An action based metaphor for description of expression in music performance

Luca Mion

CSC-SMC, Centro di Sonologia Computazionale
Department of Information Engineering
University of Padova

Workshop Toni Mian, October 26, 2007
Why modelling 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

- Expression through the auditory channel
  - Artistic applications, games and entertainment
  - Auditory Display
  - Music Information Retrieval
  - Education
Why modelling 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

- Expression through the auditory channel
  - Artistic applications, games and entertainment
  - Auditory Display
  - Music Information Retrieval
  - Education
Why modelling 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

- Expression through the auditory channel
  - Artistic applications, games and entertainment
  - Auditory Display
  - Music Information Retrieval
  - Education
Why modelling 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

- Expression through the auditory channel
  - Artistic applications, games and entertainment
  - Auditory Display
  - Music Information Retrieval
  - Education
Expressive intention is a broad concept:

- emotions $\rightarrow$ affective domain
- metaphoric aspects $\rightarrow$ sensory domain

We want to describe expression at an intermediate level between sound as waveform and music as language.

Yet another metaphor?

We propose an interpretation of the expressive intentions based on action metaphor and on ideal physical systems.
Expression in music performance

Expressive intention is a broad concept:

- emotions \(\rightarrow\) affective domain
- metaphoric aspects \(\rightarrow\) sensory domain

We want to describe expression at an intermediate level between sound as waveform and music as language

yet another metaphor?

we propose an interpretation of the expressive intentions based on action metaphor and on ideal physical systems
Objectives

- Exploring similarities of affective and sensorial expressive intentions
- Organization of the feature space
- Interpretation of affective and sensorial expressions clustering
- Perceptual experiment to study whether/how the listeners cluster expressive intentions
Methodology

- **Feature selection and validation:** Sequential Forward Selection (SFS) and Minimum Distance Classifier
- **Projection** on a 2D space by Principal Component Analysis (PCA)
- **Interpretation** of the expression clusters (*k*-means clustering)
- **Perceptual experiment**, Pearson’s $\chi^2$ test, Correspondence Analysis

**Audio data:** performances from Flute, Violin and Guitar (repeated notes, scales, excerpts)
The affective and sensorial domains

- **Affective Domain**: the Valence/Arousal space
  - Happy, Sad, Angry, Calm

- **Sensorial Domain**: the Kinetics/Energy space
  - Hard, Soft, Heavy, Light

- Neutral performance: with no artistic aims
The affective and sensorial domains

- **Affective Domain**: the Valence/Arousal space
  - Happy, Sad, Angry, Calm

- **Sensorial Domain**: the Kinetics/Energy space
  - Hard, Soft, Heavy, Light

- **Neutral performance**: with no artistic aims
The affective and sensorial domains

- **Affective Domain**: the Valence/Arousal space
  - Happy, Sad, Angry, Calm

- **Sensorial Domain**: the Kinetics/Energy space
  - Hard, Soft, Heavy, Light

- **Neutral performance**: with no artistic aims
Audio features for sound description

**Local audio features** (on 46 ms-length non overlapping frames)

- **Roughness** $R$ (cochlear filter-bank, texture perception)
- **Spectral Ratio** $SR_a = \sum_{j \in LB} |X(j)|^2 / \sum_{k=1}^{N/2-1} |X(k)|^2$
  - indicates the amount of energy in the low frequency band $LB$ ($f < 1$ kHz)
- **Residual Energy ratio**
  $$RE_h = \sum_{j \in HB} |X_R(j)|^2 / \sum_{k=1}^{N/2-1} |X(k)|^2$$
  - describes the stochastic energy in the high frequency band $HB$ ($f > 1.8$ kHz)

**Event audio features** (on 4s-duration and 3.5s-overlapping frames)

- **Peak Sound Level** $PSL$
- **Notes per Second** $NPS$
- **Attack time** $A$
Representing expressions in the feature space

\( k\)-means algorithm for unsupervised clustering (cosine distance metric)
\[
d = 1 - \cos(x, y) = 1 - \frac{\langle x, y \rangle}{||x|| \times ||y||}
\]
Interpretation of the joint space

Using the physical analogy:

- Force is considered as the cause, movement (velocity or position) as the effect
- Cause-effect relationship: represented by the **admittance** operator $Y$:

\[
\nu(t) = \int_{-\infty}^{t} f(\tau) \cdot y(t - \tau) d\tau
\]

- It describes the dynamic mapping and the qualitative behavior from force to velocity by an integral-differential equation.
Admittance as metaphor

cause \[ \rightarrow \] effect

Damping

\[ v(t) \rightarrow \frac{1}{\mu} \rightarrow f(t) \]

Friction

\[ v(t) = \frac{1}{\mu} f(t) \]

Mass

\[ v(t) \rightarrow \frac{1}{m} \rightarrow f(t) \]

Inertia

\[ v(t) = \frac{1}{m} \int f(t) \, dt \]

Spring

\[ v(t) \rightarrow \frac{1}{k} \rightarrow f(t) \]

Elasticity

\[ v(t) = \frac{1}{k} \frac{df(t)}{dt} \]

\textbf{(low) Friction: Angry - Hard, Heavy} \\
\textbf{Elasticity: Happy - Light} \\
\textbf{Inertia: Sad, Calm - Soft}
We want to find how the listeners cluster expressive intentions.

- 16 participants (musicians and not musicians)
- 9 expressive intentions (4 emotional, 4 sensory plus neutral)
- 3 performances representative for our clusters

Representative adjectives (minimizing the cosine distance): hard for friction, happy for elasticity, calm for inertia.

**Task**: associate to one representative the expressive intention that the performer wants to convey.
We want to find how the listeners cluster expressive intentions.

- 16 participants (musicians and not musicians)
- 9 expressive intentions (4 emotional, 4 sensory plus neutral)
- 3 performances representative for our clusters

Representative adjectives (minimizing the cosine distance): *hard* for friction, *happy* for elasticity, *calm* for inertia.

**Task:** associate to one representative the expressive intention that the performer wants to convey.
Pearson’s $\chi^2$ test:

- compared observed frequencies with expected frequencies
- strong dependence between clusters and adjectives
  $\chi^2 = 401.99$ (critical value = 36.12, $\alpha = 0.001$, df=14)
- neutral performance not categorized, as expected by the performer
Results from perceptual test (contd.)

Simple Correspondence Analysis on the contingency table:
- 2 factors (74.19% and the 25.81% of the total inertia)
- agreement with $k$-means clustering on the feature space
  - three stable groups
  - hard, happy, calm near to the centroid of respective cluster (distances 0.088, 0.214, 0.199 respectively)
- intrinsic similarity of affective and sensory expressive intentions

![Correspondence analysis plot (violin)](image)
Three clusters emerged, both in the feature space and in the subjects evaluation

Agreement between clusters (\(k\)-means clustering)

\[ \rightarrow \] listener evaluation in agreement with performer’s 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.?
Thanks for your attention.

Sound and Music Computing Group
http://smc.dei.unipd.it
luca.mion@dei.unipd.it