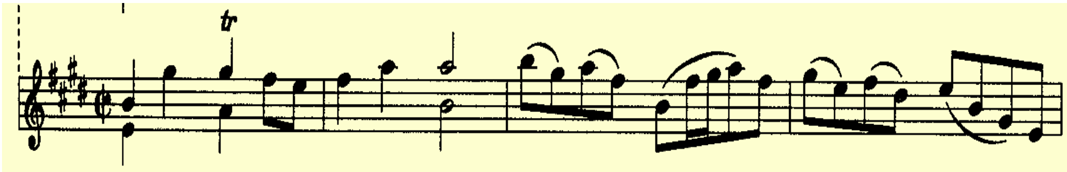
HST.725 Music Perception and Cognition, Spring 2009
Harvard-MIT Division of Health Sciences and Technology
Course Director: Dr. Peter Cariani



# Timbre perception

www.cariani.com

# Roadmap

- overview
- functions of music
- sound, ear
- loudness & pitch
- timbre
- consonance, scales & tuning
- melody & harmony
- time, rhythm, and motion
- grouping, expectation, meaning
- music & language

basic qualities of notes

interactions between notes

patterns of pitches

patterns of events

interpretations

# Wikipedia on timbre

In music, **timbre** (pronounced /ˈtæm-bər'/, /tɪm.bər/ like *tamber*, or /ˈtæm(brᵊ)/,[1] from Fr. **timbre** tɛ̃bʁ) is the quality of a musical note or sound or tone that distinguishes different types of sound production, such as voices or musical instruments. The physical characteristics of sound that mediate the perception of timbre include spectrum and envelope. Timbre is also known in psychoacoustics as *tone quality* or *tone color*.

For example, timbre is what, with a little practice, people use to distinguish the saxophone from the trumpet in a jazz group, even if both instruments are playing notes at the same pitch and loudness. Timbre has been called a "wastebasket" attribute[2] or category,[3] or "the psychoacoustician's multidimensional wastebasket category for everything that cannot be qualified as pitch or loudness."[4]

# Timbre ~ sonic texture, tone color

Paul Cezanne. "Apples, Peaches, Pears and Grapes." Courtesy of the iBilio.org WebMuseum.

*Paul Cezanne, Apples, Peaches, Pears, and Grapes* c. 1879-80); Oil on canvas, 38.5 x 46.5 cm; The Hermitage, St. Petersburg

# Timbre ~ sonic texture, tone color

"Stilleben" ("Still Life"), by Floris van Dyck, 1613.  (Public domain image, from Wikipedia.)

## Analogy to visual texture

## Roughness
## Smoothness

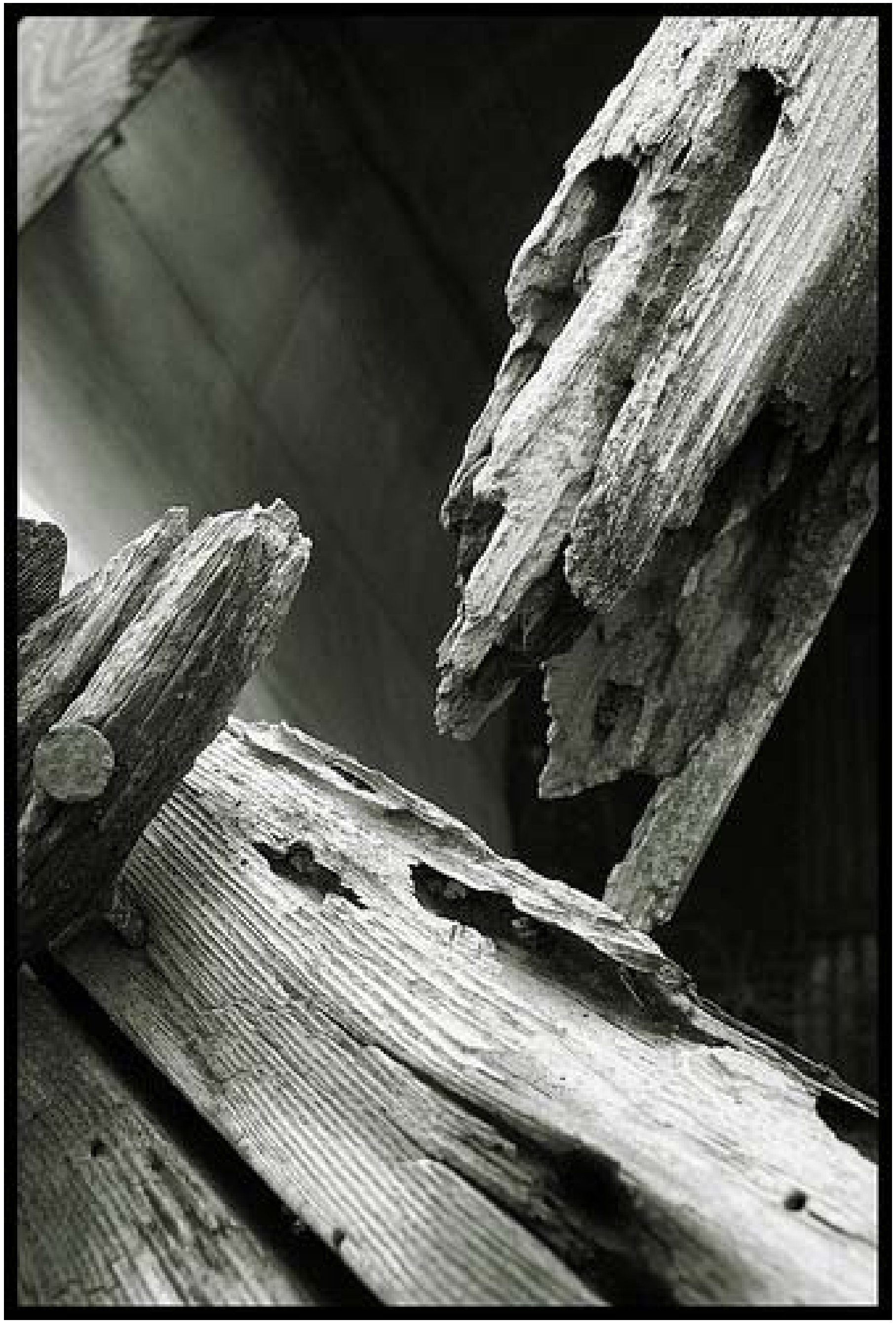


Photo courtesy of hoveringdog on Flickr.

# Timbre: a multidimensional tonal quality

uses in tonal music:
tone "color", "texture"
distinguishes instruments


Photo courtesy of Miriam Lewis. Used with permission.

important for
instrument
design


Photo courtesy of Pam Roth.
Used with permission.


Photo courtesy of Per-Åke Byström.
Used with permission.

"timbral music":
primary dimension of
sonic change

sound mass
ambient music
electronic music
lexical music

# What makes different timbres distinctive?

## Timbre: a multidimensional tonal quality

## Complicated.....but there are two basic aspects

**Stationary Aspects**

(spectrum)

**Vowels**

**Dynamic Aspects**

Δ spectrum
Δ intensity
Δ pitch
attack
decay

**Consonants**

Photo courtesy of Per-Åke Byström. Used with permission.

http://www.wikipedia.org/

# “The elusive attributes of timbre”

J.F. Schouten (1968, p.42) describes the "elusive attributes of timbre" as "determined by at least five major acoustic parameters" which Robert Erickson (1975) finds "scaled to the concerns of much contemporary music":

1. The range between tonal and noiselike character.
2. The spectral envelope
3. The time envelope in terms of rise, duration, and decay.
4. The changes both of spectral envelope (formant-glide) and fundamental frequency (micro-intonation).
5. The prefix, an onset of a sound quite dissimilar to the ensuing lasting vibration.

