TOWARDS EXPLAINING EXPRESSIVE QUALITIES IN PIANO RECORDINGS: Transfer of Explanatory Features via Acoustic Domain Adaptation

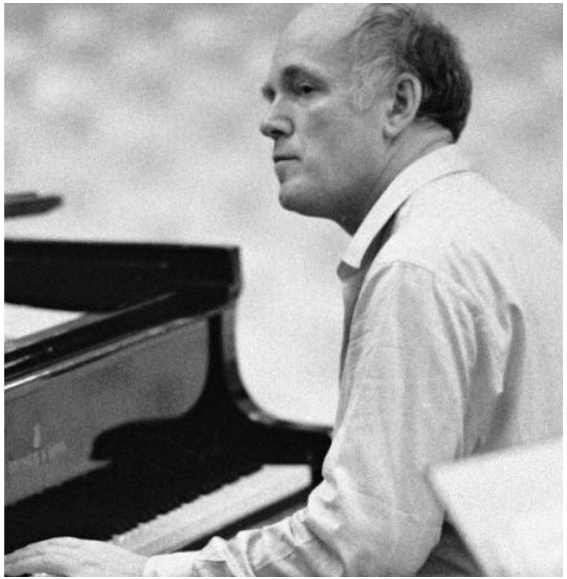
Shreyan Chowdhury, Gerhard Widmer

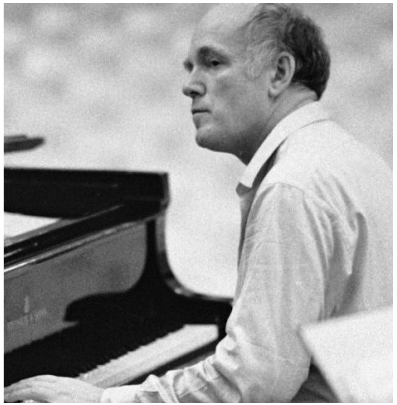
JOHANNES KEPLER UNIVERSITY LINZ

European Research Council
Established by the European Commission

Institute of Computational Perception

Prologue





“The search for audio features that capture the expressive perceptual qualities of performed music”

The Story So Far

Con Espressione Game

On the Characterization of Expressive Performance in Classical Music: First Results of the Con Espressione Game (ISMIR 2020)

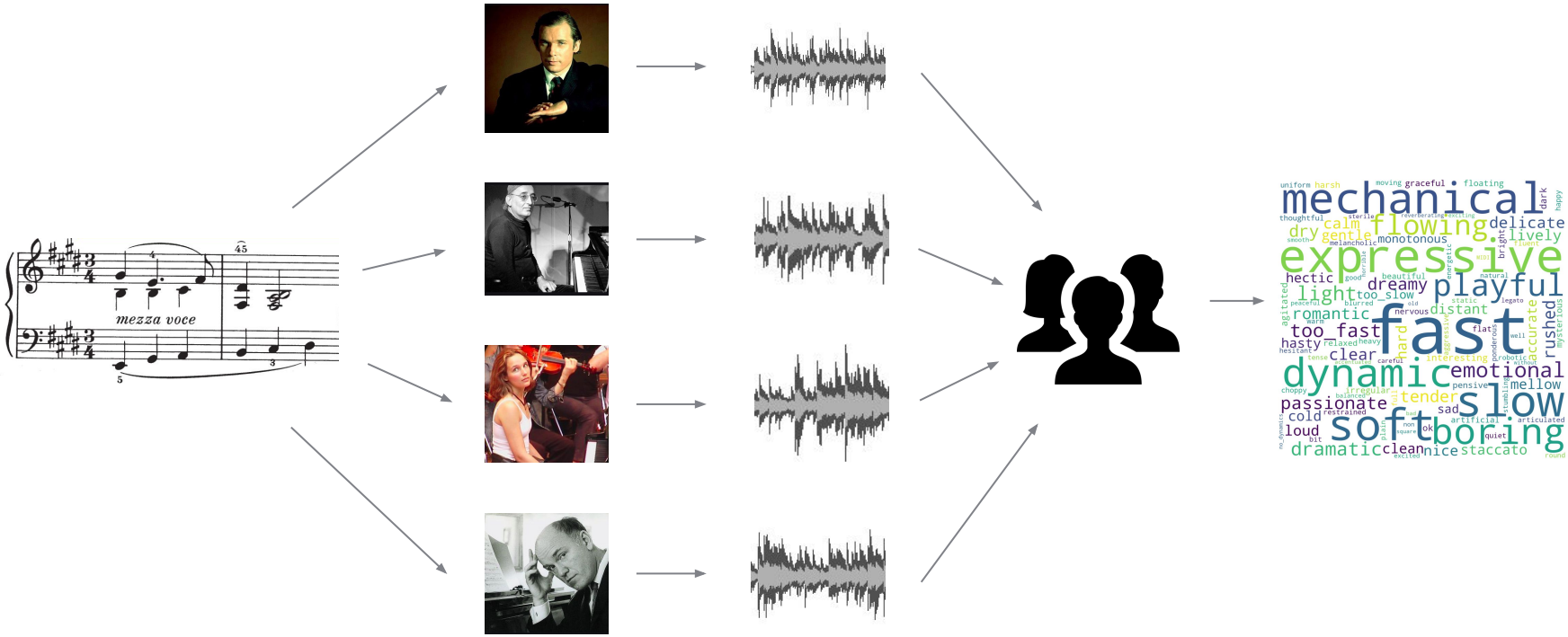
[C. Cancino-Chacón, S. Peter, S. Chowdhury, A. Aljanaki, G. Widmer]

