Timbre

Patel Chapter 1b

Timbral contrasts in music

- Timbre (sound quality): “ that aspect of a sound that distinguishes it from other sounds of the same pitch, duration, and loudness.”
- Timbral contrasts *between* instruments used in systematic ways by composers from numerous cultures.
- But organized systems of timbral contrasts *within* instruments of a culture are rare. Why should this be?

Timbral contrasts in music cont.

- Temporal profile of a sound
  - the temporal evolution of the amplitude of a sound.
  - A piano tone has a sharp attack and a rapid decay, giving it a percussive quality.
  - The onset and offset of a tone played in a legato manner by a violin are much more gradual.
  - If you play a recording of piano music backwards in time, it sounds like a different instrument.

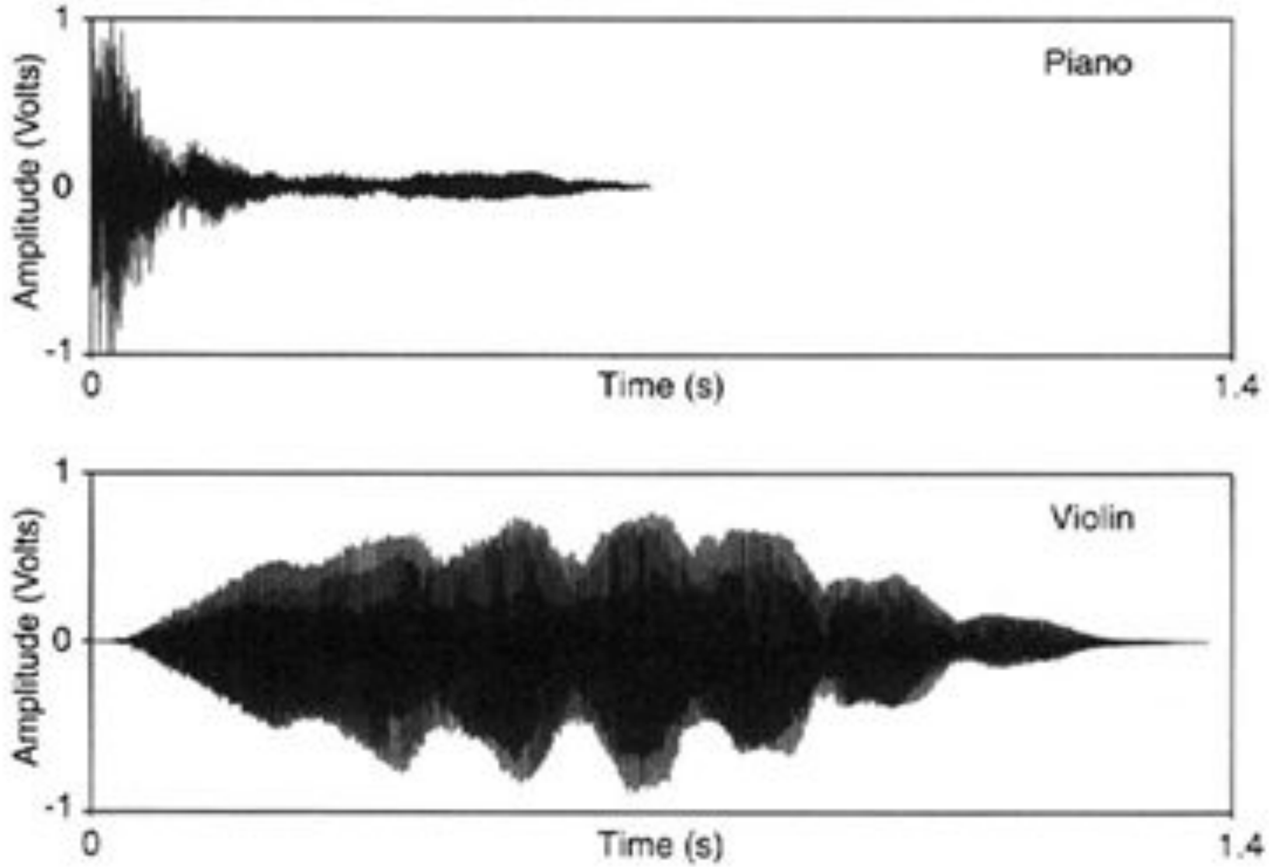

Figure 2.5 Acoustic waveform of a piano tone (top panel) and a violin tone (bottom panel). Note the sharp attack and rapid decay of the piano tone vs. the gradual attack and decay of the violin tone.

Timbral contrasts in music cont.

- Spectral profile of a sound
  - the distribution of frequencies that make up a sound, as well as their relative amplitudes.
- Clarinet
  - Spectrum is dominated by partials that are odd number multiples of the fundamental.
- Trumpet
  - Spectrum doesn't have this asymmetric frequency structure.
- The spectral profile of an instrument depends on the note being played and how loud it is being produced, and it can vary at different points in a room because of acoustic reflections.
  - Yet humans do not perceive dramatic changes in timbre because of these spectral changes.

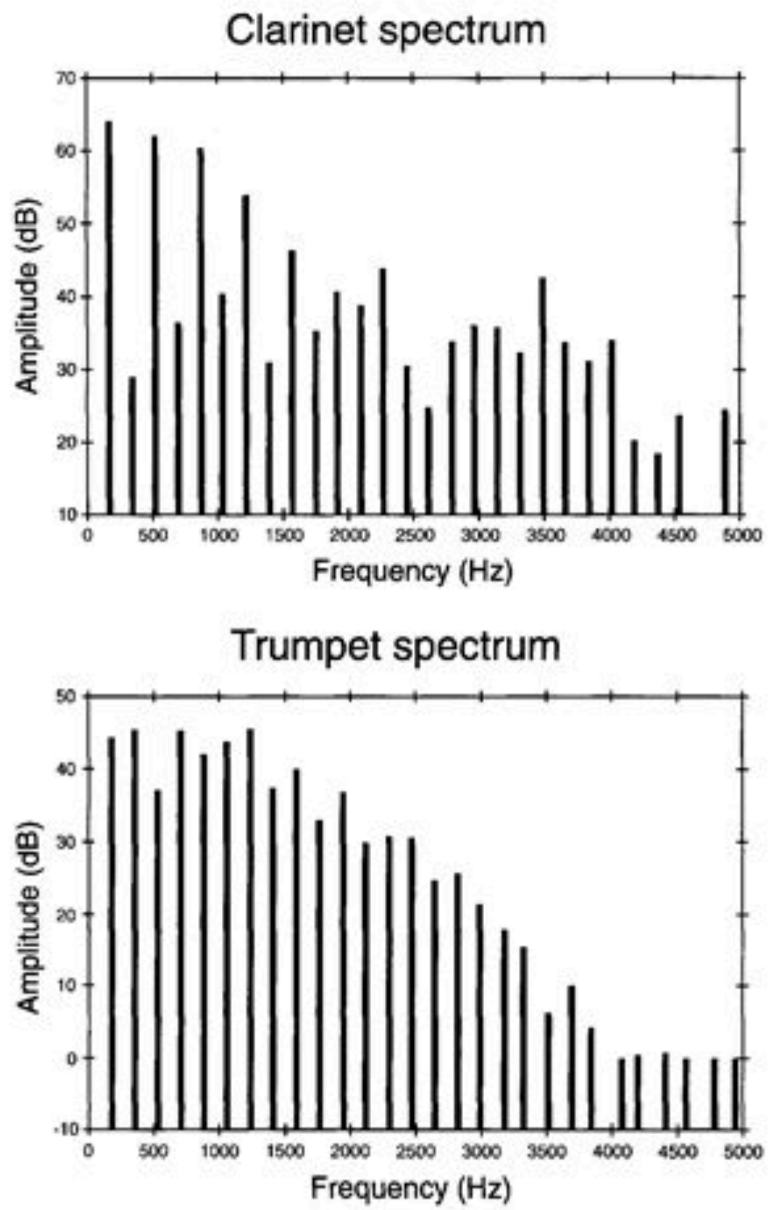

Figure 2.6 Spectrum of a clarinet (top) and a trumpet (bottom) playing the tone F3 (fundamental frequency = 175 Hz). Note the dominance of the odd-numbered partials among the first 10 partials of the clarinet spectrum. (Note: y axis is logarithmic). Courtesy of James Beauchamp.

