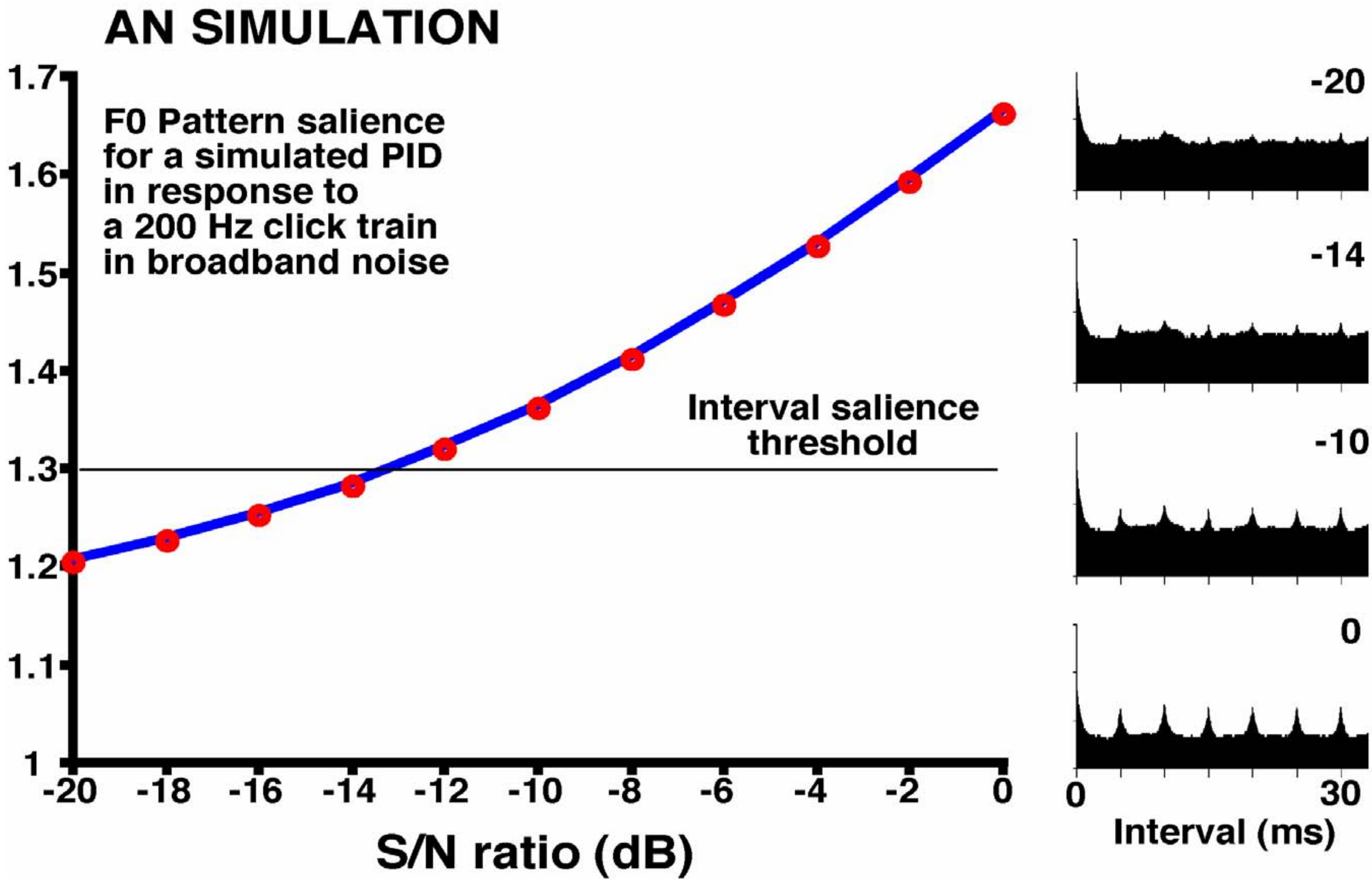
Pitch masking

Variable-F0 click train in broad-band noise
Click train: F0= 160-320 Hz, positive polarity, 80 dB SPL
Peak/background ratio: intervals @ 1/ F0 ± 150 usec/mean over all intervals
Informal pitch thresholds, s/n dB: 12 (MT), 13 (PC), 16 (BD)

What degree of temporal correlation is necessary for pitch to become audible?

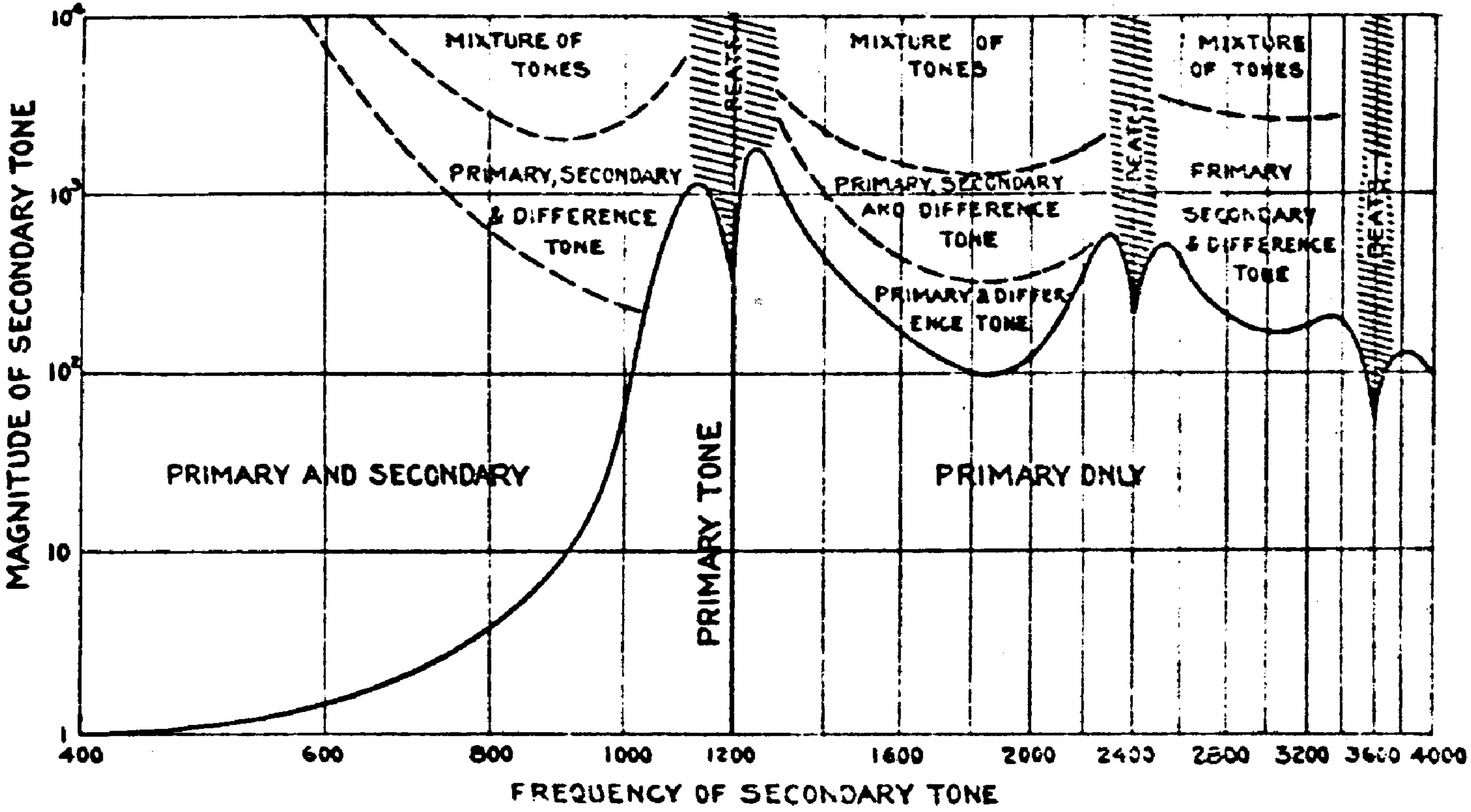


Masking of a click train in broadband noise



Click trains were presented in white noise (0-11 kHz, 70 dB total SPL) at different s/n ratios.
My threshold for pitch detection with low-isolation headphones is -6dB.

