Action and expression in music performance

Giovanni De Poli e Luca Mion
Department of Information Engineering
Centro di Sonologia Computazionale
Università di Padova
1. Why study expressiveness

- Understanding human communication strategies
  - Non-verbal communication
  - Expressiveness tells “how to take” the explicit message
  - Disambiguate language expressions (e.g. in a movie)

- To embody expressive knowledge in machines
- To adapt HCI to the basic forms of human behaviour

- Aim of this work:
  - toward an interpretation of expressive intentions based on the action metaphor
Affective and sensorial domains

- **Affective Domain**: Valence/Arousal space
- **Sensorial Domain**: Kinetics/Energy space
- **Neutral** performance: with no artistic aims
2. Audio features for sound description

- Most relevant audio features:
  - **Local audio features** (frames: 46ms)
    - Roughness (cochlear filter-bank, texture perception)
    - Spectral ratio (energy in the frequency band < 1 kHz)
    - Residual Energy ratio
  - **Event audio features** (frames: 4s)
    - Peak Sound Level
    - Notes per Second
    - Attack time
Representing expressions in the feature space

- k-means algorithm for unsupervised clustering (cosine distance metric)
- clusters = similarity in terms of physical features
  - angry, hard, heavy
  - happy, light
  - calm, sad, soft
3. Perceptual organization: exp. 1

- **Aim**: find how listeners cluster expressive intentions

- Representative adjectives (as attractors):
  - A. angry, **hard**, heavy
  - B. **happy**, light
  - C. **calm**, sad, soft

- **Task**:
  - choose the most similar representative
    - = similar expressive intention, the performer wants to convey

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excerpt 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excerpt 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excerpt 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results from experiment 1

- Pearson’s chi-square test:
  - observed frequencies vs. expected frequencies
  - strong dependence between clusters and attractors
  - neutral performance not categorized, as expected

- Simple Correspondence Analysis and k-means clustering
  - three stable groups
  - hard, happy, calm near to the centroid
Perceptual organization: exp. 2

- **Task:**
  - group the excerpts that convey similar expressive intentions
- **Analysis:**
  - PAM clustering and MDS

- **Clusters**
  - A. angry, hard, heavy
  - B. happy, light
  - C. calm, sad, soft

- **1st dim.:** NPS, A, R
  - qualitative articulation

- **2nd dim.:** PSL, R
  - energetic, quantitative

- **non musicians**
  - mainly 1st dimension
Explaining aspects of musical experience

- Music is not only score or sound;
  - it can convey also sensorial and/or affective contents.
- Musical experience can be described by **metaphors**
  - e.g. emotions, sensorial labels, tension, etc
- **Aims**
  - describe musical expression at an abstract level
    - between sound as waveform and music as language
  - an action metaphor can explain aspects of musical expressiveness?
- **Background:**
  - Enactive knowledge: learning by doing
  - clustering expressions in physical and perceptual spaces
An action metaphor (quantitative vs. qualitative)

Haptic feedback device reacts to actions
Perceptual experiment. Subject task:
- associate each expressive performance to one haptic feedback

Strong relation among expressive intentions and haptic feedbacks
A. Angry/Hard/Heavy ⇔ Friction
B. Happy/Light ⇔ Elasticity
C. Calm/Sad/Soft ⇔ Inertia

Action metaphor
- Friction [quantitative]
  - dissipates (scales) energy
- Elasticity [qualitative]
  - stores elastic energy
  - opposes changes in force
- Inertia [qualitative]
  - stores kinetics energy
  - opposes changes in movement
Action metaphor: discussion

Allows direct association without semantic mediation

Association is semantically plausible
  - e.g. when I’m angry, I feel friction with someone

Comparison of sound envelopes of expressive performances

Can action metaphor be extended to other gesture based arts, such as dance or drawing?