Mid-level features

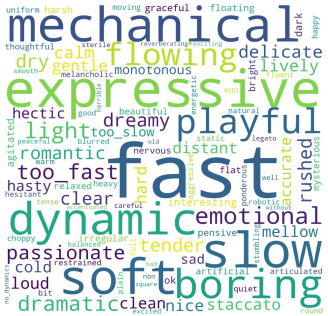
Towards Explainable Music Emotion Recognition: The Route via Mid-level Features (ISMIR 2019)

[S. Chowdhury, A. Vall, V. Haunschmid, G. Widmer]

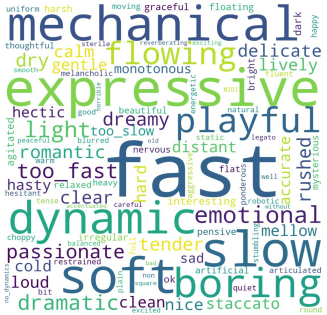
The Con Espressione Game



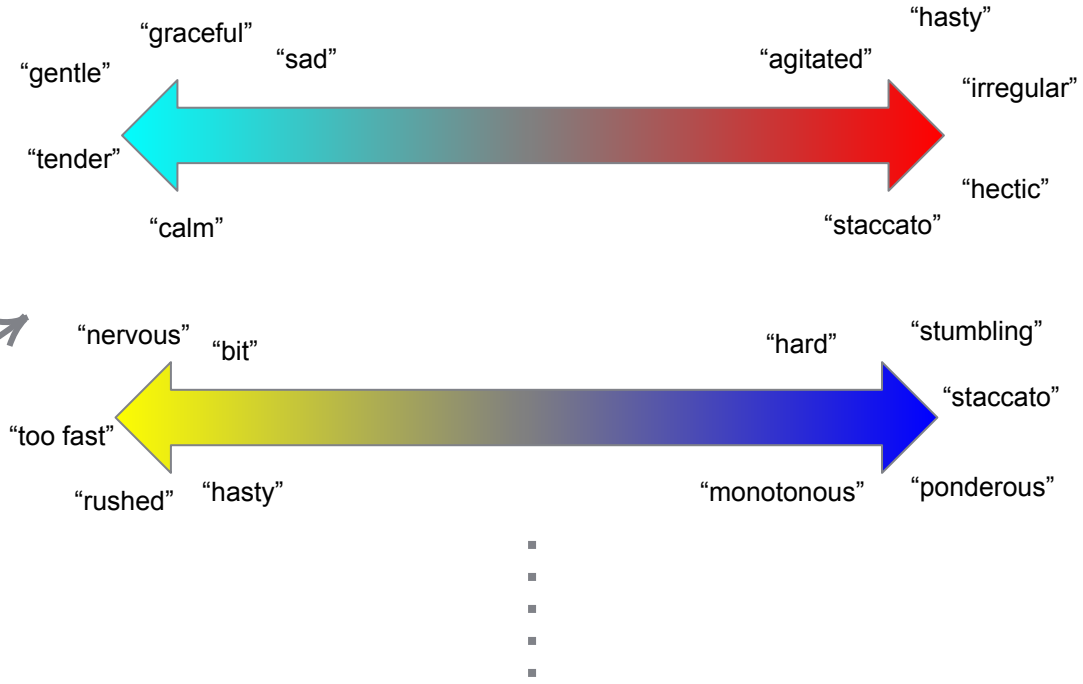
The Con Espressione Dataset



The Con Espressione Dataset



PCA



The Con Espressione Dataset

Can the embedding dimensions obtained
from free-text descriptions of expressive piano performances