Masking audiograms

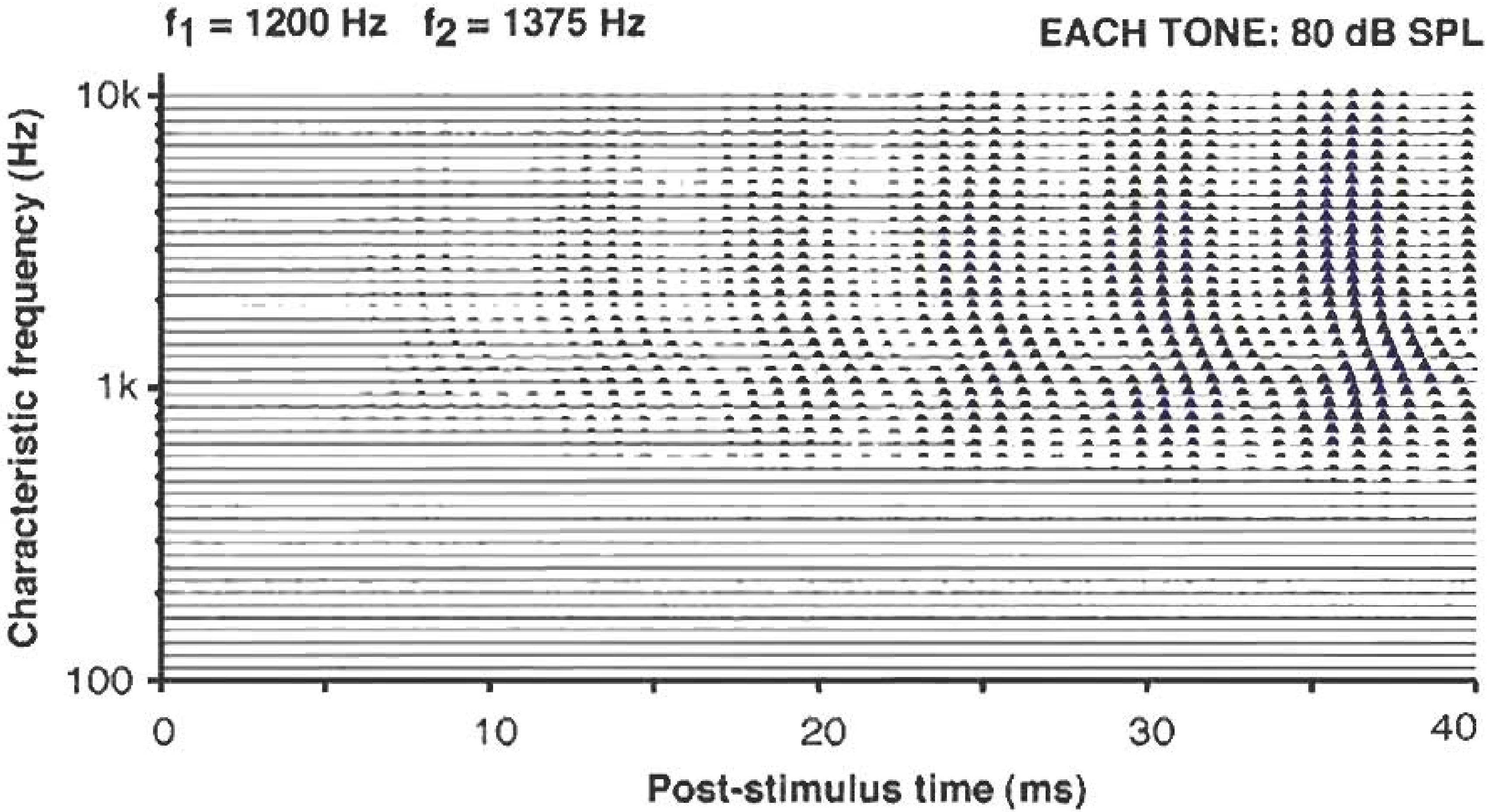


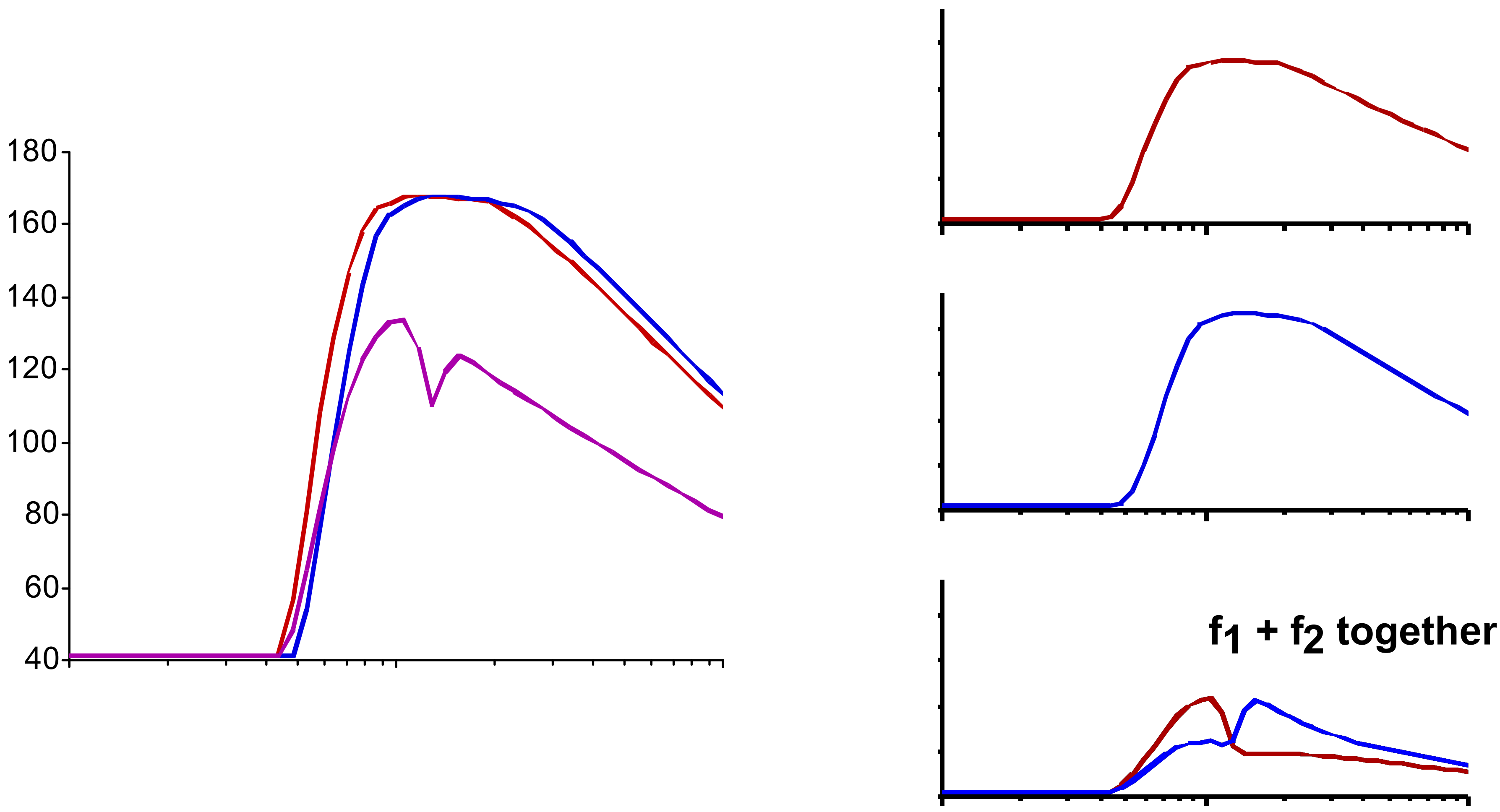
Wegel & Lane, 1924

Pitch masking: pure tones

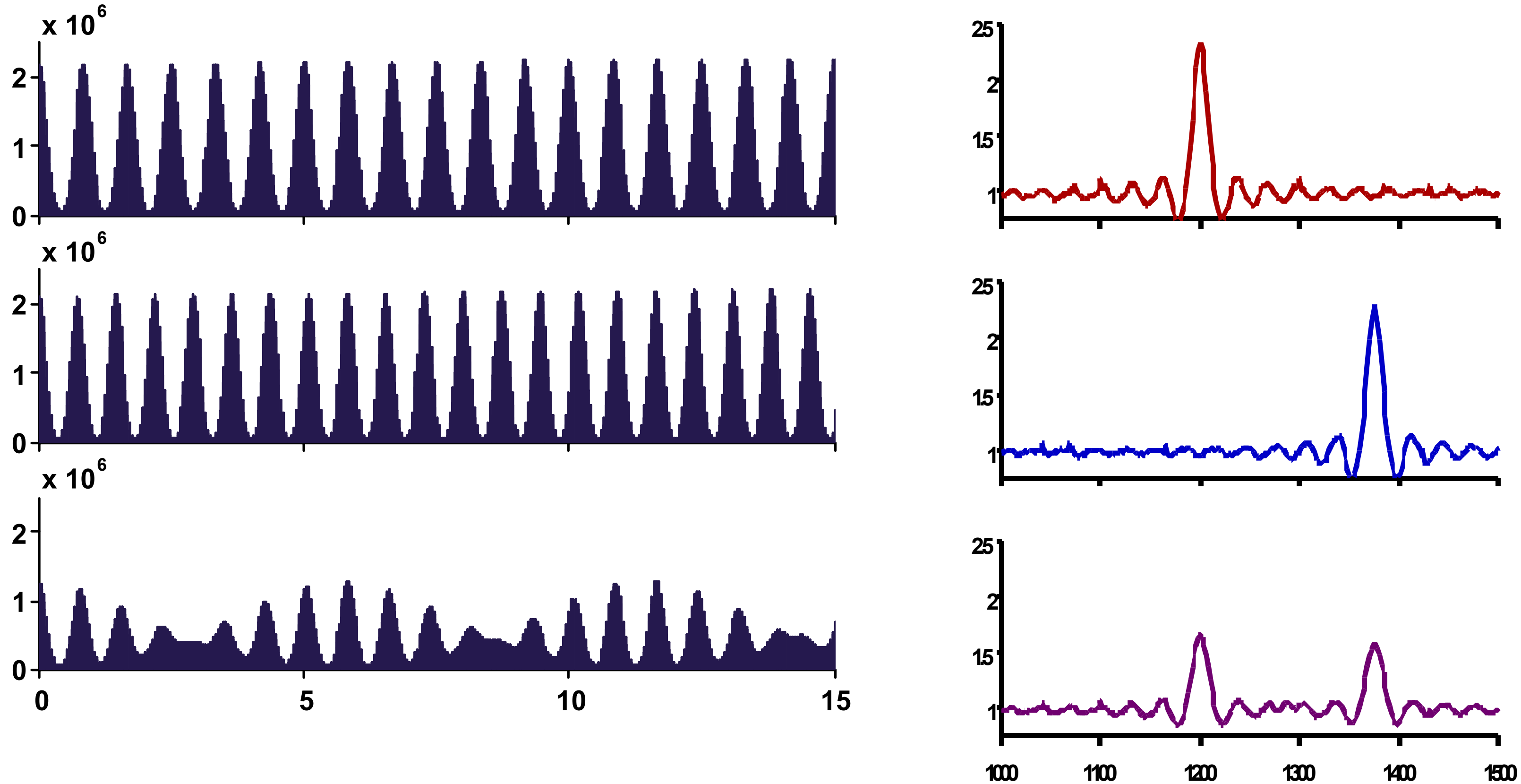
Competition for cochlear territory

Response of auditory array to two concurrent pure tones



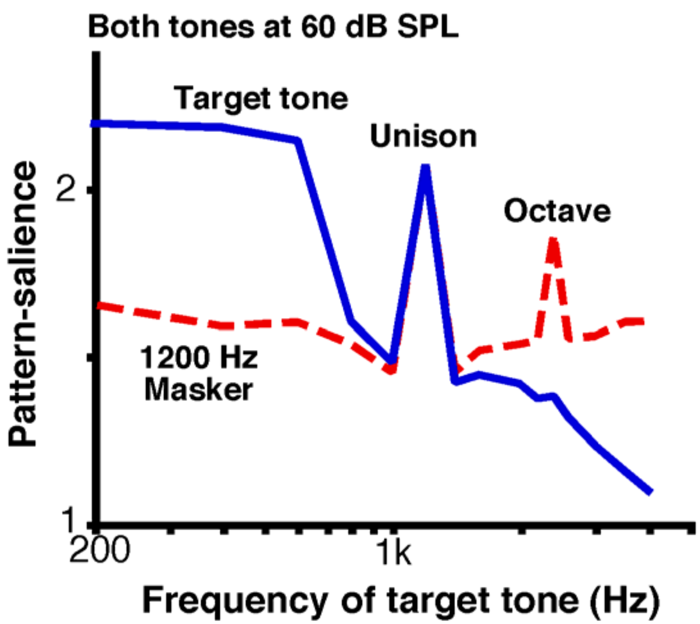


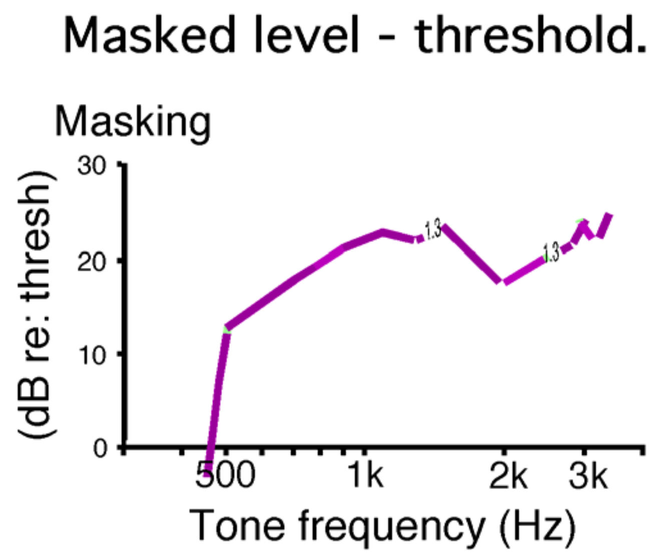
180
160
140
120
100
80
60
40
f1 + f2 together

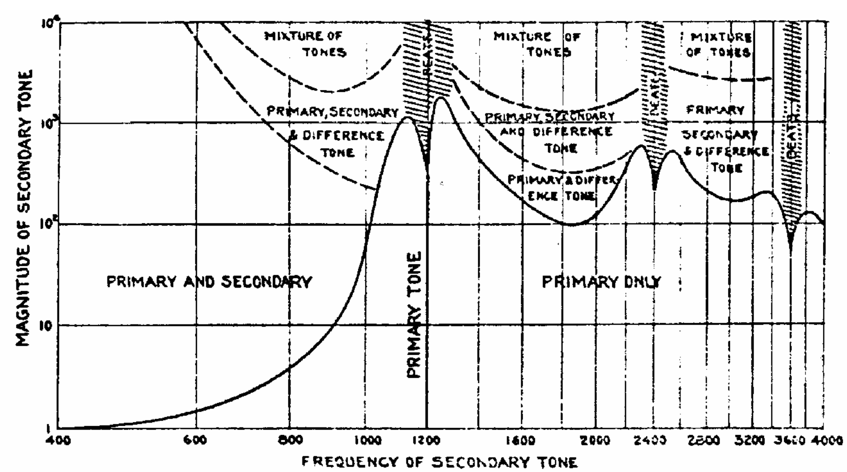


x 10^6
2
1
0
x 10^6
2
1
0
x 10^6
2
1
0
0
5
10
15
25
2
15
1
25
2
15
1
25
2
15
1
1000
1100
1200
1300
1400
1500

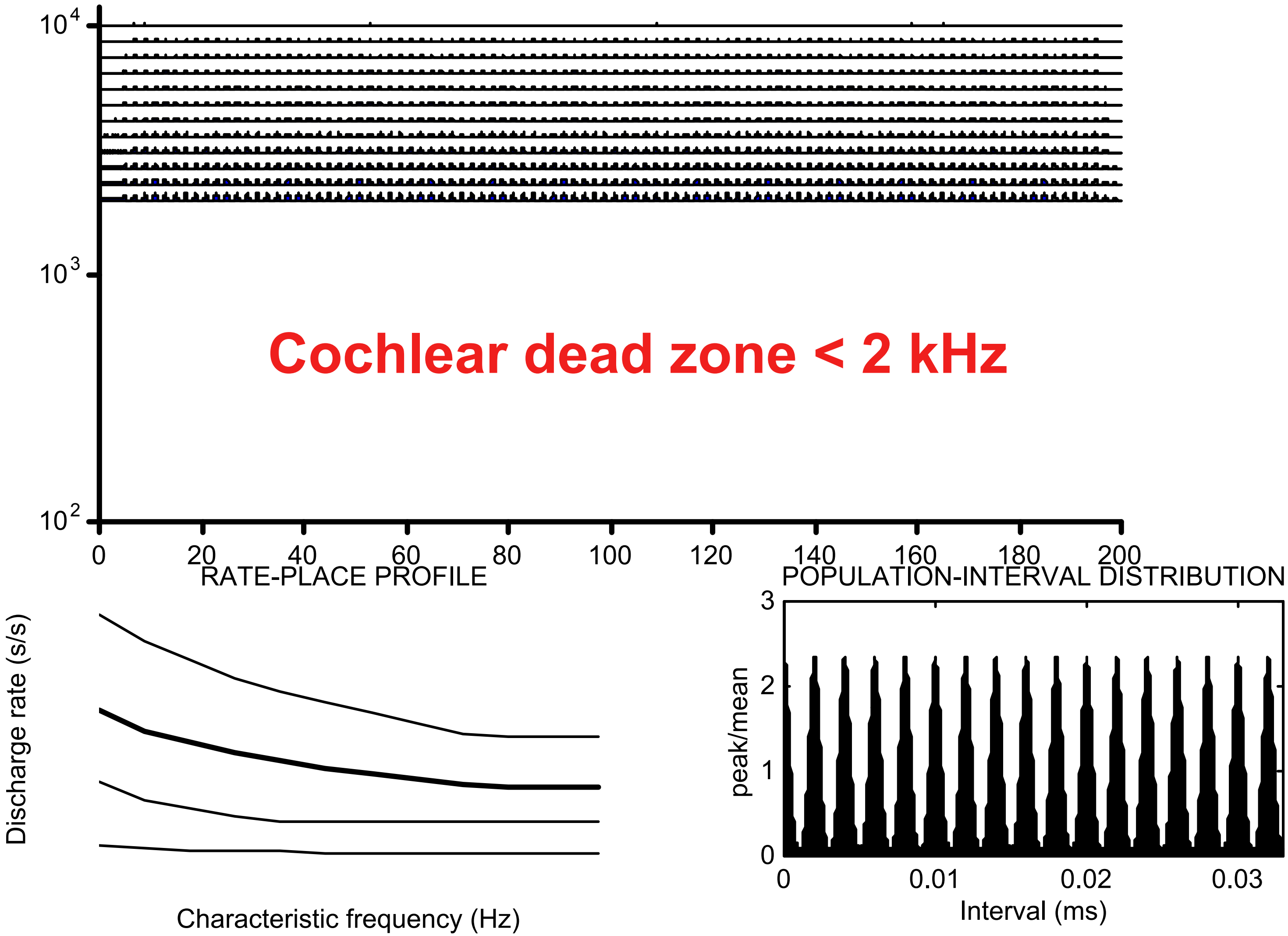
We assume that a pitch can be heard if its salience is >1.3.
Masking functions can be then estimated for those conditions that involve pitch detection (vs. Δ loudness)