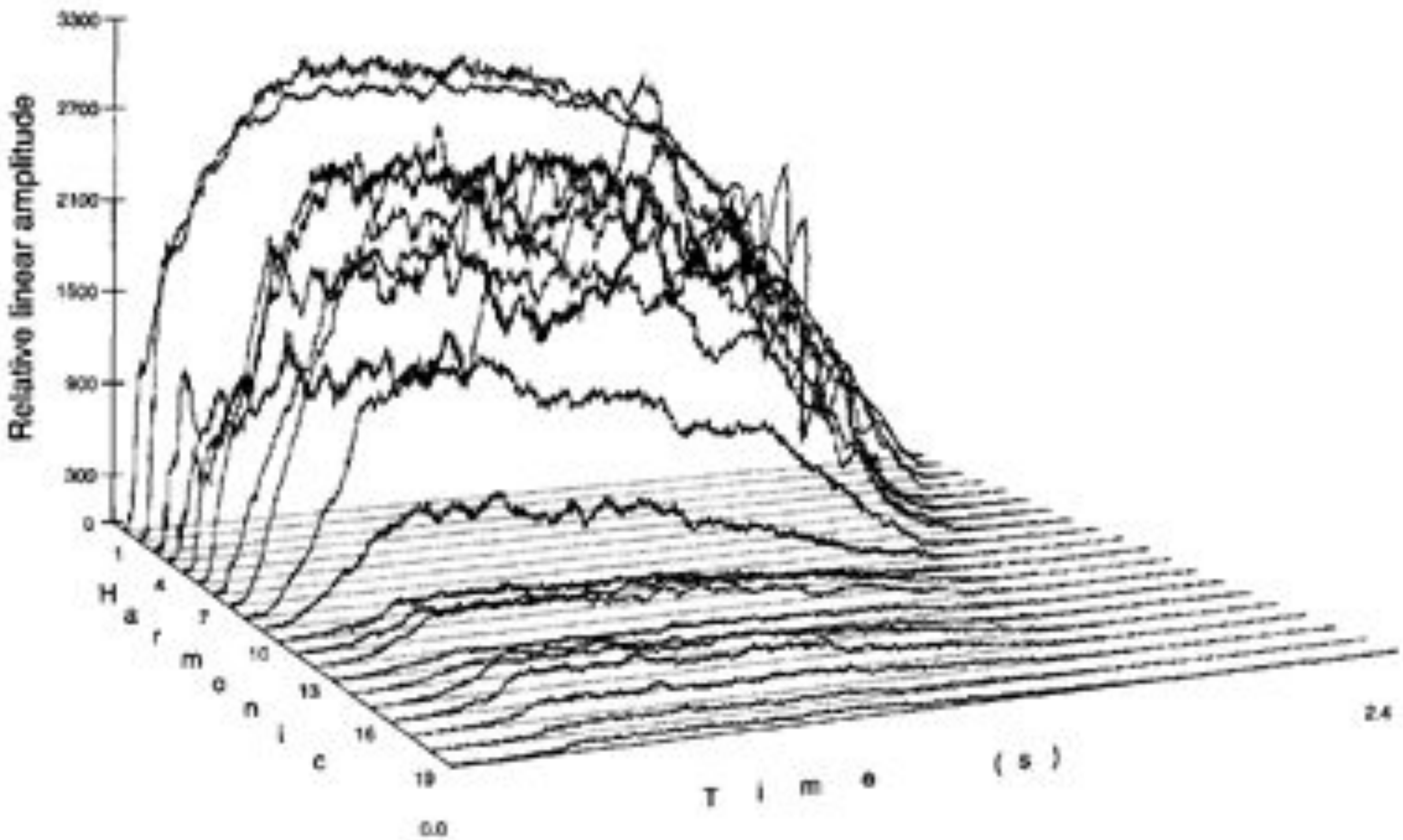

Figure 2.7 Amplitude dynamics of the first 20 partials of a trumpet tone (i.e., the fundamental and 19 harmonics). Note the faster buildup of amplitude in partials 1 and 2 versus 3–10 early in the tone, and the lack of energy in the partials 11–20 until ~200 ms. Courtesy of James Beauchamp.

The rarity of timbral contrasts as a basis for musical sound systems

- Pitch has an advantage over loudness as the basis for sound categories because it is multidimensional.
- But timbre is also multidimensional, and many instruments can produce salient timbral contrasts.
  - A cello can produce different timbres depending on how it is being bowed, by stroking the strings with the wooden part of the bow, or by plucking the strings; but normal cello music is not organized around timbre.
  - Jews harp and Australian didgeridoo use timbre.

The rarity of timbral contrasts as a basis for musical sound systems cont.

- Physical reason
  - Dramatic changes in timbre usually require some change in the way an instrument is struck or bowed, or in the geometry or resonance properties of the instrument.
    - Rapid changes in either on either of these dimensions is usually difficult or impossible.
- Cognitive reason
  - Timbral contrasts are not organized in a system of orderly perceptual distances from either other.
    - In pitch, having a system of intervals allows higher-relations to emerge: a move from C to G can be recognized as similar in size as a move from A to E—the pitches are different but the interval is the same.

Example of a timbre-based musical system

- However, rich sound systems can be based on timbral contrasts that have nothing to do with intervals or scales.
- Drumming tradition of North India, using tabla drums.
- Drum strokes are distinguished in several ways.
  - By which fingers hit the drum, the region of the membrane struck, whether other fingers damp the membrane, etc.
  - This is analogous to different manners and places of articulation in speech.
  - Each kind of stroke is associated with a speech sound (a nonsense syllable called a 'bol'), which sounds similar to how the drum stroke sounds.
  - http://www.youtube.com/watch?v=Vh3DjTbhb6o

Figure 2.9 North Indian tabla drums. The pitch of the dayan is tuned to the basic note of the scale being played on the melodic instrument that it accompanies. The pitch of the bayan is tuned lower, and can be modulated by applying pressure with the heel of the palm to the drum head.

