Comparing affective and sensorial expressions in music performance by machine learning techniques

Luca Mion and Giovanni De Poli
University of Padova
{luca.mion,depoli}@dei.unipd.it

Objectives

A paradigm for music expression understanding for:

- *Recognition* of the expressive content in audio data by machine learning techniques;
- *Organization* of the feature space;
- *Interpretation* of affective and sensorial expressions clustering.

Motivations

A few study were conducted on automated analysis of expression with reference to:

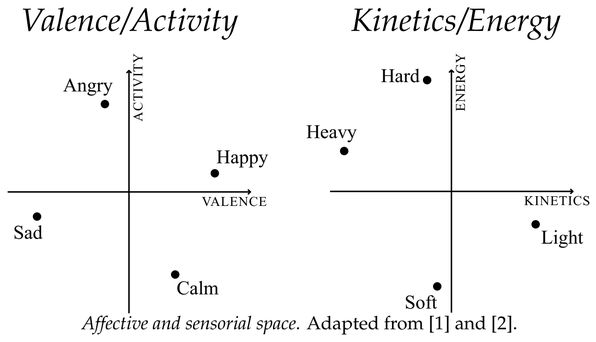
- *Scoreless* audio material;
- *Joint Space* from affective and sensorial domains.

Method

- *Feature selection* and validation: Sequential Forward Selection (SFS) and Naïve Bayesian Classifier;
- *Projection* on a 2D space by Principal Component Analysis (PCA);
- *Interpretation* of the expression clusters.

Audio data Performances from Flute, Violin and Guitar (repeated notes, scales, excerpts).

Affective and Sensorial Domains



Affective and sensorial space. Adapted from [1] and [2].

Results of Feature Selection

- *On frame:*

  Roughness R: texture perception

  Spectral Ratio SR_a: energy above/below 1 kHz, relevant for brightness

  Residual Energy RE_{sb}: related to effort

  $RE_{sb} = \sum_{j \in sb} |X_R(j)|^2 / \sum_{k=1}^{N/2-1} |X(k)|^2$, where sb indexes each subband and X_R is the spectrum of the residual component.

- *On window:*

  Notes per Second NPS, Attack time A, Peak Sound Level PSL.

- *General* features: NPS, PSL, A, R (set $\mathcal{G}$);
- *Specific* features: SR_a (flute), RE_h (violin).

Expression Recognition

Features were tested via supervised classification (Naïve Bayesian):

- Selected features well discriminate the expressive content:

  High recognition rating: 64.58% with set $\mathcal{G}$

- Extend traditional musical parameters with perceptual features

  Features at an intermediate level, no music language

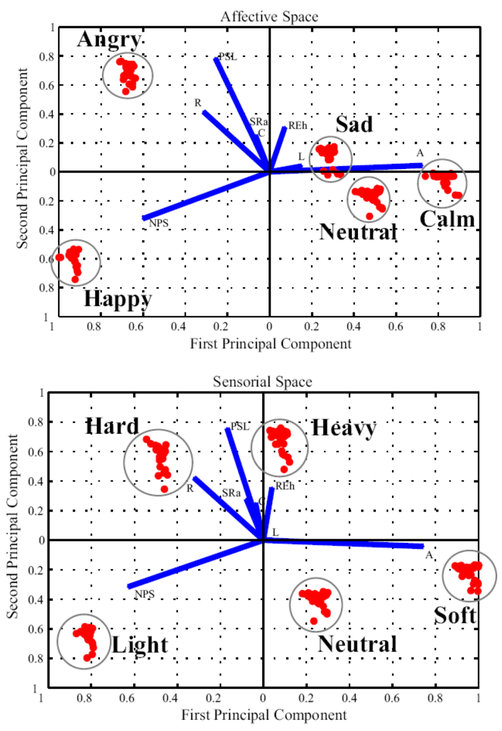
  Parameters can be mapped to physical properties

  Metaphors and physical features are possible [3]

Features for General and Specific Description

Affective and sensorial space separately

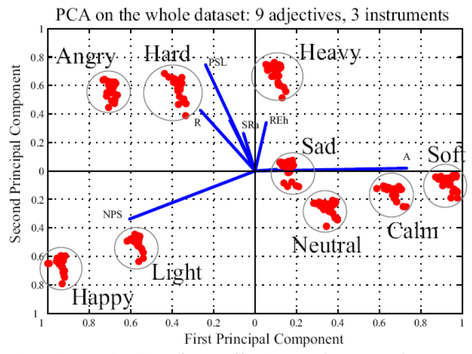
- Adjectives are placed according to the theoretical spaces;
- Set $\mathcal{G}$ differentiate the expressions;
- Features can be related to the dimensions *Valence/Activity* and *Kinetics/Energy*.



*PCA on affective adjectives (top) and sensorial adjectives (bottom).*The factor loadings of each feature contribution are shown.

Description of the Joint Space

- Main finding: performances are grouped in 3 clusters;
- Features $\mathcal{G}$ still differentiate the expressions.



PCA, 3 instruments. Data from affective and sensorial expressions.