# Timbre perception: summary of factors

- **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 - low or high, "brightness"
  - Formant structure (spectral peaks)
  - 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 (noisiness, 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

# 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**
  - Spectral peaks (formants), spectral tilt (brightness)
- **Amplitude-frequency-phase dynamics**
  - Amplitude dynamics (attack, sustain, decay)
    - amplitude modulation (roughness, tremolo)
  - Frequency dynamics
    - relative timings of onsets & 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)

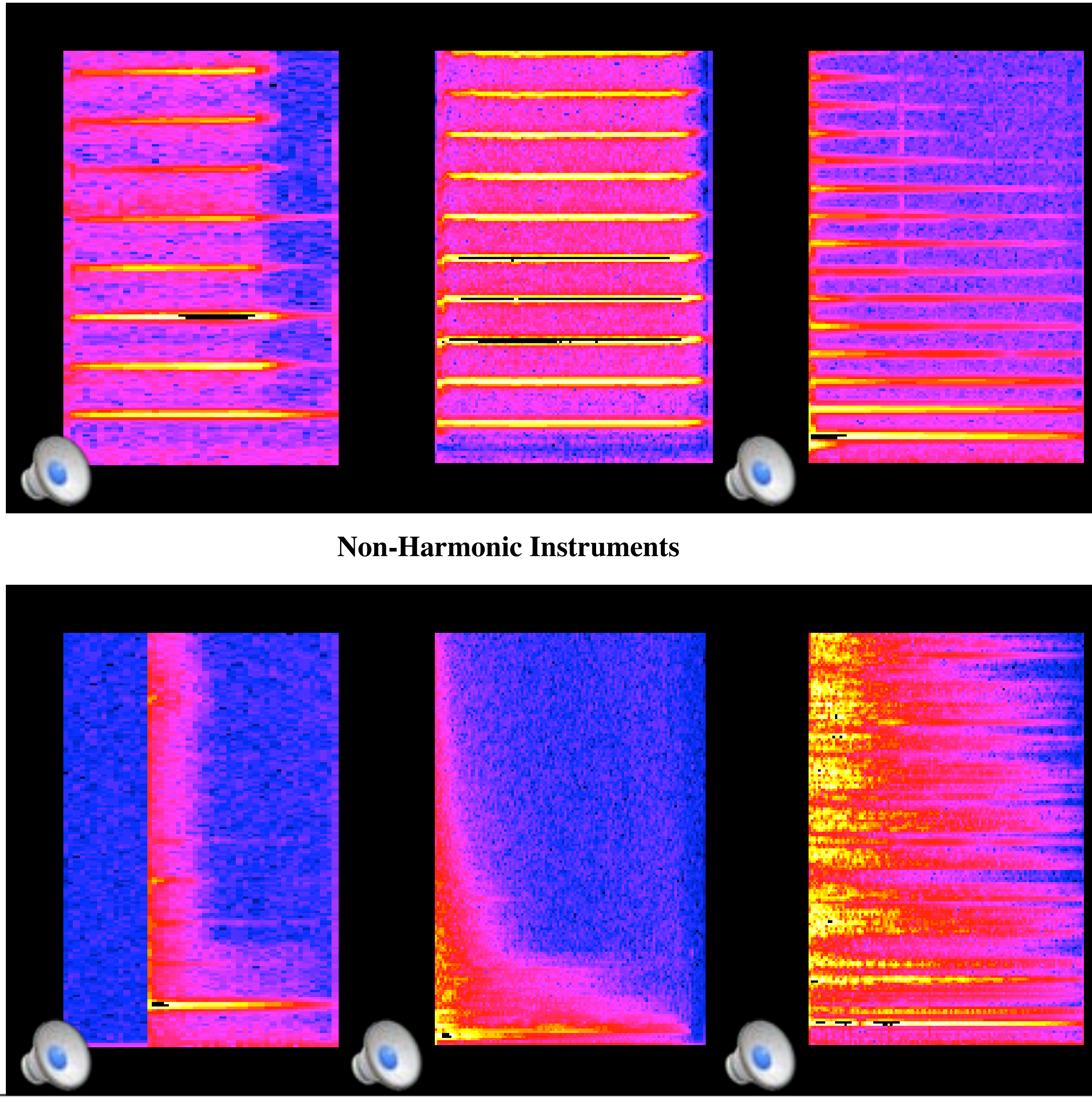
# Harmonicity
# Frequency dynamics

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

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

Rafael A. Irizarry's Music and Statistics Demo

Spectrograms of Harmonic Instruments



Non-Harmonic Instruments

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

# Timbre: a multidimensional tonal quality

## tone texture, tone color distinguishes voices, instruments

Photo courtesy of Pam Roth.
Used with permission.

Photo courtesy of Miriam Lewis.
Used with permission.

### Stationary Aspects

(spectrum)

### Vowels

### Dynamic Aspects

$\Delta$ spectrum
$\Delta$ intensity
$\Delta$ pitch
attack
decay

### Consonants

Photo courtesy of Per-Åke Byström.
Used with permission.

http://www.wikipedia.org/

## Some methods for studying the perceptual space of timbre

1. Try to derive the structure of the space from the dimensionality of listener judgments
   - Similarity magnitude estimations
   - Similarity rankings
   - Multidimensional scaling

2. "Analysis by synthesis"

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

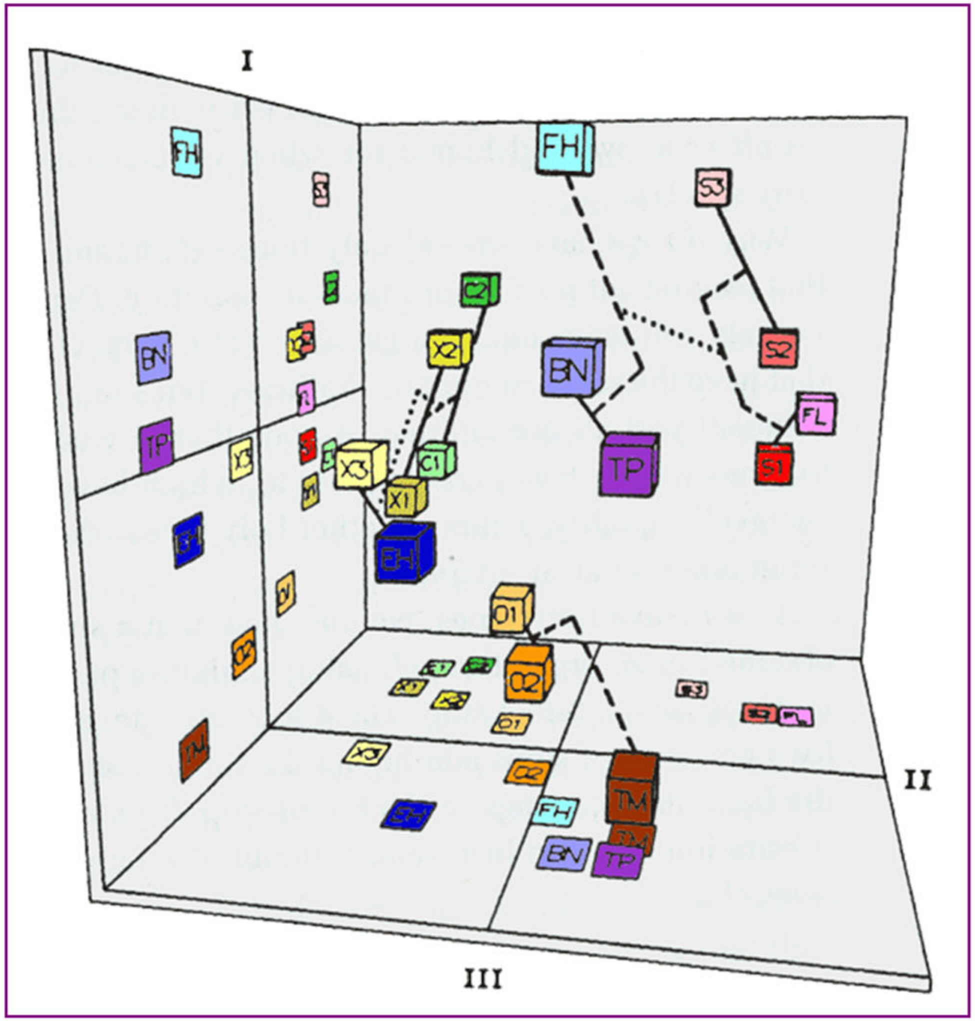
# Grey (1975) Timbre: Perceptual dimensions studied using a "confusion matrix"

Figure removed due to copyright restrictions.