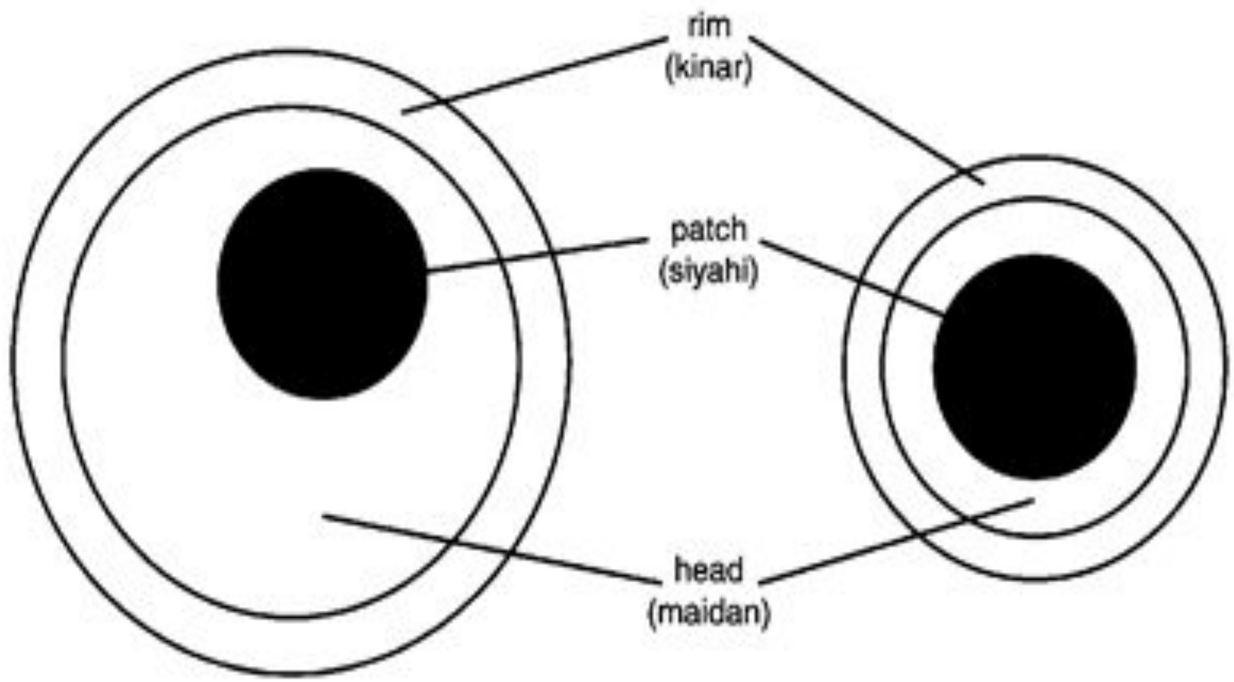

Figure 2.10 Top view of dayan (right) and bayan (left). Each drum surface is divided into three regions: a rim, a head, and a central circular patch of iron fillings and paste. Indian names of the regions are also given. Surface diameters: dayan ~14 cm, bayan ~22 cm.

Bol	Manner of Playing
Tin /ṭin/	*Dayan:* Index finger strikes head. Third finger damps fundamental by resting on head.
Tun /ṭun/	*Dayan:* Index finger strikes patch; no damping.
Kat /kət/	*Bayan:* Damped stroke played with flat hand across rim, head, and patch.
Ghe /gʰe/	*Bayan:* Index finger strikes head at far edge of patch. Heel of hand, on head, can alter pitch.
Tra /ṭrə/	*Dayan:* Damped stroke: middle and index fingers strike patch in rapid succession, and remain on drum.
Kra /krə/	*Bayan:* Damped flat-hand stroke similar to Kat. *Dayan:* Index finger strikes rim. Third finger damps fundamental by resting on head.
Ta /ṭa/	*Dayan:* Index finger strikes rim. Third finger damps fundamental by resting on head.
Dha /ḍʰa/	Simultaneous striking of Ta and Ghe.

Note that fingers bounce off drum (open stroke) unless the stroke is damped (closed). IPA symbols are between slashes.

Timbral contrasts in language

- Although pitch contrasts can be quite organized in language, the primary dimension for organized sound contrasts in language is timbre.
  - Speech is fundamentally a system of organized timbral contrasts.
- The human voice is the supreme instrument of timbral contrast.
  - There are at least 800 distinct phonemes in the world's languages
    - Ranging from 11 (5 vowels, 6 consonants in Rotokas [Papua New Guinea] to 156 (28 vowels, 128 consonants in !Xóõ [Khoisan language, South Africa]. Click song.
    - Average: 27 phonemes

Timbral contrasts in language: Overview

- The timbral contrasts of speech result from continuous changes in the shape of the vocal tract as sound is produced from various sources.
- With consonants, the vocal tract is obstructed or partially obstructed in some way temporarily, and with vowels the sound flows through the vocal tract.
- The succession of timbral contrasts is extremely rapid, e.g. 10 phonemes per second.
- Consonants and vowels tend to alternate.
- How are the timbres of language organized with respect to each other?
  - Articulatory features
    - Consonants: place and manner of articulation.
    - Vowels: position of the tongue in the mouth.

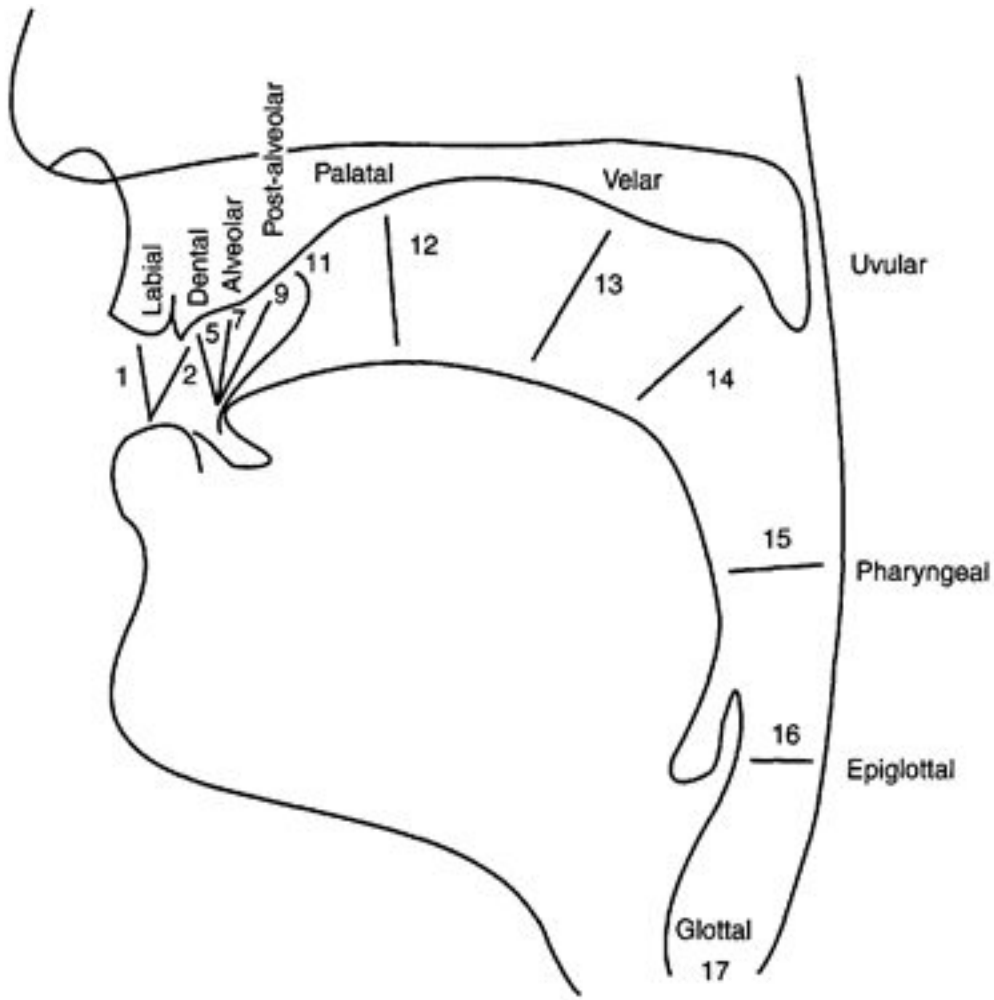

Figure 2.16 Places of articulation in a cross-sectional view of the vocal tract. Numbered lines indicate some of the 17 named articulatory gestures (e.g., 2 is a labiodental maneuver in which the lower lip touches the upper teeth). Note the concentration of places of articulation near the front of the vocal tract. From Ladefoged & Madiesson, 1996.