Mapping into physical descriptions:

- **Timing** (fast/slow) $\longleftrightarrow NPS$
- **Intensity** (loud/weak) $\longleftrightarrow PSL$
- **Attack** (sudden/loose) $\longleftrightarrow A$
- **Texture** (rough/smooth) $\longleftrightarrow R$
- **Brightness** (bright/dark) $\longleftrightarrow SRa$
- **Effort** (strong/weak) $\longleftrightarrow REh$

Physical analogy	Adjectives	**Tempo** + fast - slow	**Intensity** + loud - weak	**Attack** + loose - sudden	**Texture** + rough - smooth	**Brightness** + bright - dark	**Effort** + strong - weak
friction	Hard, Heavy – Angry	-	+++	-	++	---	++
elasticity	Happy – Light	+++	--	--	-	+	--
inertia	Sad, Calm – Soft	---	---	+++	--	++	-

Interpretation of the Joint Space

- Sensorial adjectives may have an emotional quality.
- Expression communication associated with superficial cues [4].
- How to develop a metaphor to combine different semantic spaces?

Enactive interpretation

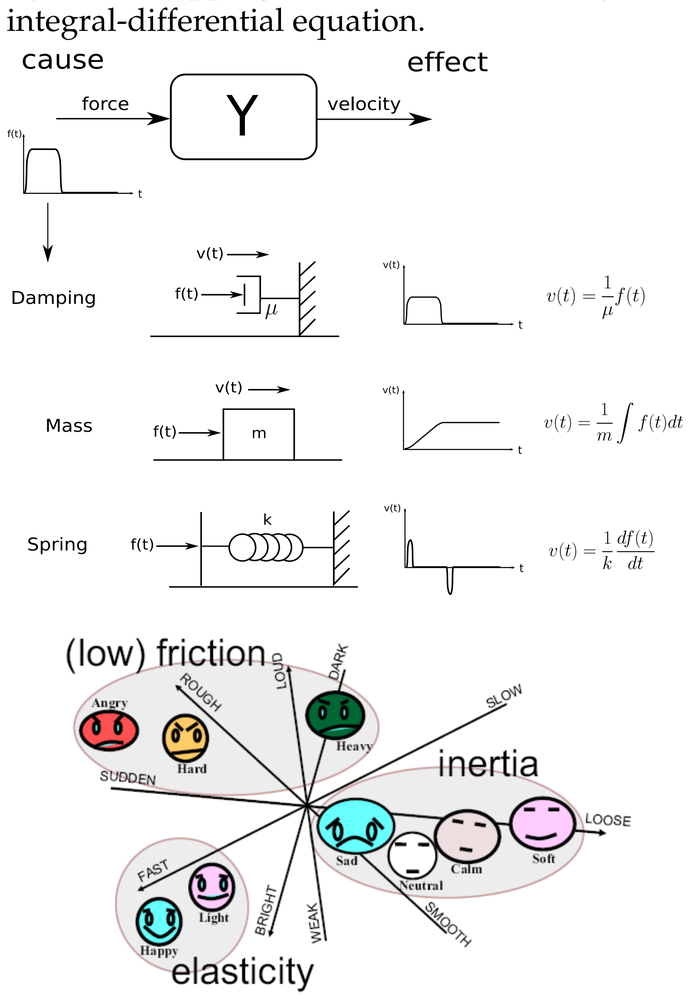
- Words as hard, light or heavy are used to describe emotional meanings.
- **Actions** can be considered as an intermediate level of representation between the semantic and superficial feature levels [3].
- Reference to human physical behavior can be a common denominator to all the multi-modal expressive actions (play fluid, dance fluid).

Action and physics analogy:

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

Admittance as Metaphor

Dynamic mapping from force to motion by an integral-differential equation.



(low) Friction: Hard, Heavy, Angry
Elasticity: Happy, Light
Inertia: Sad, Calm, Soft

Conclusions

- Selection of features by machine learning techniques;
- Projection of the feature space:

  separated: strong analogies with the theoretical spaces;

  joint: three clusters emerged.

- Action-based interpretation

Open Issues

- Perceptual experiments to confirm the three clusters.
- Can the admittance-based metaphor be extended to other expressive domains, where the **action** is a key factor (e.g. dance, drawing)?

References

[1] J. A. Russell, "A circumplex model of affect", *Journal of Personality and Social Psychology*, vol. 39, pp. 1161-1178, 1980.
[2] S. Canazza, G. De Poli, A. Rodà, and A. Vidolin, "An abstract control space for communication of sensory expressive intentions in music performance", *Journal of the New Music Research*, vol. 32, no. 3, pp. 281-294, 2003.
[3] G. De Poli, F. Avanzini, A. Rodà, L. Mion, G. DIncà, C. Trestino, D. Pirrò, A. Luciani, and N. Castagne, "Towards a multi-layer architecture for multi-modal rendering of expressive actions", In Proc. 2nd Int. Conf. on Enactive Interfaces, 2005.
[4] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau and A. Dacquet, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts", *Cognition and Emotion*, vol. 19, no. 8, pp. 1113-1139, 2005.

Acknowledgements: This work was supported by the European Network of Excellence *Enactive Interfaces*.