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

Also see:
Grey, J. & Moorer, J.
1977.
Perceptual evaluations
of synthesized musical
instrument tones.
J. Acoustical Society of America
63:1493-1500

# Timbre dimensions: spectrum, attack, decay




- BN - Bassoon
- C1 - E flat Clarinet
- C2 - B flat Bass Clarinet
- EH - English Horn
- FH - French Horn
- FL - Flute
- O1 - Oboe
- O2 - Oboe (different instrument and player)
- S1 - Cello, muted *sul ponticello*
- S2 - Cello
- S3 - Cello, muted *sul tasto*
- TM - Muted Trombone
- TP - B flat Trumpet
- X1 - Saxophone, played *mf*
- X2 - Saxophone, played *p*
- X3 - Soprano Saxophone

- Dimension I: spectral energy distribution, from broad to narrow
- Dimension II: timing of the attack and decay, synchronous to asynchronous
- Dimension III: amount of inharmonic sound in the attack, from high to none

Courtesy of Hans-Christoph Steiner. Used with permission.
After J. M. Grey, Stanford PhD Thesis (1975) and Grey and Gordon, JASA (1978)

Grey & Gordon, 1978, JASA

# Amplitude dynamics (envelope, intensity contour)
## (Garageband demonstration)

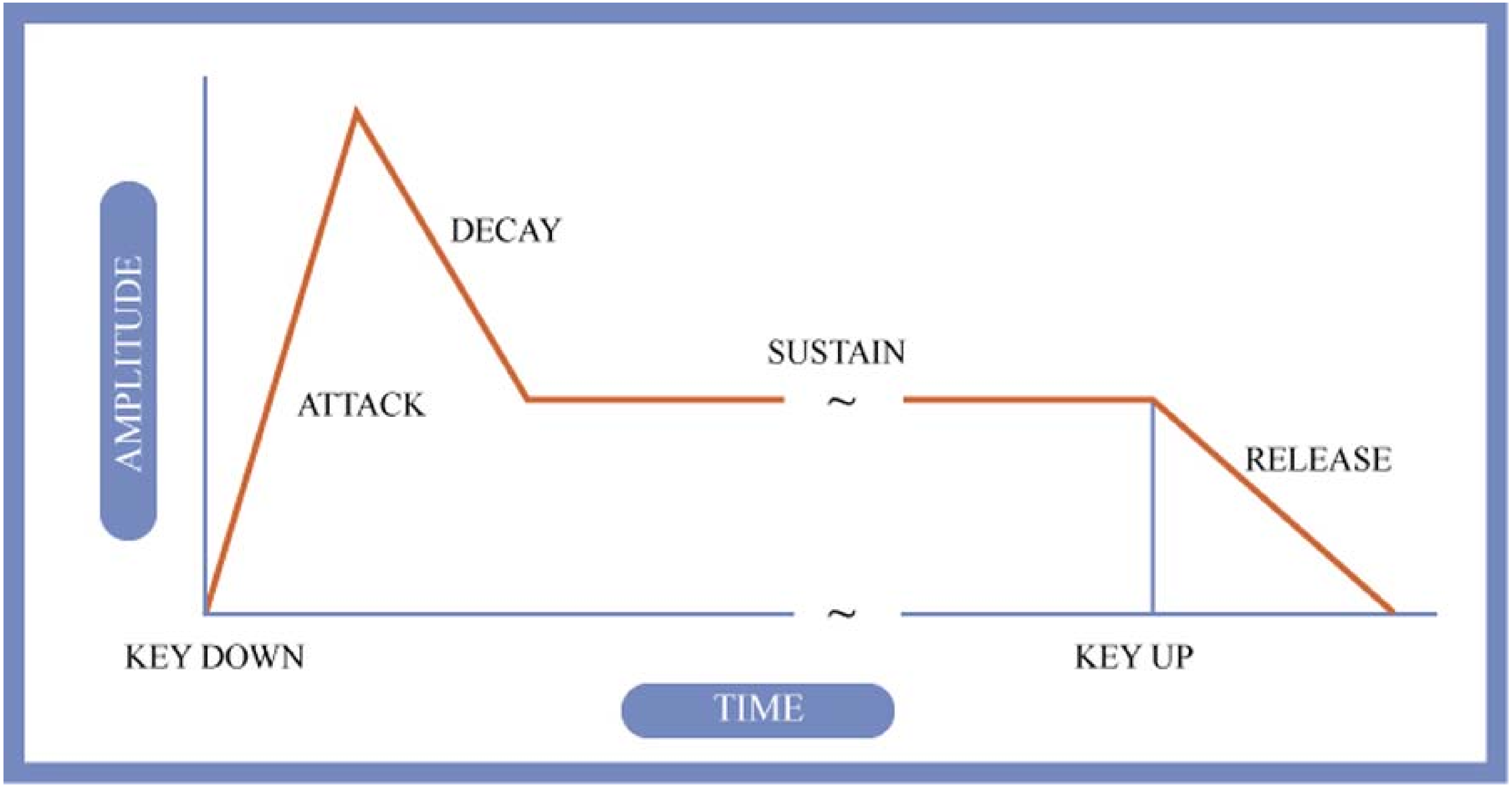
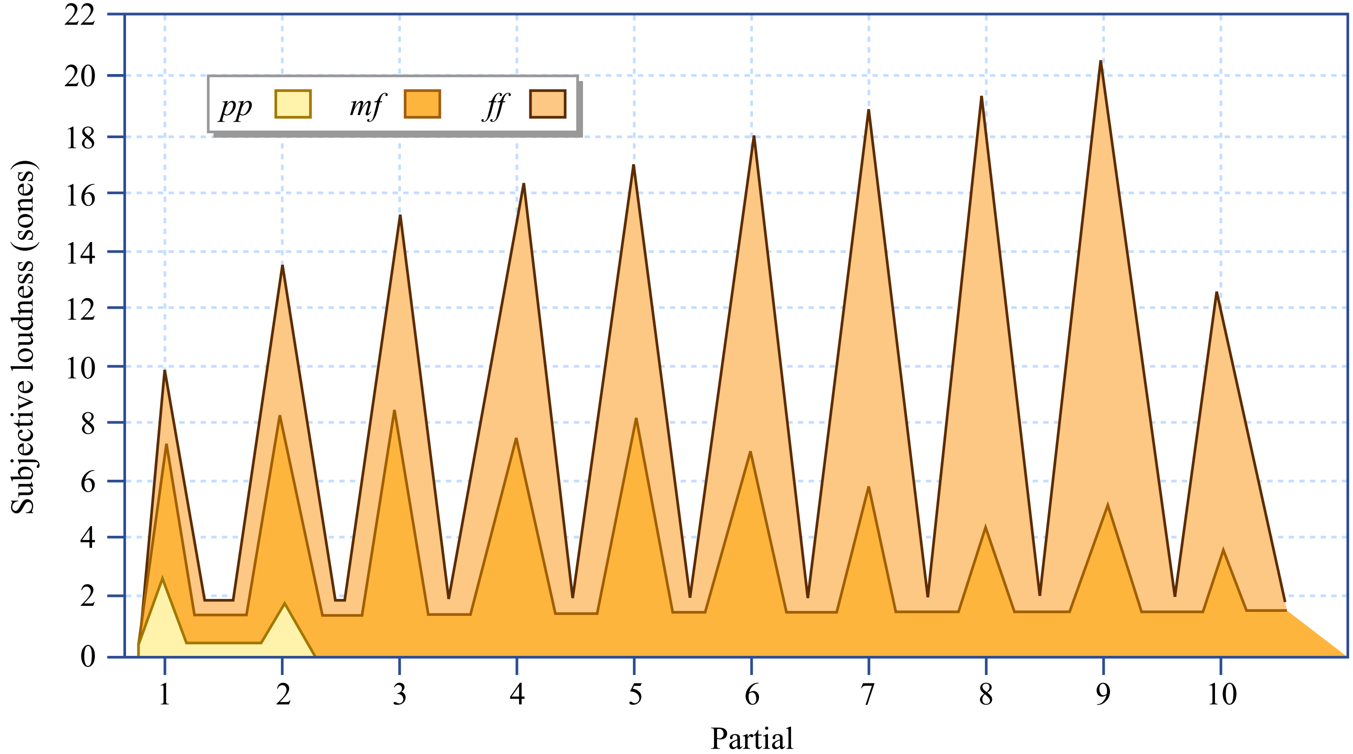



Figure by MIT OpenCourseWare.