Table 2.3 Manner Versus Place Classification for Some English Consonants

	Place					
Manner	Bilabial	Labiodental	Dental	Alveolar	Post-alveolar	Velar
Plosive: Total closure in the vocal tract, followed by a sudden release	p b			t d		k g
Nasal: Complete closure of the vocal tract, with lowering of soft palate so air is forced through the nose	m			n		ŋ
Fricative: Narrowing of the vocal tract, causing air turbulence as a jet of air hits the teeth.		f v	θ ð	s z	ʃ ʒ	

Timbral contrasts among vowels

- Vowels are the most musical of speech sounds, having a clear pitch and a rich harmonic structure.
- Number of vowels in a language ranges from 3-24, average 5. American English has about 15, depending on dialect, usually less.
- Vowels are normally voiced, resulting from the vibration of the tensed vocal folds as air from the lungs rushes past them, resulting in a buzzing source.
- The sound source has a frequency spectrum that consists of a fundamental frequency (F0, corresponding to the perceived pitch) and a large number of harmonics.
  - The strongest harmonics are in the bass, typically being the first 5 or 6 multiples of the fundamental.

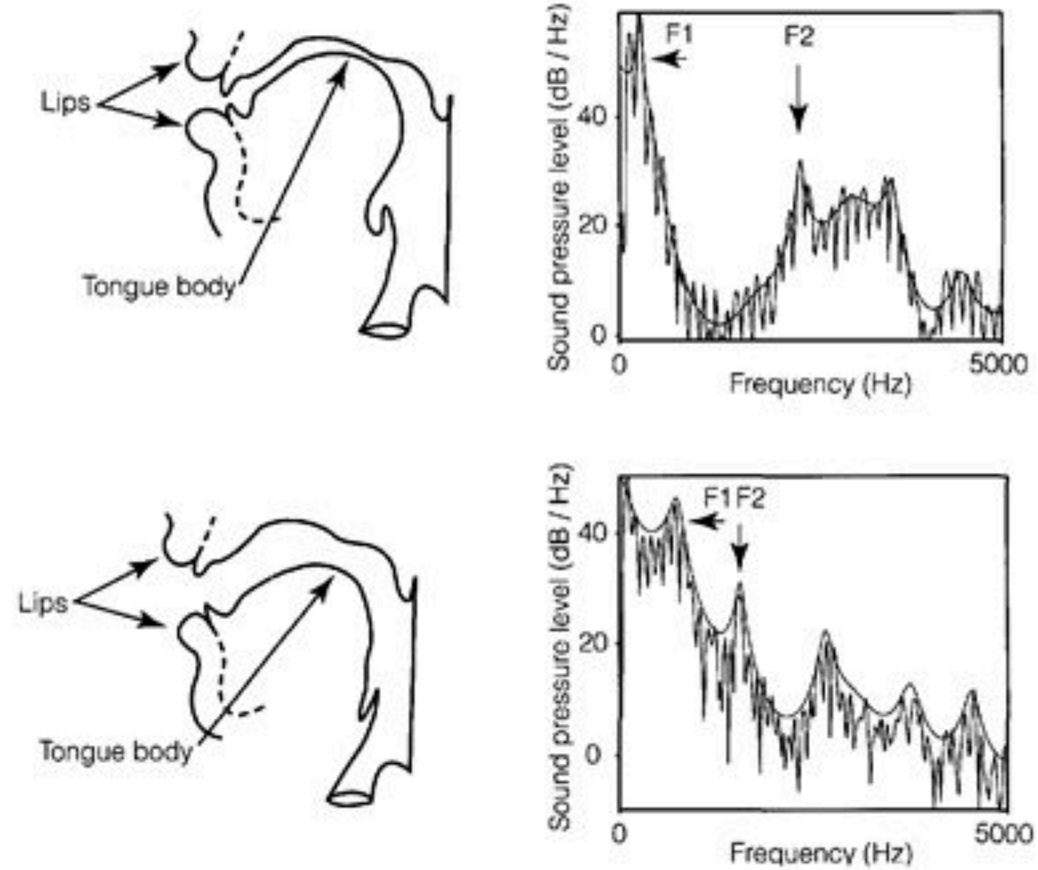

Figure 2.17 Examples of vocal tract configurations and frequency spectra for the vowels /i/ (in "beat") and /æ/ (in "bat") are shown in the top and bottom row, respectively. In the frequency spectra, the jagged lines show the harmonics of the voice, and the smooth curves shows mathematical estimates of vocal tract resonances. The formants are the peaks in the resonance curves. The first two formant peaks are indicated by arrows. The vowel /i/ has a low F1 and high F2, and /æ/ has a high F1 and low F2. These resonances result from differing positions of the tongue in the vocal tract. Vocal tract drawings courtesy of Kenneth Stevens. Spectra courtesy of Laura Dilley.

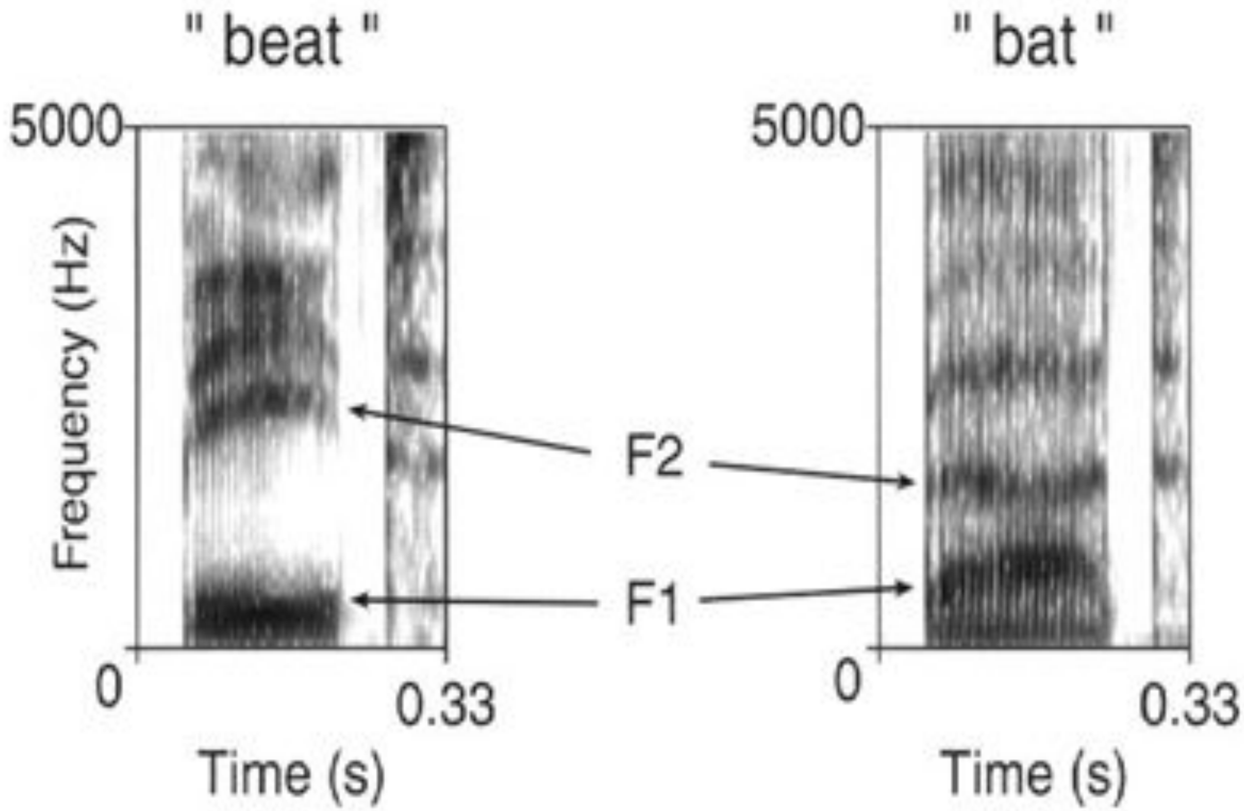

Figure 2.20 Spectrograms of a male speaker saying the words "beat" and "bat," to illustrate the vowels /i/ and /æ/. Arrows point to the first and second formants of the vowels. Note the low F1 and high F2 in /i/, and the inverse pattern in /æ/.

