Joyful for you and tender for us: the influence of individual characteristics and language on emotion labeling and classification

Juan Sebastián Gómez-Cañón, Estefanía Cano, Perfecto Herrera, Emilia Gómez

In this work, we address **the relation between the emotions perceived in pop and rock music and the language spoken by the listener**. Two main research questions are addressed:

- **RQ1** - Do personal characteristics and mother tongue have an influence on the annotation of perceived emotions for listeners?
- **RQ2** - Can this information be used to improve MER algorithms?

Problem definition: inter-rater agreement

- **Tagging a song** is vague and ambivalent
  - This also applies to **auto-tagging**, **genre recognition**, **similarity**, **chord estimation**, **beat tracking** **(!)**
- **Perceived emotions** vs. induced emotions
- Musical emotion complexity relates to **personal- and cultural-specific associations**

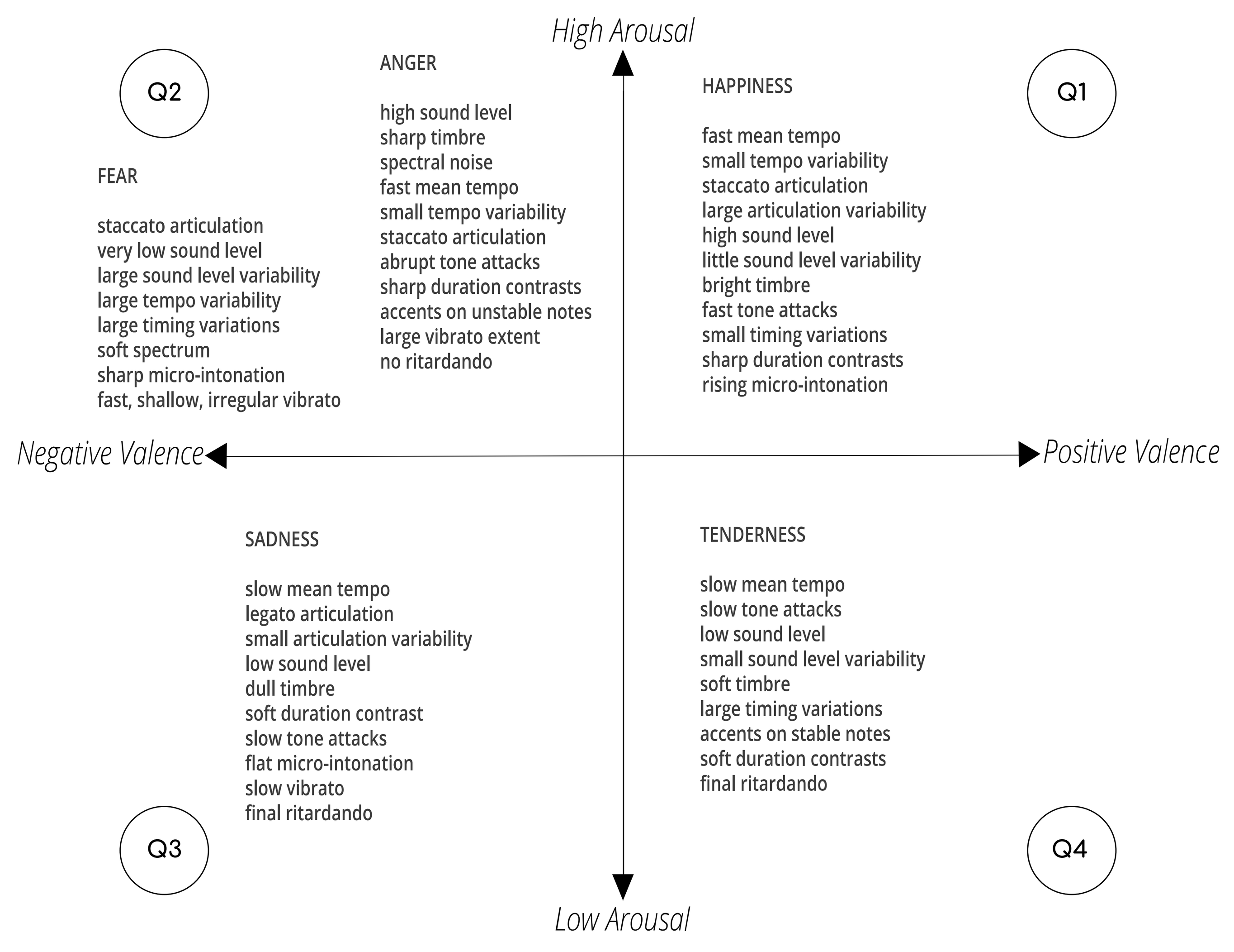

Fig 1. Two-dimensional emotion space in music, adapted from [1]

Methodology (1): Survey of pop/rock emotion perception

- Selection from the **4Q Emotions** dataset (22 songs).
- 11 emotions from the Geneva Emotion Music Scales (GEMS): *Q1 (Joy, Power, Surprise), Q2 (Anger, Tension, Fear), Q3 (Sadness, Bitterness), Q4 (Peace, Tenderness, Transcendence)* [2]
- 4 languages: English Spanish German Mandarin
- Evaluate **inter-rater agreement** (Krippendorff's alpha [3]) with respect to listeners' **familiarity (F)**, **preference (P)**, **lyrics comprehension (LC)**, and **musical sophistication (MSI)**.

