**Goals for the afternoon**

- **Music Perception and Cognition**
  - What can we hear?
  - How can we describe it?
- **Modeling Musical Intelligence**
  - Focus on pitch structures
  - [1] Modeling tonality
  - Computational music cognition
  - [2] Key Finding
  - [3] Segmentation
  - [4] Pitch Spelling

**Implications for Retrieval**

- Exploit music theoretic knowledge
  - avoid re-inventing the wheel
  - understand the subject at hand
- Create content-based description
  - context, boundaries
  - summarization, indexing
  - similarity assessment
- Build user-centered systems
  - perceptually and cognitively-inspired descriptors
  - human-level apprehension of music

**What can we hear?**

- **stable pitches**
  - [LISTEN] Simple Gifts from Copland’s Appalachian Spring
  - [LISTEN] Schubert Vier Impromptus No.3 D 935 theme
  - [LISTEN] suggestions welcome
- **ordered sets**
  - [LISTEN] Twinkle Twinkle Little Star
  - [LISTEN] Joy to the World
  - [LISTEN] starting from “doh”, octave equivalence

**What can we hear?**

- **context**
  - [LISTEN] Schubert Vier Impromptus No.3 D 935 spliced
  - [LISTEN] Mozart Rondo K.511
  - [LISTEN] random example
- **context change**
  - [LISTEN] Schubert Vier Impromptus No.3 D 935, 2nd half
  - [LISTEN] Mozart Rondo K.511, continued
  - [LISTEN] Bach Minuet in G

**What can we hear?**

- **similarity**
  - Schubert Vier Impromptus No.3 D 935 theme
  - Schubert Vier Impromptus No.3 D 935 var x
  - Mozart Var on “Ah, vous dirais-je, Maman” theme
  - Mozart Var on “Ah, vous dirais-je, Maman” var x
  - Beethoven Piano Sonata Op.79 mvt 3
  - Beethoven Piano Sonata Op.109 mvt 1
How can we describe it?

• scales of description
  – local, global
  – note, cluster, context
• frame of reference

• what is a pitch?
  – A, B, C, #, b
  – pitch class notation
• what is an interval?
  – major/minor
  – augmented/diminished
• what is a chord / triad?
  – I, IV, V
  – ii, vi, iii
• what is a key?
  – “doh” (tonic)

Modeling Tonality: from Experience to Description

• A walk through some history of tonality models
  – Shephard (psychology) [] Krumhansl [] Lerdahl
  – Euler (mathematics) [] Riemann [] Lewin [] Cohn
  – Longuet-Higgins (cognitive science) [] Steedman

• the Spiral Array

Modeling Tonality: Roger N. Shephard

• mental models (1982)
• multi-dimensional scaling

“the cognitive representation of musical pitch must have properties of great regularity, symmetry, and transformational invariance.”

Modeling Tonality: Carol Krumhansl

• the Basic Space (multidimensional scaling)
  – pitch proximity
  – chord proximity
  – key proximity
• application
  – probe tone ratings
  – Key-finding

From Krumhansl (1990) p.46

LISTEN

Mozart bar on “Ah, vous dirais-je, Maman” theme
Modeling Tonality: Carol Krumhansl

- The Basic Space (multidimensional scaling)
  - pitch proximity
  - chord proximity
  - key proximity
    - stepping by fifths
    - relative major/minor
    - parallel major/minor
- Application
  - probe tone ratings
  - Key-finding

Modeling Tonality: Fred Lerdahl

- Tonal Pitch Space
  - pitch space
  - chordal space
  - regional space
Modeling Tonality: Fred Lerdahl

- Tonal Pitch Space
  - pitch space
  - chordal space
  - regional space

Modeling Tonality: Transition

- Tonal Pitch Space (2001)
  - pitch space
  - chordal space
  - regional space

Modeling Tonality: Fred Lerdahl

- Tonal Pitch Space (2001)
  - pitch space
  - chordal space
  - regional space
  - fifths
  - relative maj/min
  - parallel maj/min
Modeling Tonality: Fred Lerdahl

- **Tonal Pitch Space** (2001)
  - pitch space
  - chordal space
  - regional space
  - fifths
  - relative maj/min
  - parallel maj/min

From Lerdahl (2001) p.65

Modeling Tonality: Hugo Riemann

- **The tonnetz** (see Cohn 1998)

Modeling Tonality: Hugo Riemann and Leonhard Euler

- **The tonnetz** (see Cohn 1998)

Modeling Tonality: David Lewin and Richard Cohn

- **Transformational (neo-Riemannian) theory**
  - Dual graph of the tonnetz

Cohn (1997)
Modeling Tonality: David Lewin and Richard Cohn

- Transformational (neo-Riemannian) theory
  - Dual graph of the tonnetz

Modeling Tonality: Hugh Christopher Longuet-Higgins

- Harmonic Network (1962a, 1962b)

Modeling Tonality: Hugh Christopher Longuet-Higgins and Mark Steedman (1971)