Timbral contrasts among vowels cont.

- The tongue's position in the vocal tract results in acoustic resonances that filter the underlying source spectrum via the emphasis of certain frequency bands.
  - The position and sharpness of these resonances, called *formants*, provide the vowel with its characteristic timbre, which in turn determines its linguistic identity.
  - The position of the first two formants (F1 and F2) are the most important due for vowel identity.

Timbral contrasts among vowels cont.

- The vowels of any language can be represented by placing them on a graph whose axes are the center frequencies of F1 and F2.
- Vowels seem to be organized into acoustic regions with better or more prototypical exemplars in the middle of the region.
- There is a strong relationship between the articulatory configuration and the acoustics of the vowel.
  - The dimension of tongue height (high, mid, low) determines F1.
  - The dimension of tongue backness (back, central, front) determines F2.
  - Roundness of the lips is a third distinctive feature, and nasalization is also used.

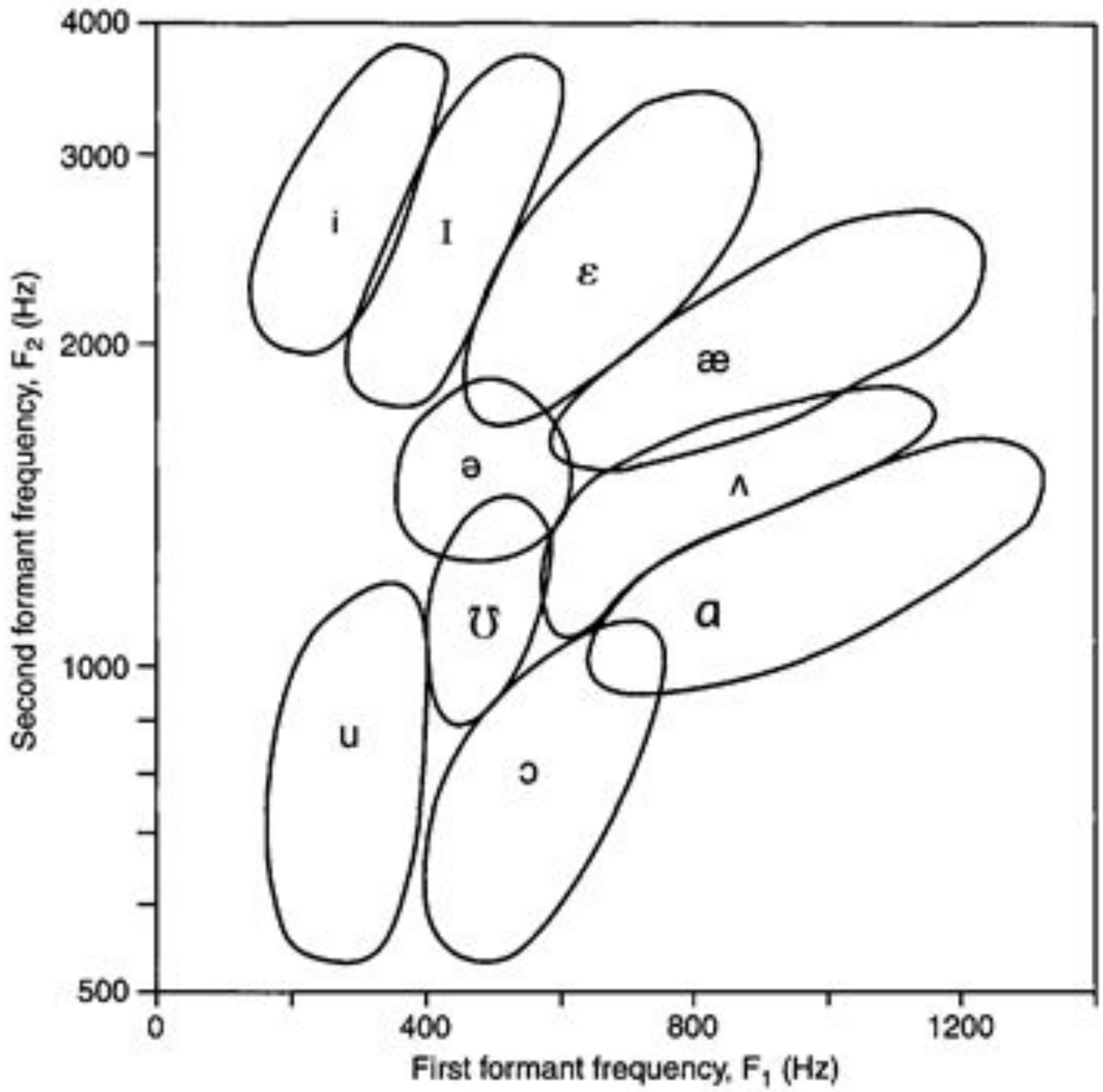

Figure 2.18 Formant frequencies of 10 English vowels as uttered by a range of speakers (x axis = F1, y axis = F2). The ovals enclose the majority of tokens of vowels in a single perceptual category. Symbols follow IPA conventions. Adapted from Peterson & Barney, 1952.

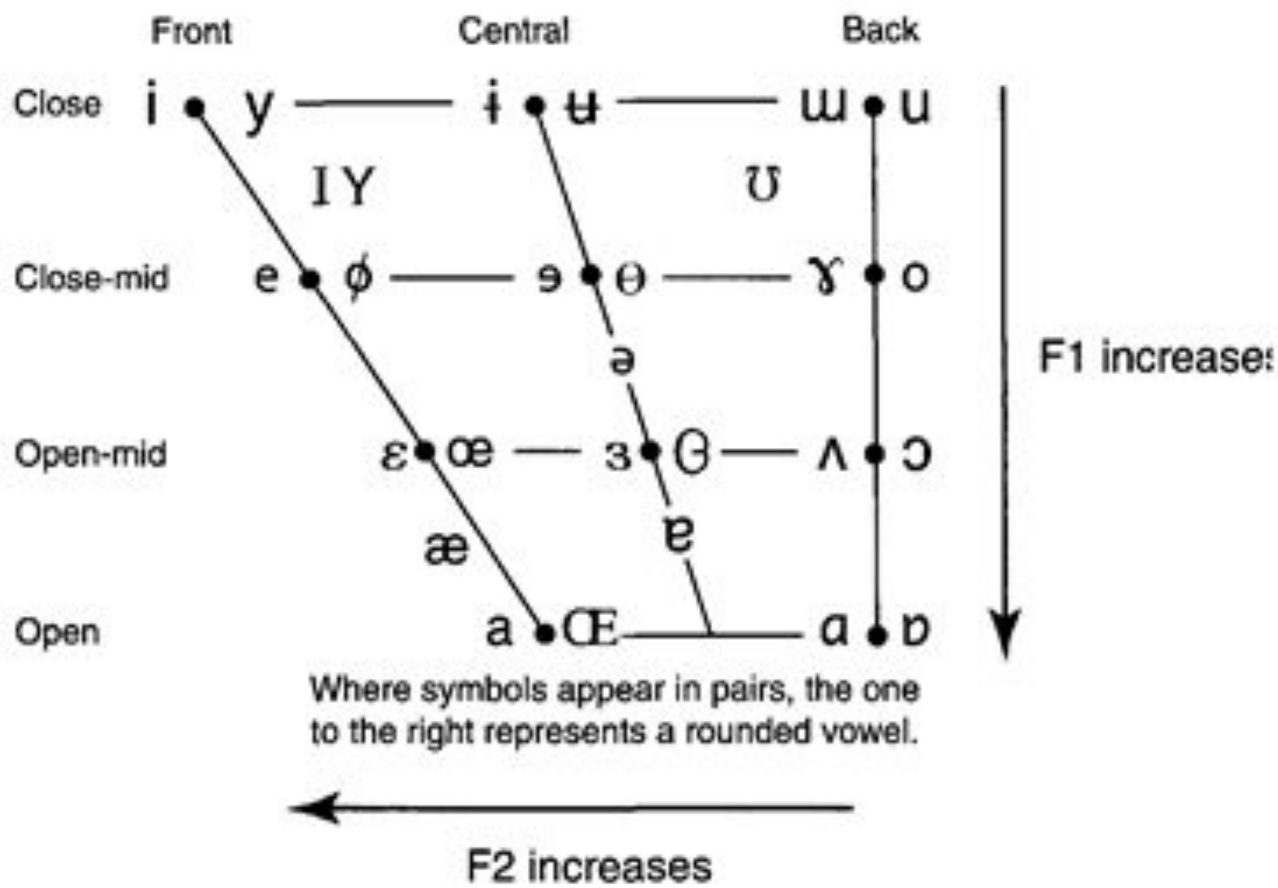

Figure 2.19 The IPA vowel chart. "Front," "back," "close," and "open" refer to the position of the tongue body in the mouth (close = high, open = low). From Ladefoged, 2006. To hear the sounds in this figure, visit: http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/course/chapter1/vowels.html

Timbral contrasts among vowels cont.

