Explainable Models and their Application in Music Emotion Recognition

Talk @ ARI (ÖAW) ~ 16.10.2019

Verena Haunschmid, Shreyan Chowdhury.

JOHANNES KEPLER
UNIVERSITY LINZ

Institute of
Computational
Perception

Agenda

- The Institute of Computational Perception
- Interpretable (Audio) Models: State of the Art
  - Invertible Neural Networks
  - Gradient-Based Methods
  - Local Interpretable Model-Agnostic Explanations
- Perceptually Explainable Emotion Recognition in Music
  - Emotion rating models
  - Perceptual features
  - Neural network architecture design
  - Explaining emotions - linear effects
  - Explaining intermediate representations

The Institute of Computational Perception

- Focus of our research and teaching:
  - Artificial Intelligence, Machine Learning and Reinforcement Learning for Intelligent Audio and Music Processing
- ~20 people (~10 PhD students)
- Thematic Focus
  - Acoustic Scene Understanding
  - Intelligent Music Processing
  - Multimedia Data Mining
  - Recommender Systems and User Modeling
  - Image Processing
  - Biometric Identification
  - Cryptography

https://www.jku.at/en/institute-of-computational-perception/

The Institute of Computational Perception

- Video: CP.JKU Symphomaniacs and Friends present “Torturing Mozart”

- 20th anniversary of ISMIR
  - Call for Music: https://ismir2019.ewi.tudelft.nl/?q=node/15

Projects

New Frontiers in Music Information Processing

CON ESPRESSIONE: Towards Expressivity-aware Computer Systems in Music

Computer-assisted Analysis of Herbert von Karajan's Musical Conducting Style

Fine-grained Culture-aware Music Recommender Systems

Deep Learning for Symbiotic Mechatronics

Deep Learning and Sensor Fusion

https://www.jku.at/en/institute-of-computational-perception/research/projects/

New Frontiers in Music Information Processing

CON ESPRESSIONE: Towards Expressivity-aware Computer Systems in Music

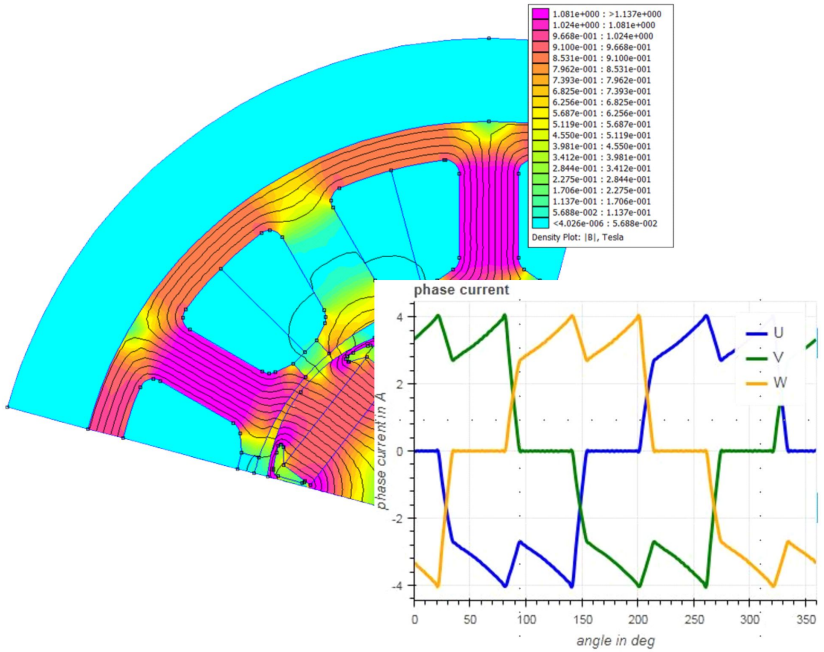
European Research Council

ACCompanion: Ravel "Menuet (Op. posth.)" (4 hands version)