# Spectrum as a function of intensity (trumpet)
# Timbre can change with intensity




Loudness spectra of a trumpet playing the same pitch (concert $B^{\flat}_3$) at three dynamic levels (*pp, mf* and *ff*). Sones are linear increments of subjective loudness.

Figure by MIT OpenCourseWare. After Hanson (1988).

## Vocal Ring, or The Singer's Formant

One seemingly mysterious property of the singing voice is its ability to be heard even over a very loud orchestra. At first glance, this is counter-intuitive, since the orchestra is perceived by us to be so much *louder* than a single singer. The answer to this mystery lies in the way the sound energy of the operatic voice is distributed across various frequencies.



http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/singer.html

# Vocal Ring, or The Singer's Formant

# Singer's formant

Cook, Perry, ed.
Music, Cognition &
Computerized Sound
MIT Press 2001

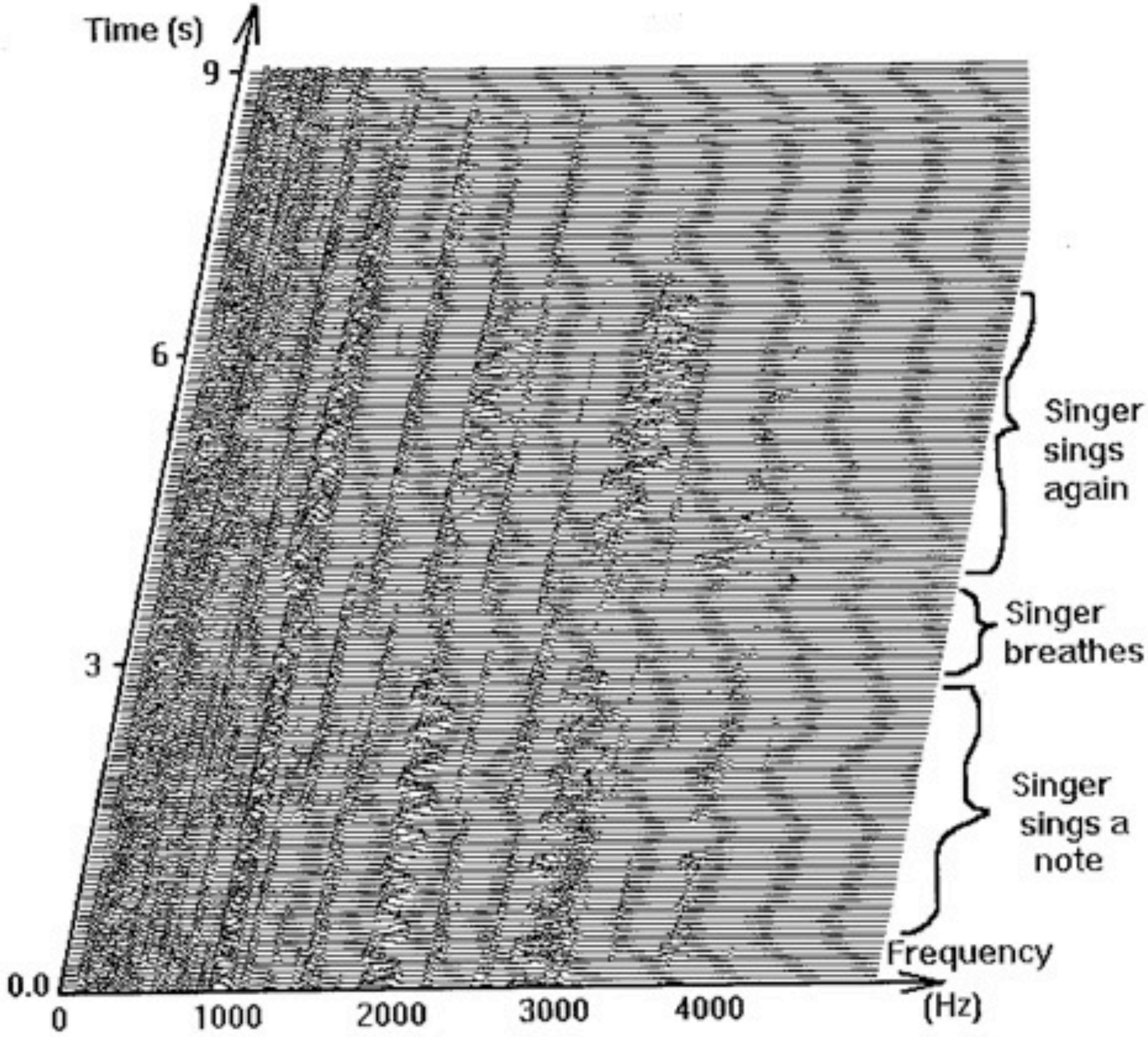



**Figure 11.12** The singer's formant is evident in this waterfall plot of the last two notes of the soprano aria "Un Bel Di Vedremo," from Puccini's *Madame Butterfly.* Common frequency modulation of the first three partials allows the fundamental to be picked out visually.

# Frequency dynamics of note onsets (clarinet)

Image removed due to copyright restrictions.
Figure 4-4 in Butler, David. The Musician's Guide to Perception and Cognition.
New York, NY: Schirmer/Macmillan, 1992. ISBN: 9780028703411.

# Time-course of harmonics

Figure 3 (p. 119) in Risset, J-C., and Wessel, D. L. "Exploration of Timbre by Analysis and Synthesis." Chapter 5 in Deutsch, D., ed. *The Psychology of Music*. 2nd ed. San Diego, CA: Academic Press, 1998. ISBN: 9780122135651. [View this image in Google Books]

## Time-window for timbral integration

**Appears to be similar to that for pitch (~30 ms)**

**Evidence:**

**Indistinguishability of ramps vs. damps < 30 ms (Patterson)**

**Reversal of 30 ms speech segments - no effect**

**Timbral fusion of 2 single-formant vowels**

- **(L.A. Chistovich, 1985)**
- **50 Hz alternating double vowels did not fuse (20 ms offset)**

**Common onset grouping windows (~25-30 ms)**

# Voice qualities

## another description of aspects of timbral space outside phonetic distinctions

Table removed due to copyright restrictions.

See http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/quality.html

# Music timbre space and phonetic space

Human speech communications systems are mostly built on timbral distinctions, although there are tonal languages in which pitch contour conveys distinctions.......

This could be because of the different voice pitches of human speakers, or it could be due to the relative ease of rapidly changing vocal resonances rather than changing voice pitch (harder to sing than to talk)

Vowels = sustained notes = spectral differences (formants)

Consonants = onset patterns = amplitude & frequency fluxes

I believe that we will eventually come to a unified theory of both musical timbral distinctions and phonetic distinctions that is grounded in how the auditory system encodes spectrum and rapid changes.......