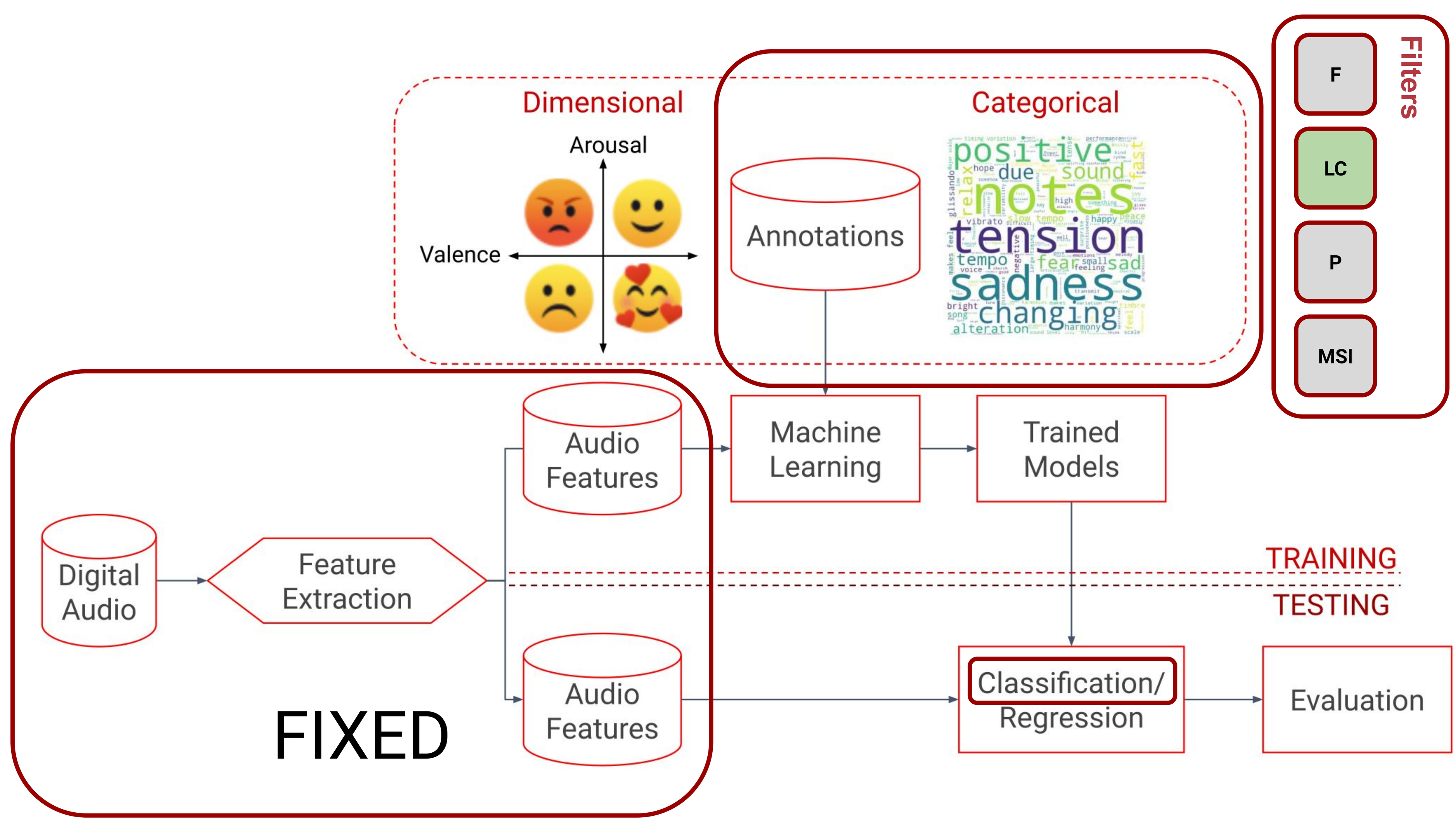

Fig 2. Proposed methodology: keep input fixed while modifying the "ground truth", adapted from [4]. We introduce filters that select annotations from groups of raters to modify the "ground truth" to our models.

