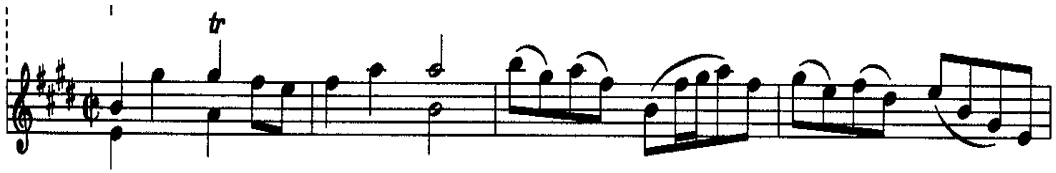
Harvard-MIT Division of Health Sciences and Technology
HST.725: Music Perception and Cognition
Prof. Peter Cariani



Timbre perception

www.cariani.com

Timbre perception

- **Timbre: tonal quality** (≠ pitch, loudness, duration or location)
- Defines separate voices, musical coloration
- Multidimensional space: not completely well understood
- Two general aspects: spectrum & dynamics
- Stationary spectrum
  - Spectral center of gravity - "brightness"
  - Formant structure;
  - Harmonicity
- Amplitude-frequency-phase dynamics
  - Amplitude dynamics (attack, decay)
    - amplitude modulation (roughness)
  - Frequency dynamics
    - relative timings of onsetsand offsets of partials
    - frequency modulation (vibrato)
  - Phase dynamics (noise, phase coherence, chorus effect)
- Analogy with phonetic distinctions in speech
    - Vowels (stationary spectra; formant structure)
    - Consonants (dynamic contrasts: amplitude, frequency & noise)
- Temporal integration windows and timbral fusion
- Some neural correlates

Timbre ~ sonic texture, tone color

Please see Paul Cezanne, Apples, Peaches, Pears, and Grapes c. (1879-80); Oil on canvas, 38.5 x 46.5 cm; The Hermitage, St. Petersburg at (http://www.ibiblio.org/wm/paint/auth/cezanne/sl/)

Timbre ~ sonic texture, tone color

Texture
Roughness

Stationary and dynamic factors in timbre perception

- **Periodicity (noise-like or tone-like)**
  - Harmonicity (is this properly an aspect of timbre?)
  - Phase coherence (noise-incoherent; tones-coherent)
  - Smoothness or roughness
- **Stationary spectrum**
  - Formant structure;
- **Amplitude-frequency-phase dynamics**
  - Amplitude dynamics (attack, decay)
    - amplitude modulation (roughness)
  - Frequency dynamics
    - relative timings of onsetsand offsets of partials
    - frequency modulation (vibrato)
  - Phase dynamics (phase shifts, chorus effect)
- **Analogy with phonetic distinctions in speech**
    - Vowels (stationary spectra; formant structure)
    - Consonants (dynamic contrasts: amplitude, frequency & noise)

Stationary and dynamic factors in timbre perception

- **Stationary spectrum**
  - **Formant structure; Harmonicity**
- **Amplitude-frequency-phase dynamics**
  - **Amplitude dynamics (attack, decay)**
    - **amplitude modulation (roughness)**
  - **Frequency dynamics**
    - **relative timings of onsetsand offsets of partials**
    - **frequency modulation (vibrato)**
  - **Phase dynamics (noise, phase coherence, chorus effect)**
- **Analogy with phonetic distinctions in speech**
    - **Vowels (stationary spectra; formant structure)**
    - **Consonants (dynamic contrasts: amplitude, frequency & noise)**

Timbre: a multidimensional tonal quality

tone texture, tone color distinguishes voices, instruments

Photo Courtesy of Pam Roth.

Stationary Aspects

(spectrum)

Dynamic Aspects

Δ spectrum
Δ intensity
Δ pitch
attack
decay

Photo Courtesy of Per-Ake Bystrom.

Vowels

Consonants

Photo Courtesy of Miriam Lewis.

http://www.wikipedia.org/

Harmonicity
Frequency dynamics

violin, trumpet, guitar
(more harmonic,
stationary spectra)

Rafael A. Irizarry's Music and Statistics Demo

Spectrograms of Harmonic Instruments

Non-Harmonic Instruments

marimba, timpani, gong
(more inharmonic,
time-varying spectra)

http://www.biostat.jhsph.edu/~ririzarr/Demo/demo.html

Some methods for studying the perceptual space

1. Try to derive the structure of the space from the dimensionality of responses

    - Similarity magnitude estimations
    - Similarity rankings
    - Multidimensional scaling

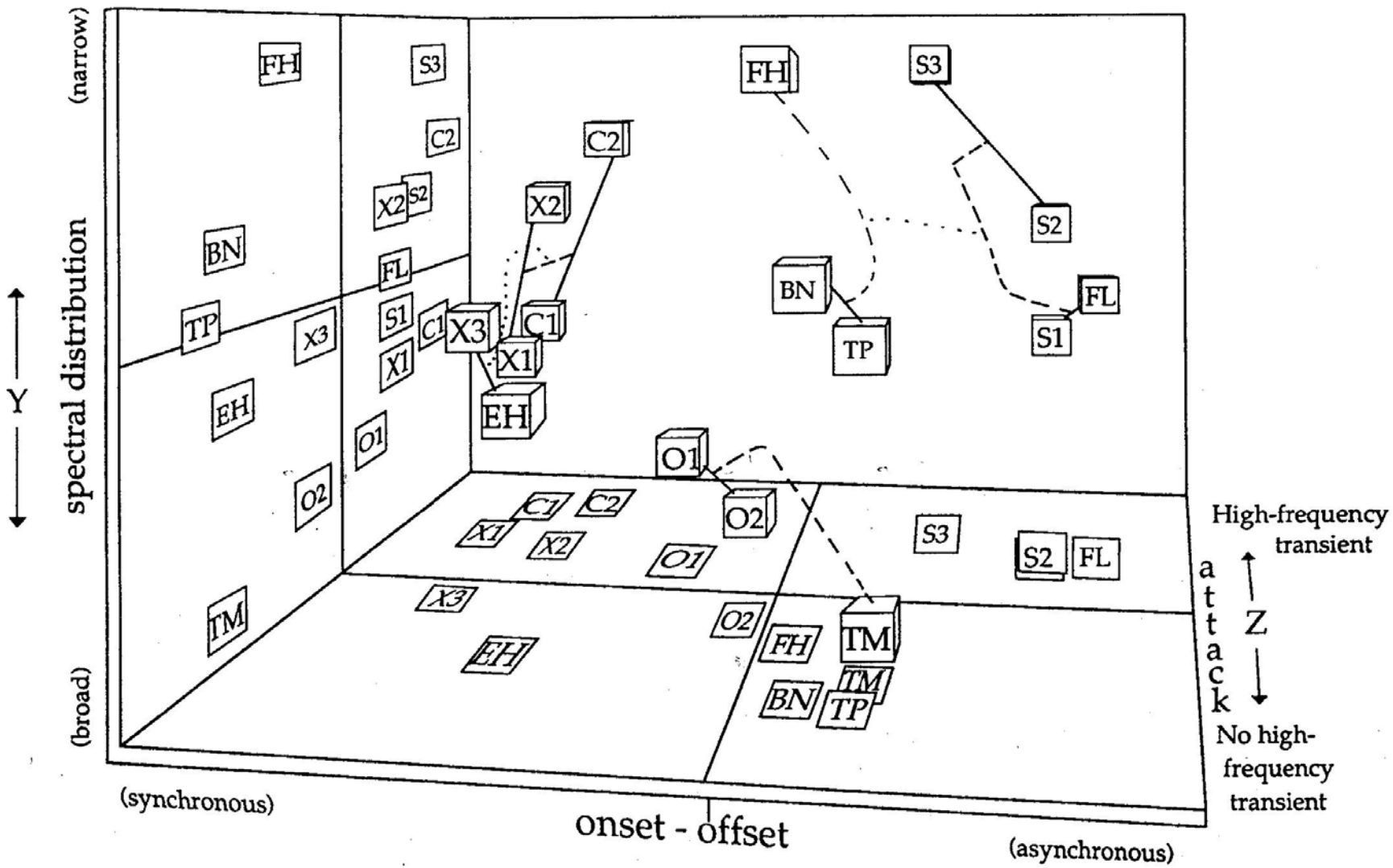
2. Systematically vary acoustic parameters known to influence timbre to find acoustic correlates of perceptual dimensions, e.g.
    - Formant structure
    - Attack and decay parameters

Grey (1975) Timbre: Perceptual dimensions

Figure from Butler, David. *The Musician's Guide to Perception and Cognition.* Schirmer, 1992.

Also see: Grey, J., and J. Moorer. "Perceptual Evaluations of Synthesized Musical Instrument Tones." *J. Acoustical Society of America* 63 (1977): 1493-1500.