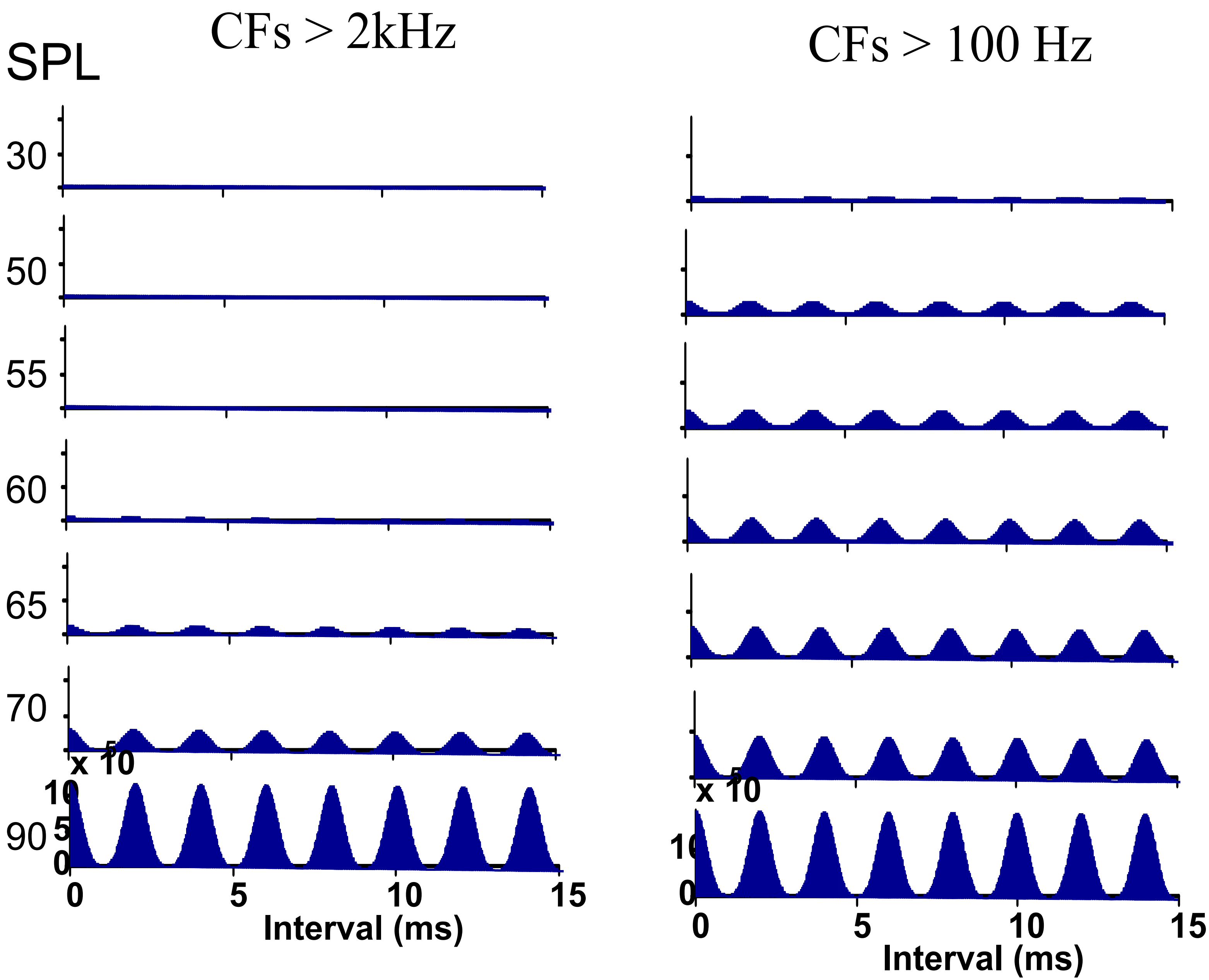




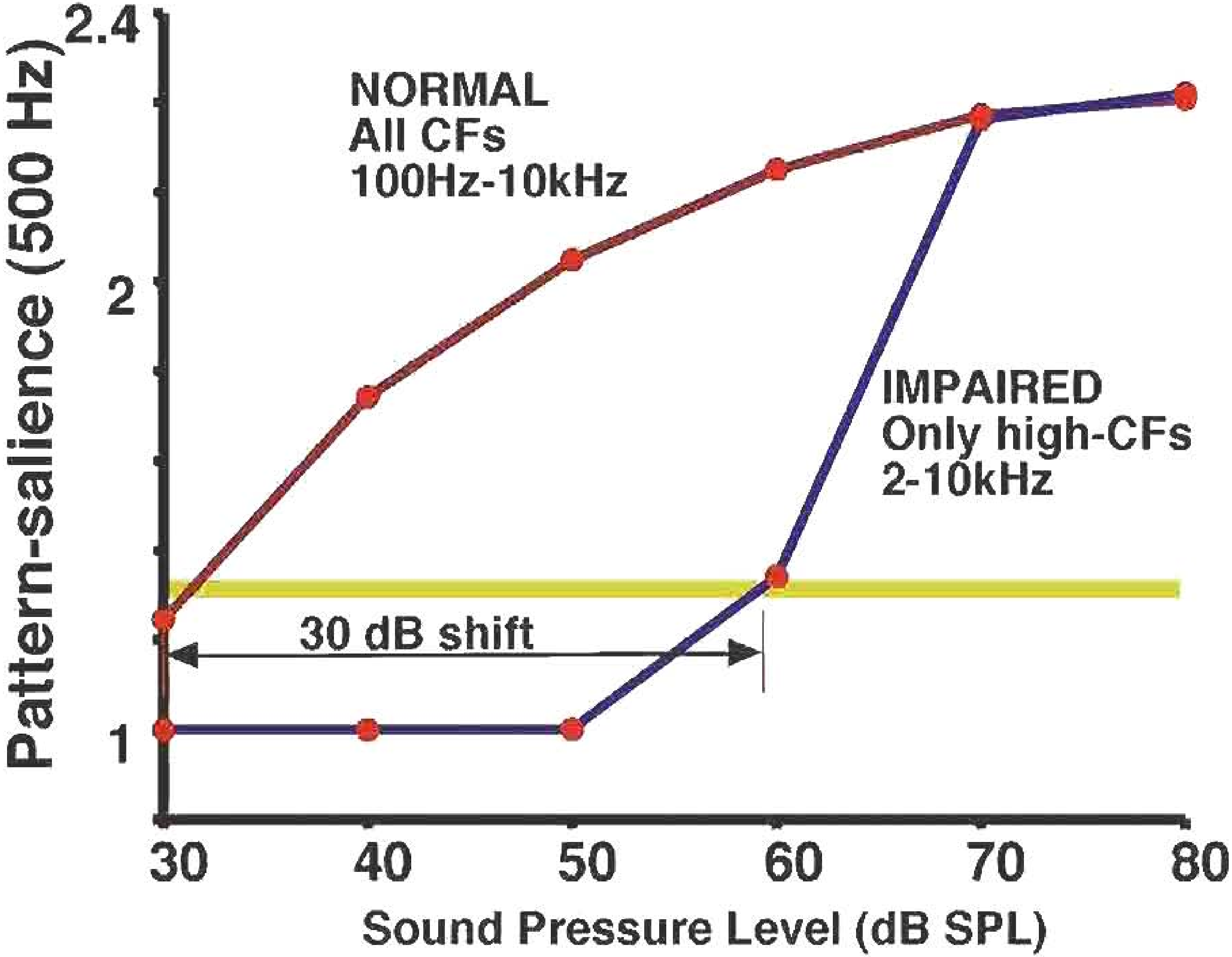


500 Hz tone 65 dB SPL




CFs > 2kHz
CFs > 100 Hz
SPL
30
50
55
60
65
70
90
x 10
5
10
5
0
0
5
10
15
Interval (ms)
x 10
5
10
0
0
5
10
15
Interval (ms)

Effect of "cochlear dead zone" on pitch thresholds (e.g. Florentine & Houtsma, 1983)

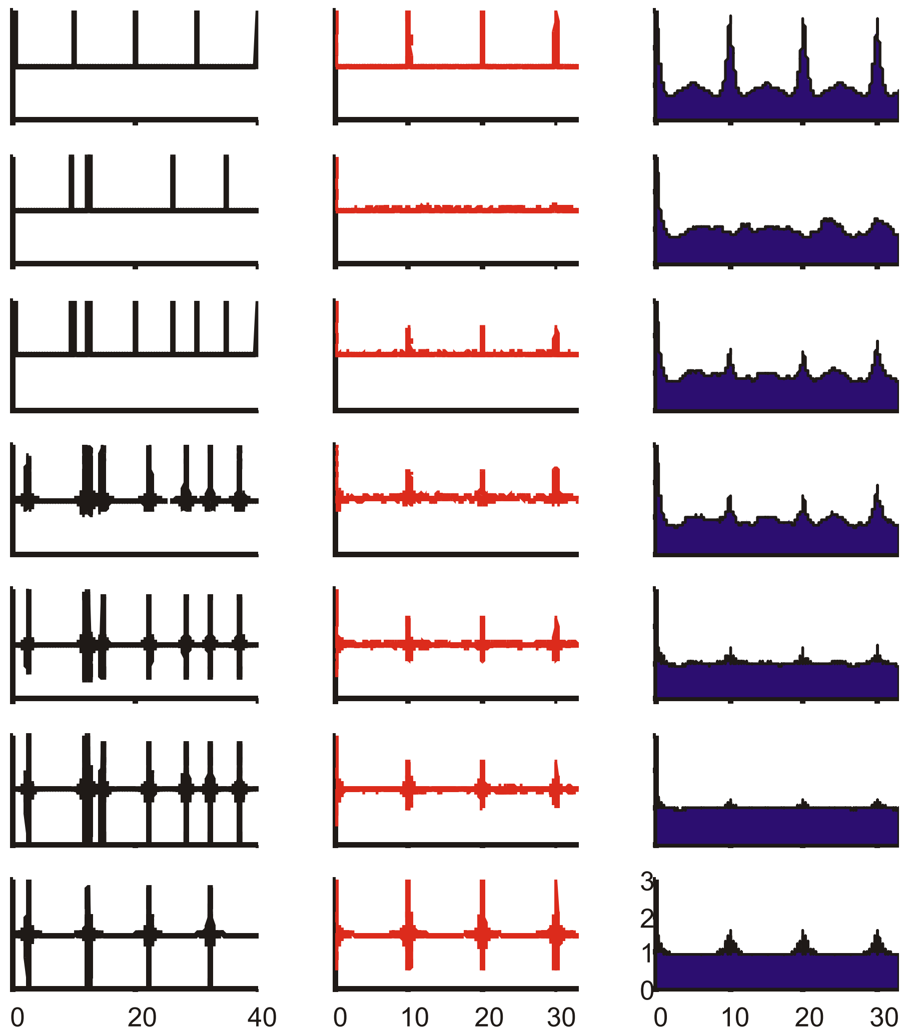


Evolution of population-interval models

- Analysis of peaks -> whole histograms
- Comparisons between whole histograms
  - Pattern similarity (octave, probe-tone:chord, FF timing nets)
- Importance of limited temporal windows
  - Finite frequency resolution of autocorrelation analysis
  - Carlyon's experiments: low F0's of high harmonics
  - Lower limits of F0-pitch (30 Hz, Pressnitzer & Patterson)
- Pitch multiplicity and salience estimation
- CF-dependent temporal analysis windows
- Recurrent timing nets - periodicity & object formation/sep.
  are intimately linked mechanisms

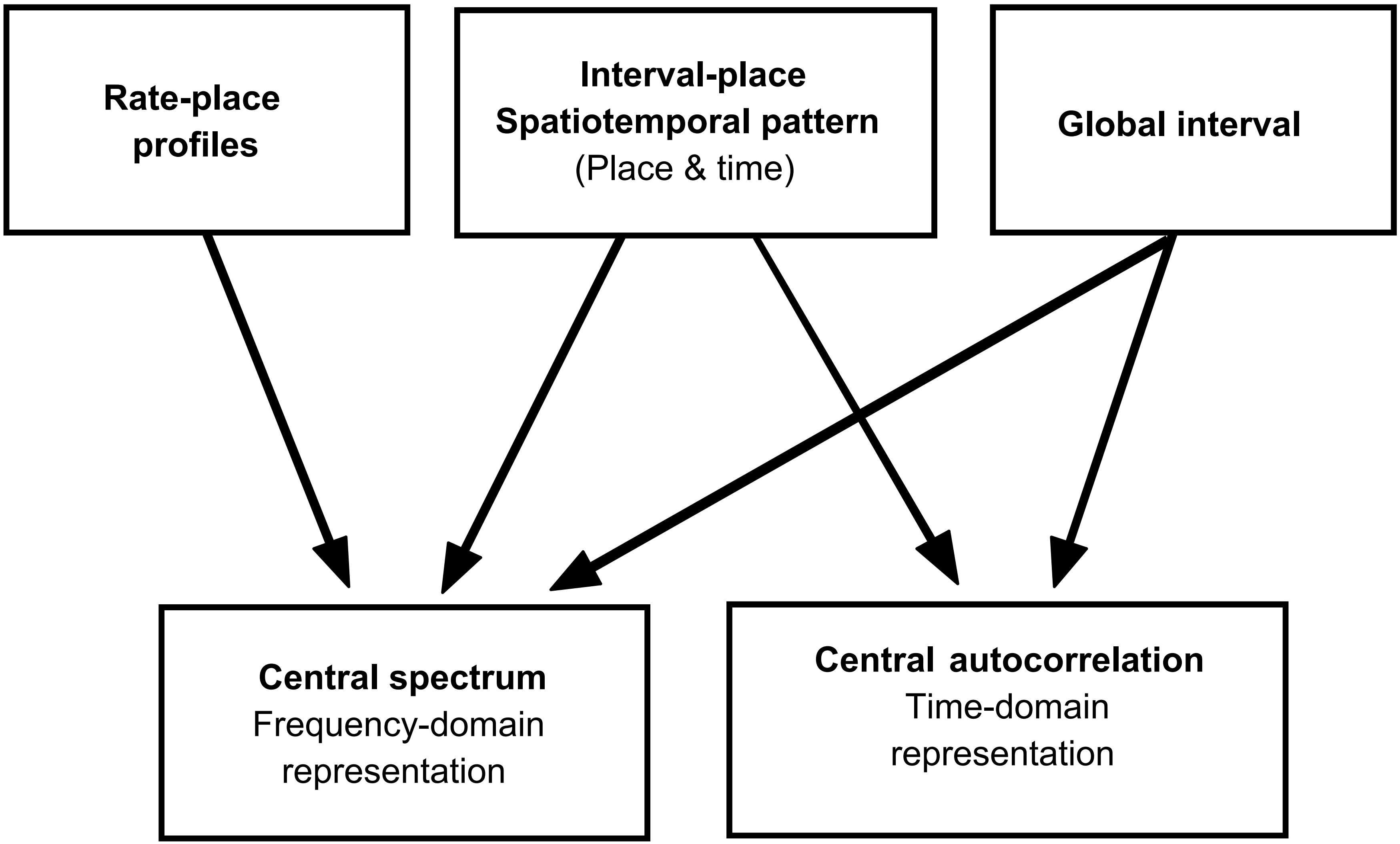
Divergences between autocorrelation & interval models

Masking of click patterns by random clicks

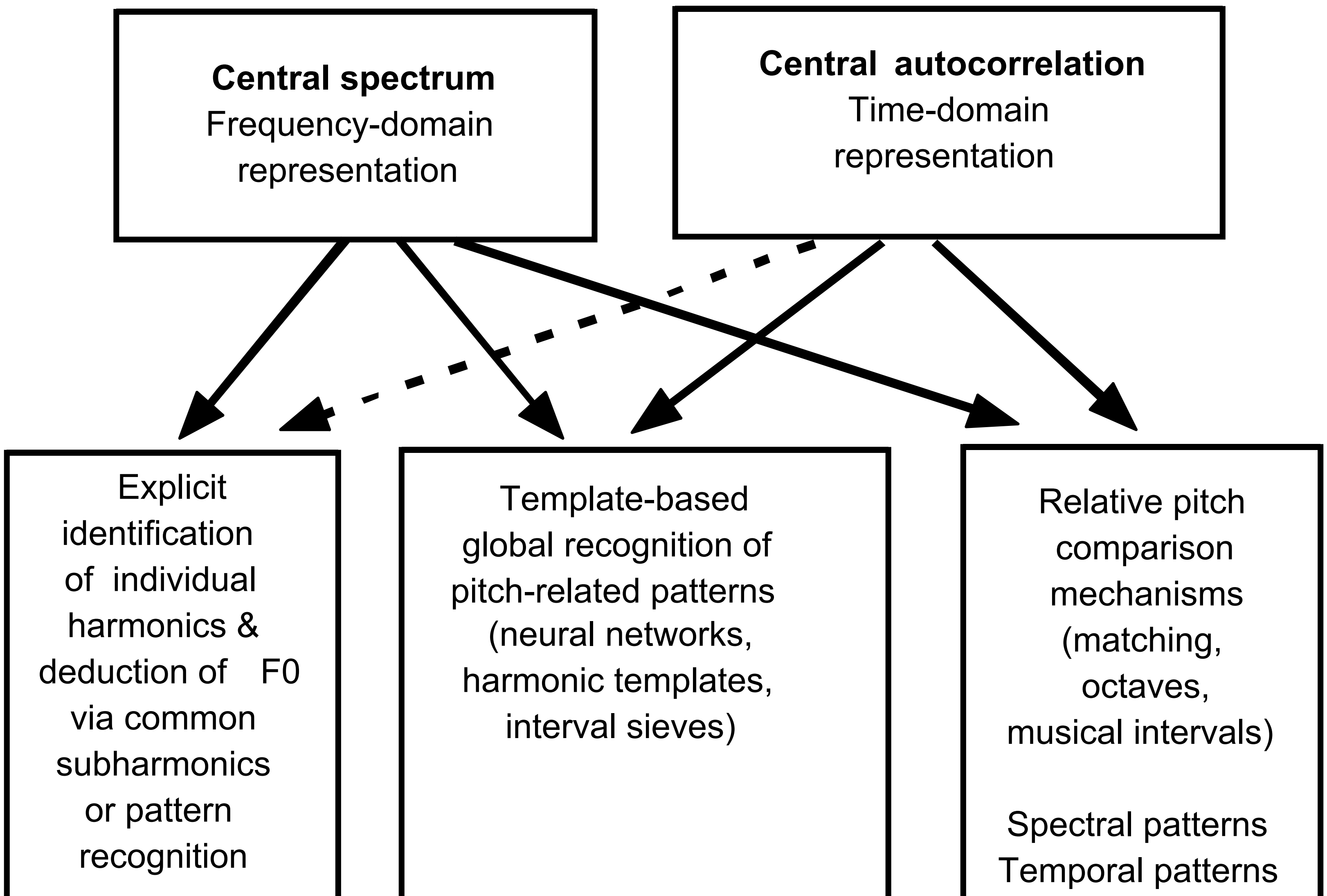


ABX	1.28
KX	1.30
KXX	1.21
KXXX	1.20

Physiological and functional representations



Different representations can support analogous strategies for pitch extraction, recognition, and comparison