# Speech Neurogram

**(cat auditory nerve, Delgutttte, 1996)**

Figure 16.1 (p. 511) in Delgutte, B. "Auditory Neural Processing of Speech."
Chapter 16 in *The Handbook of Phonetic Sciences*. Edited by W. Hardcastle and J. Laver.
Malden, MA: Wiley-Blackwell, 1999. ISBN: 9780631214786.
[View this image in Google Books]

# Possible interval-based neural correlates for basic phonetic distinctions

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

# Music based on timbral contrasts

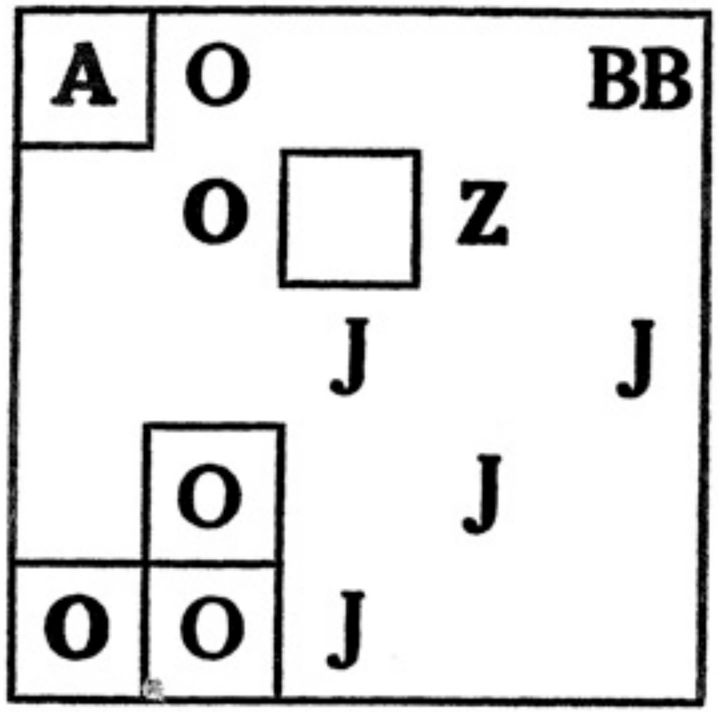
## Kurt Schwitters,
## Ur Sonata (1932)
## perf. George Melly, Miniatures

01
Fümms bö wö tää zää Uu, pögiff, kwiiee. 1
Dedesnn nn rrrrrr, Ii Ee, mpiff tillff toooo? Till, Jüü-Kaa. 2
*(gesungen)*
Rinnzekete bee bee nnz krr müü? ziiu ennze ziiu rinnzkrrmüüüü; 3
Rakete bee bee. 3a
Rummpff tillff toooo? 4
Ziiuu ennze ziiuu nnskrrmüüü, ziiuu ennze ziiuu rinnzkrrmüüüü; 03
Rakete bee bee, 0 3a
Rakete bee zee.

Fümmsbö wö tää zää Uu, 1
Uu zee tee wee bee
zee tee wee bee
zee tee wee bee
zee tee wee bee
zee tee wee bee
zee tee wee bee Fümms.

**schluss:**

Fümms bö fümms bö wö fümmes bö wö täääă? 1
Fümms bö fümms bö wö fümms bö wö tää zää Uuuu? 1
Rattatata tattatata tattatata
Rinnzekete bee bee nnz krr müüüü? 3
Fümms bö 1
Fümms böwö
Fümmes bö wö tääă???? *(gekreischt)*

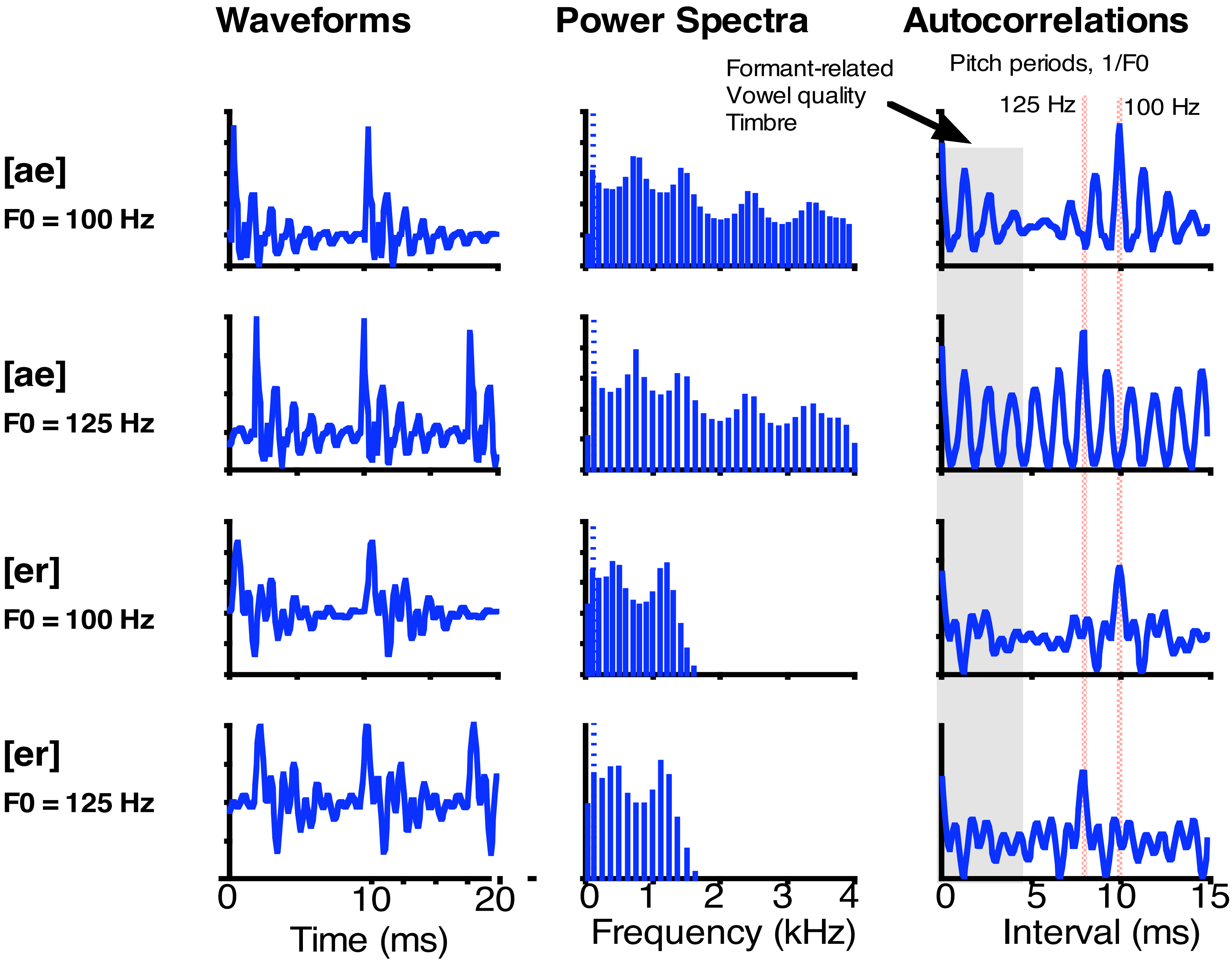


# Music based on timbral contrasts

## Kurt Schwitters,
## Ur Sonata (aka "Ursonate") (1932)
## perf. George Melly, Miniatures

Images of score and photos of Schwitters performing Ur Sonata removed due to copyright restrictions.
See http://writing.upenn.edu/pennsound/x/Schwitters.html.

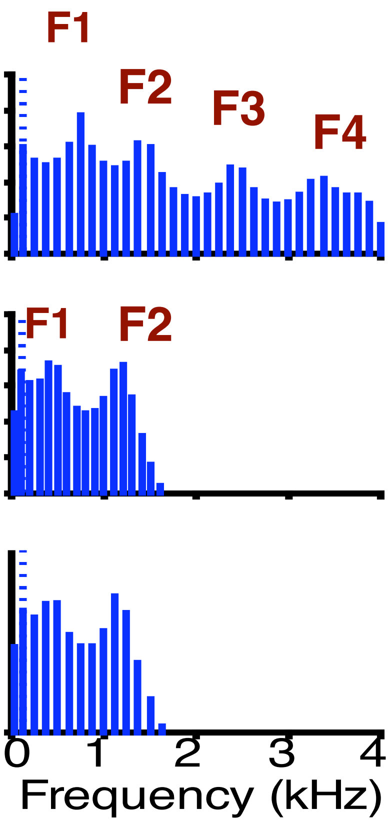
# Stationary spectral aspects of timbre



# Formants and the vocal tract

Image removed due to copyright restrictions.
Diagram of eight vocal tract positions for some english vowels:
heed, hid, head, had, hod, hawed, hood, who'd. (Source unknown.)