Methodology (2): Manifold learning

- **Intuition:** find annotations that are similar amongst them by using the proposed filter.
- Multi-dimensional Scaling **(MDS)**, t-distributed Stochastic Neighbor Embedding **(t-SNE)**, and Uniform Manifold Approximation and Projection for Dimension Reduction **(UMAP)**

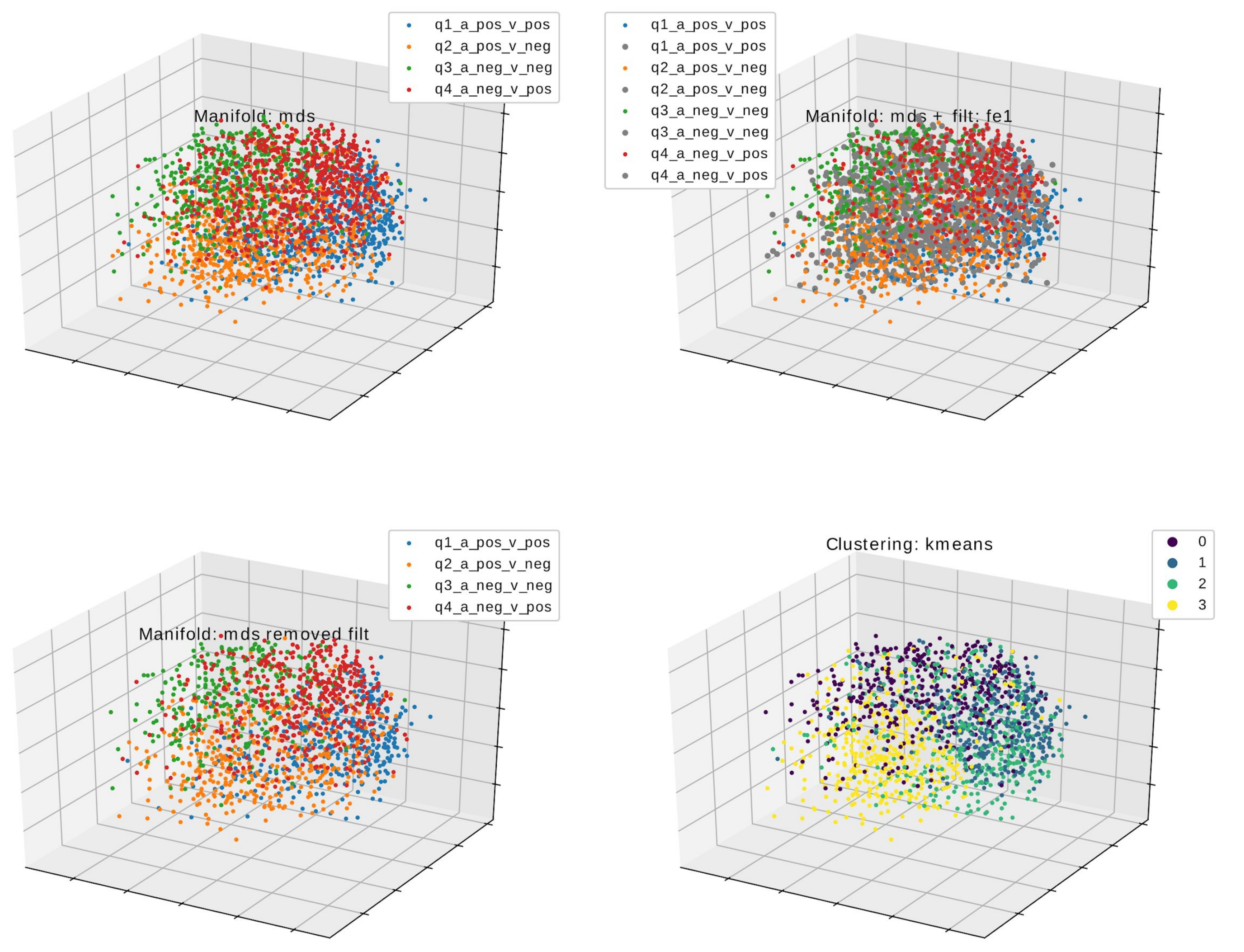

Fig 3. MDS + Filter MSI Emotion + K-Means.

Methodology (3): Classification

- **Intuition:** given different filters to assemble group-based annotations, we can train models with different "ground truths"
- Support Vector Machine: **group-based** and **multi-label**
- For example, excerpt 0 (originally labeled as **anger** is also labeled with **bitterness**, **fear**, **power**, and **tension** when considering our annotations (top-right plot).