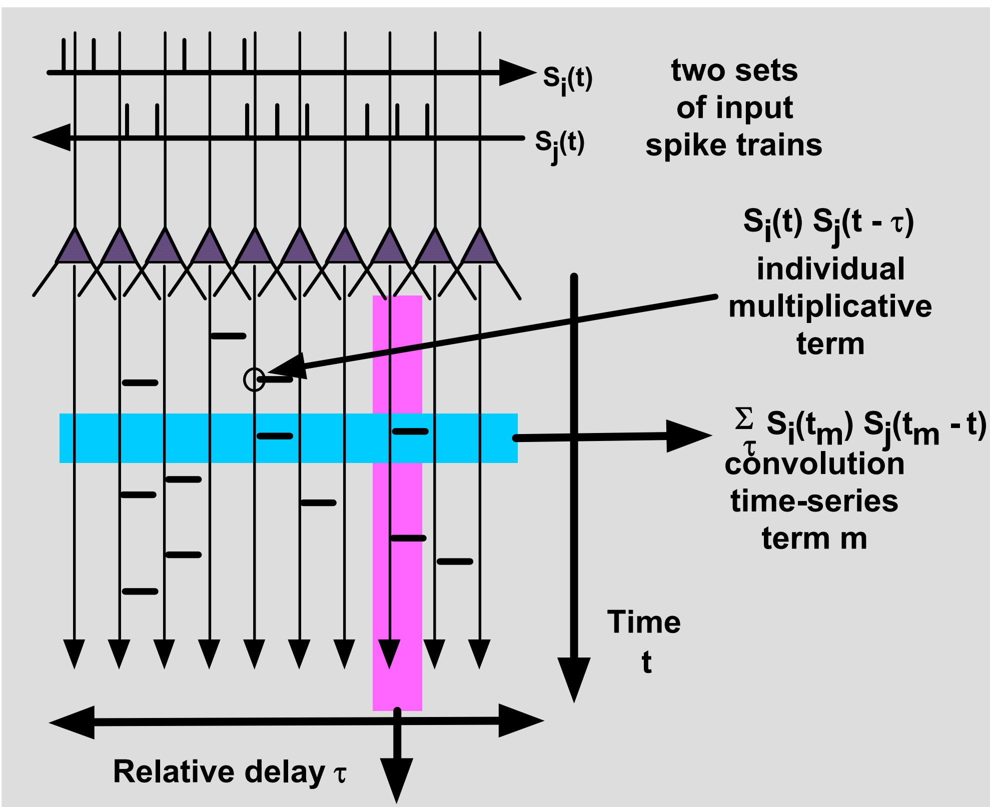
Neural timing nets

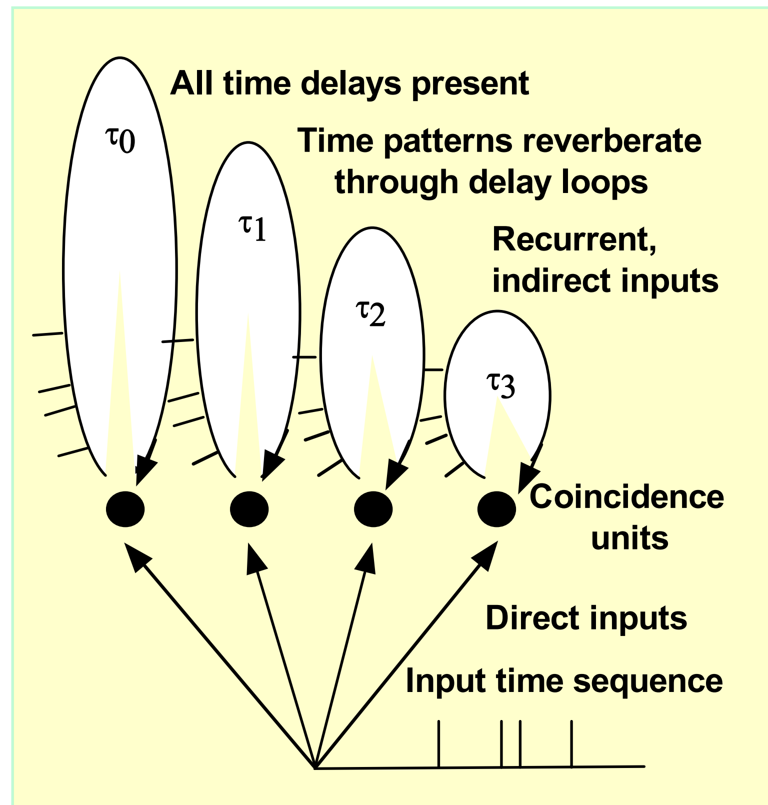
FEED-FORWARD TIMING NETS

- Temporal sieves
- Extract (embedded) similarities
- Multiply autocorrelations

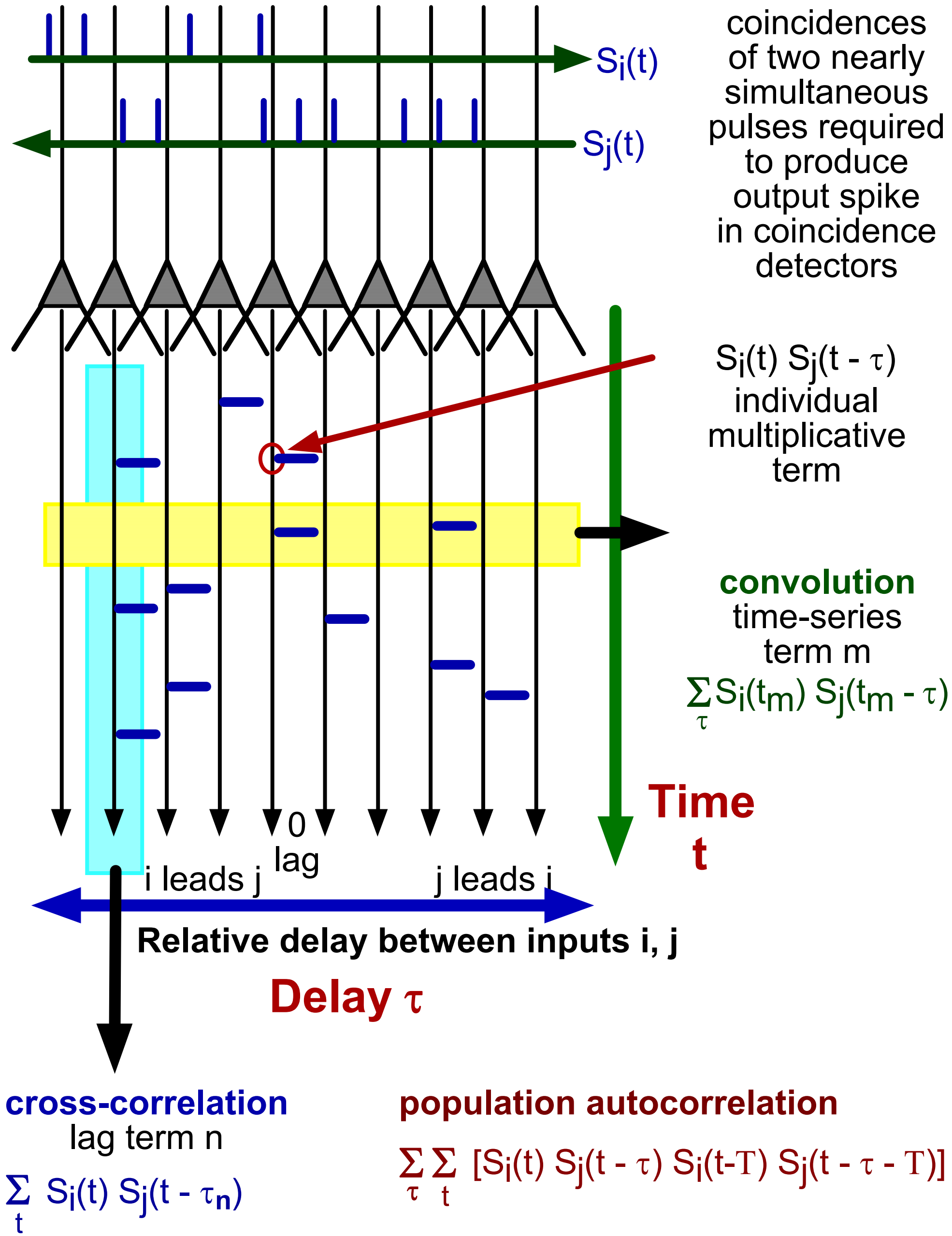
RECURRENT TIMING NETS

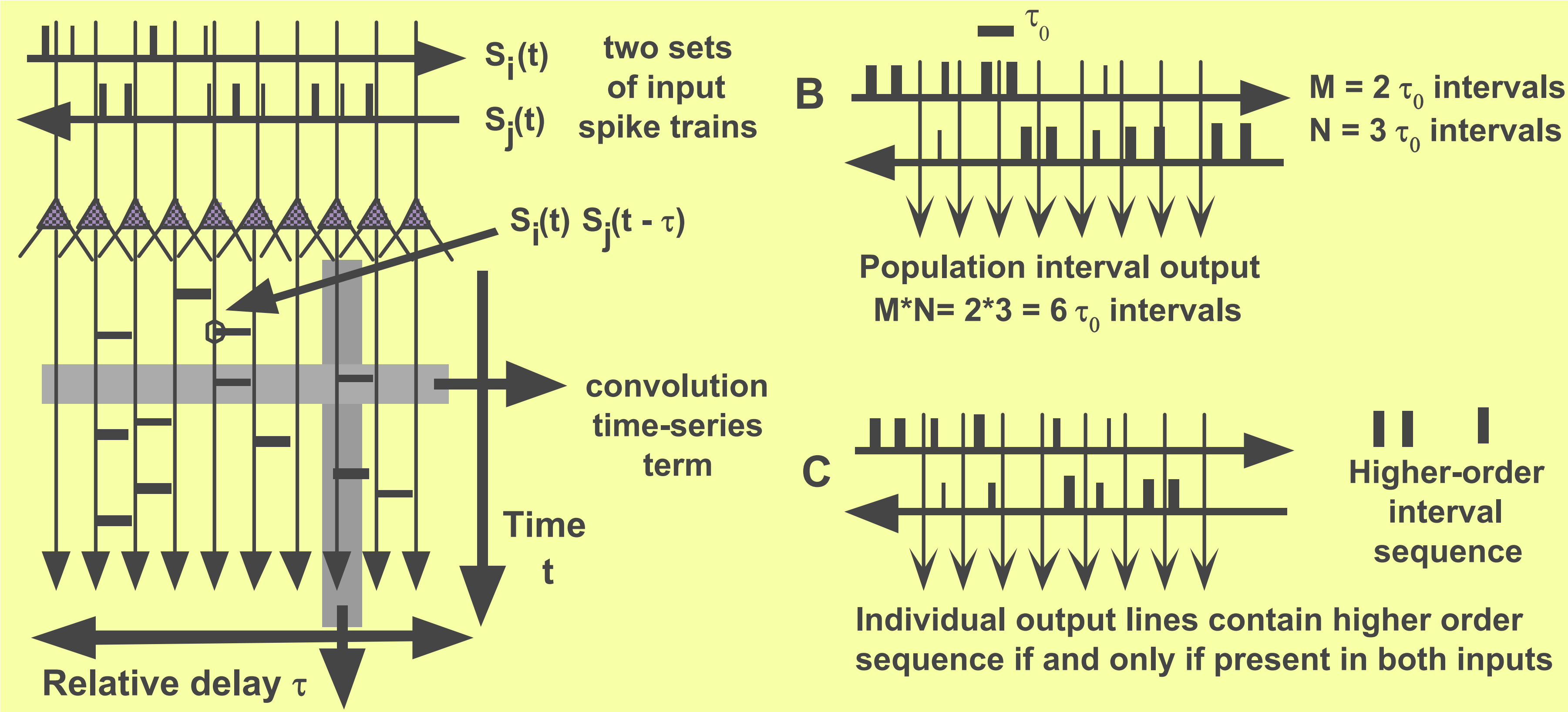
- Build up pattern invariances
- Detect periodic patterns
- Separate auditory objects





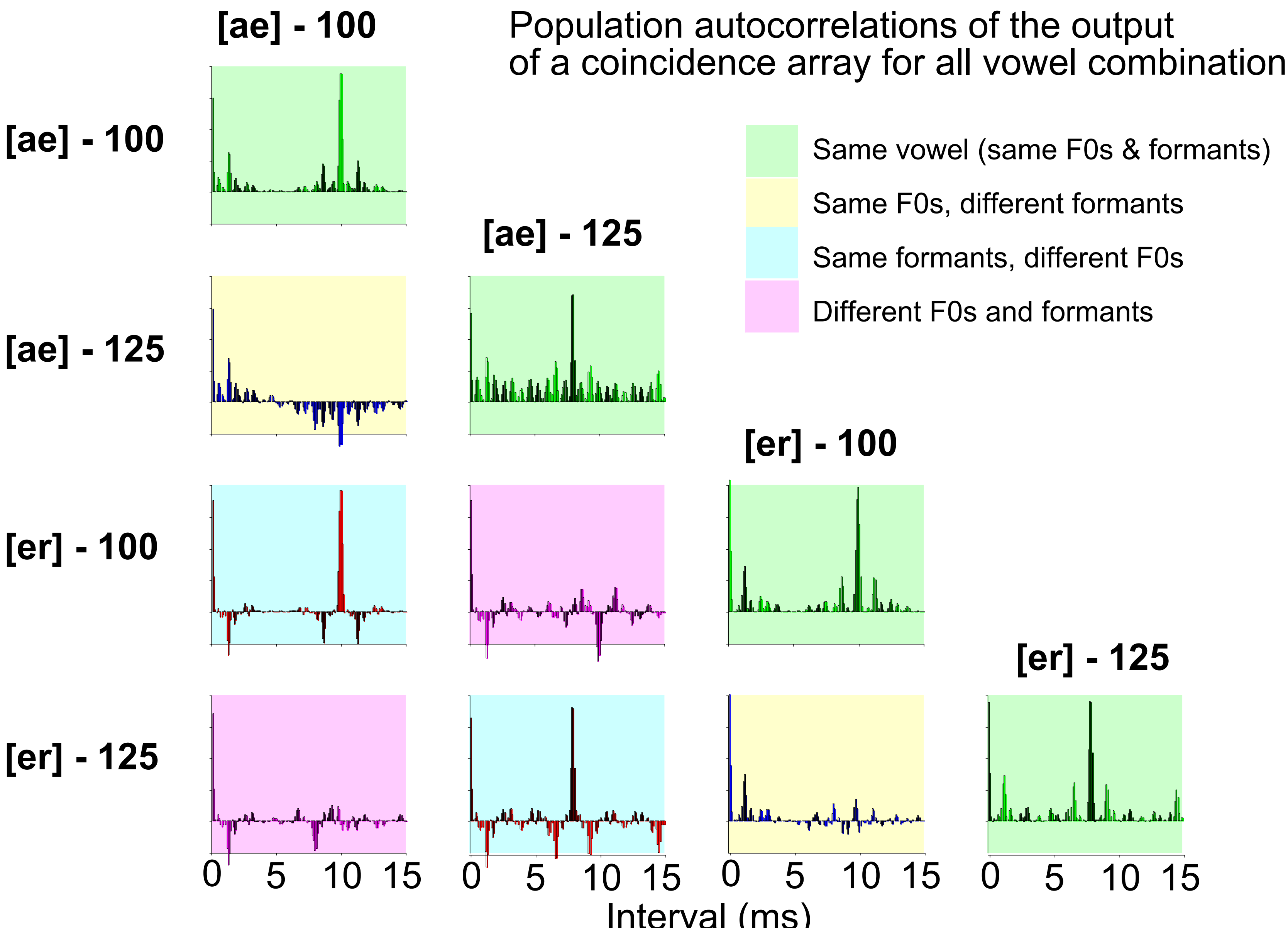
Feedforward coincidence net





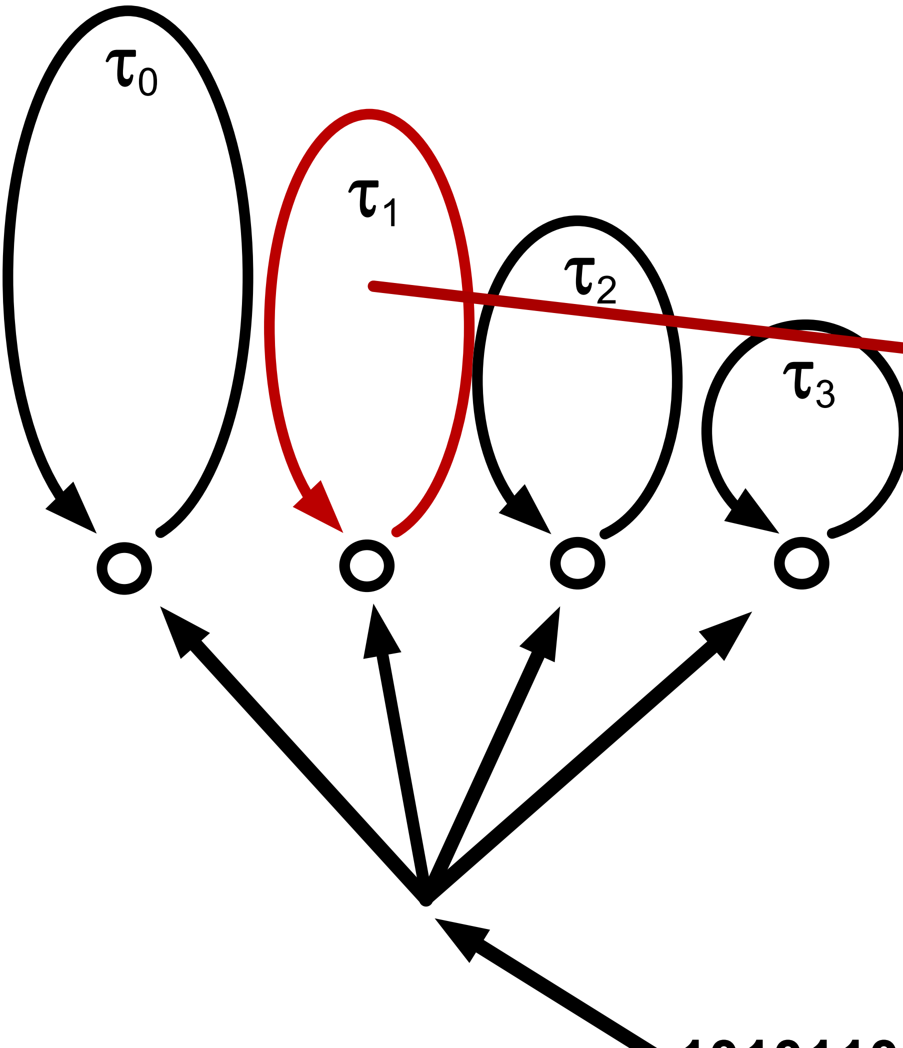
- Networks act as temporal sieves, passing only common patterns
- These are the only networks for which it is irrelevant which particular neurons are activated
- Signals are no longer tied to particular transmission lines
- Processing on the level of ensemble mass-statistics is made possible
- Extraction of complex, embedded patterns is made possible
- Time-domain multiplexing is greatly facilitated