# Timbre and spectrum
# Vowel space

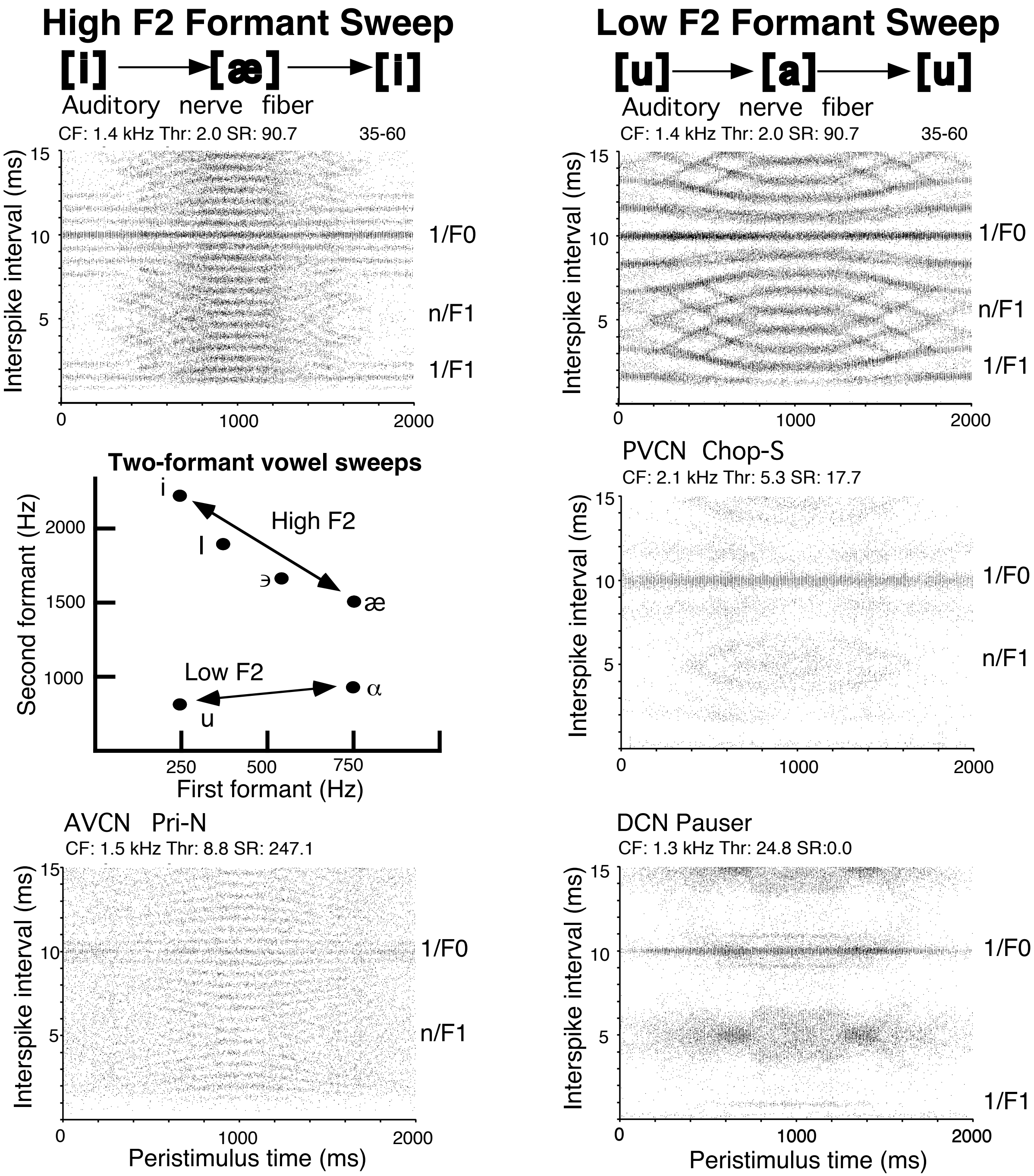



[ae]

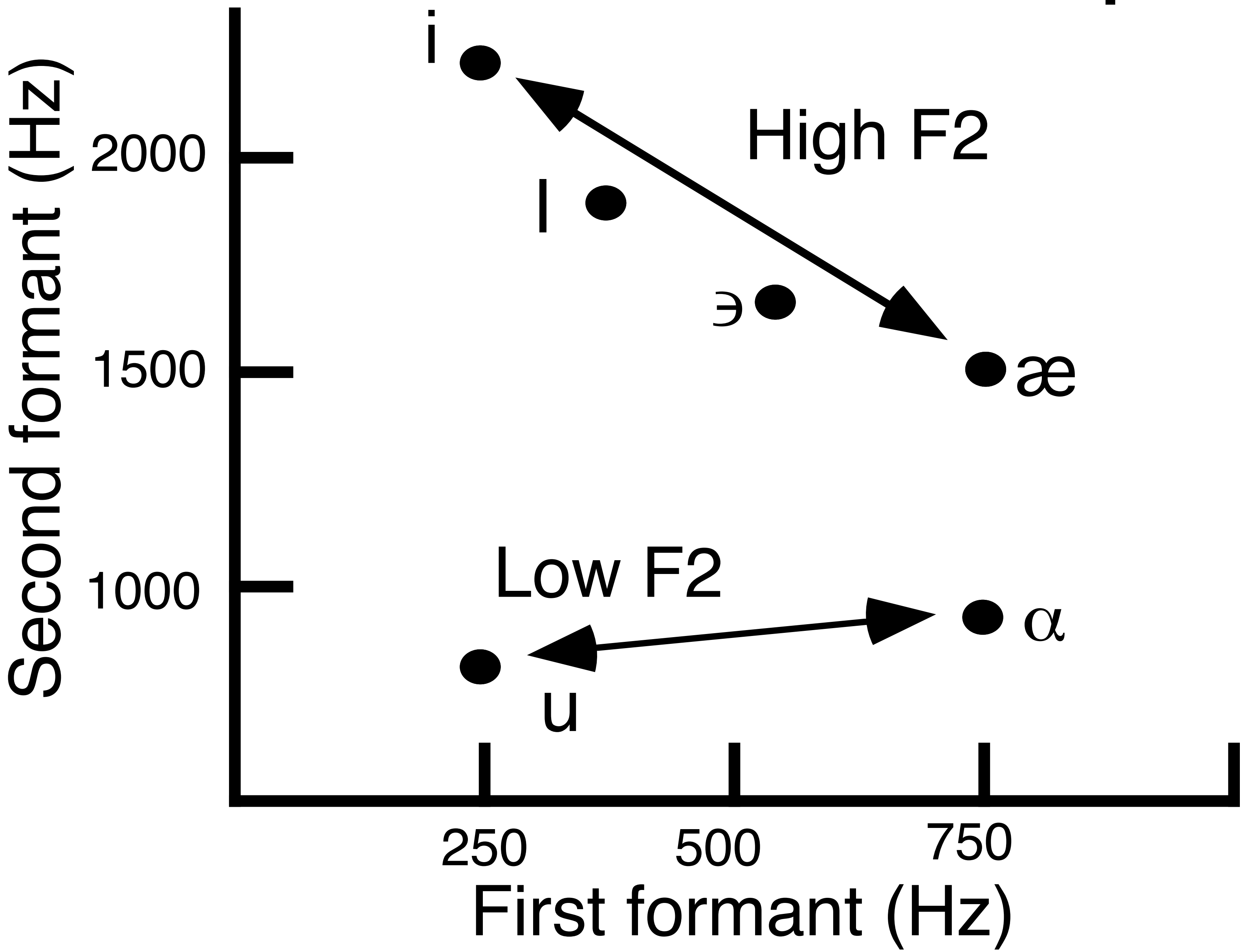
[ɛ], “er”

[ɛ], “er”

Plot of vowel space (first vs. second formant frequencies) removed due to copyright restrictions. See http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/filter.html. Adapted from Peterson, G.E., and H.L. Barney. "Control Methods Used in a Study of the Vowels." J *Acoust Soc* Am 24, no. 2 (1952): 175-184.