Timbre dimensions: spectrum, attack, decay



Music based on timbral contrasts

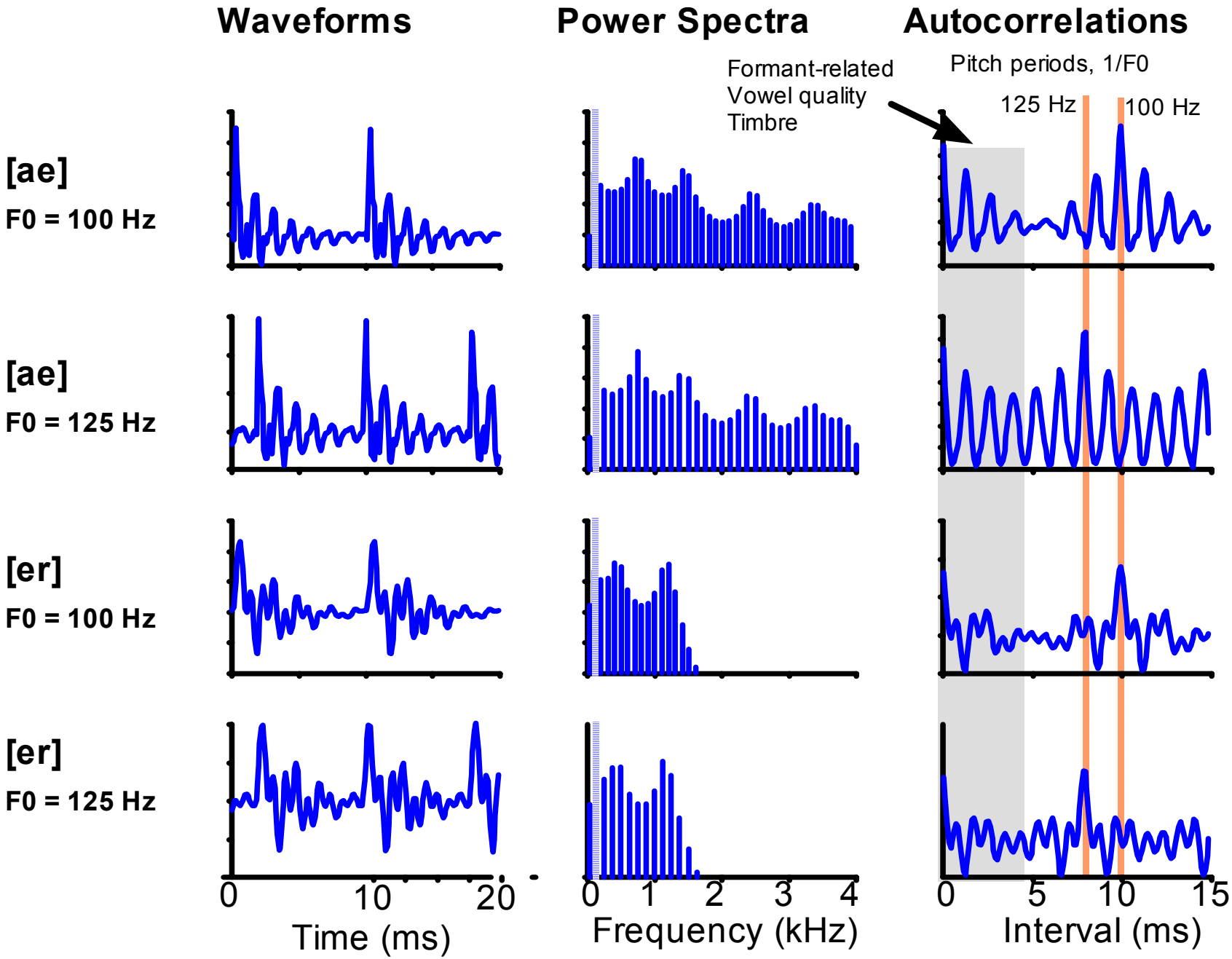
Kurt Schwitters,
Ur Sonata (1932)
perf. George Melly, Miniatures

Miniatures [Pipe/Cherry Red]
Performer(s): Various Artists

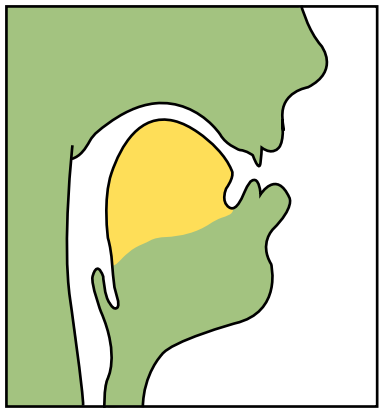
Label: Cbc Records/Musica Viva (Can)
Catalog: #1043