![Normalized envelopes](image)

\[ f(t) \xrightarrow{\text{KID}} v(t) \]
3C. Categorizing expression

- Music and emotions: Expressiveness score dependent
- Clustering musical excerpts
- Valence and Arousal space
  - Focus on affective response
- Main acoustic/musical cues:
  - Valence $\rightarrow$ Modality
  - Arousal $\rightarrow$ Tempo
- Further dimensions:
  - not clear
b. Action metaphor: score dependent

Can action metaphor be applied to other musical contexts?
- 27 musical excerpts from Bigand emotion study
- task: haptic feedback association

Results:
- 5 groups
- 2 dimensions
  - qualitative
  - quantitative

Main acoustic cues:
- Tempo
- Rolloff, ZeroCross

<table>
<thead>
<tr>
<th></th>
<th>Tempo</th>
<th>Rolloff</th>
<th>ZC</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>E</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IF</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>EF</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
c. Looking for secondary factors

- Material: excerpts only in major mode
  - 11 pieces from Bigand’s experiment
  - 12 new pieces, chosen from the Western music repertoire

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R. Strauss - Also sprach Zarathustra</td>
<td>13</td>
<td>G. Faure - nocturne op 84 no 8</td>
</tr>
<tr>
<td>2</td>
<td>A. Vivaldi - Trio Sonata Do Mayor, RV82 allegro</td>
<td>14</td>
<td>L. Beethoven - Piano, Sonata 32, mvt 2</td>
</tr>
<tr>
<td>3</td>
<td>H. Berlioz - La dannazione di Faust - Ballet des Sylphes</td>
<td>15</td>
<td>F. Mendelssohn - Italian Symphony, mvt 1</td>
</tr>
<tr>
<td>4</td>
<td>J. Brahms - Violin Concerto, Adagio.</td>
<td>16</td>
<td>F. Handel - Concerto Grosso Op6</td>
</tr>
<tr>
<td>5</td>
<td>D. Scarlatti, Sonata A for Harpsichord, K208</td>
<td>17</td>
<td>B. Marcello - Sonata No.1 in F Major</td>
</tr>
<tr>
<td>6</td>
<td>R. Schumann - Traumerei, op. 15, n. 7</td>
<td>18</td>
<td>C. Monteverdi - Prologo - Toccata</td>
</tr>
<tr>
<td>7</td>
<td>G. Bizet - Symphony No.1 in C</td>
<td>19</td>
<td>J. Haydn. Symphony Bdur - Hob I 105, Andante</td>
</tr>
<tr>
<td>8</td>
<td>L. Boccherini Luigi - Minuetto</td>
<td>20</td>
<td>W.-F. Bach - Duetto for two flutes in G, Allegro</td>
</tr>
<tr>
<td>9</td>
<td>W. Byrd - Galliard</td>
<td>21</td>
<td>F. Schubert - Valse no. 3, op. 50, D779</td>
</tr>
<tr>
<td>10</td>
<td>C. Debussy - Claire de lune</td>
<td>22</td>
<td>L. Beethoven - Symphony. 7, Vivace</td>
</tr>
<tr>
<td>11</td>
<td>F. Liszt - Poeme symphonique</td>
<td>23</td>
<td>C. Monteverdi - Sonata sopra Sancta Maria</td>
</tr>
<tr>
<td>12</td>
<td>A. Corelli - violin sonata</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Only major mode

- **Subject task:**
  - group the excerpts that convey *similar* expressive intentions

- **Results:**
  - 3 clusters
  - 2 dimensions
    - qualitative
    - quantitative

- **Main acoustic cues:**
  - Tempo
  - Rolloff, ZeroCross

<table>
<thead>
<tr>
<th></th>
<th>Tempo</th>
<th>Rolloff</th>
<th>ZC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HA1</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HA2</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Feature mean values (major-only excerpts)

