



# Deep Learning Approaches for Predicting and Explaining Emotions in Music

---

Shreyan Chowdhury

Supervisor: Prof. Gerhard Widmer

30.09.2021

Dissertation Colloquium

**JKU**



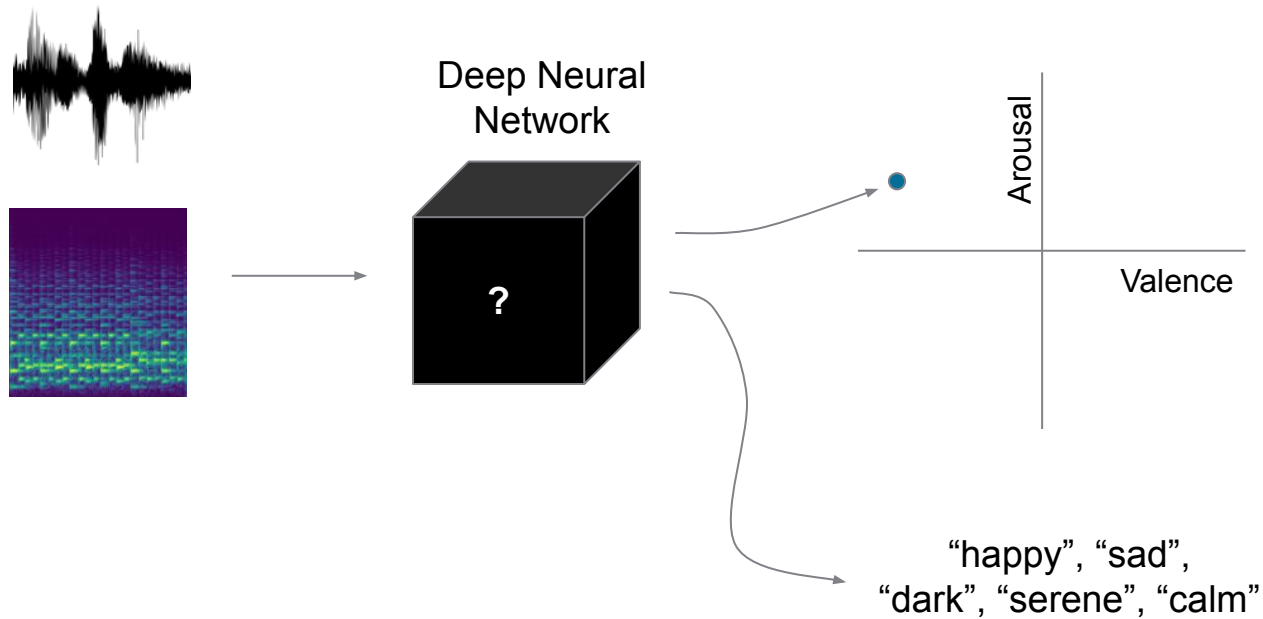
JOHANNES KEPLER  
UNIVERSITY LINZ



Institute of  
Computational  
Perception

# Background

# Music Emotion Recognition and Explainability



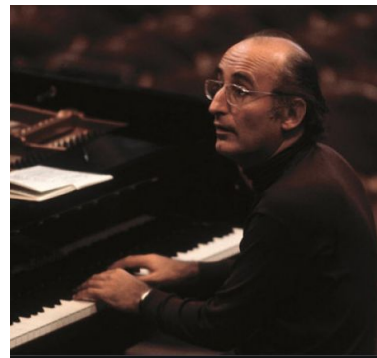
# Con Espressione!

Bach Prelude No.2 in C minor

Glenn Gould



Friedrich Gulda



# Approaches

- ➔ 1. Mid-level Features as Explanatory Variables
- 2. Transferring Mid-level Features to Solo Piano Music
- 3. Modeling Emotion in Bach's *Well-Tempered Clavier*
- 4. Two-level Explanations