Audio CD (January 19, 1997)
Number of Discs: 1
ASIN: B000003WYQ

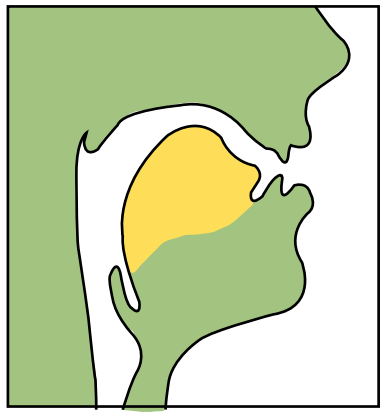
Stationary spectral aspects of timbre



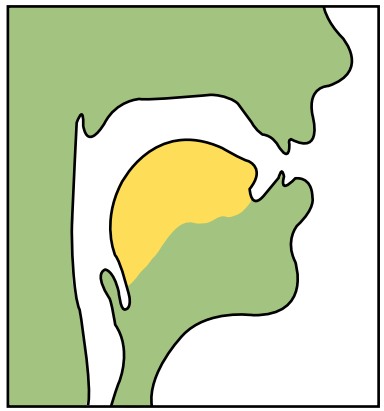
Formants and the vocal tract



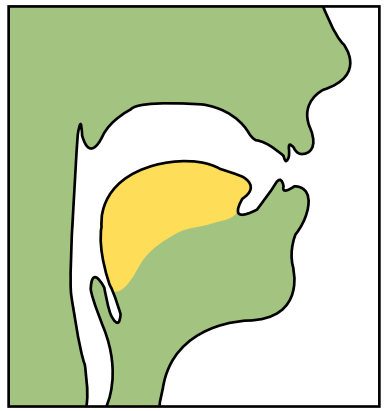
Heed



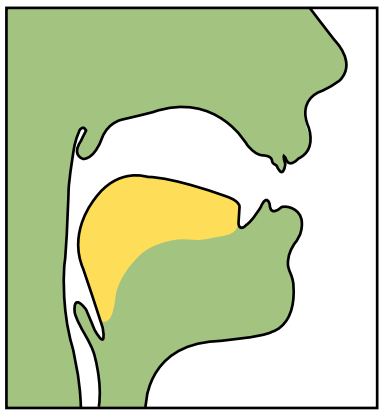
Hid



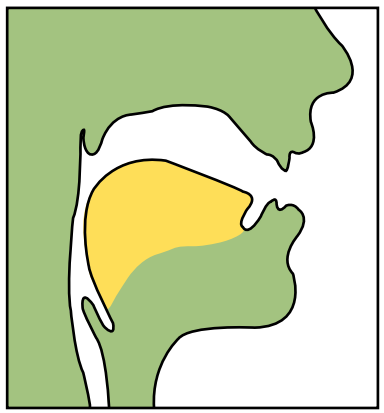
Head



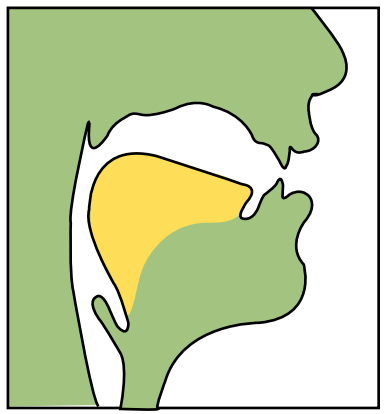
Had



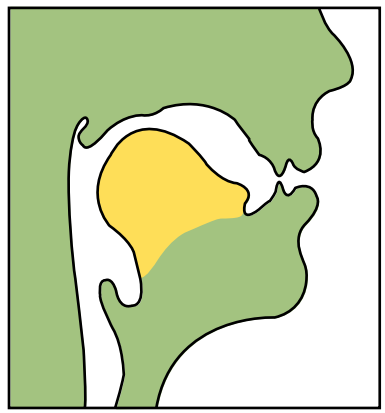
Hod



Hawed

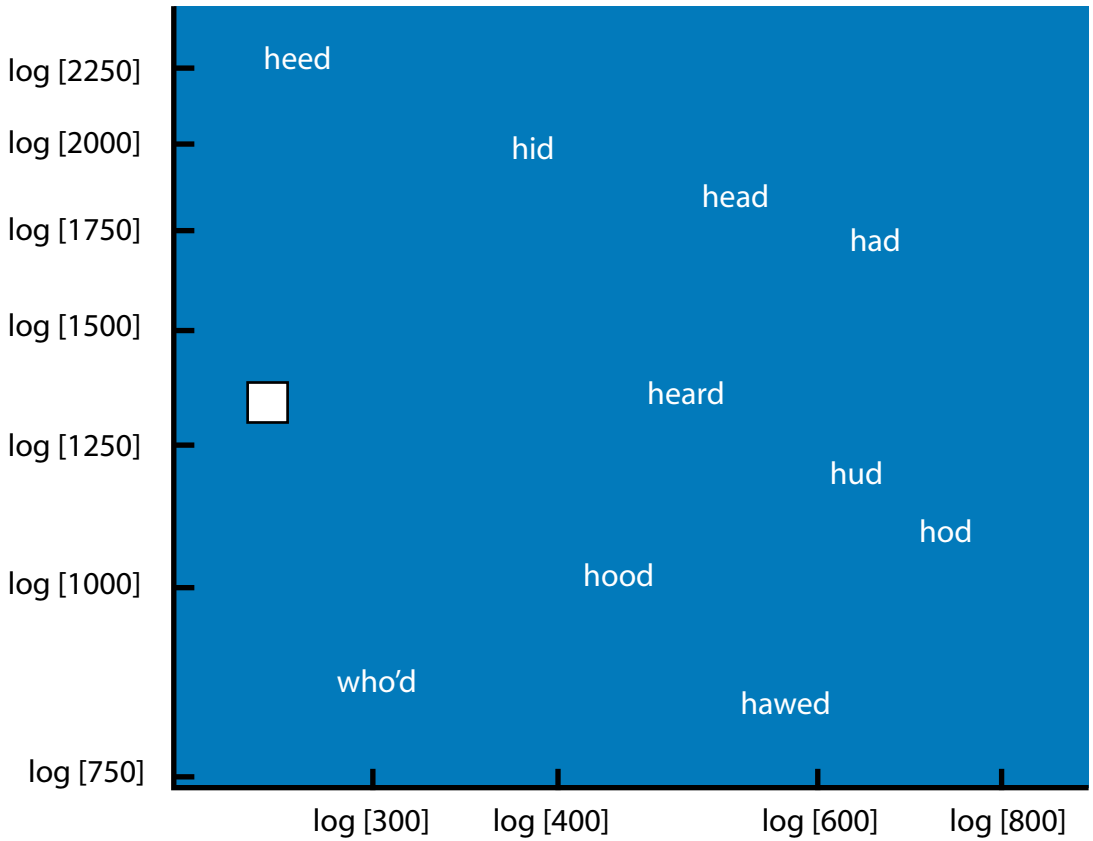


Hood

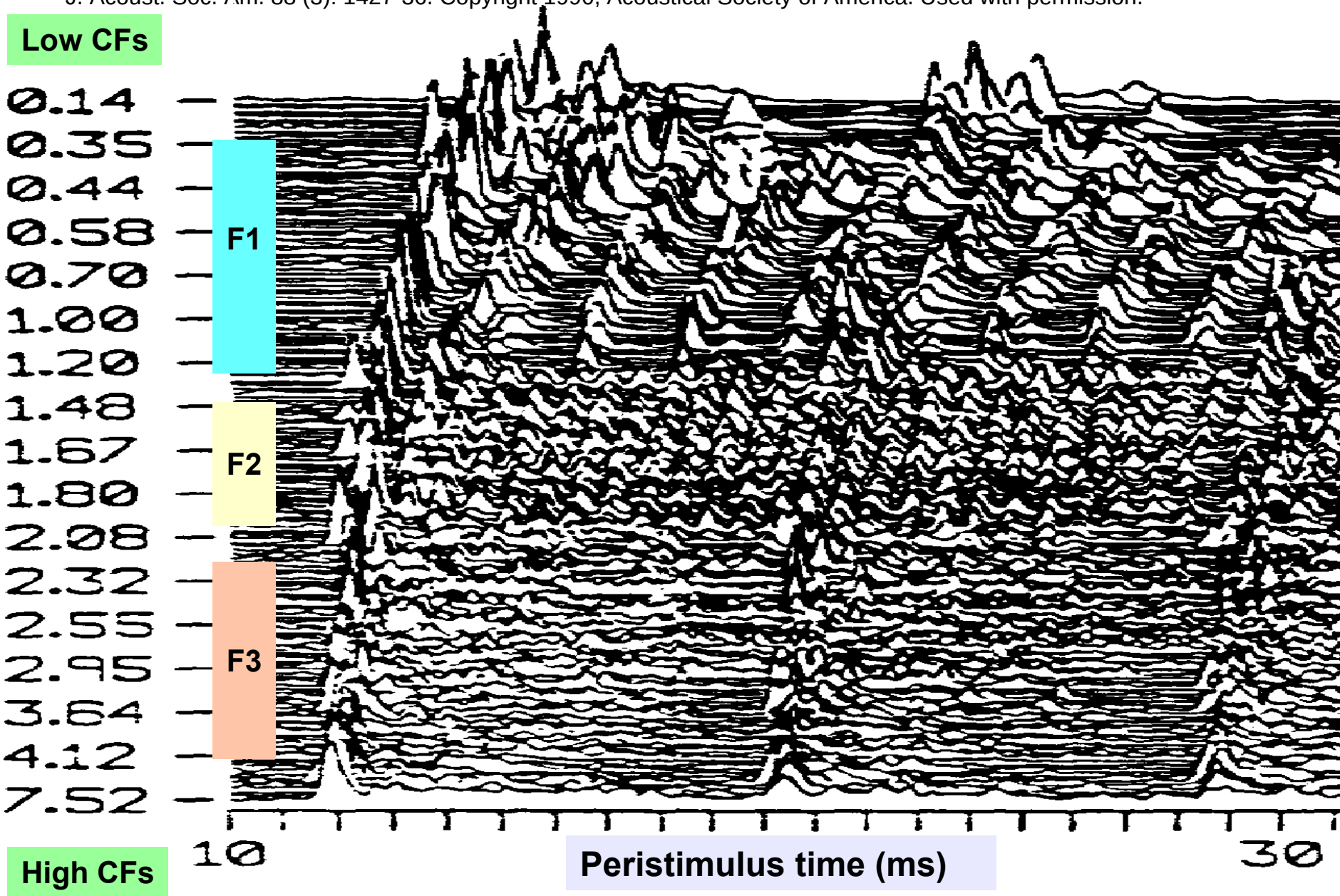


Who’d

Vowel F1-F2 space

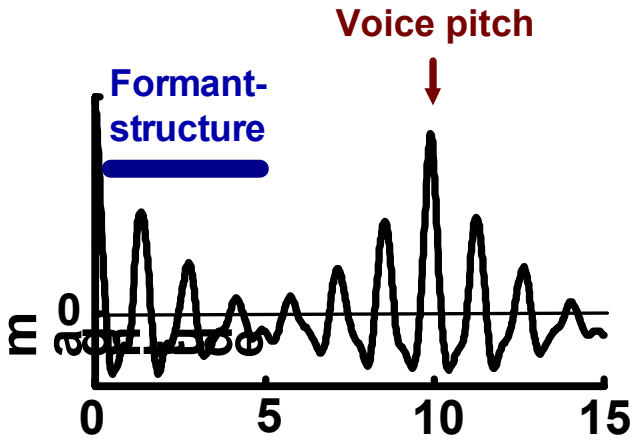


Time domain analysis of auditory-nerve fiber firing rates.
Hugh Secker-Walker & Campbell Searle, J. Acoust. Soc. 88(3), 1990
Neural responses to /da/ @ 69 dB SPL from Miller and Sachs (1983)

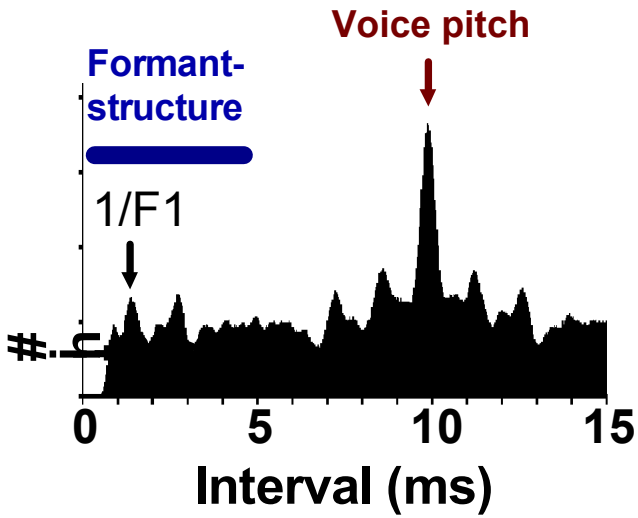


Population-interval coding of timbre (vowel formant structure)