<table>
<thead>
<tr>
<th>cluster</th>
<th>excerpt</th>
<th>Tempo [bpm]</th>
<th>Rolloff [Hz]</th>
<th>Zerocross</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>B1</td>
<td>61</td>
<td>2372</td>
<td>521</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>52</td>
<td>1868</td>
<td>516</td>
</tr>
<tr>
<td></td>
<td>B4</td>
<td>53</td>
<td>1747</td>
<td>938</td>
</tr>
<tr>
<td></td>
<td>B5</td>
<td>53</td>
<td>1106</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td>B6</td>
<td>50</td>
<td>1028</td>
<td>443</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>78</td>
<td>2289</td>
<td>576</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>54</td>
<td>2707</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>56</td>
<td>1069</td>
<td>449</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>60</td>
<td>2560</td>
<td>713</td>
</tr>
<tr>
<td></td>
<td>B21</td>
<td>76</td>
<td>2487</td>
<td>852</td>
</tr>
<tr>
<td>HA1</td>
<td>2</td>
<td>120</td>
<td>3210</td>
<td>784</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>98</td>
<td>3234</td>
<td>1044</td>
</tr>
<tr>
<td></td>
<td>B14</td>
<td>102</td>
<td>1799</td>
<td>655</td>
</tr>
<tr>
<td></td>
<td>B20</td>
<td>103</td>
<td>1582</td>
<td>735</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>76</td>
<td>1741</td>
<td>498</td>
</tr>
<tr>
<td></td>
<td>B23</td>
<td>59</td>
<td>1786</td>
<td>675</td>
</tr>
<tr>
<td>HA2</td>
<td>7</td>
<td>84</td>
<td>2714</td>
<td>817</td>
</tr>
<tr>
<td></td>
<td>B11</td>
<td>53</td>
<td>4972</td>
<td>650</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>104</td>
<td>4367</td>
<td>1083</td>
</tr>
<tr>
<td></td>
<td>B13</td>
<td>103</td>
<td>3177</td>
<td>972</td>
</tr>
<tr>
<td></td>
<td>B15</td>
<td>145</td>
<td>2495</td>
<td>827</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>72</td>
<td>3229</td>
<td>1121</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>116</td>
<td>5841</td>
<td>1579</td>
</tr>
</tbody>
</table>

Mean value of the selected features, calculated on the 23 excerpts of the major-only experiment

Test of variance:

- **Tempo** (F = 8.3 on 2 and 20 df, p < 0.01)
- **Rolloff** (F = 9.8 on 2 and 20 df, p < 0.01)
- **Zerocross** (F = 11.5 on 2 and 20 df, p < 0.01)
First: Tempo
Second: Spectral energy distribution and zero-crossing
d. Toward a third dimension

First dimension
- mode

Second dimension
- tempo

Other dimension?

Material
- major mode
- MM = 104 bpm

MDS and cluster analysis
Toward a third dimension (cont.)

- Excerpts:
  - major mode
  - MM = 104 bpm

- Preliminary results:
  - 2D MDS space
  - 3 clusters

- Kinematics – Energy?

- Features:

<table>
<thead>
<tr>
<th>#</th>
<th>ZeroCross</th>
<th>LowEnergy</th>
<th>RMS</th>
<th>Roughness</th>
<th>SRh</th>
<th>SRm</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim1</td>
<td>-.30</td>
<td>.16</td>
<td>-.51</td>
<td>-.54</td>
<td>-.09</td>
<td>-.01</td>
<td>-.55</td>
</tr>
<tr>
<td>dim2</td>
<td>.17</td>
<td>-.23</td>
<td>.20</td>
<td>.30</td>
<td>.07</td>
<td>-.17</td>
<td>.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Centroid</th>
<th>Brightness</th>
<th>Rolloff</th>
<th>Onsets</th>
<th>Attacks</th>
<th>BS</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim1</td>
<td>-.36</td>
<td>-.32</td>
<td>-.41</td>
<td>.16</td>
<td>.37</td>
<td>.40</td>
<td>-.56</td>
</tr>
<tr>
<td>dim2</td>
<td>.02</td>
<td>.20</td>
<td>.04</td>
<td>-.15</td>
<td>.16</td>
<td>.31</td>
<td>.19</td>
</tr>
</tbody>
</table>
Discussion and conclusions

- **Three clusters** emerged, both in the feature space and in the subjects evaluation
- **Agreement** between clusters:
  - listener evaluation in agreement with performer intention
- **Correspondence** of affective and sensorial expressive intentions

- Perceptual tests with visual and haptic representatives
- **Can action metaphor be extended to other gesture based arts, such as dance, drawing etc.?**