Modeling Tonality: Elaine Chew

- Spiral Array (2000)
Modeling Tonality: Elaine Chew

- Spiral Array (2000)

Elaine Chew, University of Southern California
Goals for the afternoon

- Music Perception and Cognition
  - What can we hear?
  - How can we describe it?
- Modeling Musical Intelligence
  - Focus on pitch structures
  1. Modeling tonality
  2. Key Finding
  3. Segmentation
  4. Pitch Spelling

Gehry’s Walt Disney Concert Hall, Los Angeles
Key-Finding

- Krumhansl & Schmuckler (1990)
- Longuet-Higgins & Steedman (1971)
- Chew (2001)

Key-Finding: Krumhansl & Schmuckler

- Probe tone profiles
  - probe tone ratings (Krumhansl & Kessler, 1982)
  - Key-finding (Krumhansl & Schuckler, 1990)

\[ \text{Input vector, } I = [0.375, 0, 0, 0, 0, 0, 0, 0.25, 0.25, 0, 0, 0.125] \]

Calculate correlation coefficient

Key-Finding: Longuet-Higgins & Steedman

- Shape-matching (Longuet-Higgins & Steedman, 1971)
  - successively eliminate options
  - tonic-dominant rule

Key-Finding: Chew

- Center of Effect Generator (Chew, 2001)
  - Clustering of pitches in a key
  - generate center of effect
  - perform nearest neighbor search for closest key
Key-Finding: Chew

- Center of Effect Generator (Chew, 2001)

From Chew 2000 p.104.

Key-Finding: Chew

- Center of Effect Generator (Chew, 2001)

From Chew 2000 p.105.

Key-Finding: Chew

- Center of Effect Generator (Chew, 2001)

From Chew 2000 p.106.

Key-Finding: Comparisons

J.S. Bach’s Well-Tempered Clavier Bk 1


Key-Finding: Comparisons

J.S. Bach’s Well-Tempered Clavier Bk 1

From Chew 2000 p.108.

LISTEN

Elaine Chew, University of Southern California
Key-Finding: Comparisons

J.S. Bach's Well-Tempered Clavier Bk 1

<table>
<thead>
<tr>
<th>Page</th>
<th>Key</th>
<th>Steps to Key</th>
<th>CEG</th>
<th>PEPSI</th>
<th>PARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>183</td>
</tr>
<tr>
<td>2</td>
<td>B♭</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>E♭</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>B♭</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>B♭</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>B♭</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>B♭</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>B♭</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td>B♭</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Average: 5.00 (CEG), 7.00 (PEPSI), 9.00 (PARS)

Segmentation

- Extensions of the CEG algorithm
  - Segmentation Algorithm 1 (Chew 2002)
  - Segmentation Algorithm 2: Argus (Chew 2004)
- Extension of Krumhansl & Schmuckler
  - Dynamic programming approach (Temperley 1999)
Segmentation Algorithm 1: Chew (2002)

Objective: Minimize sum of distances to nearest keys

\[
\sum_{i=0}^{m} \left( d_i + \frac{d_i^2}{2L} \right) = \sum_{i=0}^{m} \left( d_i + \frac{d_i^2}{2L} \right)
\]

Example 1: J.S. Bach’s Minuet in G

Example 2: J.S. Bach’s Minuet in D

Drawbacks:
- Need to know number of boundaries, else need to try all reasonable numbers
- An off-line algorithm

Towards Real-Time Segmentation
Segmentation Algorithm 2: Argus (Chew 2004)
Real-time segmentation algorithm...

Advantages:
• Computes in real-time, O(n)
• Eliminates dependence on key representations
• Segments by pitch collection (more general)
Segmentation Algorithm 2: Argus (Chew 2004b)

Example 1: Schubert’s D780 No.6

Example 1: Schubert’s D780 No.6 (results when w=9)

Example 1: Schubert’s D780 No.6 (results when w=18)

Example 2: Schubert’s D935 No.3

Example 2: Schubert’s D935 No.3 (results when w=48)

Example 2: Schubert’s D935 No.3 (results when w=64)
Goals for the afternoon

- Music Perception and Cognition
  - What can we hear?
  - How can we describe it?
- Modeling Musical Intelligence
  - Focus on pitch structures
    - [1] Modeling tonality
    - Computational music cognition
      - [2] Key Finding
      - [3] Segmentation
      - [4] Pitch Spelling

Why is the example hard?

- What is spelling? Why spell?
- Three algorithms:
  - Cumulative c.e. (Chew & Chen 2003a)
  - Sliding window c.e. (Chew & Chen 2003b)
  - Bootstrapping (Chew & Chen 2003b, in press)
- Joint work with Yun-Ching Chen

Pitch Spelling

Gehry’s Pritzker Pavilion and BP bridge, Chicago

Transcription example

Beethoven Piano Sonata Op.109

LISTEN

Transcription example

Beethoven Piano Sonata Op.109

LISTEN

LISTEN

LISTEN
Recall: Key-Finding

- Center of Effect Generator (Chew, 2001)
  - clustering of pitches in a key
  - generate center of effect
  - perform nearest neighbor search for closest key

Algorithm 1: cumulative c.e.