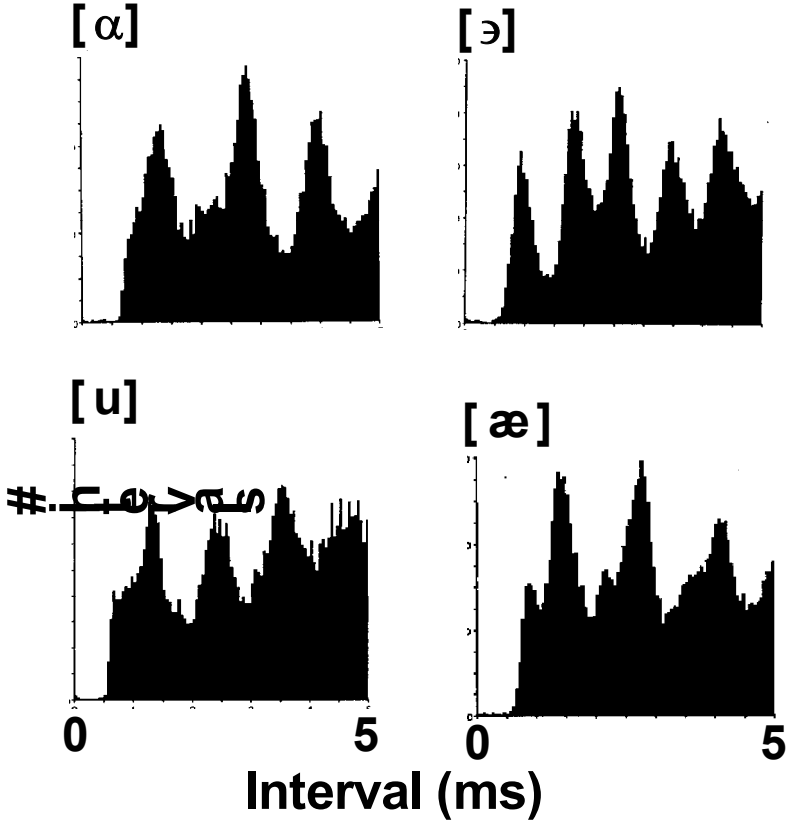
Signal autocorrelation [ae]



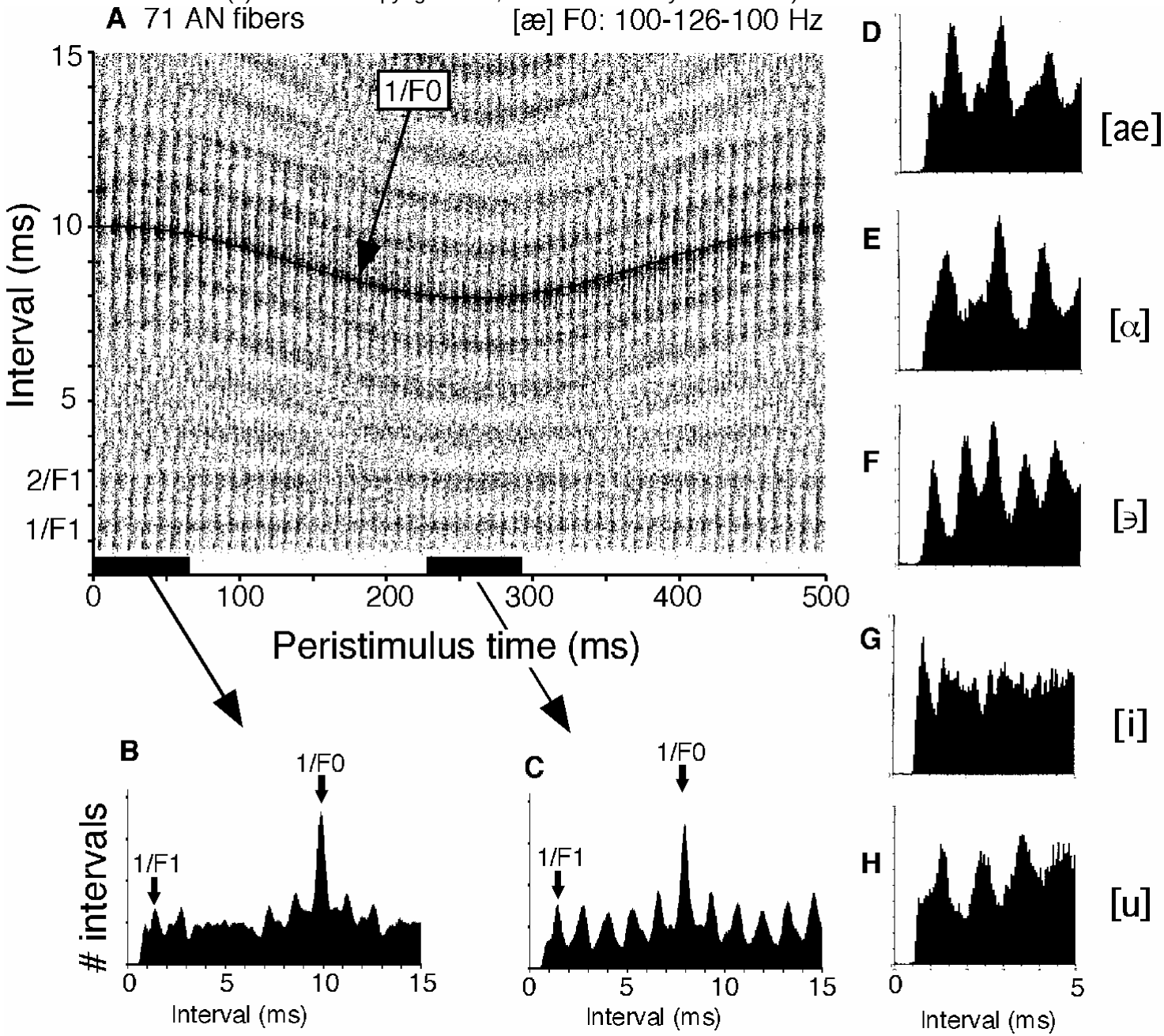
Population interval histogram



Population-wide distributions of short intervals for 4 vowels



Coding of vowel quality (timbre)



Please See Hirahara, Cariani, Delgutte (1996)

Please see Figures 6, and 7 in Hirahara, Cariani, Delgutte (1996)

Spectrum as a function of intensity (trumpet)

Please see Figure 4-3 in Butler, David. *The Musician's Guide to Perception and Cognition*. New York: Schirmer Books ; Toronto: Maxwell Macmillan Canada, New York: Maxwell Macmillan International, c1992.