https://www.youtube.com/watch?v=t0hyC5xYD40

Computer-assisted Analysis of Herbert von Karajan's Musical Conducting Style

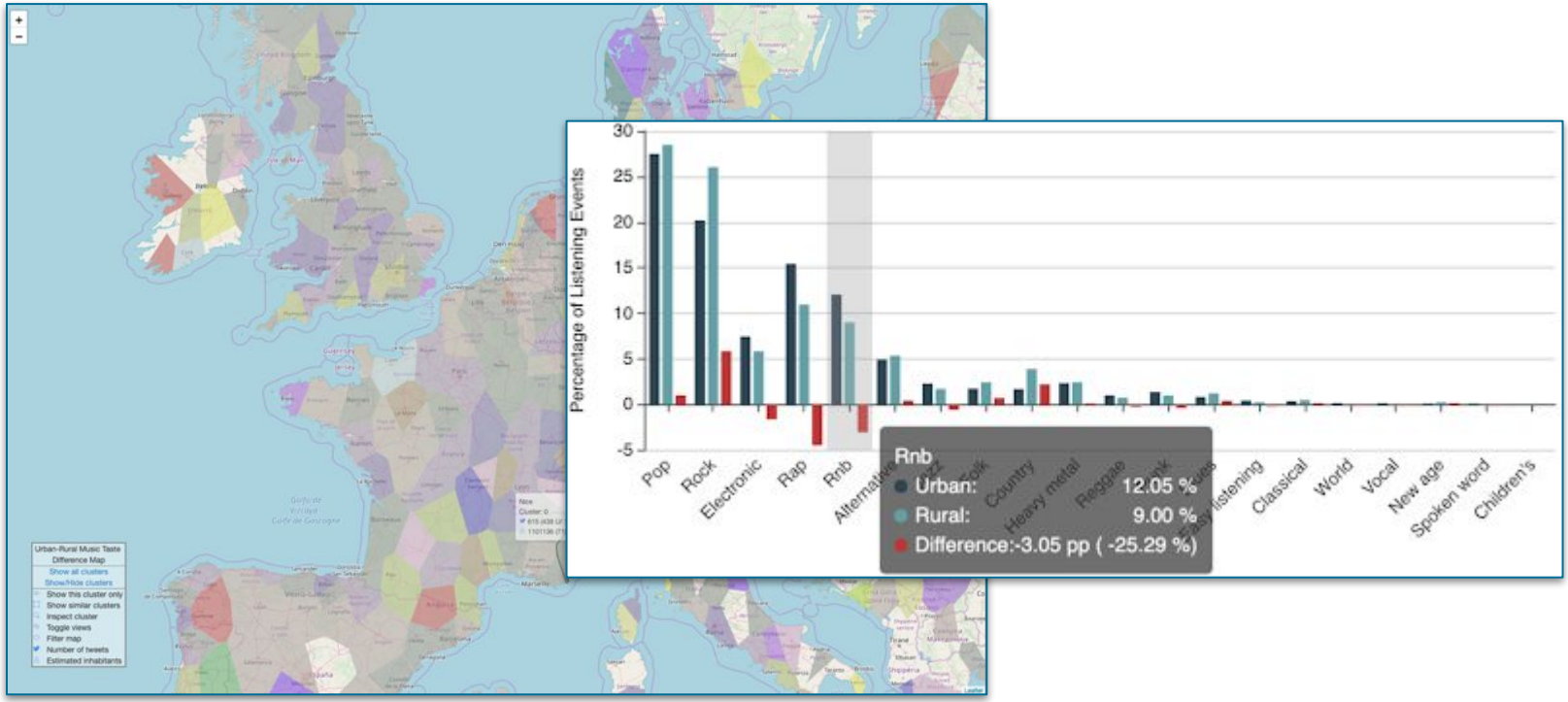
Deep Learning for Symbiotic Mechatronics



Deep Learning and Sensor Fusion

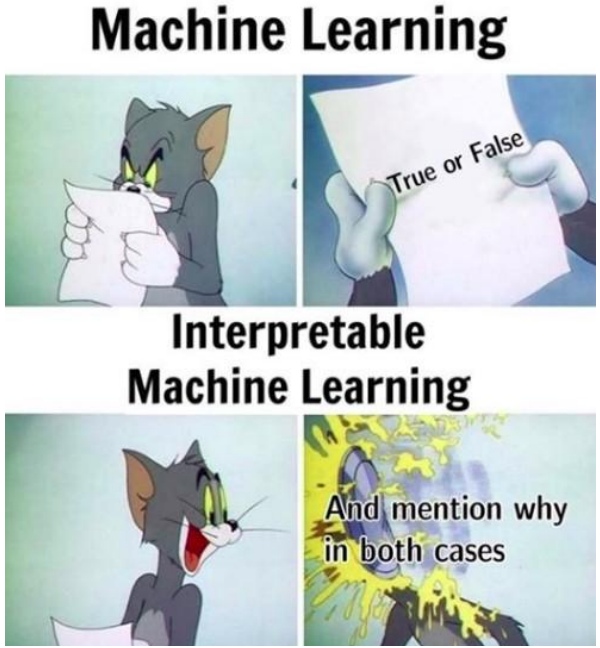
Copyright: voestalpine AG, Quelle: voestalpine.com

Fine-grained Culture-aware Music Recommender Systems



Interpretable (Audio) Models: State of the Art

What is Interpretability Machine Learning?