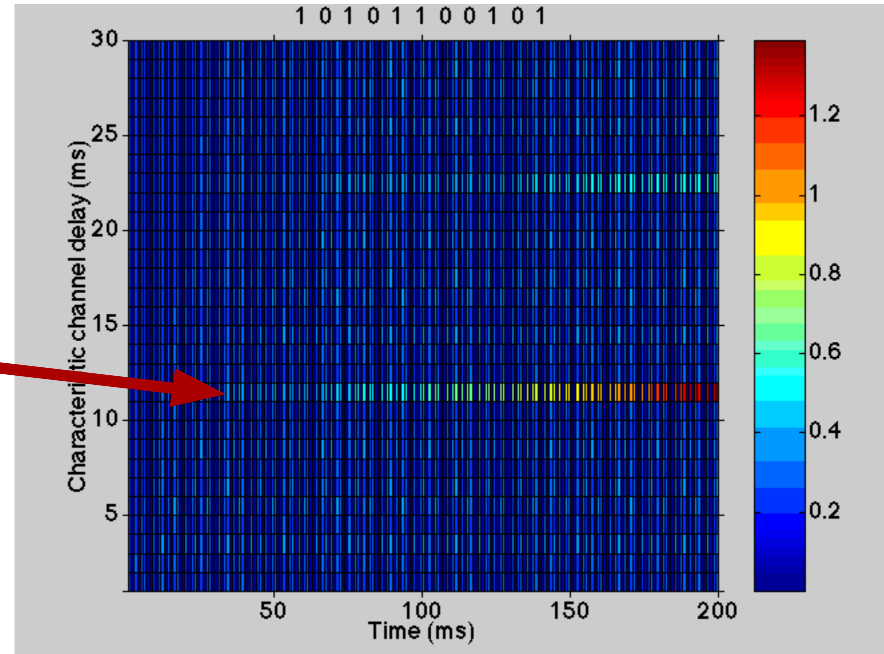
Common timbre



Detection of arbitrary periodic patterns

Periodic patterns invariably build up in delay loops whose recurrence times equals the period of the pattern and its multiples.

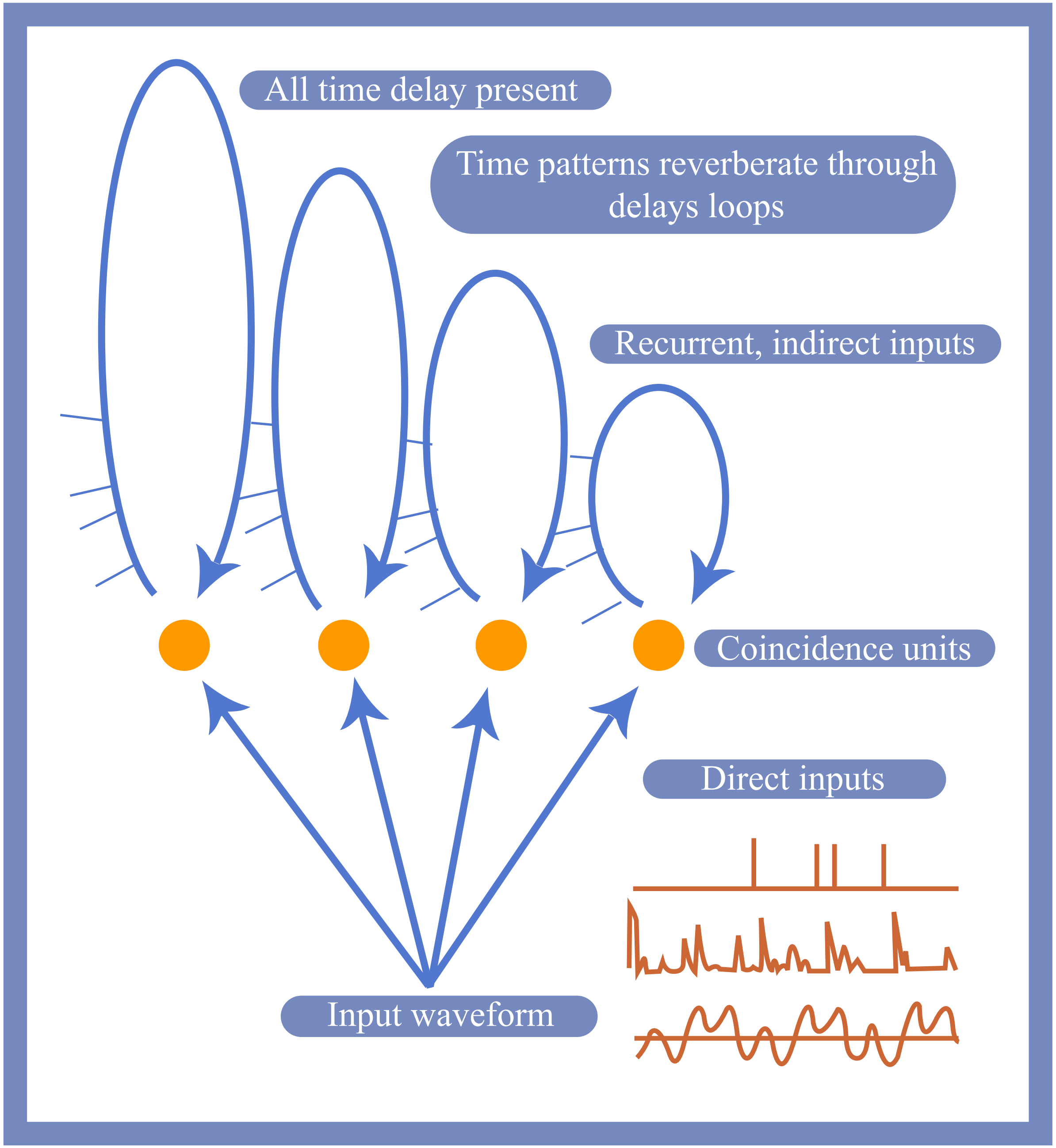




τ_1 = 11 ms = recurrence time of input pattern10101100101

Input pattern

1010110010110101100101**10101100101**1010...



All time delay present
Time patterns reverberate through delays loops
Recurrent, indirect inputs
Coincidence units
Direct inputs
Input waveform

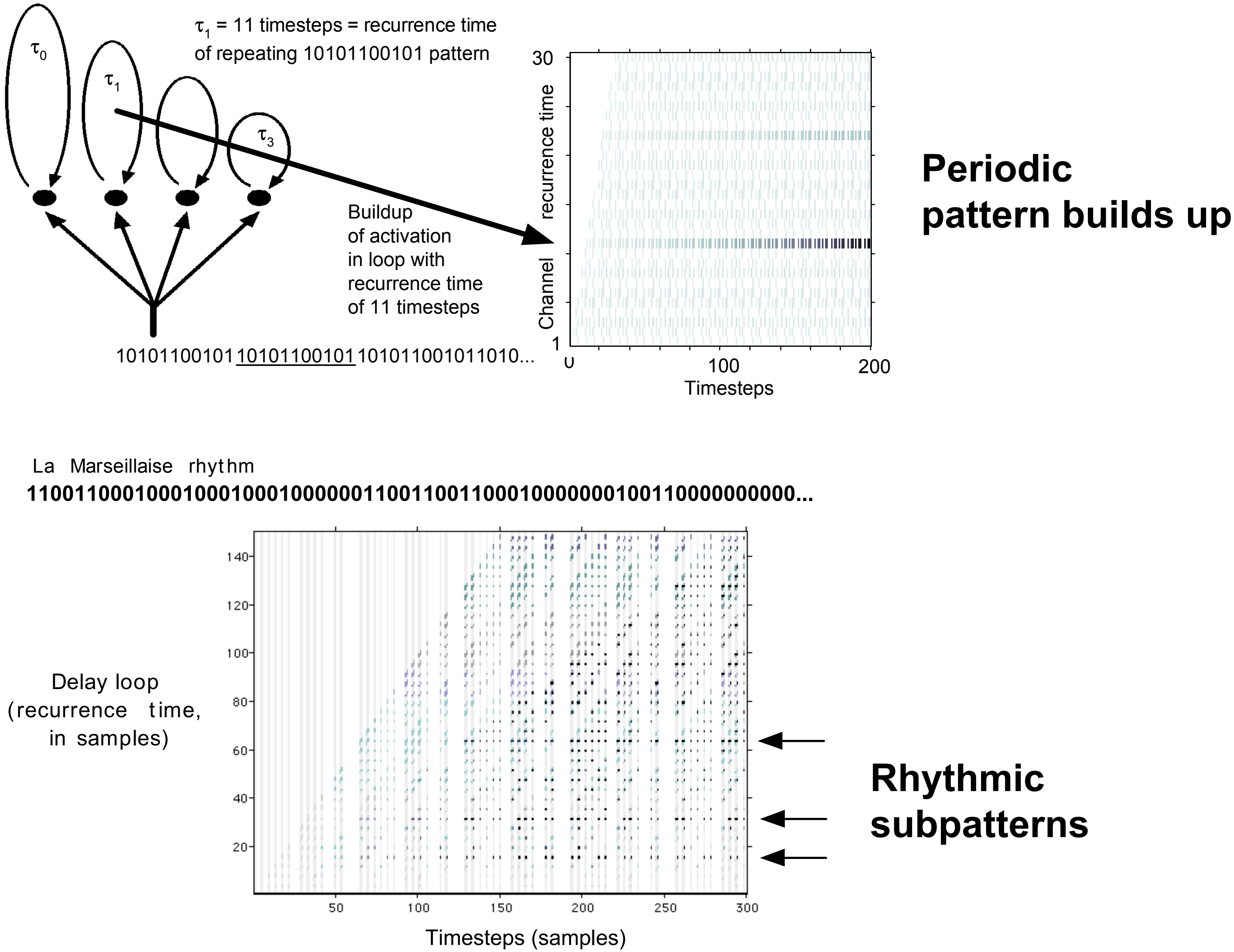
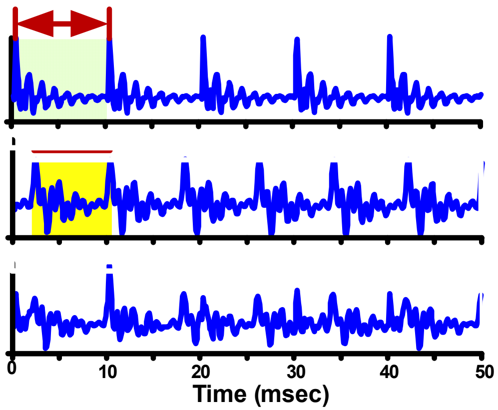


Figure 8. Recurrent tming nets. Top. Behavior of a simple recurrent timing net for periodic pulse train patterns. The network generates many sets of expectations. The delay loop whose recurrence time equals the period of the pattern builds up that pattern. Below. Response of a recurrent timing net to the beat pattern of La Marseillaise. Arrows indicate periodic subpatterns at 16, 32, and 64 timesteps that are built up by the network. The example points to potential applications of recurrent timing nets for rhythm induction and analysis. From Cariani, 1999, working paper on timing nets and rhythm.

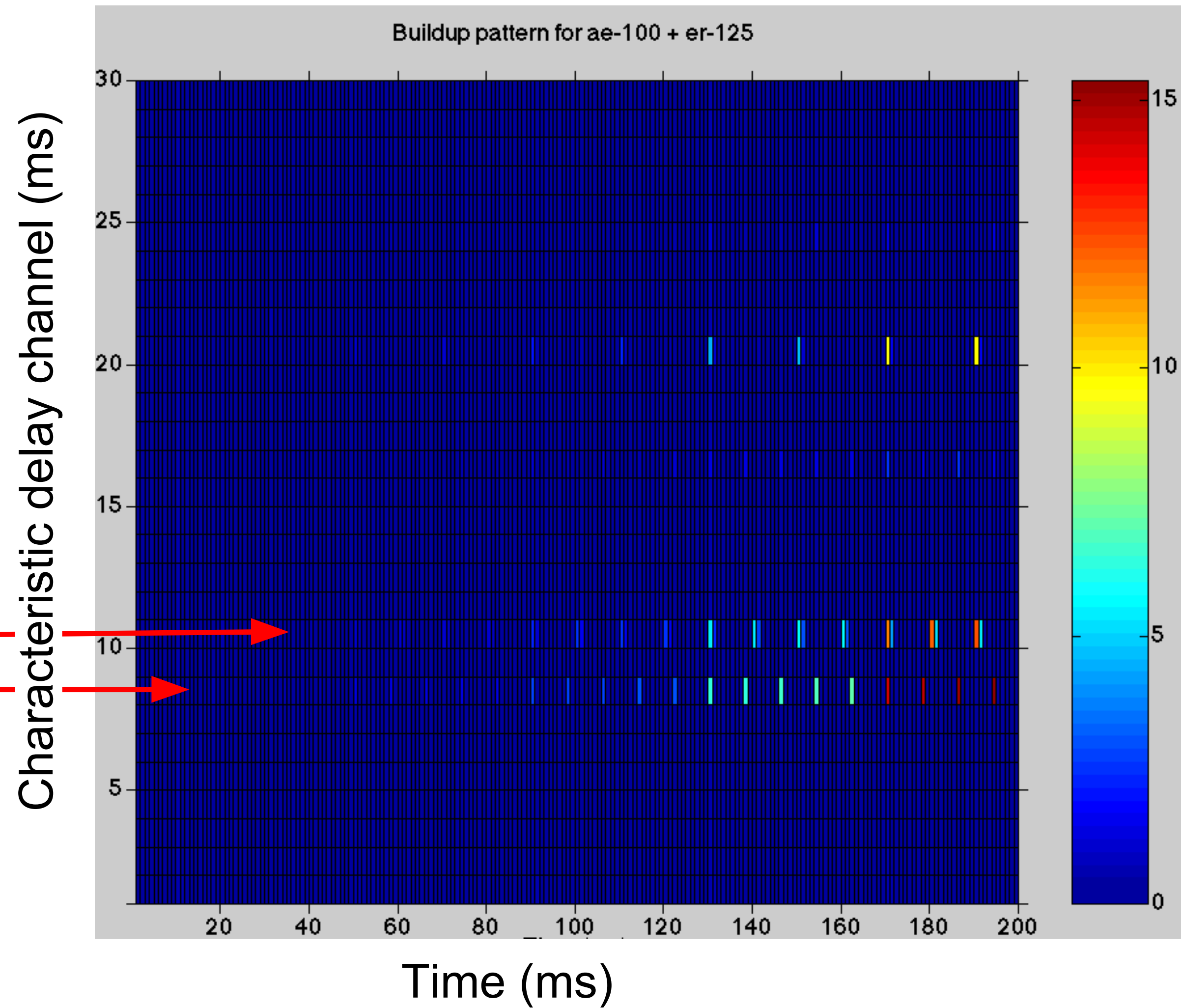
Build-up and separation of two auditory objects

Two vowels with different fundamental frequencies (F0s) are added together and passed through the simple recurrent timing net. The two patterns build up In the delay loops that have recurrence times that correspond to their periods.