Singer's formant

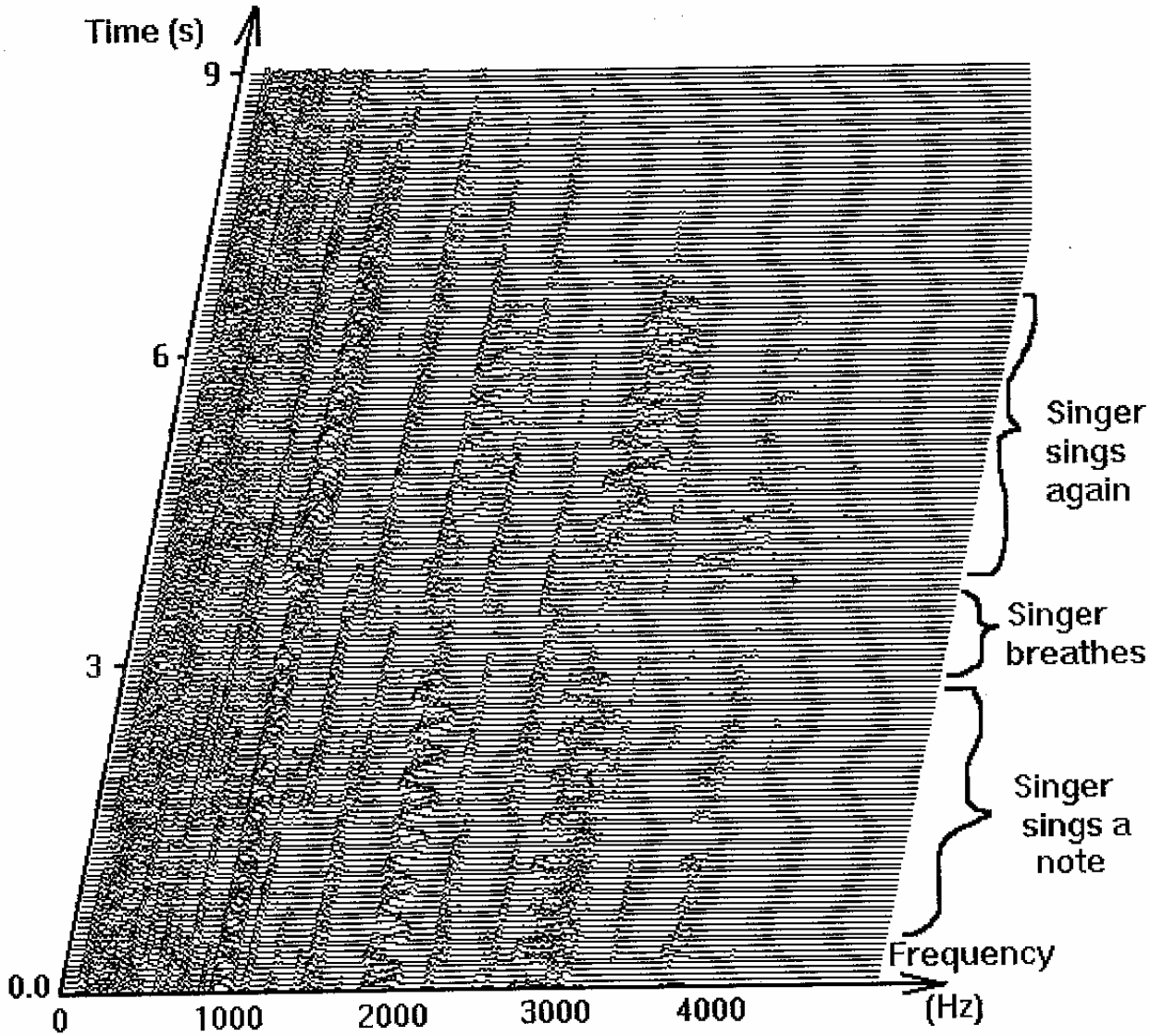
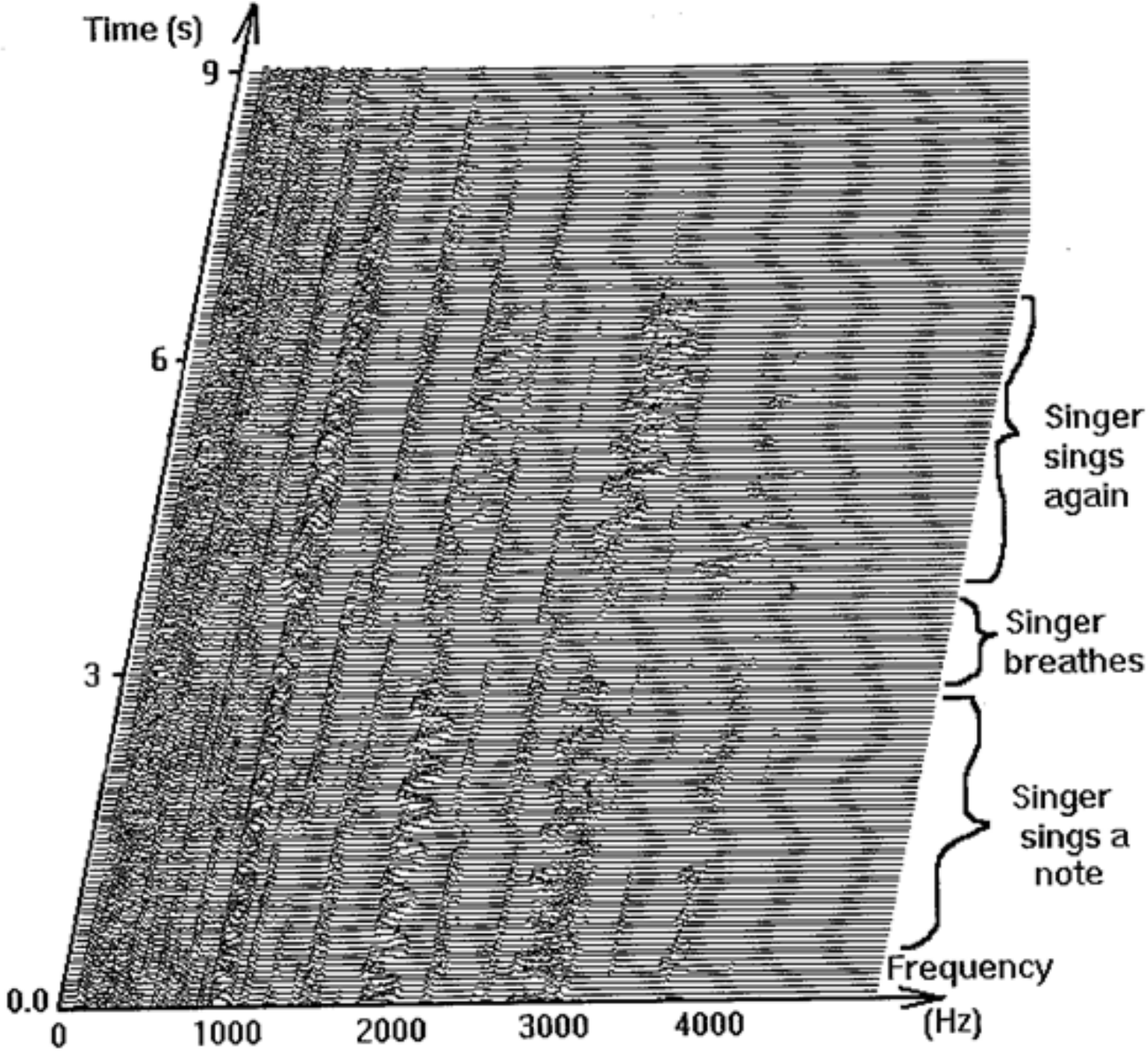


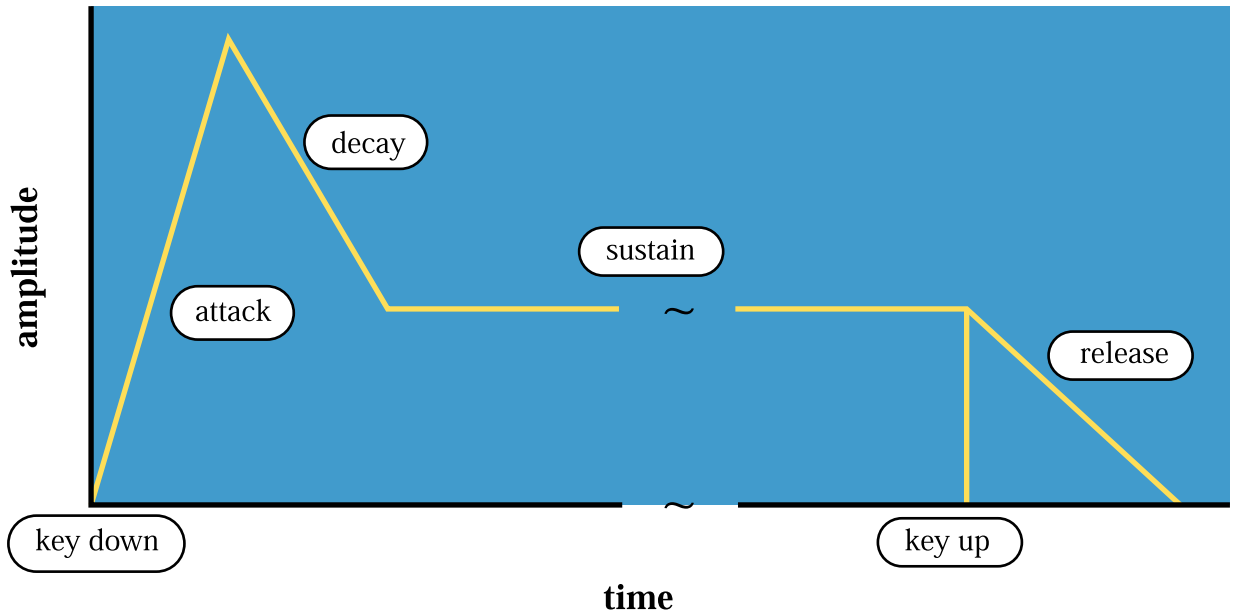
Figure 11.12 The singer's formant is evident in this waterfall plot of the last two notes of the soprano aria "Un Bel Di Vedremo," from Puccini's *Madame Butterfly.* Common frequency modulation of the first three partials allows the fundamental to be picked out visually.

Singer's formant



Graph, Fig 11.12 on page 138 of Cook, Perry, ed. Music, Cognition & Computerized Sound MIT Press 2001. Used with permission.

Amplitude dynamics (envelope, intensity contour)



Frequency dynamics of note onsets (clarinet)

Please see Figure 4-4 in Butler, David. *The Musician's Guide to Perception and Cognition*. New York: Schirmer Books ; Toronto: Maxwell Macmillan Canada, New York: Maxwell Macmillan International, c1992. ISBN: 0028703413.