# What are Mid-level Features?

Low-level features

Unambiguously defined and  
objectively verifiable

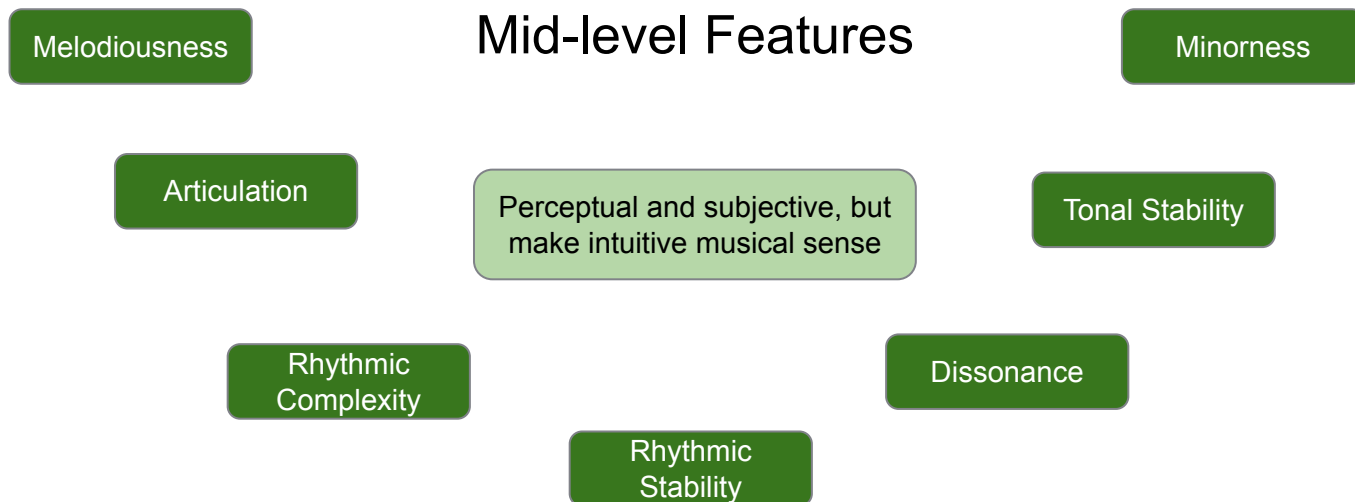
Mid-level Features

Perceptual and subjective, but  
make intuitive, musical sense

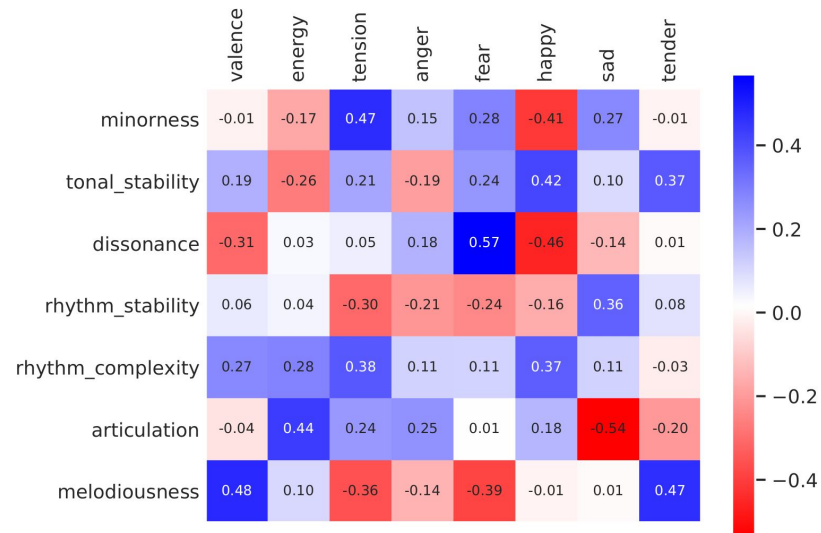
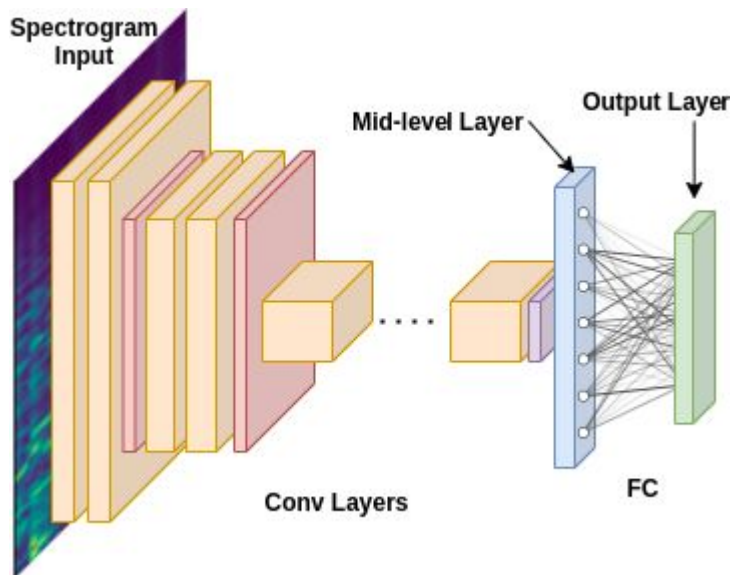
High-level features