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

# Two-formant vowel sweeps

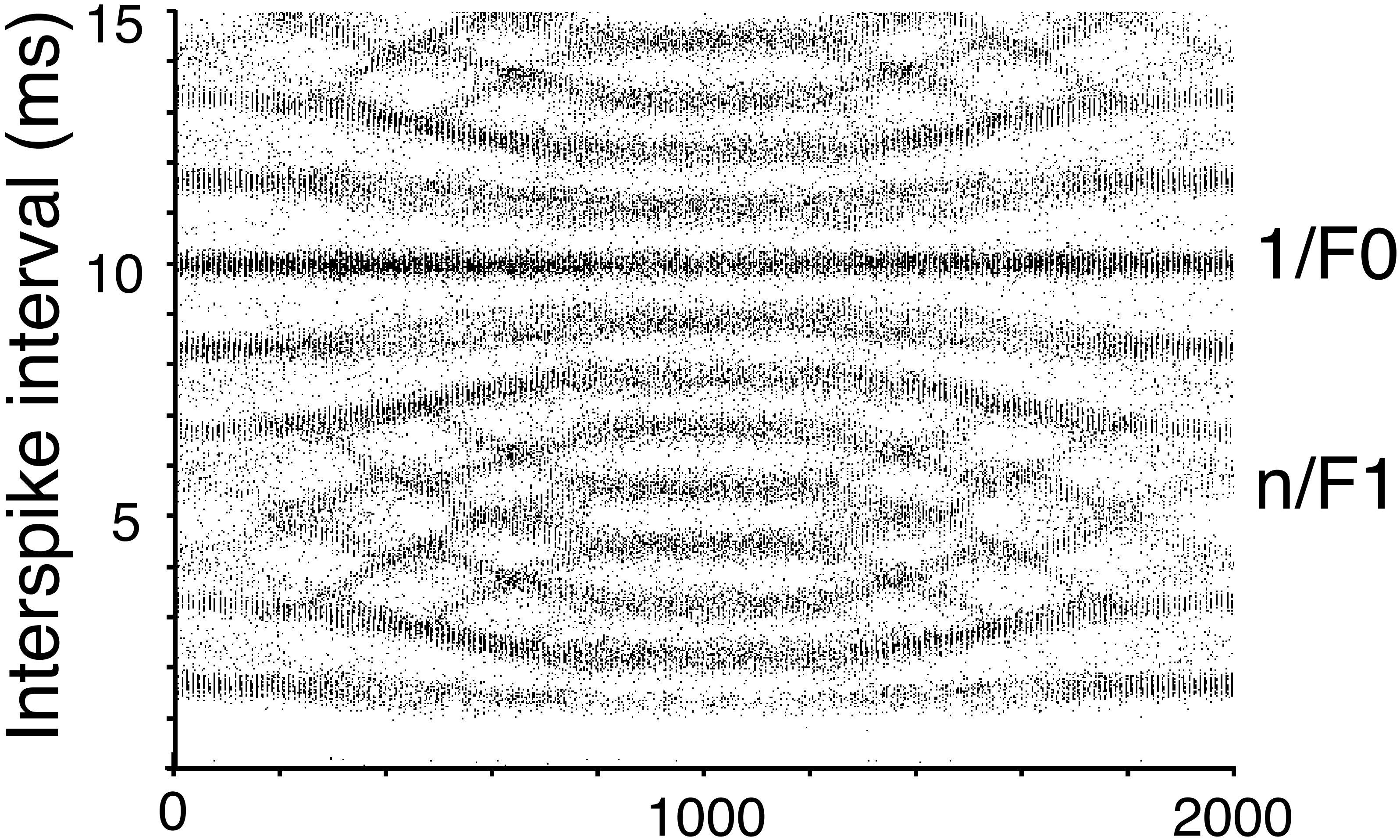


# Low F2 Formant Sweep

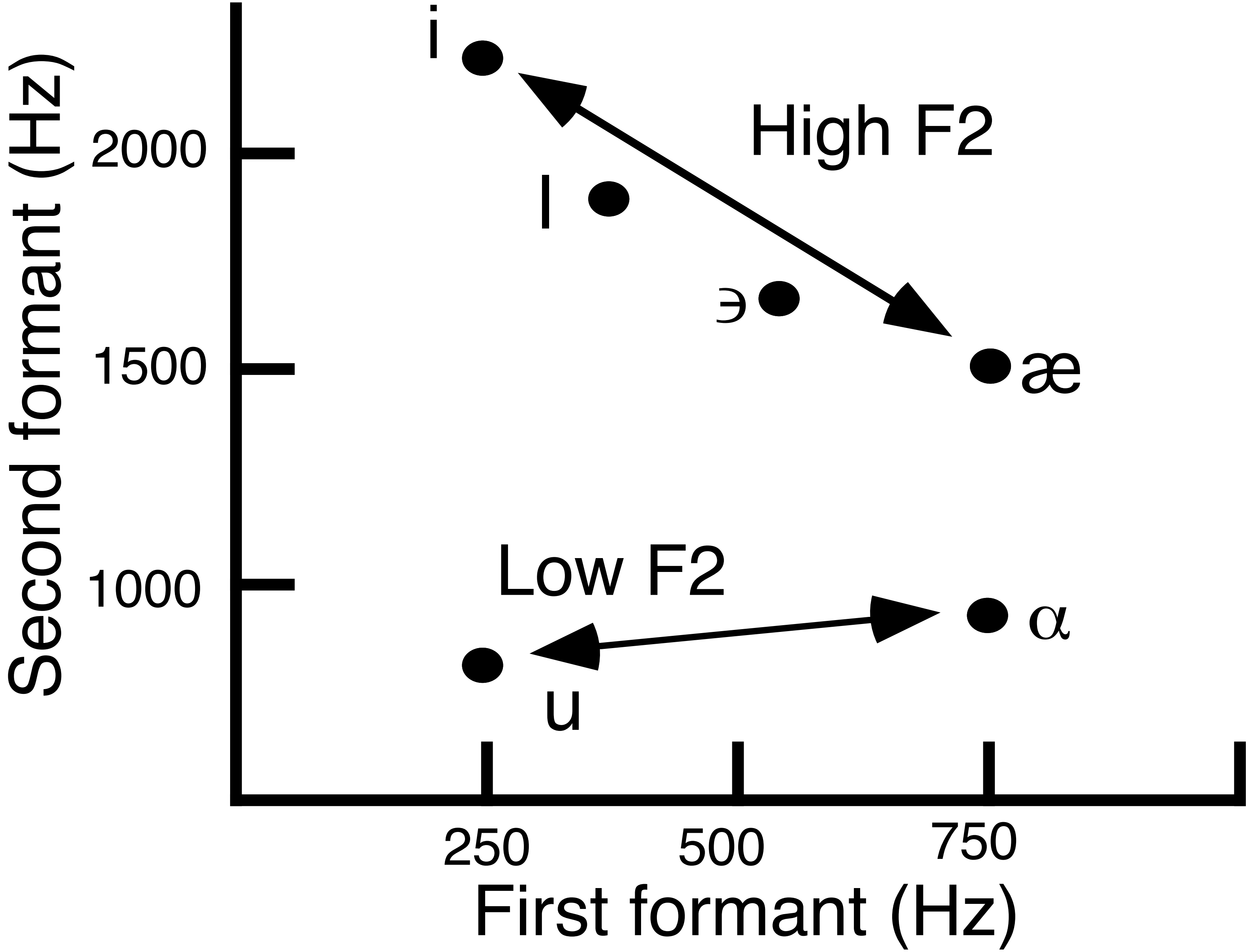
[u] ⟶ [a] ⟶ [u]