Assigning pitch names

Algorithm 1: cumulative c.e.

REMARKS

- Insufficient sensitivity to key change.
- No knowledge of voice-leading conventions.
Algorithm 2: sliding window c.e.

Algorithm 3: bootstrapping

REMARKS ... Improved sensitivity to local key changes.

Insufficient sensitivity to sudden changes.

Algorithm 2: sliding window c.e.

Algorithm 3: bootstrapping

REMARKS ... Improved sensitivity to sudden changes.

Combines Algorithms 1 and 2.
### Computational Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
<th>Errors</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beethoven Op.109 (1st movement); 1516 notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative</td>
<td>73</td>
<td>0</td>
<td>99.18</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 4</td>
<td>31</td>
<td>98.00</td>
</tr>
<tr>
<td></td>
<td>w = 8</td>
<td>47</td>
<td>96.00</td>
</tr>
<tr>
<td>Bootstrapping</td>
<td>w=4, r = 2, f = 0.6</td>
<td>28</td>
<td>98.15</td>
</tr>
<tr>
<td></td>
<td>w=4, r = 3, f = 0.8</td>
<td>27</td>
<td>98.22</td>
</tr>
<tr>
<td></td>
<td>w=8, r = 2, f = 0.9</td>
<td>31</td>
<td>97.96</td>
</tr>
<tr>
<td></td>
<td>w=8, r = 6, f = 0.9</td>
<td>27</td>
<td>98.22</td>
</tr>
<tr>
<td>Beethoven Op.79 (3rd movement); 1375 notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative</td>
<td>1</td>
<td>0</td>
<td>99.93</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 4</td>
<td>1</td>
<td>99.93</td>
</tr>
</tbody>
</table>

---

### Op.79 Movement 3: one error

---

### Insensitivity to key change

---

### No knowledge of voice leading
### Related Work 2

**Meredith’s 2004 comparison of pitch spelling algorithms.**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Method</th>
<th>Parameters</th>
<th>Errors</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun, Temerley (1982)</td>
<td>Vaticin</td>
<td>w = 4</td>
<td>73</td>
<td>96.18</td>
</tr>
<tr>
<td>Temerley (1982)</td>
<td>Vaticin</td>
<td>w = 2</td>
<td>47</td>
<td>98.00</td>
</tr>
<tr>
<td>Longuet-Higgins (1962b)</td>
<td>Sliding Window</td>
<td>w = 6, r = 0.6</td>
<td>28</td>
<td>98.15</td>
</tr>
<tr>
<td>Longuet-Higgins (1962a)</td>
<td>Sliding Window</td>
<td>w = 6, r = 0.9</td>
<td>31</td>
<td>97.95</td>
</tr>
<tr>
<td>Meredith (2003)</td>
<td>Sliding Window</td>
<td>w = 4</td>
<td>1</td>
<td>99.93</td>
</tr>
</tbody>
</table>

**Meredith’s 2004 comparison of pitch spelling algorithms.**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Method</th>
<th>Parameters</th>
<th>Errors</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun, Temerley (1982)</td>
<td>Vaticin</td>
<td>w = 4</td>
<td>73</td>
<td>96.18</td>
</tr>
<tr>
<td>Temerley (1982)</td>
<td>Vaticin</td>
<td>w = 2</td>
<td>47</td>
<td>98.00</td>
</tr>
<tr>
<td>Longuet-Higgins (1962b)</td>
<td>Sliding Window</td>
<td>w = 6, r = 0.6</td>
<td>28</td>
<td>98.15</td>
</tr>
<tr>
<td>Longuet-Higgins (1962a)</td>
<td>Sliding Window</td>
<td>w = 6, r = 0.9</td>
<td>31</td>
<td>97.95</td>
</tr>
<tr>
<td>Meredith (2003)</td>
<td>Sliding Window</td>
<td>w = 4</td>
<td>1</td>
<td>99.93</td>
</tr>
</tbody>
</table>

### Computational Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
<th>Errors</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beethoven Op.109 (1st movement), 1516 notes</td>
<td>Cumulative</td>
<td>73</td>
<td>96.18</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 4</td>
<td>31</td>
<td>98.00</td>
</tr>
<tr>
<td>Beethoven Op.109 (1st movement), 1516 notes</td>
<td>Sliding Window</td>
<td>w = 4, r = 0.6</td>
<td>28</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 4, r = 0.9</td>
<td>31</td>
<td>97.95</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 6, r = 0.6</td>
<td>27</td>
<td>98.15</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 6, r = 0.9</td>
<td>31</td>
<td>97.95</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>w = 6</td>
<td>1</td>
<td>99.93</td>
</tr>
</tbody>
</table>

### Transcription example

**[ Spiral Array Pitch Spelling (blue) ]**

### References-1


### References-2

NEXT: Gehry’s Guggenheim Museum, Bilbao