Timbre perception
Roadmap

overview

functions of music

sound, ear

loudness & pitch

basic qualities of notes

timbre

consonance, scales & tuning

interactions between notes

melody & harmony

patterns of pitches

time, rhythm, and motion

patterns of events

grouping, expectation, meaning

interpretations

music & language
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 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

important for instrument design

“timbral music”:
primary dimension of sonic change

sound mass
ambient music
electronic music
lexical music

Photo courtesy of Miriam Lewis. Used with permission.

Photo courtesy of Pam Roth. Used with permission.

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

Friday, March 13, 2009
What makes different timbres distinctive?

Timbre: a multidimensional tonal quality

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

Stationary Aspects
(spectrum)

Dynamic Aspects

Δ spectrum
Δ intensity
Δ pitch
attack
decay

Vowels

Consonants

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

http://www.wikipedia.org/
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 onsets and 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)

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

Friday, March 13, 2009

Courtesy of Rafael A. Irizarry. Used with permission.
Timbre: a multidimensional tonal quality

tone texture, tone color distinguishes voices, instruments

Stationary Aspects
(spectrum)

Dynamic Aspects
Δ spectrum
Δ intensity
Δ pitch
attack
decay

Vowels

Consonants

Photo courtesy of Pam Roth. Used with permission.

Photo courtesy of Miriam Lewis. Used with permission.

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

Also see:
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.

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♭₃) at three dynamic levels (pp, mf, and ff). Sones are linear increments of subjective loudness.

Figure by MIT OpenCourseWare. After Hanson (1988).
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

Text and images removed due to copyright restrictions.
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.
Time-course of harmonics

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, Delgutte, 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.
[View this image in Google Books]
### Possible interval-based neural correlates for basic phonetic distinctions

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

Friday, March 13, 2009
Music based on timbral contrasts

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

Friday, March 13, 2009
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

Waveforms

Power Spectra

Autocorrelations

[ae]
F0 = 100 Hz

[ae]
F0 = 125 Hz

[er]
F0 = 100 Hz

[er]
F0 = 125 Hz

Formant-related
Vowel quality
Timbre

Pitch periods, 1/F0

0 10 20

0 1 2 3 4

0 5 10 15

Time (ms)

Frequency (kHz)

Interval (ms)

Friday, March 13, 2009
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.)
Two-formant vowel sweeps

First formant (Hz)

Second formant (Hz)

250 500 750

High F2

æ

Low F2

α
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

15
10
5
0

0  1000  2000

Friday, March 13, 2009
Two-formant vowel sweeps

Second formant (Hz)

First formant (Hz)

High F2

æ

Low F2

α

u
High F2 Formant Sweep

[i] → [æ] → [i]

Auditory nerve fiber

CF: 1.4 kHz Thr: 2.0 SR: 90.7

Interspike interval (ms)

0 5 10 15

1/F0

n/F1

1/F1

Friday, March 13, 2009
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

http://www.wikipedia.org/

Photo courtesy of Pam Roth.
Used with permission.

Photo courtesy of Miriam Lewis.
Used with permission.

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

Friday, March 13, 2009
http://en.wikipedia.org/wiki/Timbre

Next up: consonance and scales

Any questions?
HST.725 Music Perception and Cognition
Spring 2009

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.