Vowel [ae]
F0 = 100 Hz
Period = 10 ms

Vowel [er]
F0 = 125 Hz
Period = 8 ms

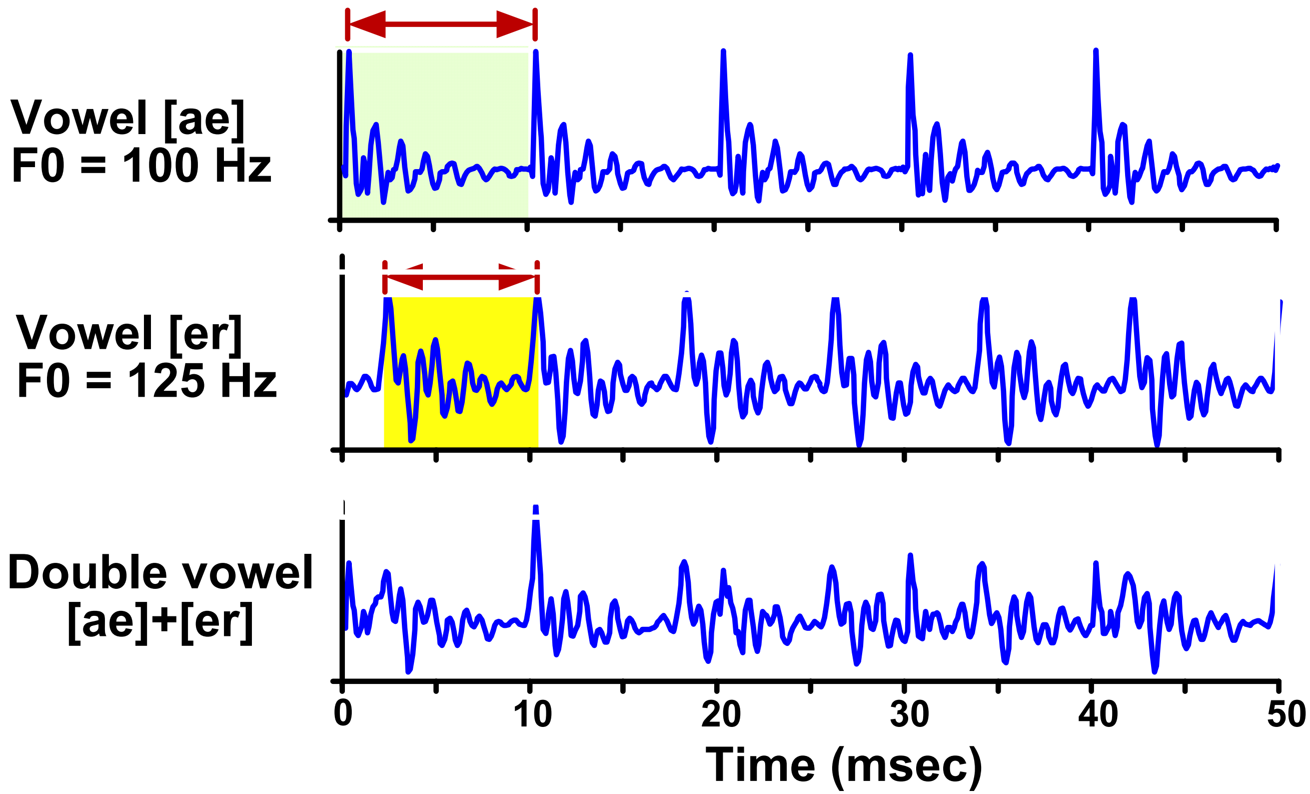


Pitch and auditory scene analysis

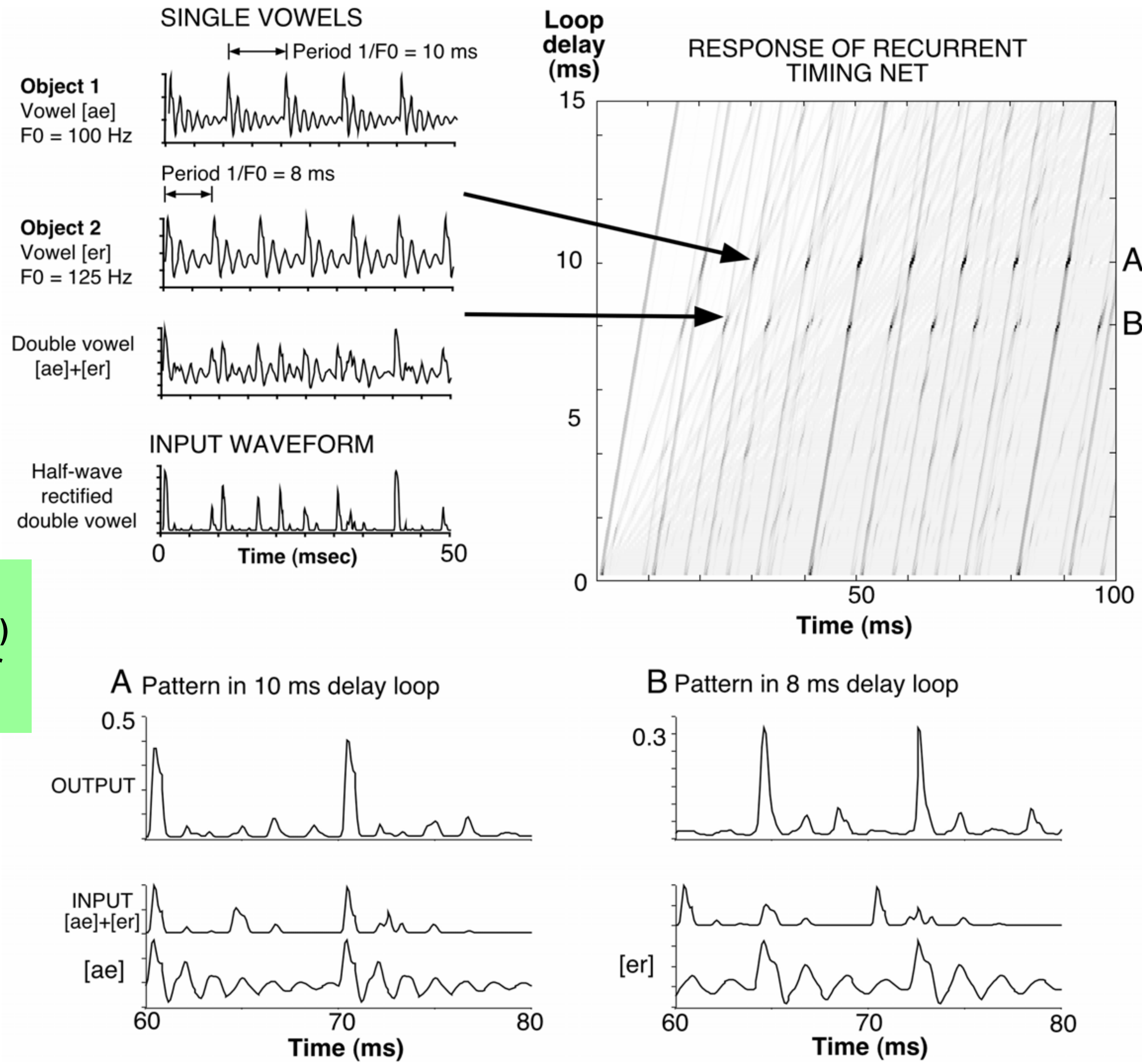
Two different strategies for scene analysis:

1. Group channels by common properties (local features)
2. Build up invariant temporal patterns as objects (invariant relations)

Separation of multiple recurrent time patterns:
two vowels with different F0's

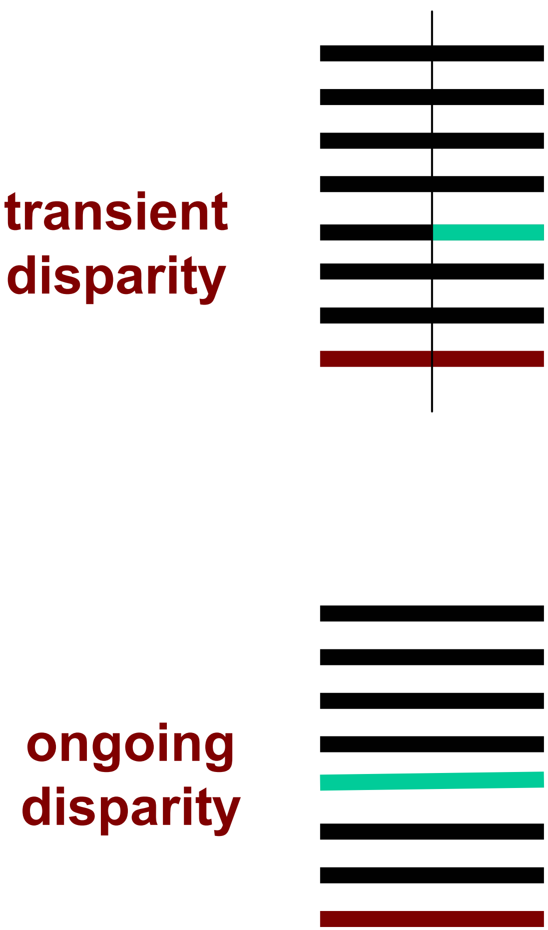


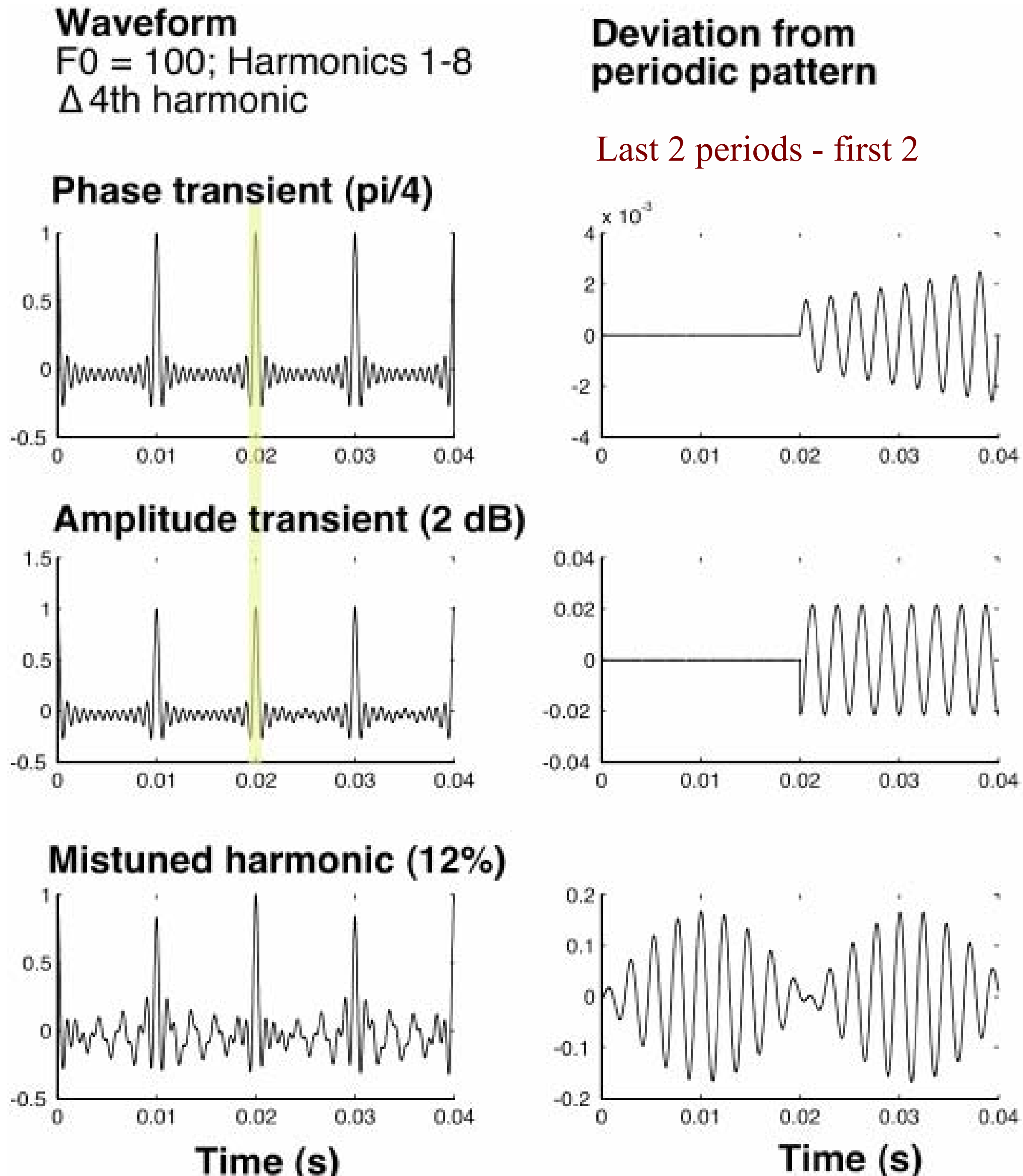
Recurrent timing net with single input channel



Revised buildup rule: Min(direct, circulating) plus a fraction of their absolute difference