Time-course of harmonics

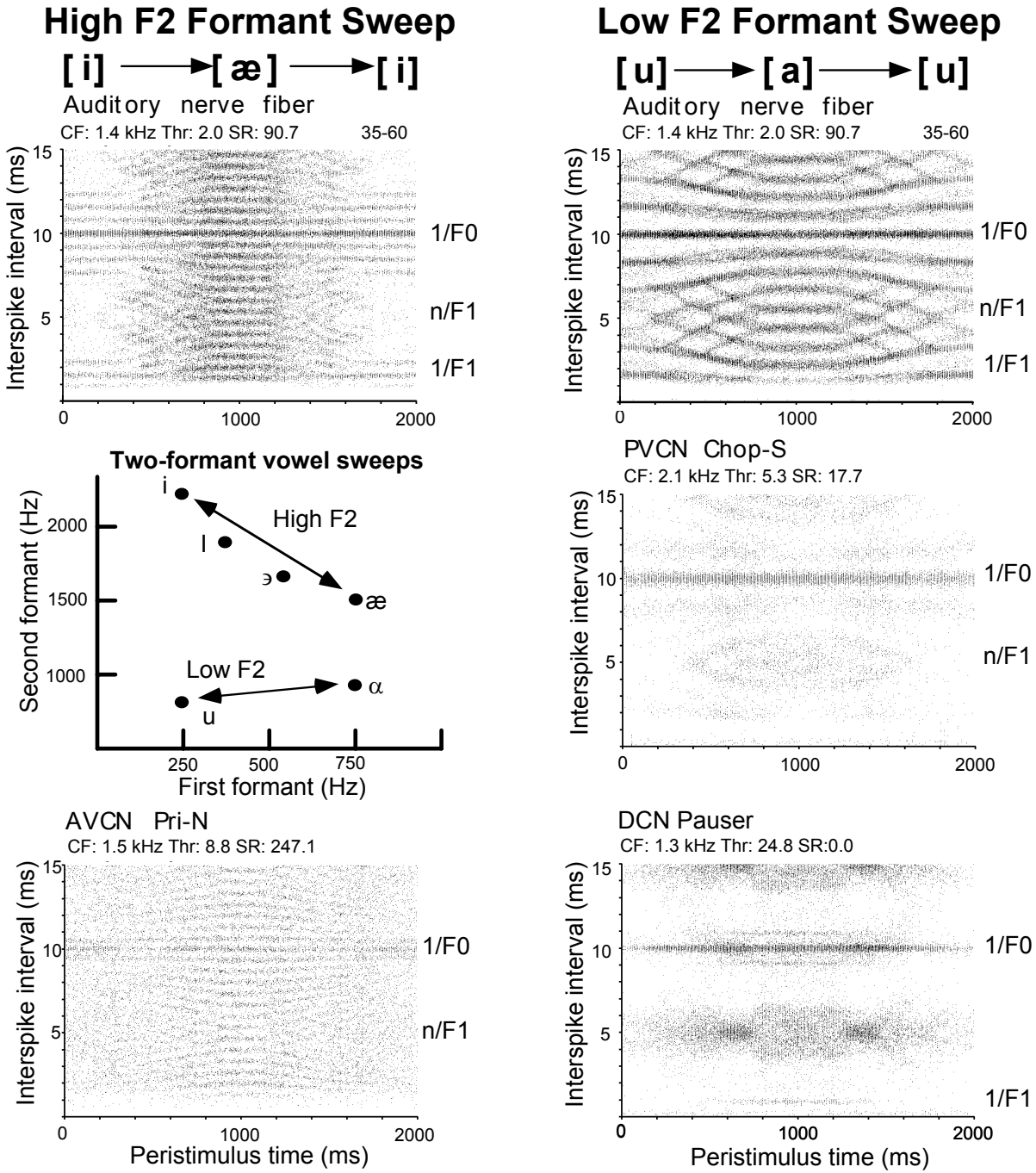
Please see Figure 3 in Deutsch, D., ed. *The Psychology of Music*. San Diego: Academic Press, 1999.

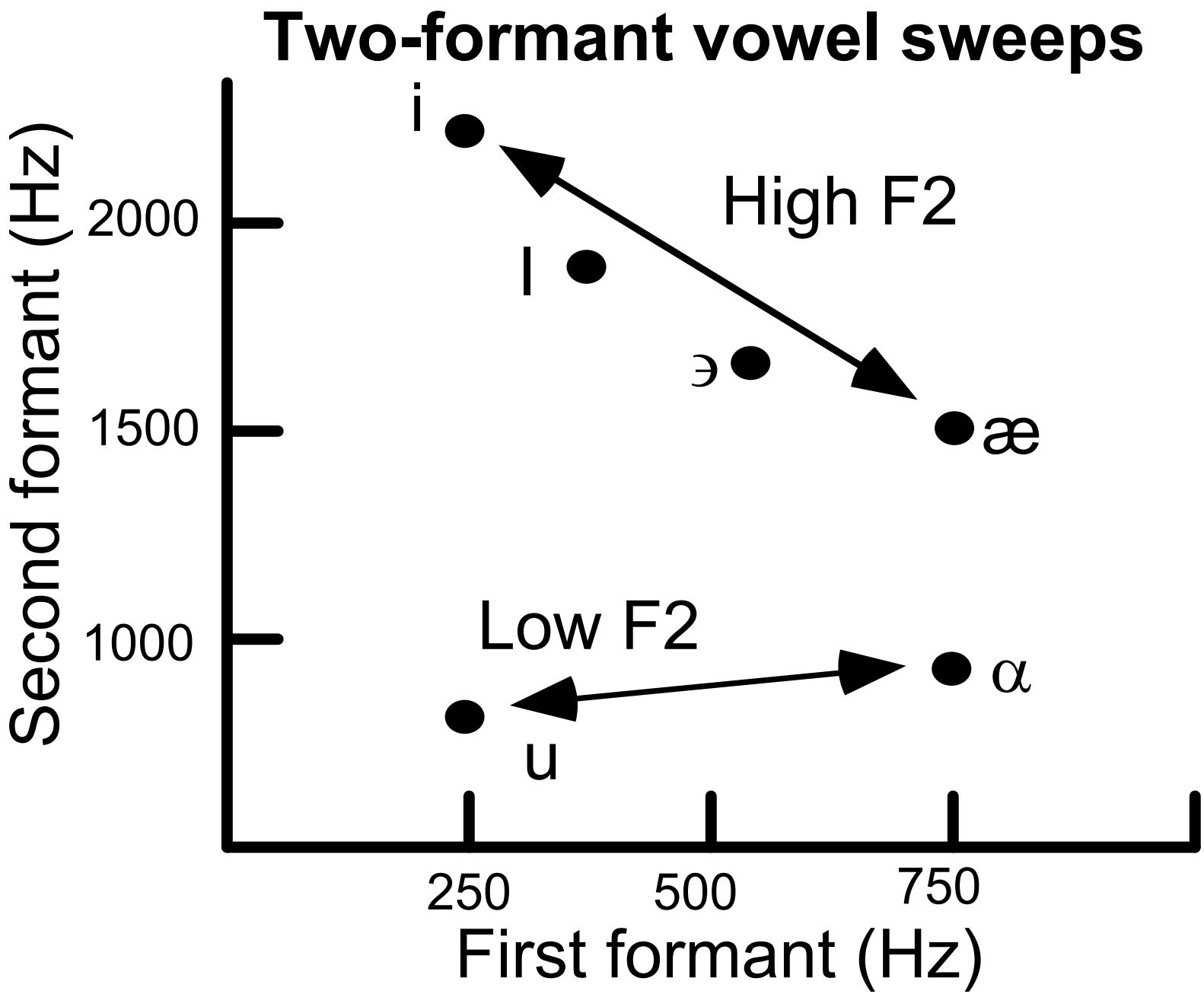
Speech Neurogram

Please see Delgutte, B. "Auditory Neural Processing of Speech." In *The Handbook of Phonetic Sciences*. Edited by W. J. Hardcastle, and j. Laver. London: Blackwell, 1995.