Concepts that can  
only be defined by  
considering multiple  
aspects of music

# What are Mid-level Features?



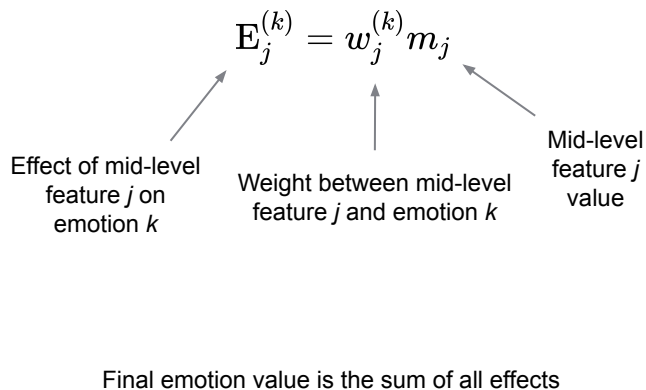
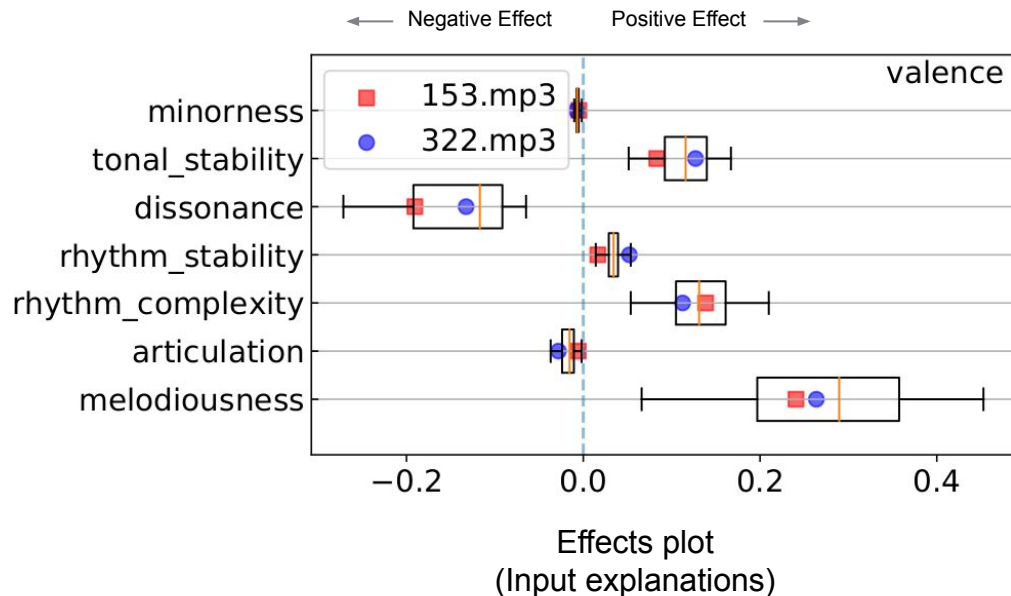
# Learning Mid-level Features as Intermediate Variables



FC Weights  
(Model explanations)