Auditory "pop-out" phenomena suggest a period-by-period waveform comparison

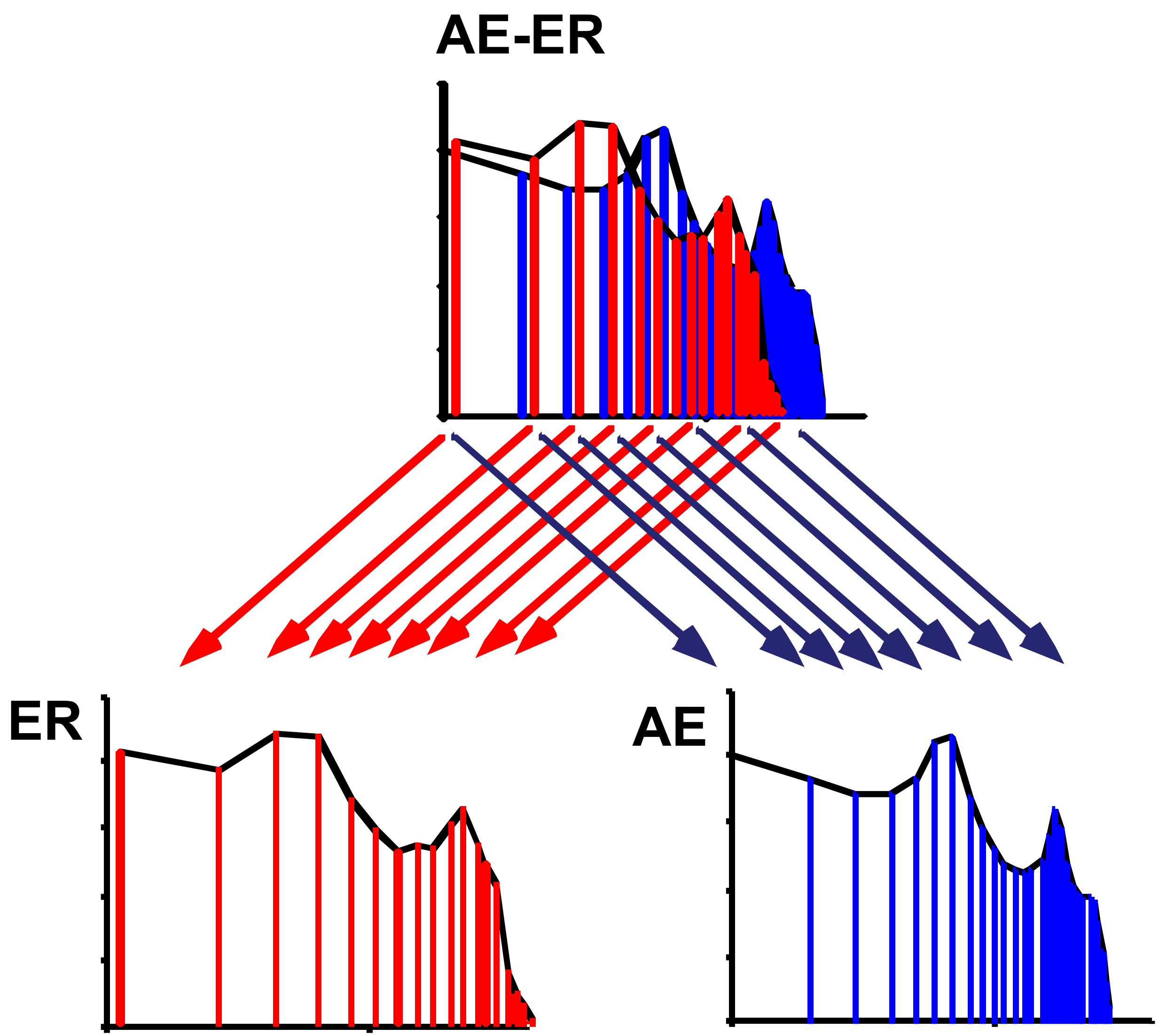




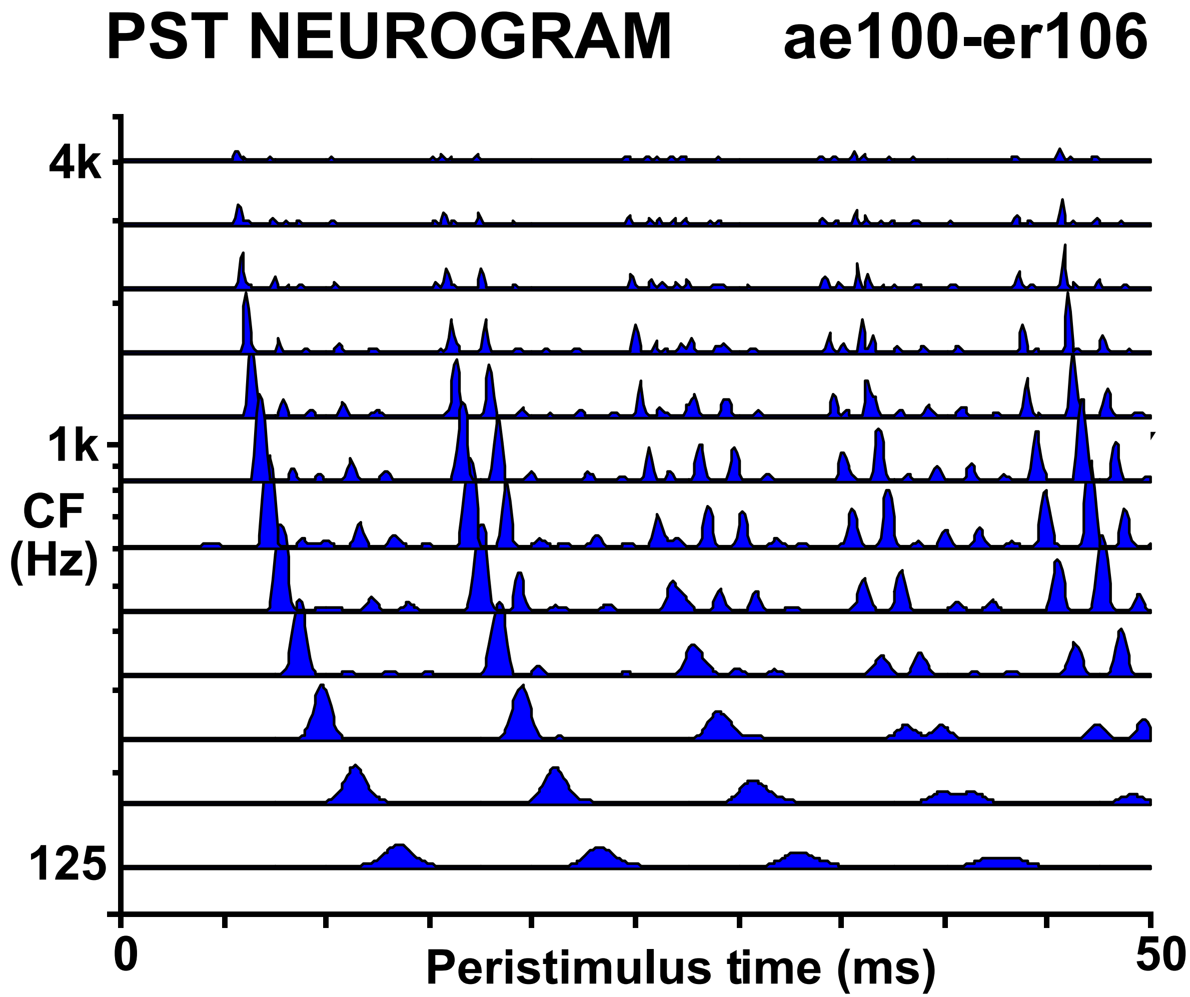
Traditional approach (Frequency domain)

Segregate frequency channels

Assign channels to objects



Double vowels drive the same frequency channels in the auditory nerve

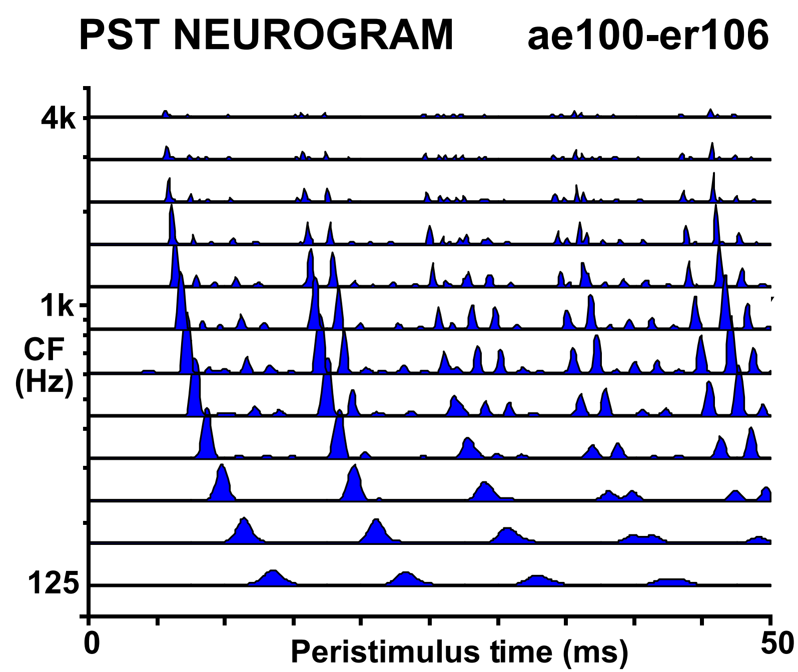


The information needed to separate sounds on the basis of F0 differences is multiplexed in the phase-locked fine time structure of the neural firing patterns

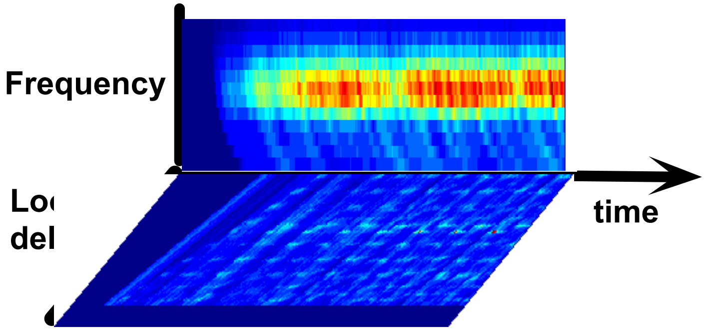
(Traditional spectrographic approaches throw away this information at an early stage.)

How might a neural system (e.g. the auditory system) exploit this information?

Strongly agree that we shouldn't be limited by ignorance of auditory system, but we need to reverse-engineer that system if we are to find new functional principles

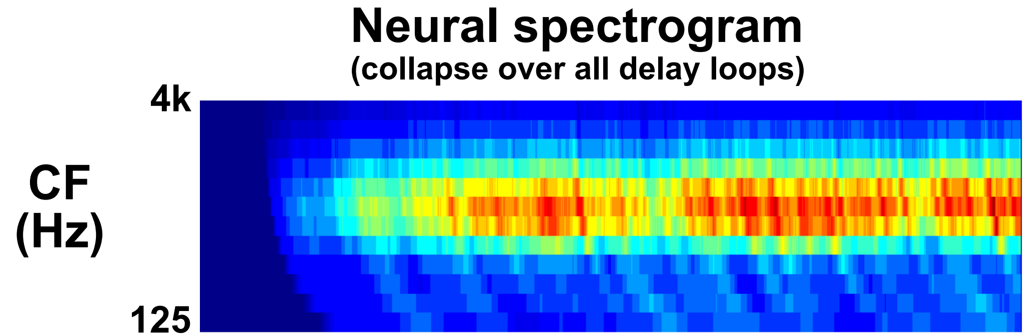


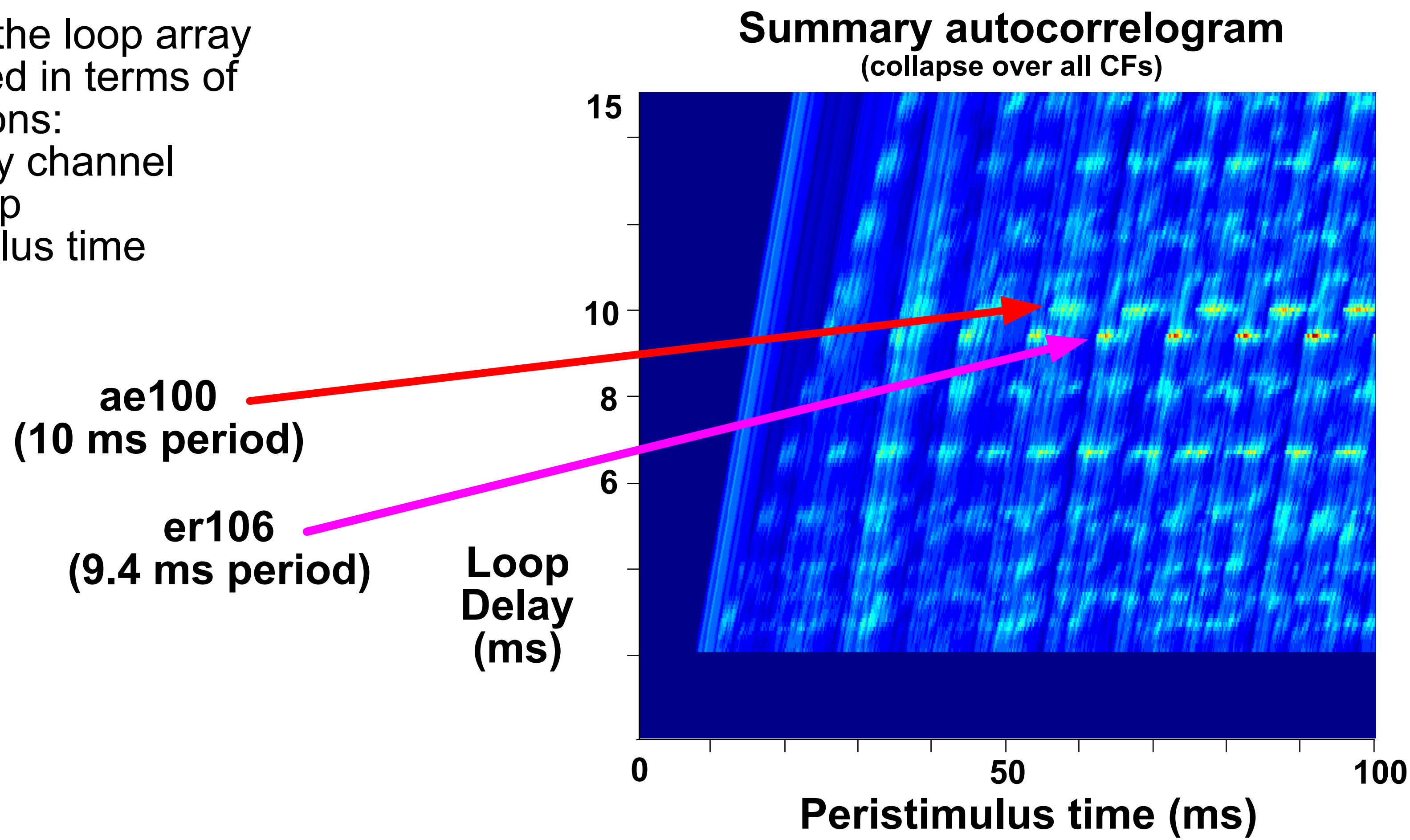
Behavior of the loop array



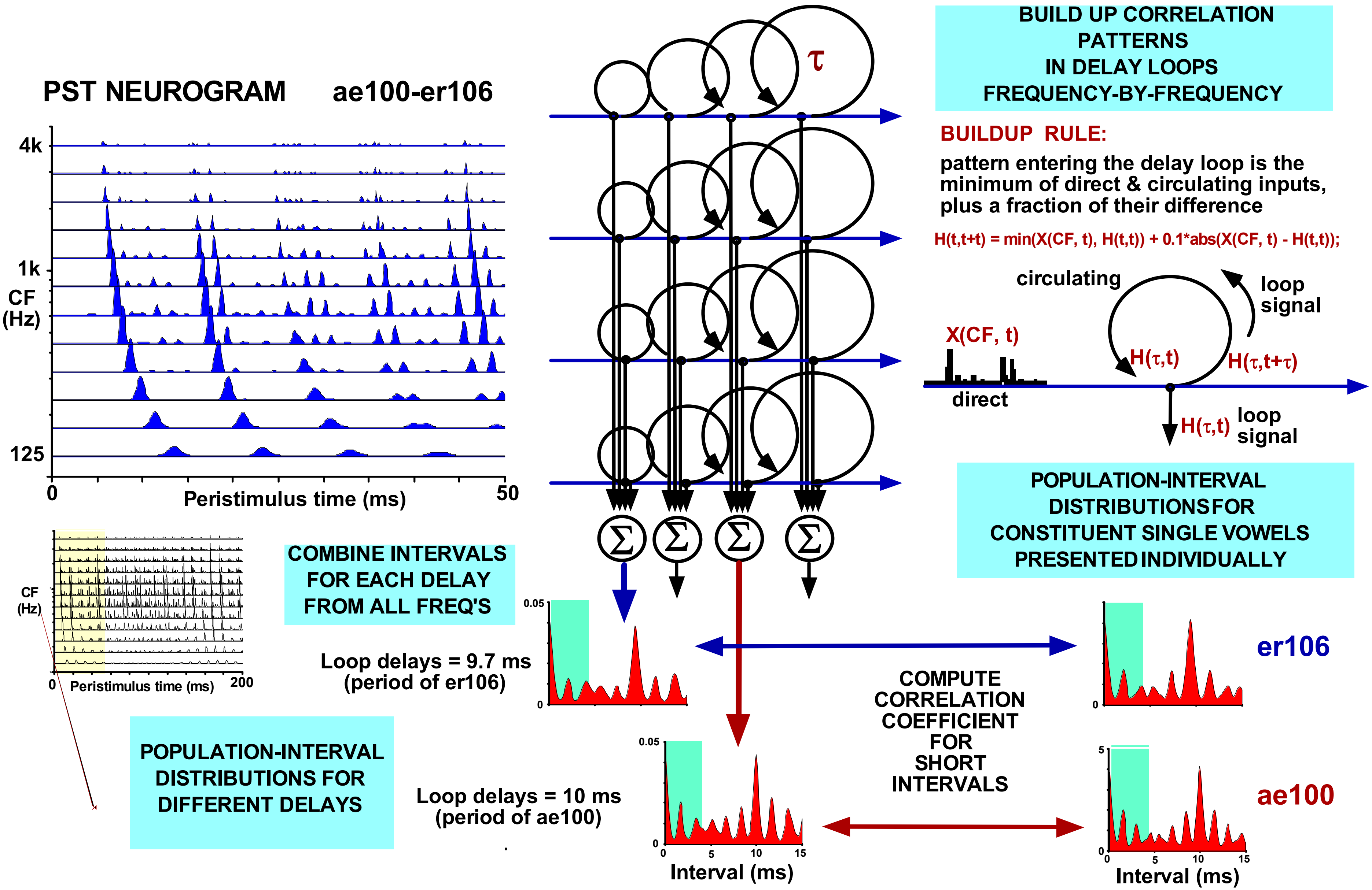
Activity in the loop array is described in terms of 3 dimensions:

- frequency channel
- delay loop
- peristimulus time

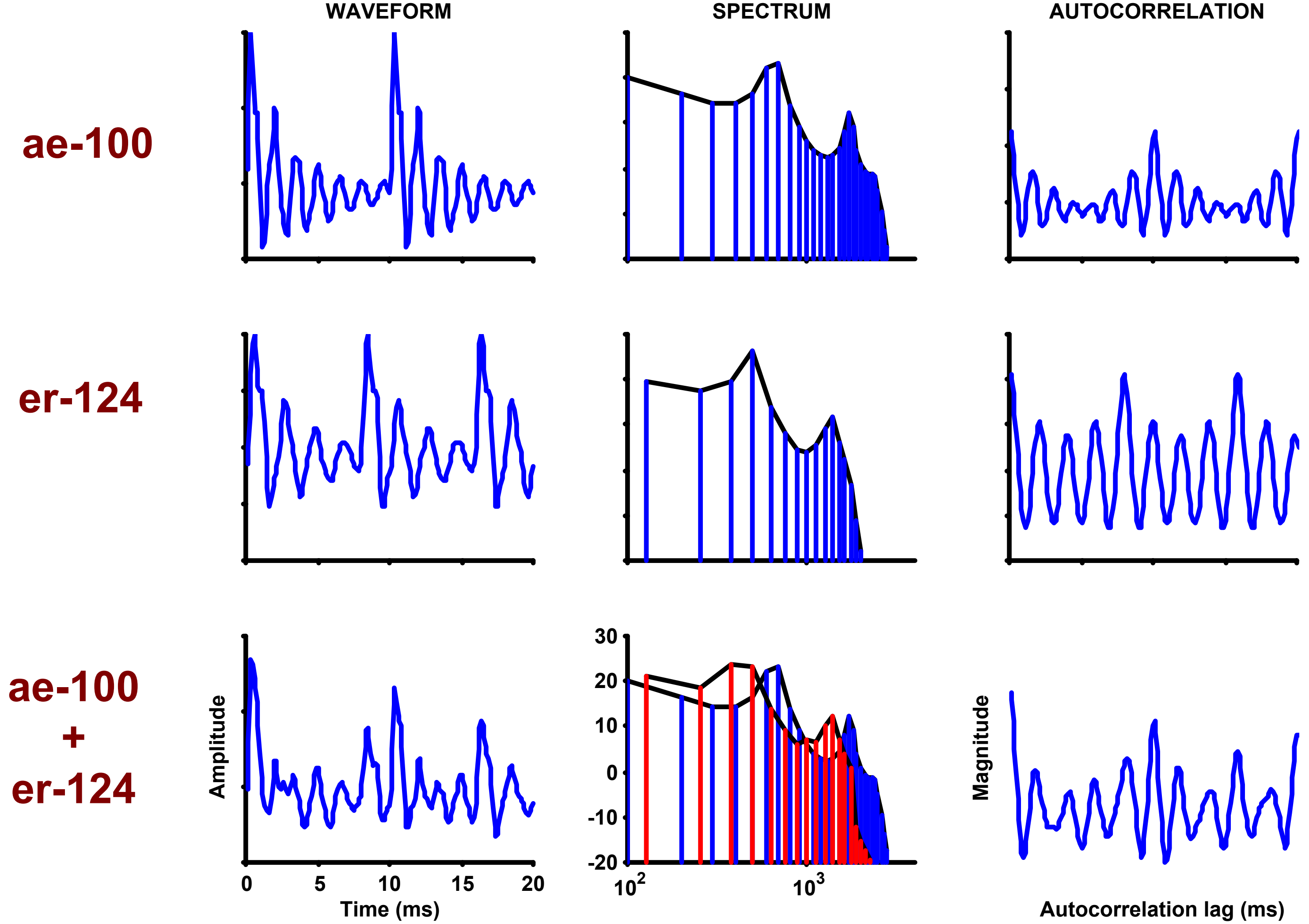




Processing of tonotopically-organized inputs



Concurrent vowels



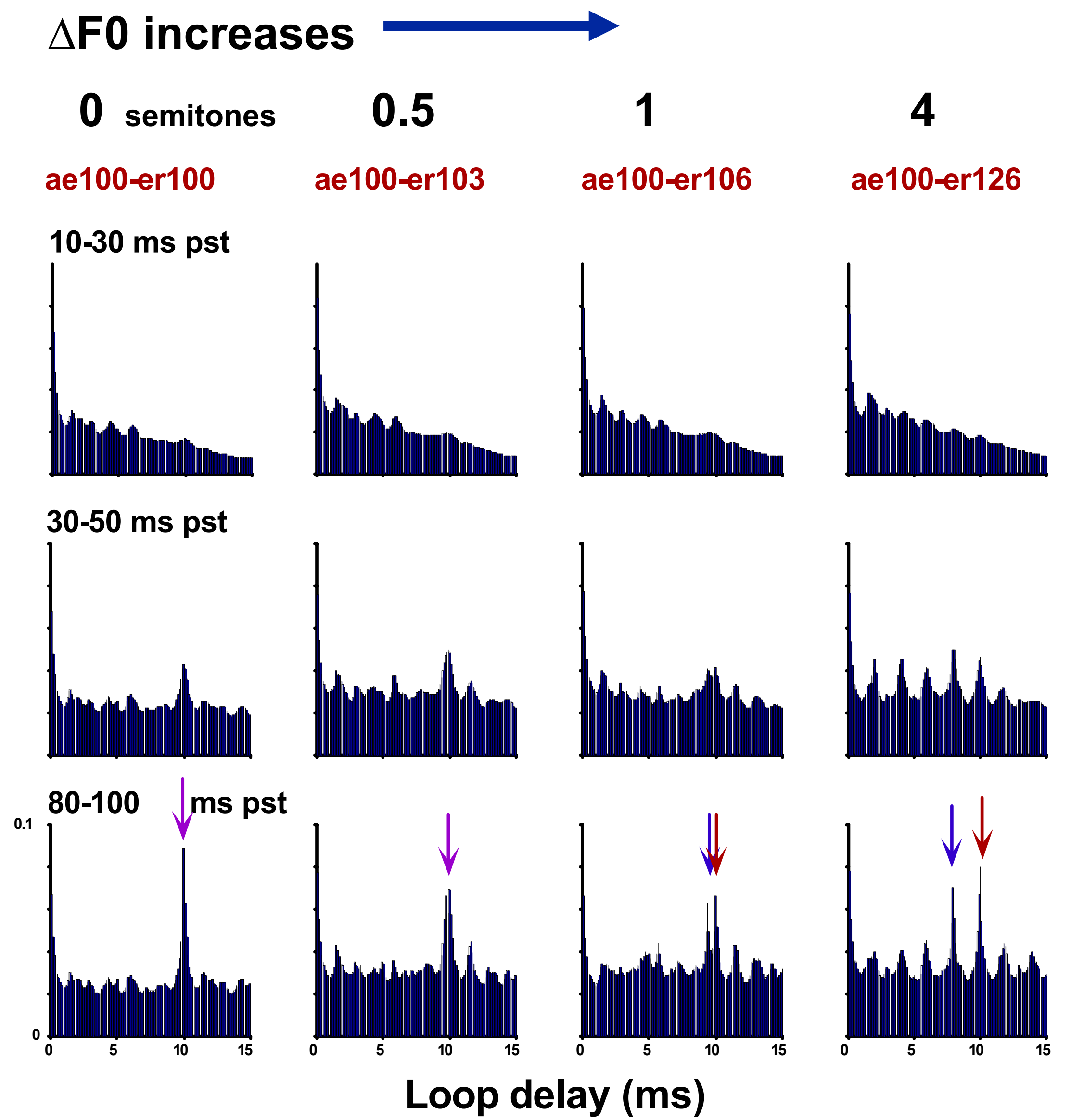
Separation of objects with increasing ΔF0

Those sets of delay loops whose recurrence times equal the fundamental periods of the two vowels are invariably activated the most.

As time goes on, the signals present in these sets of delay loops are amplified.

When the two vowels have the same fundamental (100 Hz), only one set of delay loops is activated (10 ms delay).

When the two vowels have fundamentals separated by a semitone or more (ae100-er106 & ae100-er126) two sets of delay loops are activated. The recurrence times of these sets of delay loops equal the respective vowel periods (10, 9.4, 7.9 ms) .



Simulation performance vs. psychophysics

Simulations results

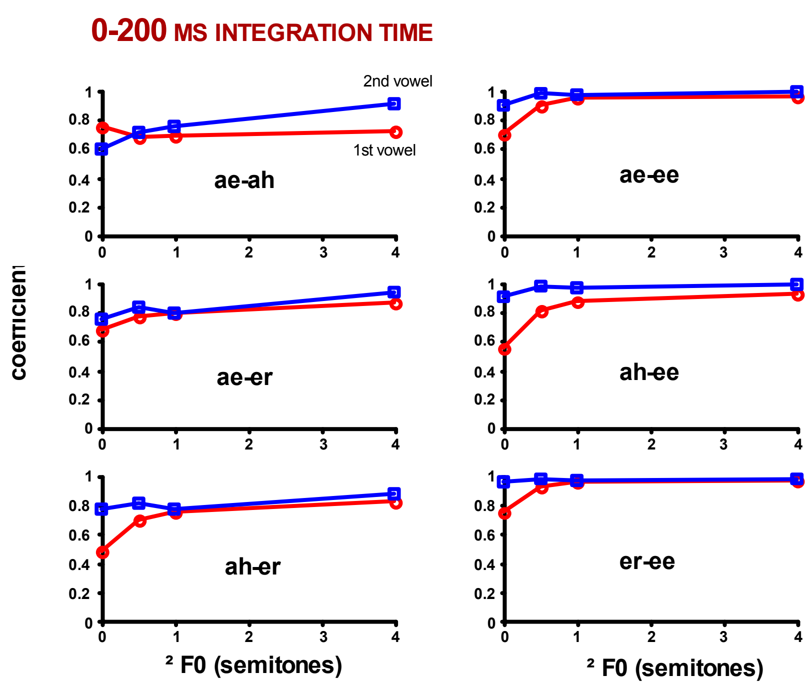
For 200 ms integrations, one (dominant) vowel produced high correlations for all ²F0's.

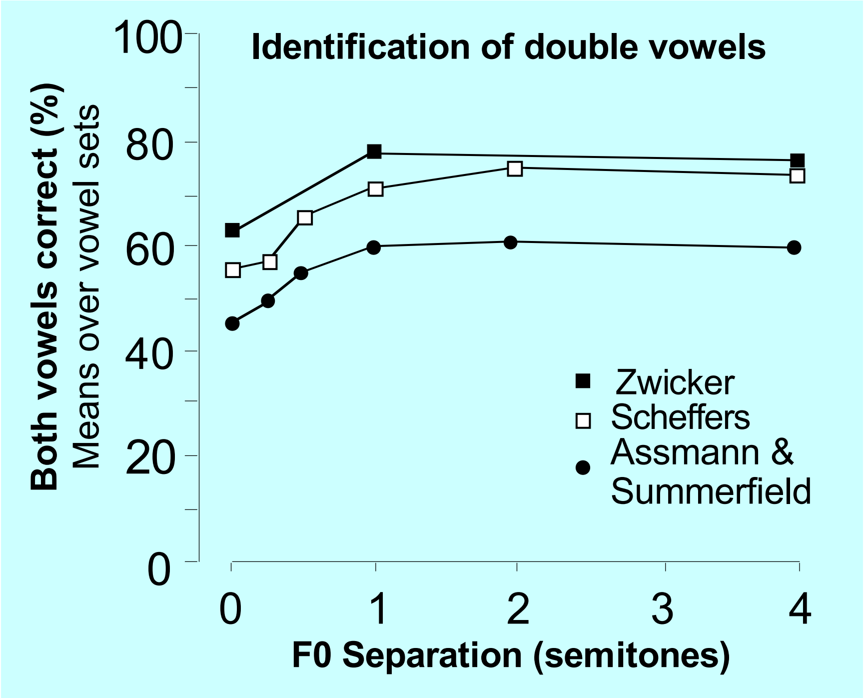
Correlations of nondominant vowels increased as ²F0's increased to a semitone, then levelling off.

Psychophysical observations

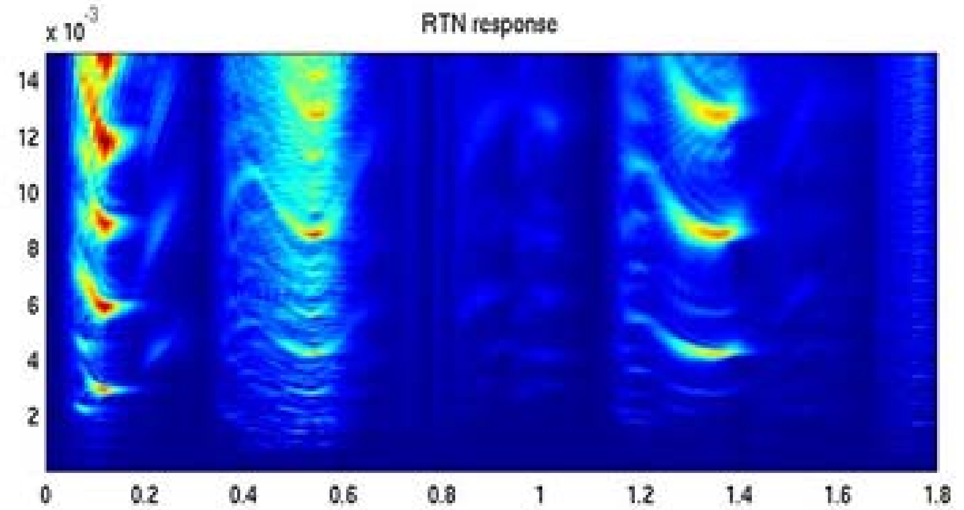
Listeners almost invariably get one vowel correct.

Correct identification of both vowels is slightly enhanced (~15%) by F0 differences of a semitone or more, where performance saturates.





Potential applications



- Pitch tracking of speakers
  & recovery of waveforms
- Separation of multiple speakers for speech recognition
- Enhance voiced speech segments in background noise
- Separation of tonal music from background noise
- Analogous 2-D nets for image stabilization and object
  formation from correlational invariants
- Reverberating temporal memories via transiently-
  facilitated synapses: periodic stimuli
  dynamically create loops on the fly

Conclusions

- Population-wide distributions of all-order interspike intervals in early stations of the auditory pathway explain many diverse aspects of pitch and timbre perception: level-invariance, discriminatory precision, pitch equivalence, octave similarity, pitch fusion.

- These distributions form autocorrelation-like temporal representations that convey spectral information up to the limits of strong phase-locking (~5 kHz).

- The mechanisms by which the central auditory system might make use of this information remain enigmatic.

- Neural timing nets can be envisioned that operate entirely in the time domain to carry out musical interval processing and scene analysis.

Reading/assignment for next meeting

- **Thursday. Feb. 26**
- **Timbre**

- **Reading:**

  Moore Chapter 8, pp. 269-273

  Chapter in Deutsch by Risset & Wessel on Timbre

  (first sections up to p. 113-118)

  Handel, chapter on Identification