Possible interval-based neural correlates for basic phonetic distinctions

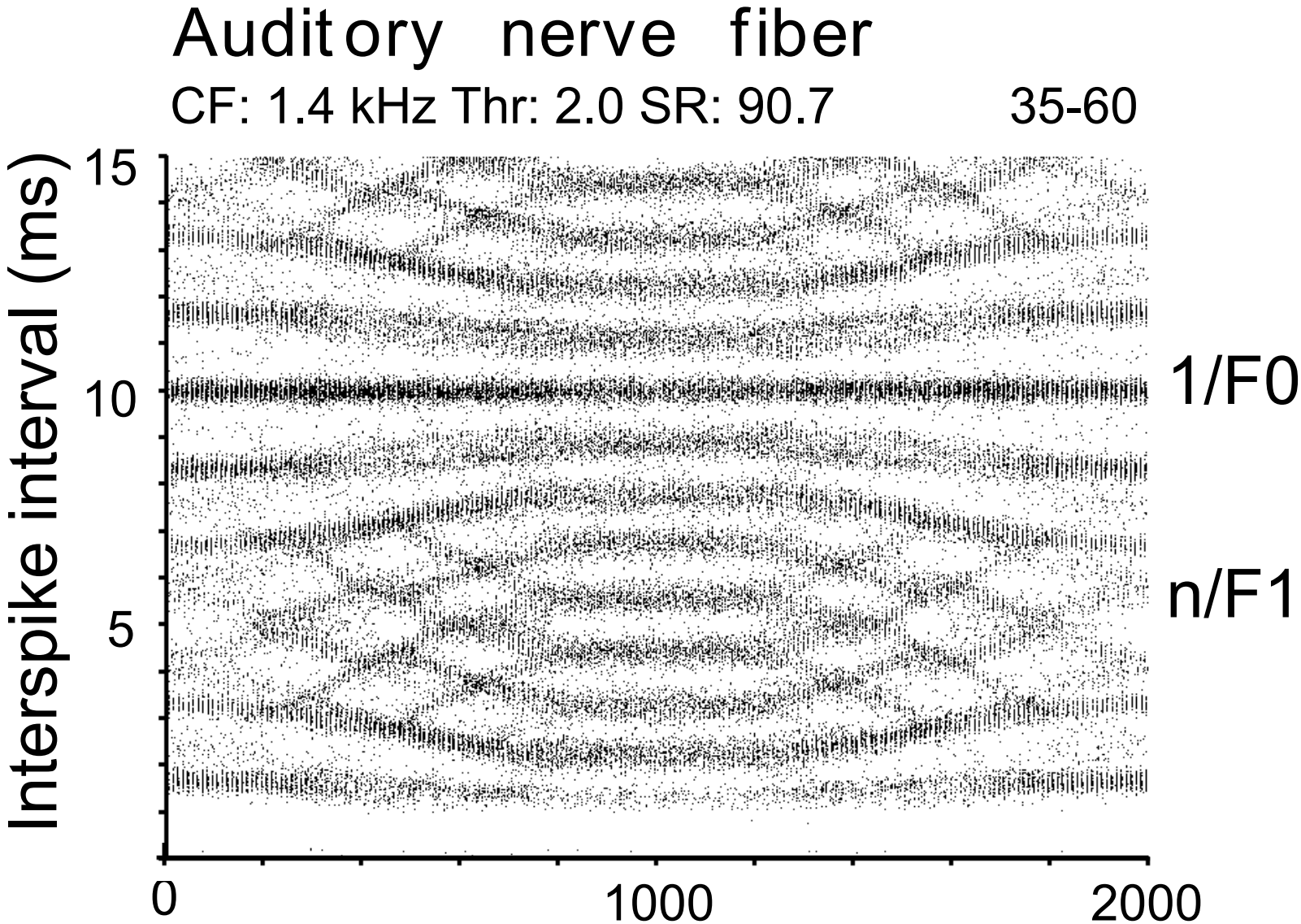
CHARACTERISTIC	ACOUSTIC DISTINCTION	PHONETIC CLASS	EXAMPLES	INTERVAL CORRELATES
Voice Pitch (80-400 Hz) pitch contours, ² over time	voice pitch, F0 prosody			most common interval running interval ²
Voice onset time	VOT			prominent interval between onset/offset responses
Spectral Pattern stationary low frequency	 formant pattern nasal resonances	 vowels nasals	 [u], [ae], [i] [m], [n]	 intervals for periodicities 50-5000 Hz
Spectro-temporal pattern fast transition slow transition	 formant transitions	 consonants semivowels dipthongs	 [b], [d], [g] [w], [r], [y] [aʸ], [aʷ],[eʸ]	cross-BF intervals (?) timing of FM responses (?) slow ² in interval distr. low freq modulations interactions
Spectral Dispersion	noise-excitation (frication)	fricative consonants	/f/, /s/,/ʃ/,/v/,/θ/	semi periodic temporal struct. ;phase incoherence
Voiced-unvoiced	voiced/unvoiced	stop consonants fricatives whispered/voiced	[b]/[p] [v]/[f]	presence of harmonic structure in intervals degree interval dispersion
Dynamic Amplitude Patterns amplitude time profiles	abrupt/gradual ² (buildup / decay)	affricative/fricative	/tʃ/ vs /ʃ/ *chip* vs *ship*	adaptation + running interval buildup patterns (Autocorrelations ² shape)
Rhythm		metrical aspects word rhythm speaking rate		Longer interval patterns (50-500 msec)
Duration	duration			prominent interval between onset & offset responders
Suprasegmental structure	word time pattern	whole word patterns		longer time structures

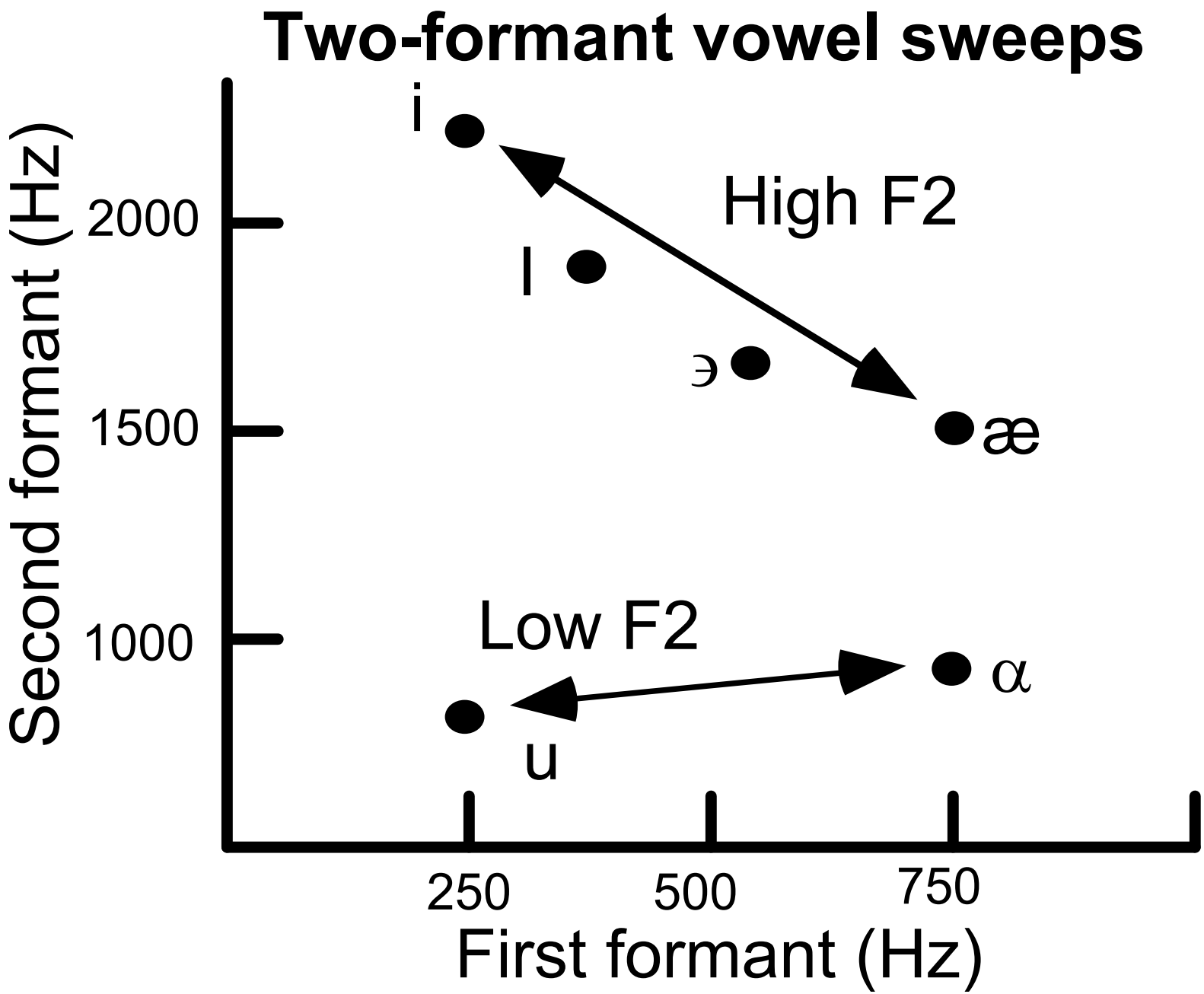

High F2 Formant Sweep
[i] → [æ] → [i]
Auditory nerve fiber
CF: 1.4 kHz Thr: 2.0 SR: 90.7 35-60
Interspike interval (ms)
15
10
5
0
1000
2000
1/F0
n/F1
1/F1
Two-formant vowel sweeps
Second formant (Hz)
2000
1500
1000
250
500
750
First formant (Hz)
i
I
ɘ
æ
α
u
High F2
Low F2
AVCN Pri-N
CF: 1.5 kHz Thr: 8.8 SR: 247.1
Peristimulus time (ms)
Low F2 Formant Sweep
[u] → [a] → [u]
Auditory nerve fiber
CF: 1.4 kHz Thr: 2.0 SR: 90.7 35-60
PVCN Chop-S
CF: 2.1 kHz Thr: 5.3 SR: 17.7
DCN Pauser
CF: 1.3 kHz Thr: 24.8 SR:0.0
Peristimulus time (ms)


Two-formant vowel sweeps
i
I
ɘ
æ
High F2
Low F2
ɑ
u
2000
1500
1000
Second formant (Hz)
250
500
750
First formant (Hz)