# Interpreting Individual Mid-level Predictions

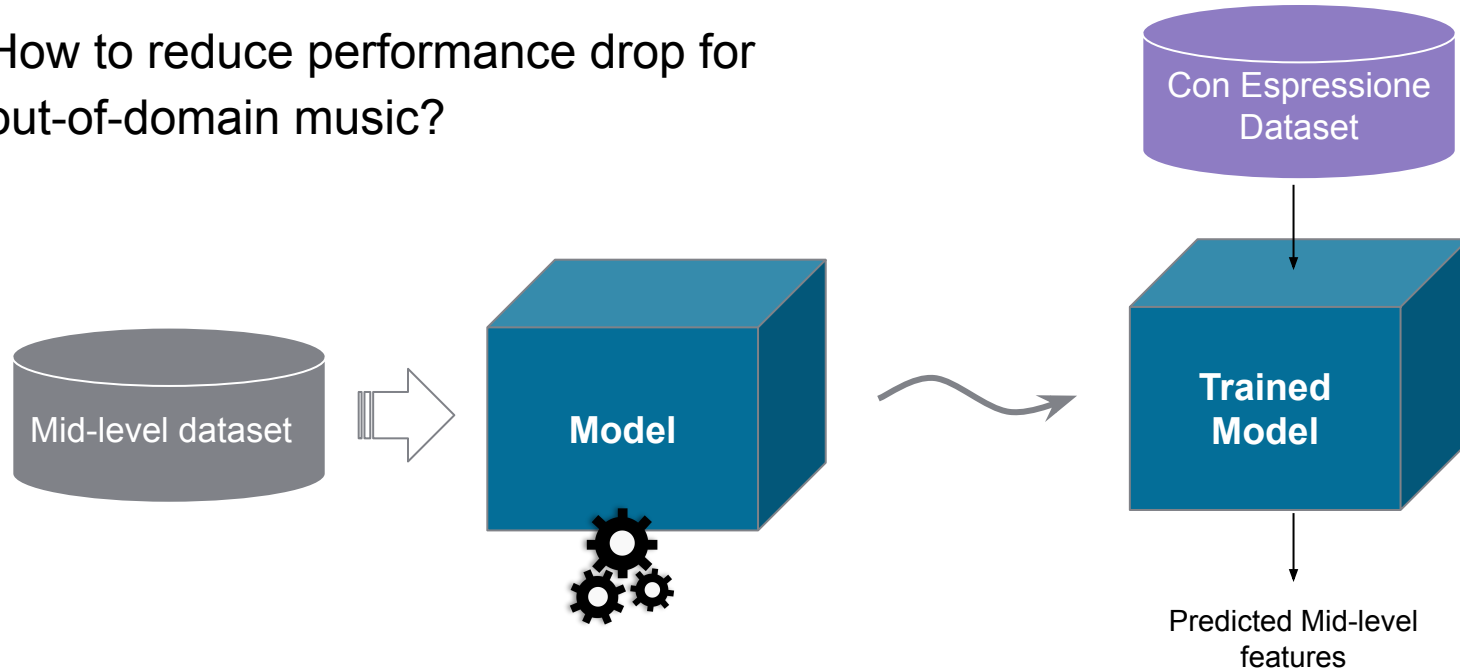


# Approaches

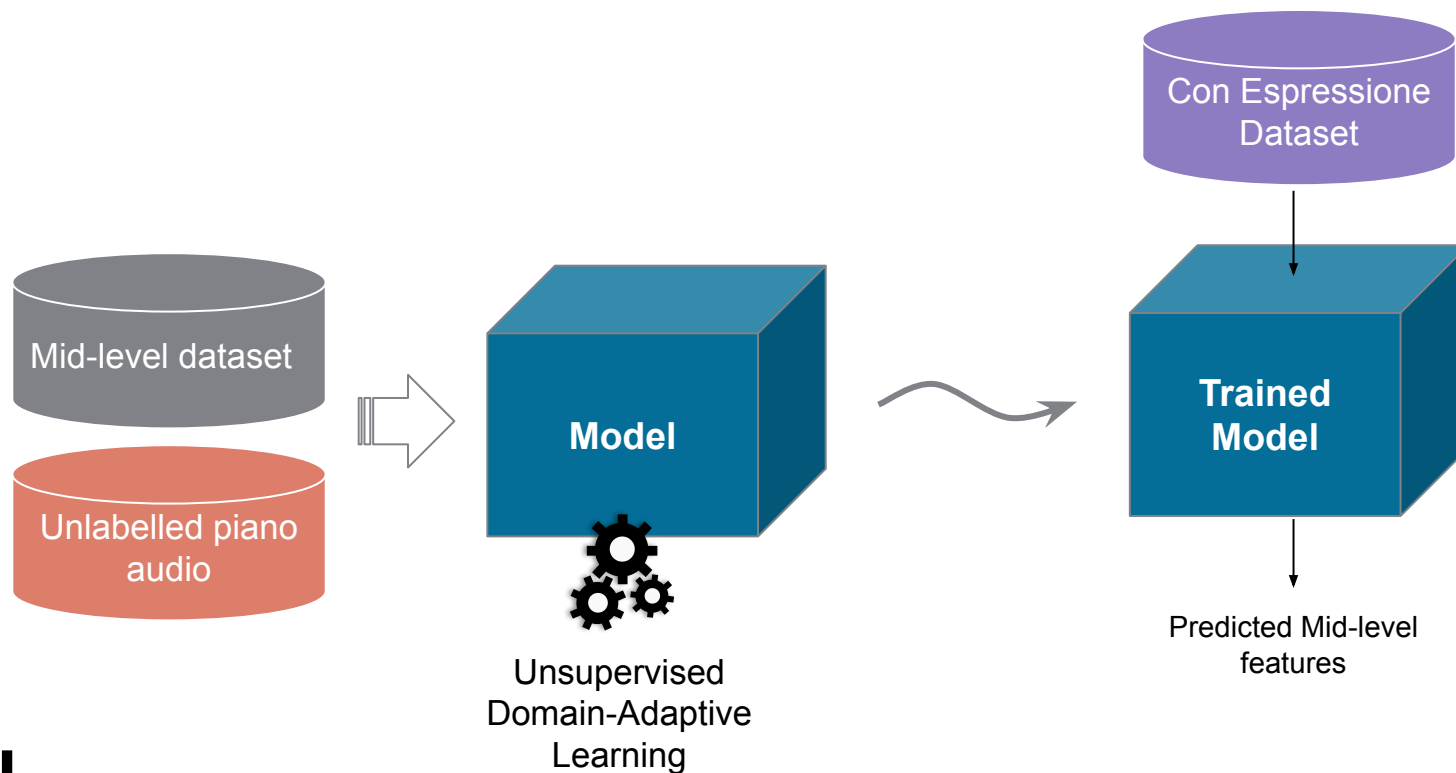
1. Mid-level Features as Explanatory Variables
- ➔ 2. Transferring Mid-level Features to Solo Piano Music
3. Modeling Emotion in Bach's *Well-Tempered Clavier*
4. Two-level Explanations