be modeled using audio features?

Mid-level Features[1]

Low-level features, such as pitch

Building blocks of musical signals

Melodiousness

Articulation

Rhythm complexity

Rhythm stability

Dissonance

Tonal stability

Minorness

Perceptual and subjective, but make intuitive musical sense

High-level features, such as emotion

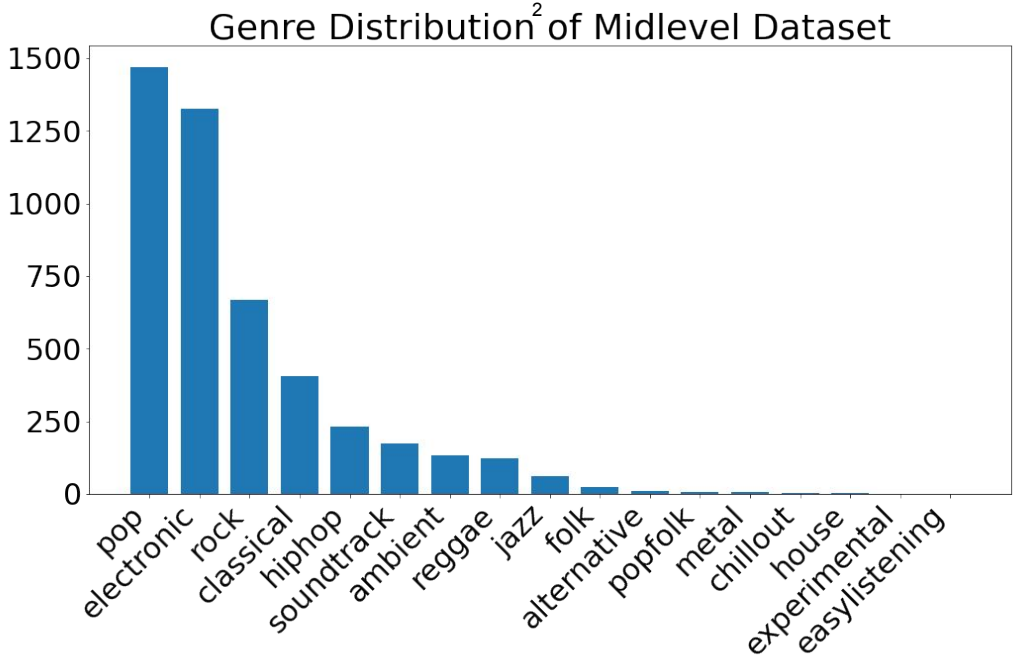
Subjective, abstract descriptions

[1] A. Aljanaki and M. Soleymani, *A Data-driven Approach to Mid-level Perceptual Musical Feature Modeling*, ISMIR 2018