Low F2 Formant Sweep

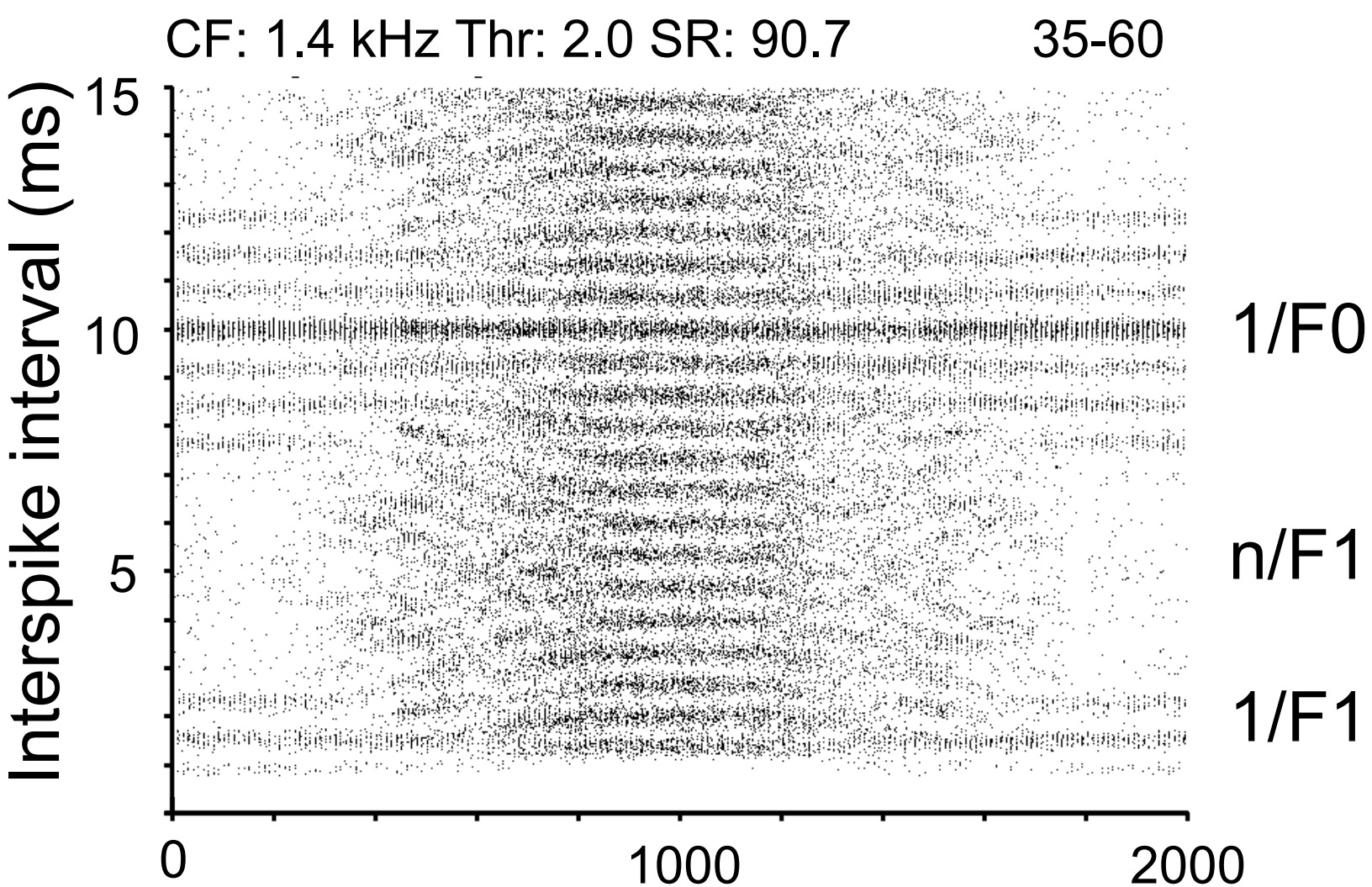
[u] ⟶ [a] ⟶ [u]

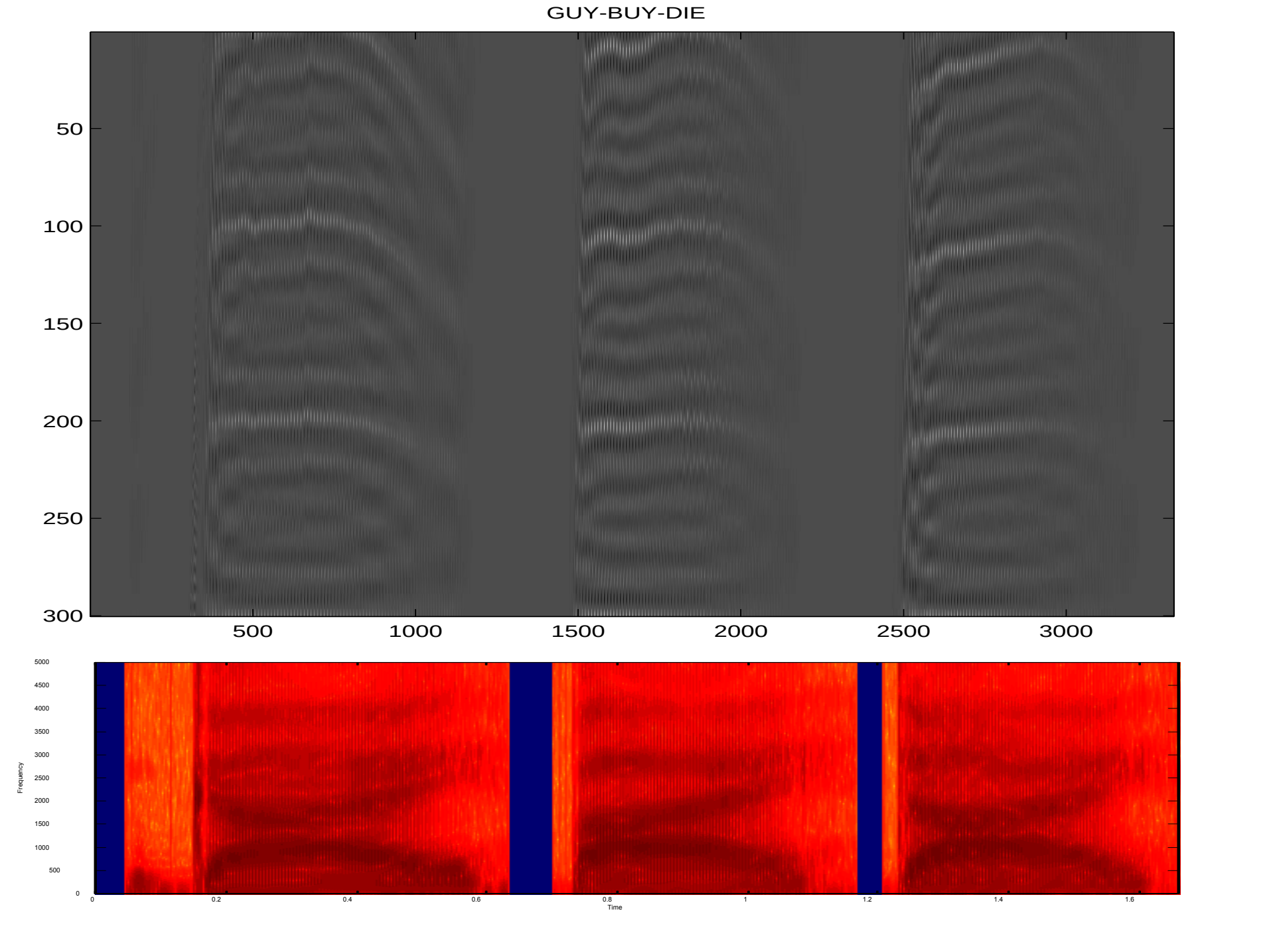



Two-formant vowel sweeps
i
I
ɘ
æ
High F2
Low F2
ɑ
u
2000
1500
1000
Second formant (Hz)
250
500
750
First formant (Hz)

High F2 Formant Sweep

[i] → [æ] → [i]




GUY-BUY-DIE
50
100
150
200
250
300
500
1000
1500
2000
2500
3000
Frequency
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Time

Timbre: summary

- Time-invariant properties (static)
  - Stationary spectrum
  - Relatively well-understood & characterized

- Time-varying properties (dynamic)
  - Amplitude dynamics (envelope)
  - Frequency dynamics (spectral changes, vibrato)
  - Phase shifts (chorus effect & electronic contexts)
  - Relatively poorly understood & characterized

Reading/assignment for next meeting

- **Tuesday, March 2**
- **Consonance, dissonance, and roughness**

- **<u>Reading</u>:**
  - **Deutsch, Rasch & Plomp chapter re: beats, combination tones, and consonance**
  - **also Burns chapter on intervals & scales**

  (look at section on consonance)