Auditory nerve fiber

CF: 1.4 kHz Thr: 2.0 SR: 90.7 35-60



# Two-formant vowel sweeps

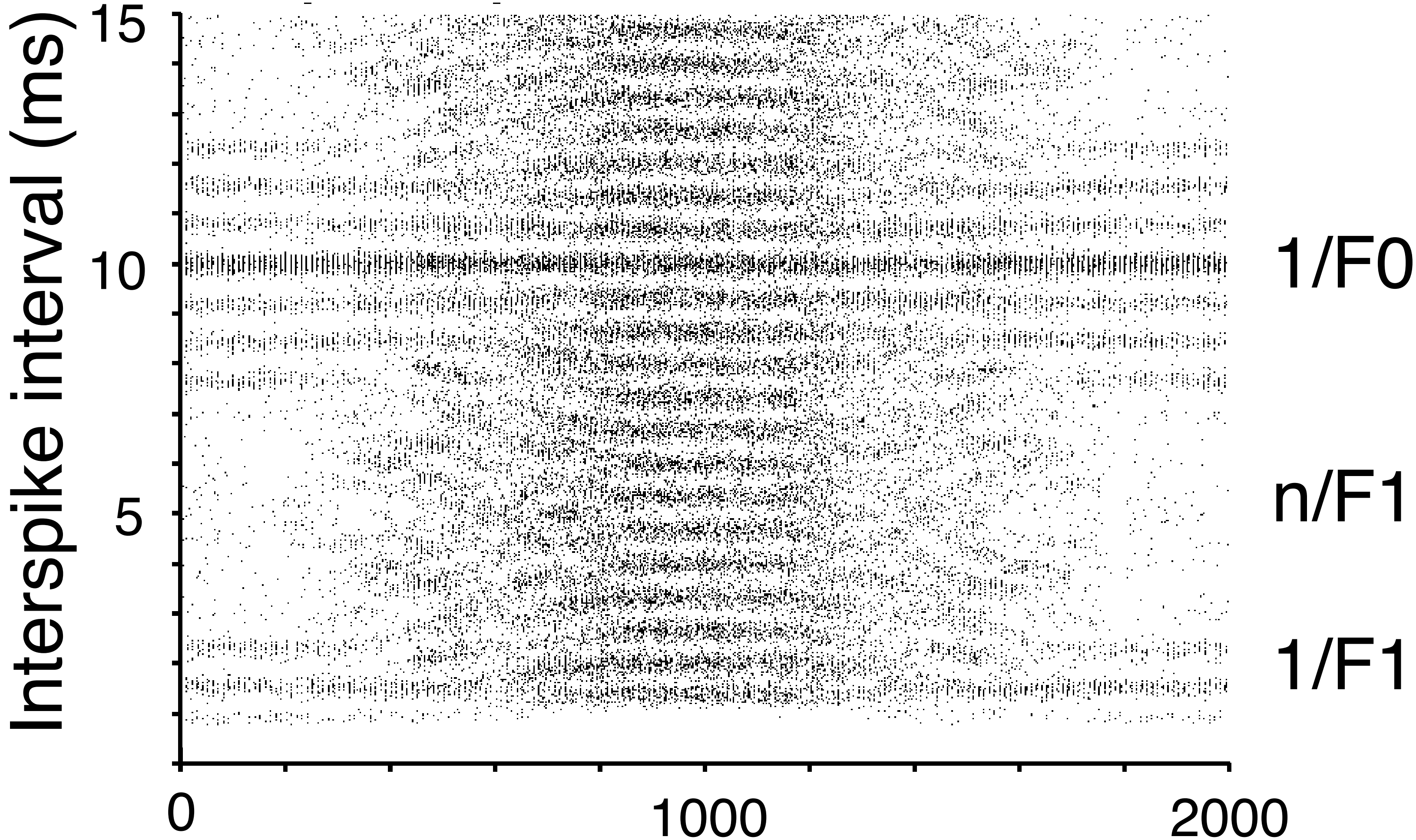


# High F2 Formant Sweep

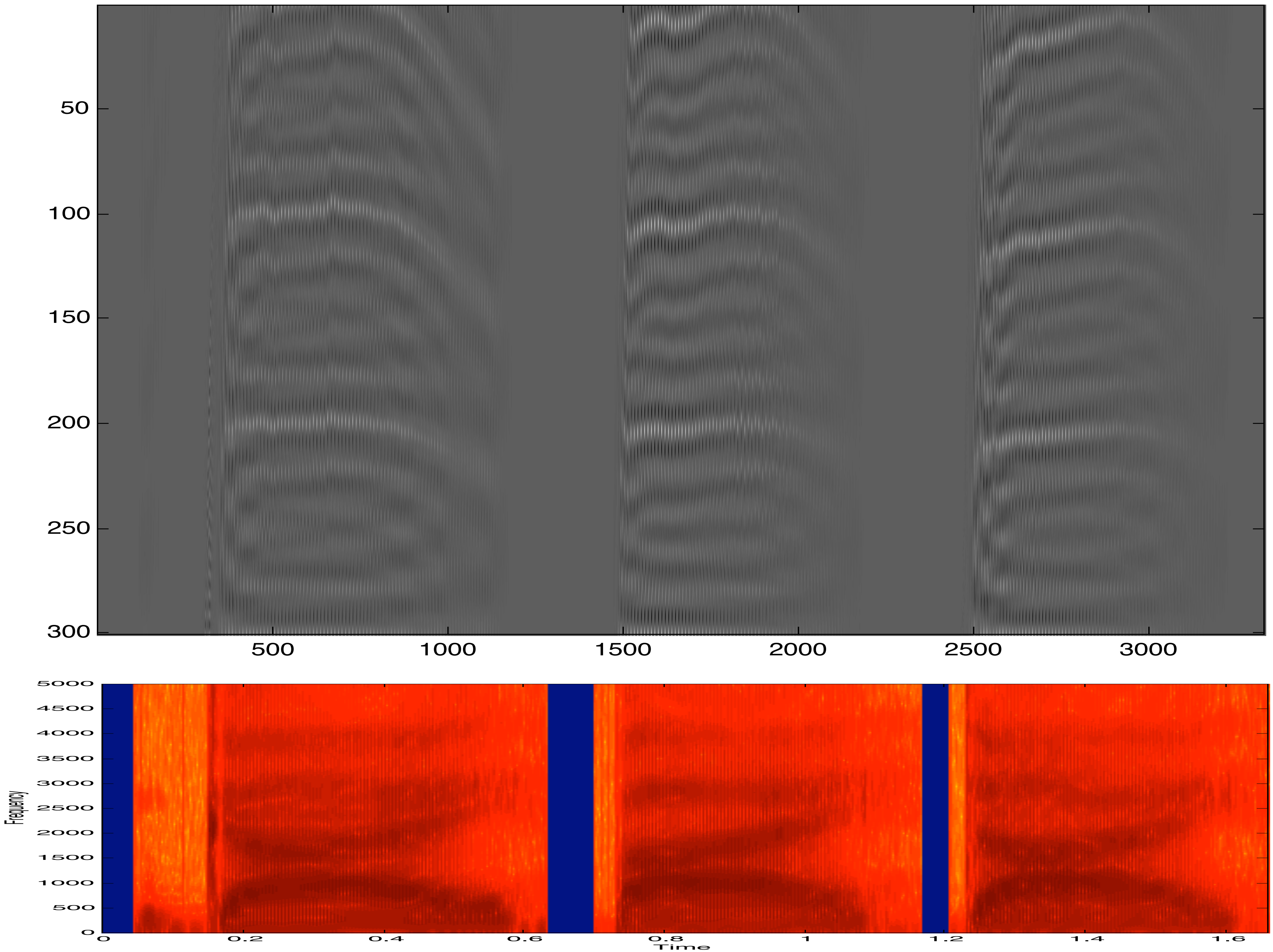
[i] → [æ] → [i]

Auditory nerve fiber

CF: 1.4 kHz Thr: 2.0 SR: 90.7 35-60



# 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
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Time

# Summary I: Uses of timbre in music

- Distinguishes musical instruments
- Tone coloration (Western tonal music)
- Primary dimension of auditory contrast in some music (electronic, ambient)

# Summary II: Acoustical correlates of timbre

- **Time-invariant properties** (static sounds)
  - Stationary spectrum (sustained notes)
  - Speech: vowels
  - Relatively well-understood & characterized

- **Time-varying properties** (rapidly changing sounds)
  - Onsets & offsets of notes
  - Amplitude dynamics (envelope, attack, decay)
  - Frequency dynamics (spectral changes, vibrato)
  - Speech: consonants
  - Phase shifts (chorus effect & electronic contexts)
  - Relatively poorly understood & characterized

# Timbre: a multidimensional tonal quality

**uses in tonal music:**
**tone "color", "texture"**
**distinguishes instruments**

**"timbral music": primary dimension of change**

**Stationary Aspects**

(spectrum)

**Vowels**

**Dynamic Aspects**

Δ spectrum
Δ intensity
Δ pitch
attack
decay

**Consonants**

**important for instrument design**

Photo courtesy of Miriam Lewis.
Used with permission.

Photo courtesy of Pam Roth.
Used with permission.

http://www.wikipedia.org/

Photo courtesy of Per-Åke Byström.
Used with permission.

http://en.wikipedia.org/wiki/Timbre

**Next up: consonance and scales**

**Any questions?**

HST.725 Music Perception and Cognition
Spring 2009