- The acoustic structure of a vowel can vary a great deal depending on context.
  - Rapid and informal speech to adults
  - Speech in isolation.
  - Hyper-clear speech to infants (Motherese).
- The amount of acoustic variation between different tokens of vowels in speech exceeds the variation between different tokens of the same note or pitch interval in music.

A brief introduction to the spectogram

- Spectogram: a series of spectra taken at successive points in time that allow us to view the temporal evolution of the frequency structure of a signal.
  - Time: horizontal axis.
  - Frequency: vertical axis.
  - Spectral amplitude: higher amplitudes are darker.
- The individual harmonics can no longer be seen.
  - Instead one sees broad energy bands that correspond to the formants, as well as thin vertical striations corresponding to the fundamental period of the vocal fold vibration.

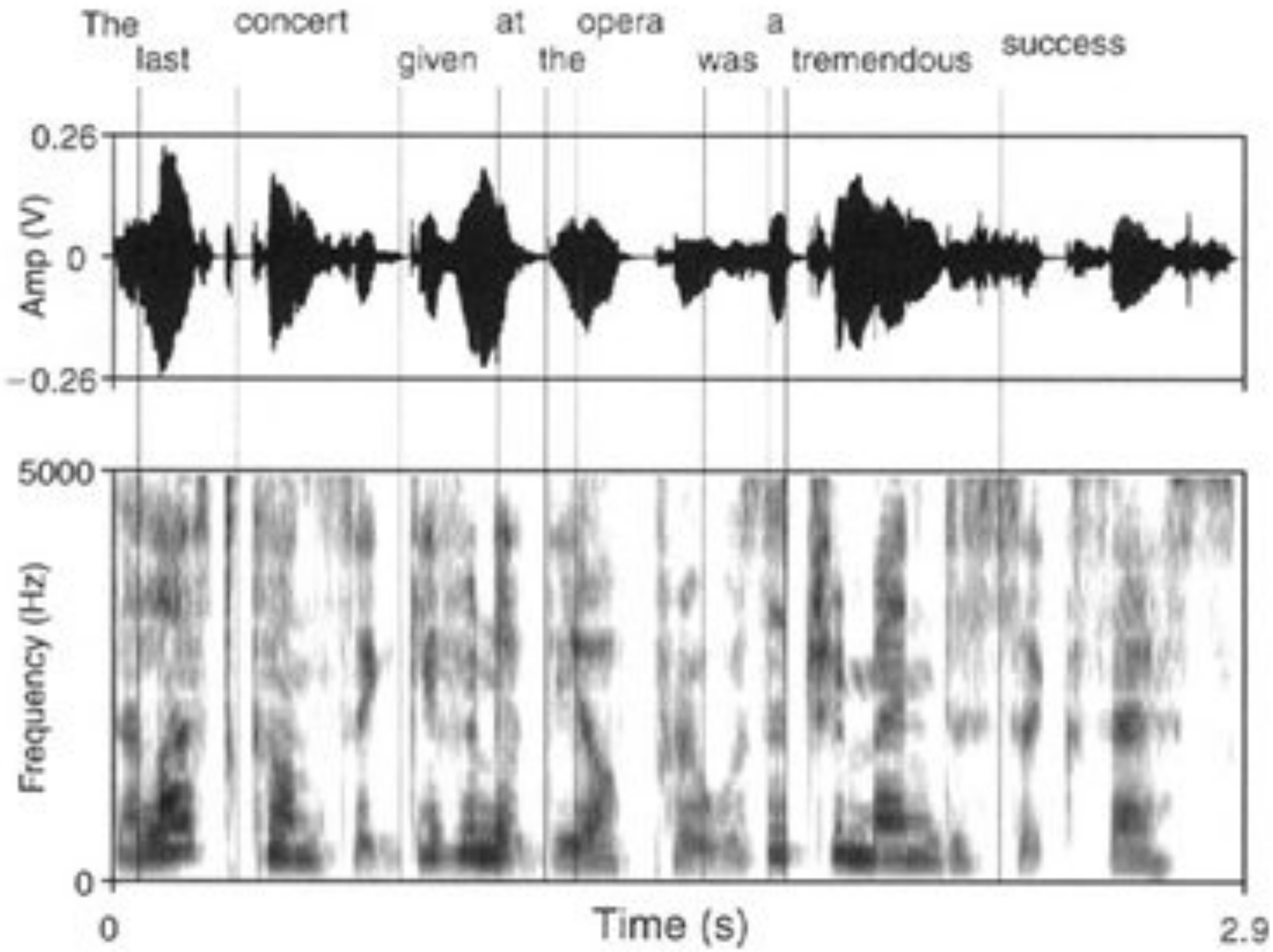

Figure 2.21c Spectrogram of the same sentence as in 2.21a, with word onsets marked.

A brief introduction to the spectogram cont.

- The formants of speech are in almost constant motion.
- Vowels are not the only phonemes with formants.
- The boundaries of words do not necessarily correspond to obvious acoustic discontinuities in the waveform or the spectrogram.

Mapping linguistic timbral contrasts onto musical sounds

- Vocables (nonsense syllables) are used to represent musical sounds in many cultures.
  - We have the soflége system: *do*, *re*, *mi*, *fa*, *so*, *la*, *ti*, *do.*
  - More interesting are cases where there seems to be some acoustic similarity between the vocables and the musical sounds.

Mapping linguistic timbral contrasts onto musical sounds cont.

- Players of the tabla have the intuition that the vocables in their tradition resemble the drum sounds.
  - Patel & Iverson 2003 investigated this.
  - /dha/ is a combination of a /ta/ stroke on the bayan and a /ghe/ stroke on the dayan.
  - The drummed /dha/ composite sound has a brief period of frequency modulation (FM) at the beginning followed by a longer period of stability.
  - The word /dha/ starts with an aspirated voiced stop and then moves into a steady state vowel. Very similar to the drum stroke.
  - Both show a higher ratio of low frequency to high frequency energy during the initial segment.