# Research Question

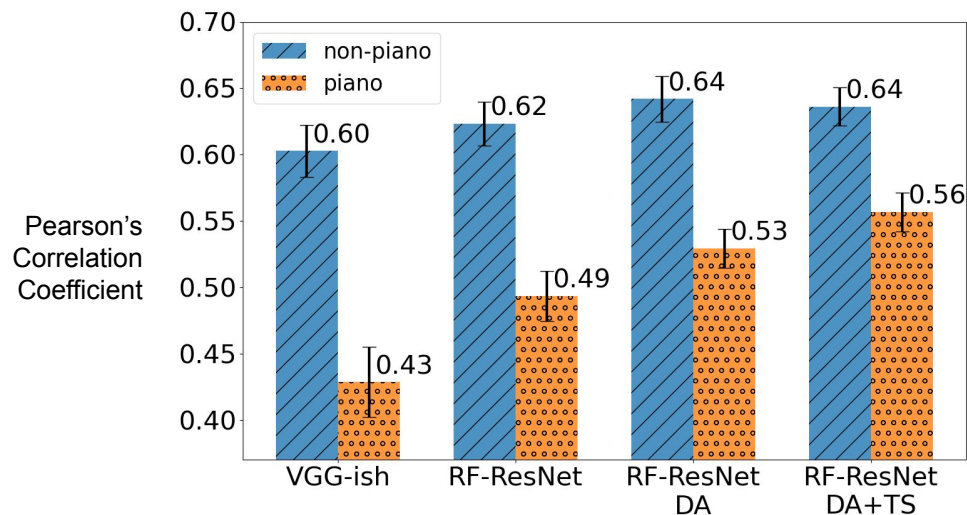
- How to reduce performance drop for out-of-domain music?



# Transfer Scenario: Domain Adaptation



# Domain Adaptation Results



# Approaches

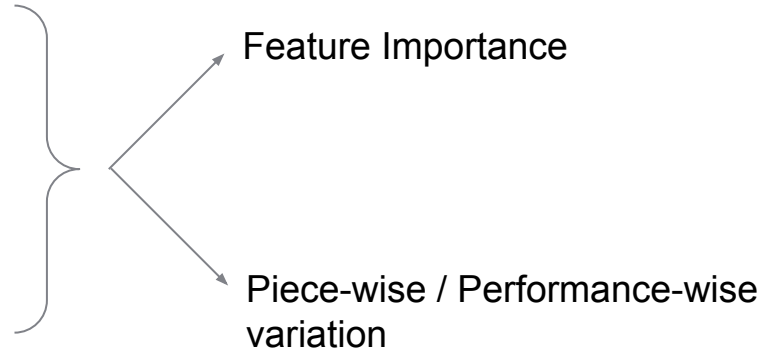
1. Mid-level Features as Explanatory Variables
2. Transferring Mid-level Features to Solo Piano Music
- ➔ 3. Modeling Emotion in Bach's *Well-Tempered Clavier*
4. Two-level Explanations

# Research Questions

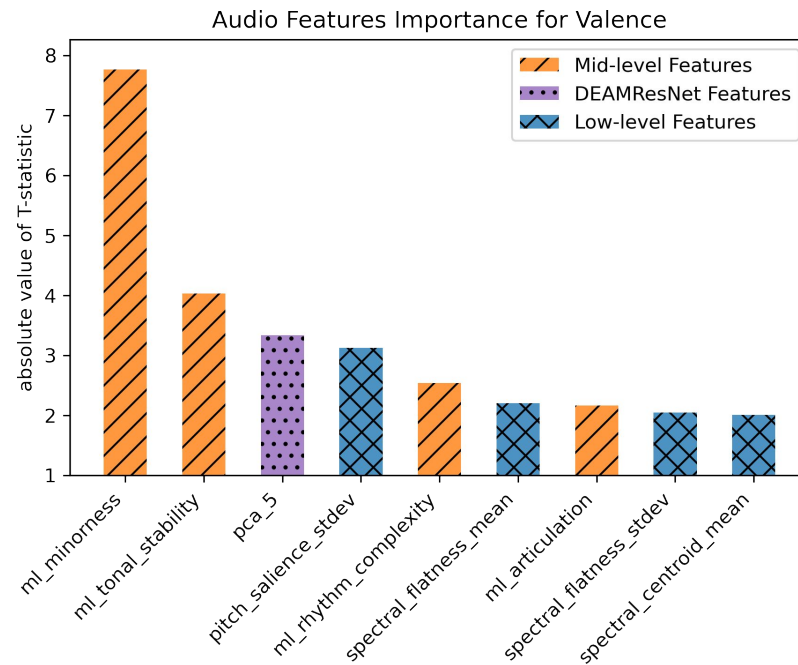
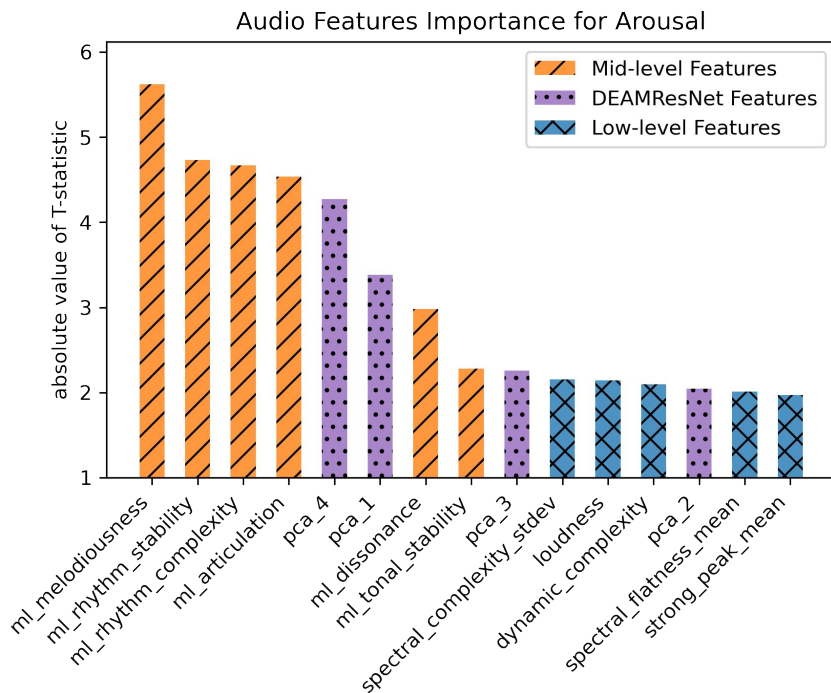
- 48 pieces, 6 performances of Bach's Well Tempered Clavier Book 1