Learning and Transferring Mid-level Features

Mid-level Dataset

- 5000 snippets
- 15-second clips
- Crowdsourced annotation



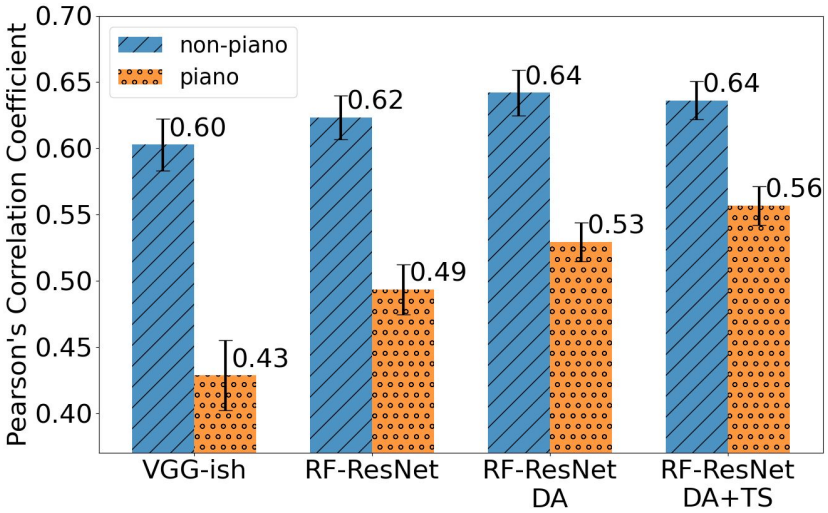
[2] M. Won, A. Ferraro, D. Bogdanov, X. Serra *Evaluation of CNN-based automatic music tagging models*. In Proc. of 17th Sound and Music Computing 2020

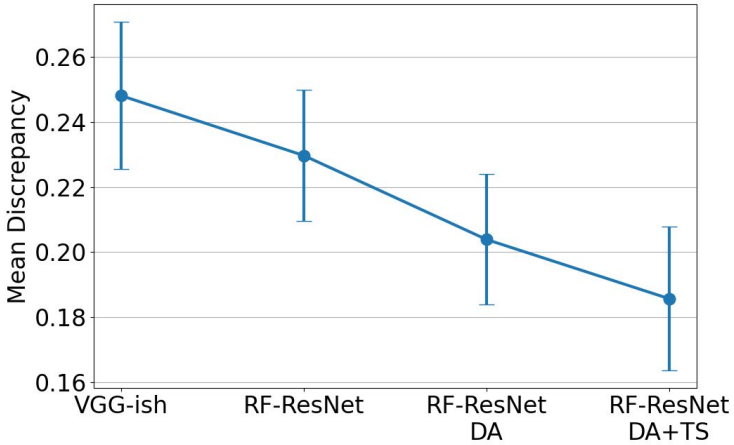
Transferring to our Domain of Interest (Piano)

- Setup
  - Test-set: manually created from piano recordings in the Mid-level Dataset
  - RF-ResNet architecture
- Learn Mid-level Features with Domain Adaptation
  - Unsupervised Domain Adaptation (UDA) by Backpropagation[3]
- Teacher-Student Refinement
  - Multiple DA teachers
  - Student learns from Mid-level combined with pseudo-labeled piano dataset

[3] Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." International conference on machine learning. PMLR, 2015.

Results on the Test-set





Results – Transferring to Con Espressione Recordings

	Dim 1	Dim 2	Dim 3	Dim 4
VGG-ish	0.35	0.10	0.22	0.32
RF-ResNet	0.36	0.07	0.28	0.33
RF-ResNet DA	**0.40**	0.09	**0.29**	0.32
RF-ResNet DA+TS	0.35	**0.15**	**0.29**	**0.34**

Coefficient of determination (R2-score)

RF-ResNet		RF-ResNet DA+TS	
Feature	r	Feature	r
articulation	0.47	melodiousness	− 0.39
rhythmic complexity	0.41	articulation	0.46
		rhythmic complexity	0.41
		dissonance	0.40

Pearson's correlation (r) for mid-level features with description embedding dimension 1.
Features with $p < 0.05$ and $|r| > 0.20$ are selected.

This dimension has positive loadings for words like "hectic", "irregular", and negative loadings for words like "sad", "gentle", "tender".

Looking forward to see you at the poster session **AUD-20**!

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 670035 ("Con Espressione")