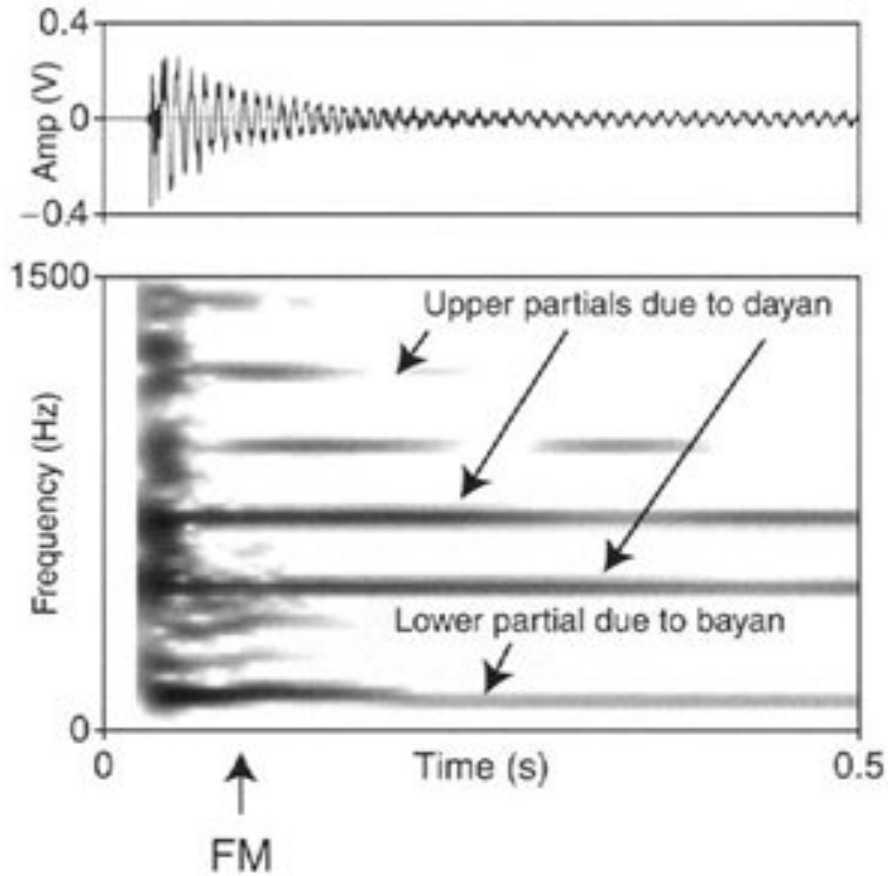

Figure 2.22a Waveform and spectrogram of drummed /dha/, a combination stroke on the dayan and bayan. Note the brief period of frequency modulation (FM) in the bayan sound.

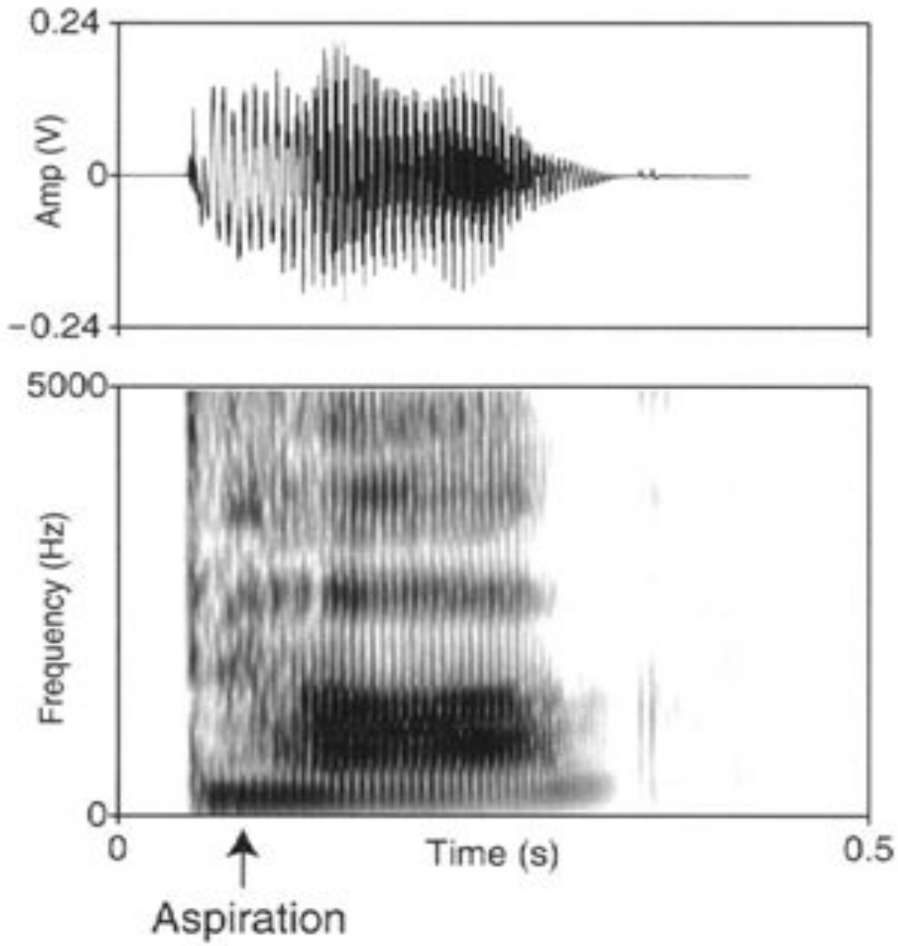

Figure 2.22b Waveform and spectrogram of spoken /dha/. Note the period of heavy aspiration between the onset of the syllable (the release of the /d/) and the onset of stable formant structure in the vowel.

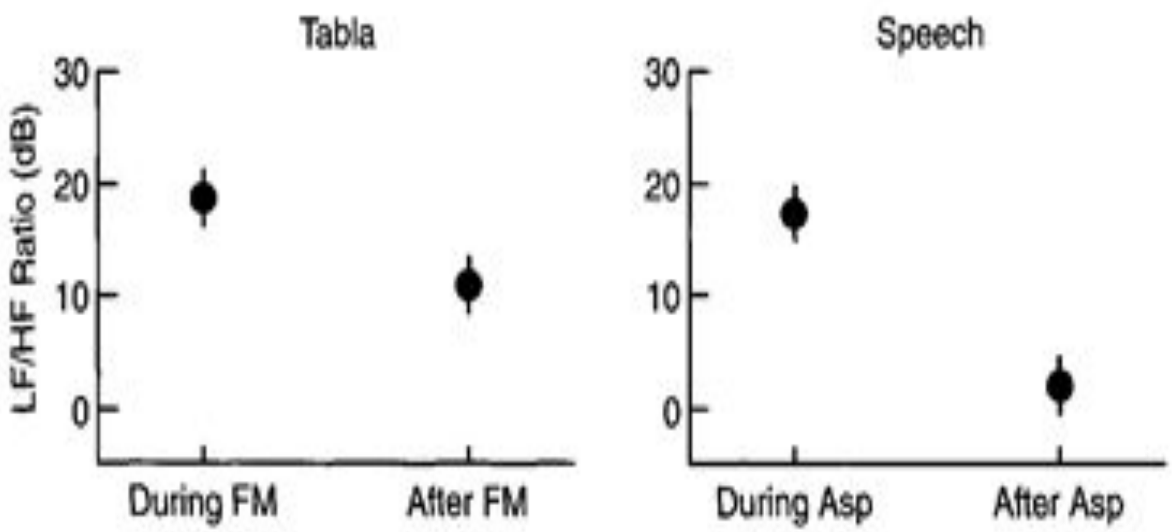

Figure 2.22c Measurements of the ratio of low frequency (LF) to high frequency (HF) energy during the two acoustic stages of the drummed /dha/ and spoken /dha/ (mean and standard error across 30 tokens played/spoken by 6 speakers are shown). In both cases, the early portion of the sound is dominated by low frequency energy, though by completely different mechanisms.

Consonants and vowels as learned sound categories in language

- Evidence abounds that experience with speech produces a mental framework of sound categories that influences the perception of linguistic sound.
- The framework allows a listener to transform acoustically variable tokens into stable mental categories.
- We will look at behavioral evidence pertaining to consonants and then neural evidence pertaining to vowels.