Comparison of feature sets:

- Low-level audio features
- Score-based features
- Mid-level features
- Emotion features



# Feature Importance among Audio-based Features





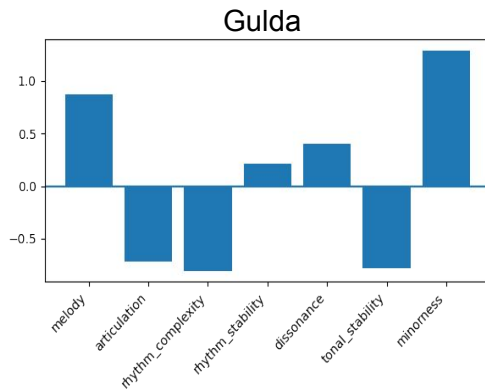
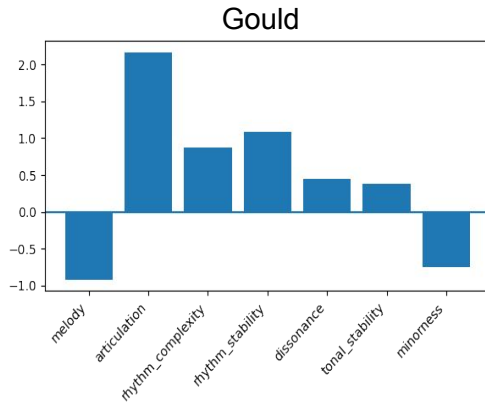
# Modeling Piece-wise and Performance-wise Variation

Feature Set	Arousal	Valence
Mid-level	0.50	0.86
DEAMResNet	<b>0.47</b>	0.89
Low-level	0.66	0.90
Score	0.63	<b>0.68</b>

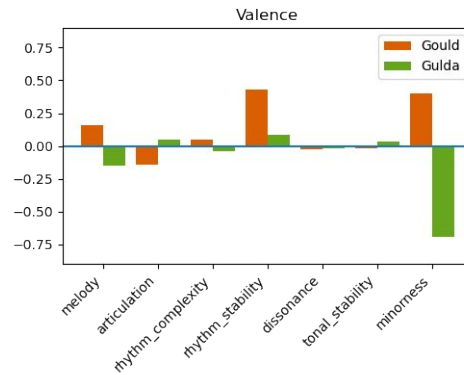
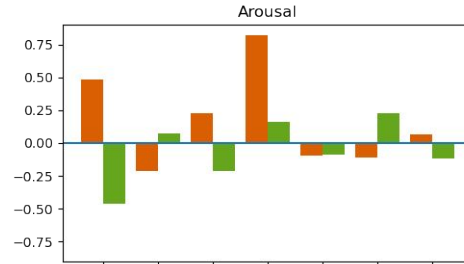
Fraction of residual variance explained by the random effect of “piece id”.