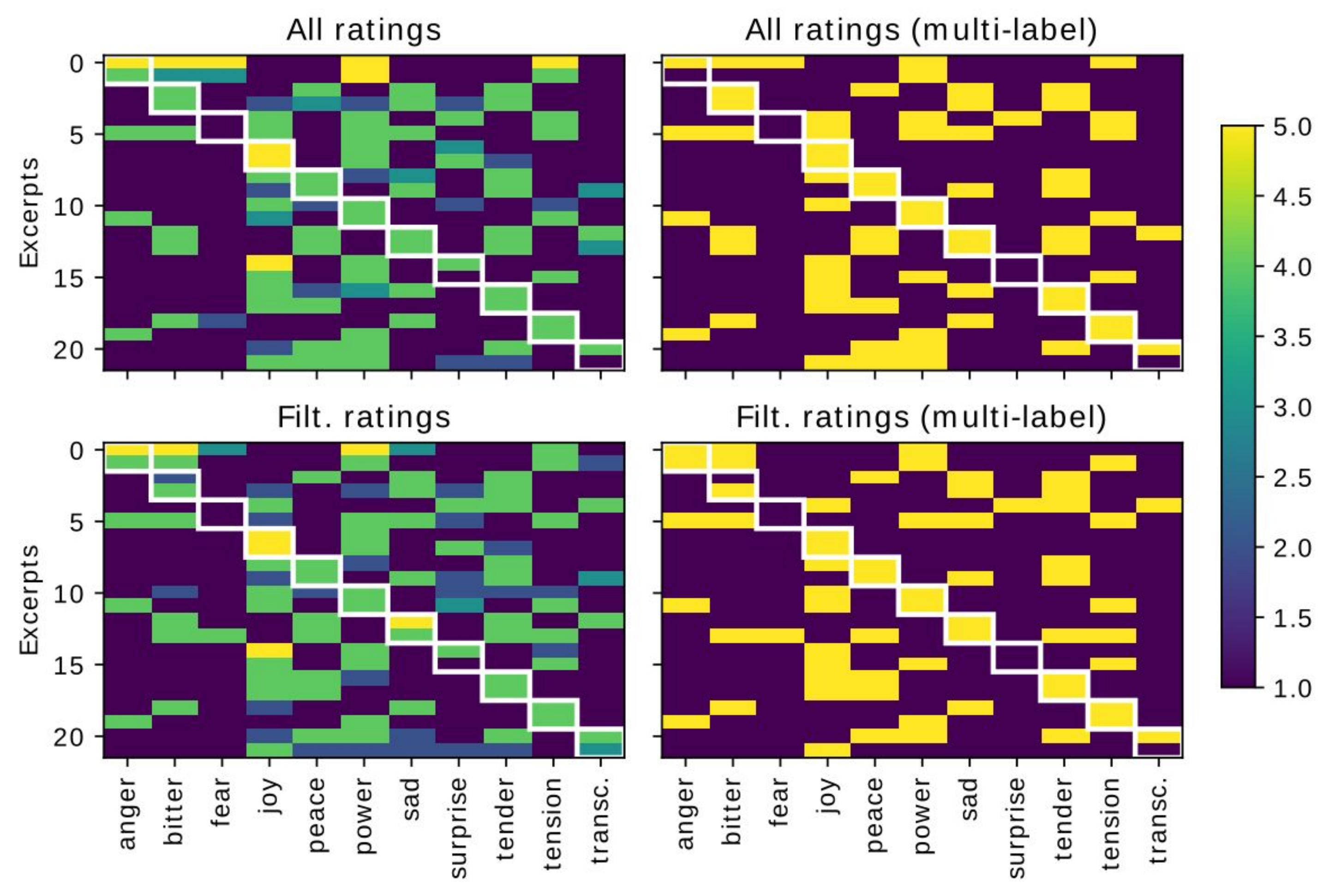

Fig 4. Multi-label annotations ALL vs. Filt. ratings.

Findings

- Overall low agreement
- Ratings from different surveys have different distributions
- Group-based annotations with LC consistently improve classification
- **Future work:** test this methodology on benchmark datasets

References

[1] P. N. Juslin, *Musical Emotions Explained*, 1st ed. Oxford: Oxford University Press, 2019.

[2] M. Schedl, E. Gómez, E. S. Trent, M. Tkalcic, H. Eghbal-Zadeh, A. Martorell, "*On the Interrelation Between Listener Characteristics and the Perception of Emotions in Classical Orchestra Music*", IEEE Transactions of Affective Computing, vol. 9, no. 4, pp. 507-525, 2018.

[3] K. H. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 2nd ed. SAGE Publications, 2004

[4] M. Sordo, "*Semantic Annotation of Music Collections: A Computational Approach*", PhD Dissertation, Barcelona, Spain.

ISMIR MTL2020

upf. Music Technology Group

TROMPA

This work is partially supported by the European Commission under the TROMPA project (H2020 770376).