Consonants: perception of nonnative contrasts

- English: /r/ and /l/ are distinct phonemes, but in Japanese the sounds for a single /r/-like phoneme.
- Iverson et al 2003 created an acoustic matrix of /la/ and /ra/ sounds.
  - English speakers labeled the left half as /ra/ and the right half as /la/, but Japanese speakers labeled all but one as /ra/.
  - English speakers showed a peak in discrimination when the two tokens straddled the boundary, but Japanese speakers showed no such peak.

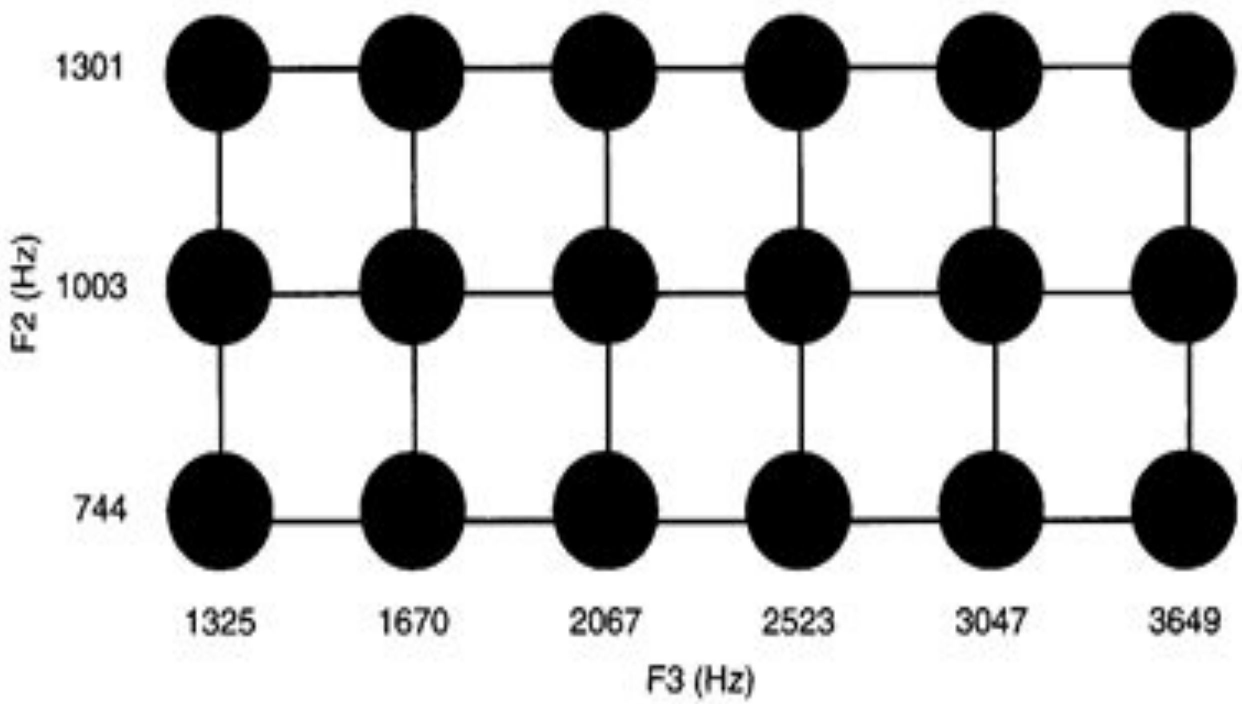

Figure 2.23a Schematic of stimulus grid used by Iverson et al. (2003) to explore the perception of "L" and "R" by American and Japanese listeners. F2 and F3 refer to the second and third speech formant, respectively. The frequencies of F2 and F3 are equally spaced on the mel scale (a psychoacoustic scale for pitch perception).

Consonants: perception of nonnative contrasts cont.

– Mappings on a similarity-rating task showed a completely different way of perceptually "warping" the acoustic space.
  - English speakers were most sensitive to F3 difference.
  - Japanese speakers were more sensitive to variation in F2.

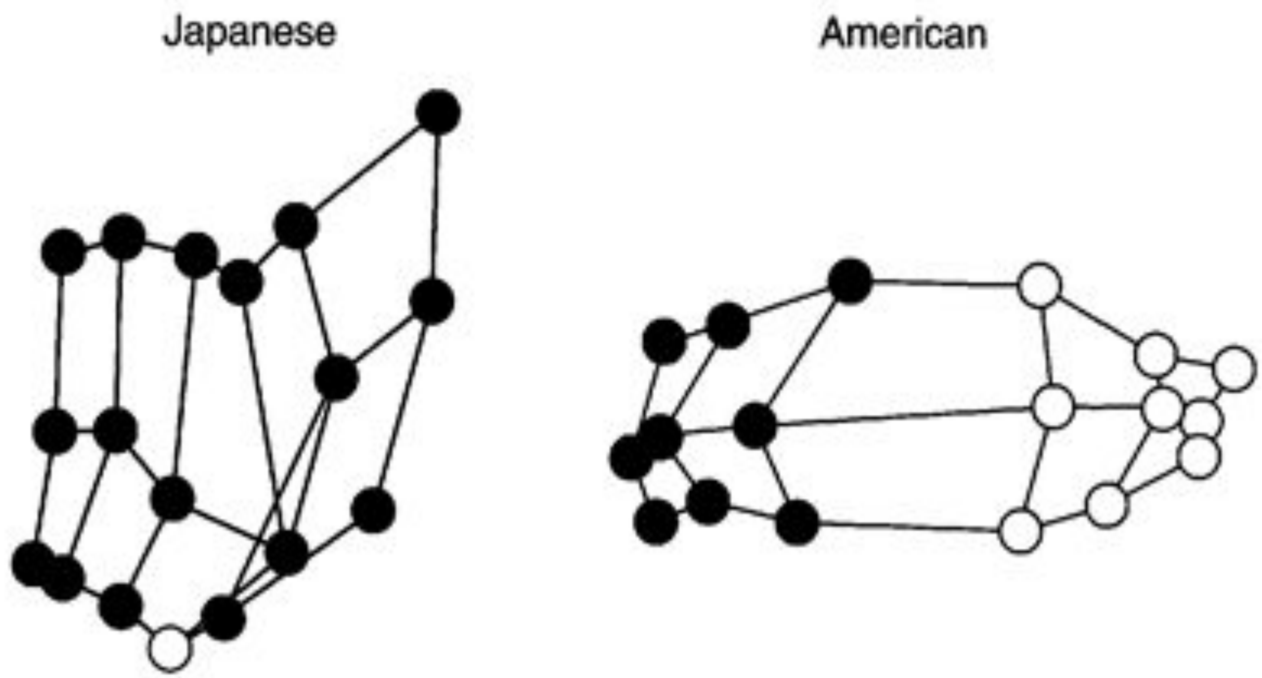

Figure 2.23b Representation of perceived similarity between tokens of the stimulus grid in Figure 2.23a, based on multidimensional scaling. See text for details. Black circles were identified as "R," and white circles were identified as "L" by the American listeners (the white circle in the Japanese data was identified as a "W").

Consonants: perception of nonnative contrasts cont.

- Werker et al 1981 showed that prior to 6 months of age Canadian infants could discriminate two different forms of /t/ in Hindi, but Canadian adults could not.
  - This ability was gone by 1 year of age.
- Best et al. 1988 showed that both American infants and adults could discriminate between click sounds from Zulu.
  - This shows that adults can distinguish sounds that do not resemble anything in the native language.

Vowels: Brain responses to native versus nonnative contrasts

- Mismatch negativity (MMN) technique can be applied to speech perception.
- Näätänen et al 1997 used as a standard a vowel /e/ that occurs in both Finnish and Estonian.
- The deviants were four vowels that ranged between /e/ and /o/ via manipulation of F2.
  - Three deviants corresponded to vowels found in both languages, but one /ö/ only occurs in Estonian.
- Only for /ö/ did the MMN differ between groups.
  - Estonians showed a significantly larger MMN to /ö/, suggesting that a learned sound category was modulating the brain's response.