Feature Set	Arousal		Valence	
	FVU	Corr (p<0.1)	FVU	Corr (p<0.1)
Mid-level	<b>0.31</b>	<b>0.58</b> (47.9%)	<b>0.36</b>	0.42 (27.0%)
DEAMResNet	0.32	0.54 (43.8%)	0.61	<b>0.47</b> (37.5%)
Low-level	0.43	0.56 (54.2%)	0.75	0.38 (22.9%)

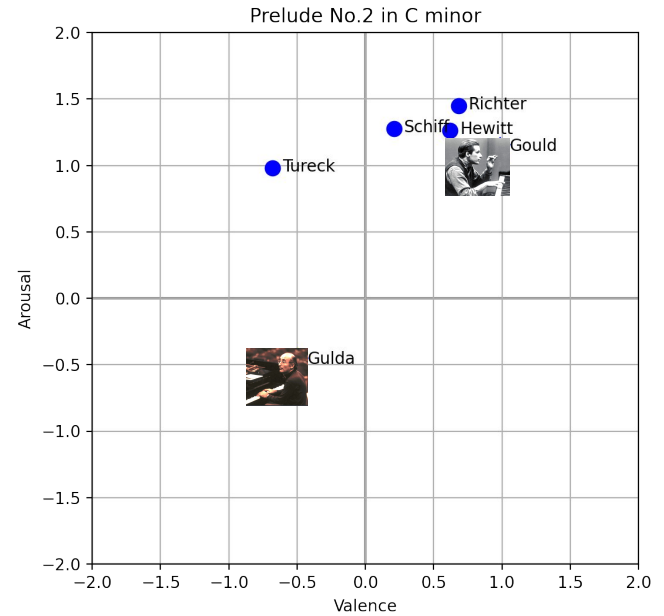
Fraction of variance between different performances of a piece unexplained by the feature set



Mid-level Predictions



Effects



Emotion Predictions

# Approaches

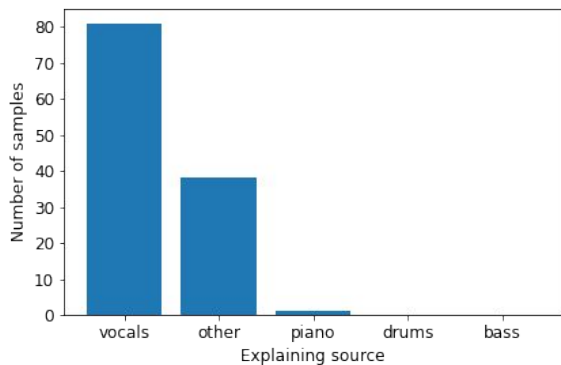
1. Mid-level Features as Explanatory Variables
2. Transferring Mid-level Features to Solo Piano Music
3. Modeling Emotion in Bach's *Well-Tempered Clavier*
- ➔ 4. Two-level Explanations

# Research Questions

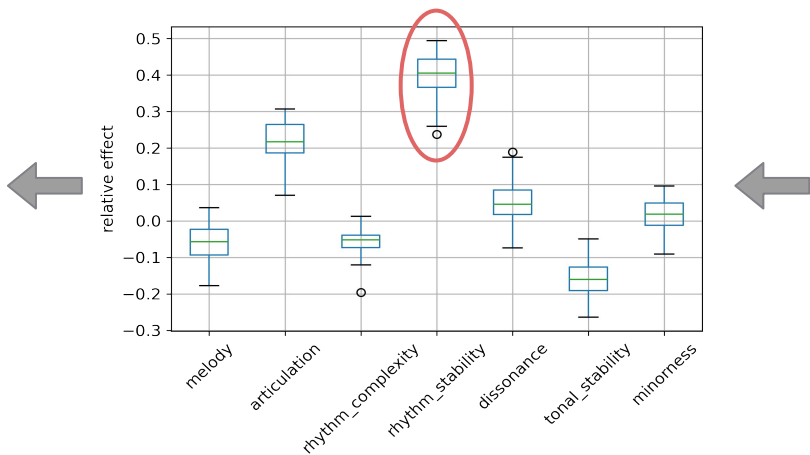
- How to explain the black-box between the input and Mid-level features?
- Can we use explanations for model debugging?



# Two-level Explanations for Valence in Hiphop Songs



Which sound source explains high rhythmic stability?



Which Mid-level feature explains high valence predictions?

Problem:  
Valence  
Overestimation

# Progress

# Publications

1. Shreyan Chowdhury, Gerhard Widmer  
“On Perceived Emotion in Expressive Piano Performance: Further Experimental Evidence for the Relevance of Mid-level Features”  
(*ISMIR 2021*)
2. Shreyan Chowdhury, Verena Praher, Gerhard Widmer  
“Tracing Back Music Emotion Predictions to Sound Sources and Intuitive Perceptual Qualities”  
(*SMC 2021*)
3. Shreyan Chowdhury, Gerhard Widmer  
“Towards Explaining Expressive Qualities in Piano Recordings: Transfer of Explanatory Features via Acoustic Domain Adaptation” (*ICASSP 2021*)
4. Shreyan Chowdhury, Andreu Vall, Verena Haunschmid, Gerhard Widmer  
“Towards Explainable Music Emotion Recognition: The Route via Mid-level Features” (*ISMIR 2019*)

# Publications

5. Carlos Cancino-Chacon, Silvan Peter, Shreyan Chowdhury, Anna Aljanaki, Gerhard Widmer  
“On the Characterization of Expressive Performance in Classical Music: First Results of the Con Espressione Game” (*ISMIR 2020*)
6. Khaled Koutini, Shreyan Chowdhury, Verena Haunschmid, Hamid Eghbal-zadeh, Gerhard Widmer  
“Emotion and Theme Recognition in Music with Frequency-Aware RF-Regularized CNNs”  
(*MediaEval Multimedia Benchmark 2019*)
7. Verena Haunschmid, Shreyan Chowdhury, Gerhard Widmer  
“Two-level Explanations in Music Emotion Recognition”  
(*ML4MD Workshop, ICML 2019*)



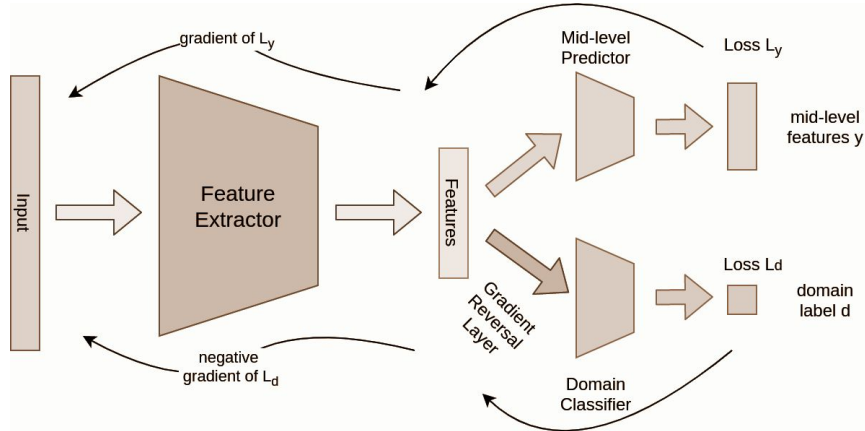
# Other Output

1. Invited talk at MAPLE Lab McMaster University, Hamilton, Canada – “Towards Better Features for Music Emotion Recognition: A Machine Learning Approach” (August 2021).
2. First rank in MediaEval Emotion and Mood recognition challenge (October 2019).
3. Invited talk at Acoustics Research Institute, OAW, Vienna, Austria – “Explainable Models and their Application in Music Emotion Recognition” (October 2018).
4. Reviewer for ISMIR 2020 and ISMIR 2021

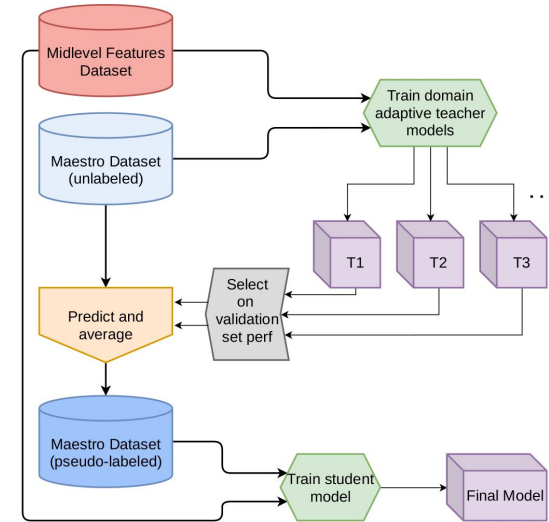
**Thank you!**

# Extra slides ...

# Domain Adaptation Procedure

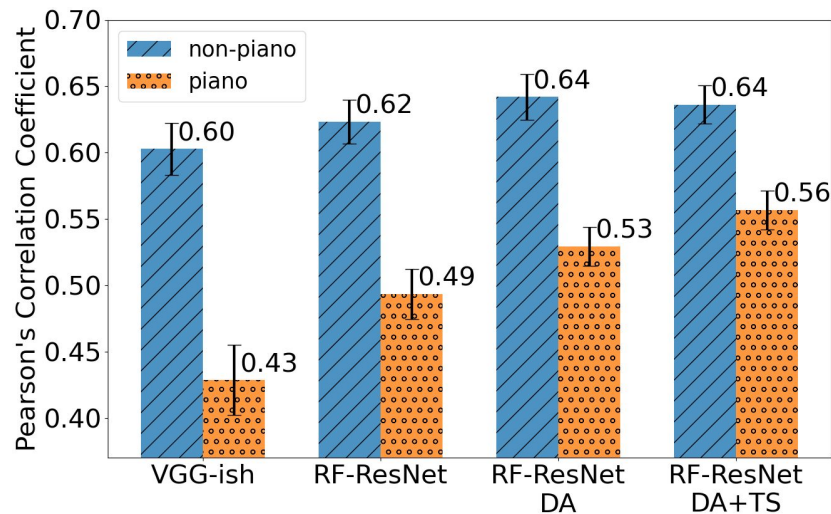


Unsupervised Domain Adaptation



Teacher-Student Refining

# Domain Adaptation Results



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