Motivation for Interpretable Machine Learning

- Penalizing model when it has incorrectly assigned importance to some features (Rieger 2019)
- Debugging models

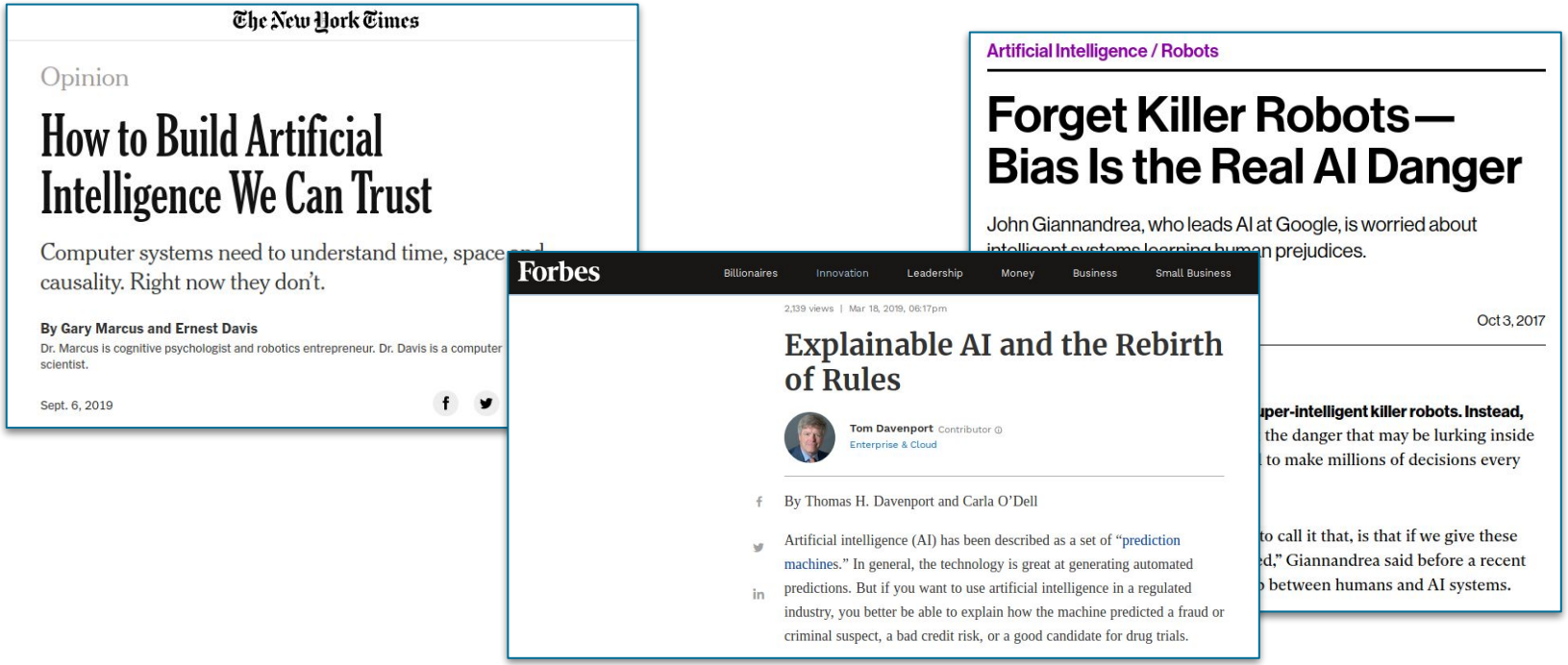
- Machine Teaching [Goyal 2019]
- Right to an explanation (Credit score in the US, EU GDPR)
  - "Credit bureau X reports that you declared bankruptcy last year; this is the main factor in considering you too likely to default, and thus we will not give you the loan you applied for."
- ...

Motivation for Interpretable Machine Learning

(a) Husky classified as wolf (b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

Motivation for Interpretable Machine Learning



The New York Times

Opinion

How to Build Artificial Intelligence We Can Trust

Computer systems need to understand time, space and causality. Right now they don't.

By Gary Marcus and Ernest Davis

Dr. Marcus is cognitive psychologist and robotics entrepreneur. Dr. Davis is a computer scientist.

Sept. 6, 2019

Artificial Intelligence / Robots

Forget Killer Robots—Bias Is the Real AI Danger

John Giannandrea, who leads AI at Google, is worried about intelligent systems learning human prejudices.

Oct 3, 2017

...uper-intelligent killer robots. Instead, ... the danger that may be lurking inside ... to make millions of decisions every ...

...to call it that, is that if we give these ...d," Giannandrea said before a recent ... between humans and AI systems.

Forbes

Billionaires Innovation Leadership Money Business Small Business

2,139 views | Mar 18, 2019, 06:17pm

Explainable AI and the Rebirth of Rules

Tom Davenport Contributor

Enterprise & Cloud

By Thomas H. Davenport and Carla O'Dell

Artificial intelligence (AI) has been described as a set of "prediction machines." In general, the technology is great at generating automated predictions. But if you want to use artificial intelligence in a regulated industry, you better be able to explain how the machine predicted a fraud or criminal suspect, a bad credit risk, or a good candidate for drug trials.

Making Models Interpretable/Explainable

- Interpretable models:
  - Simple models (linear models, decision trees, …)
  - Invertible models ⬅
- Explaining models:
  - Example Based Explanations
  - Gradient Based Methods ⬅
  - Model-Agnostic Explanations ⬅

Towards Interpretable Polyphonic Transcription with Invertible Neural Networks

Polyphonic Transcription: Predicting notes from a polyphonic music piece

Invertible Neural Networks

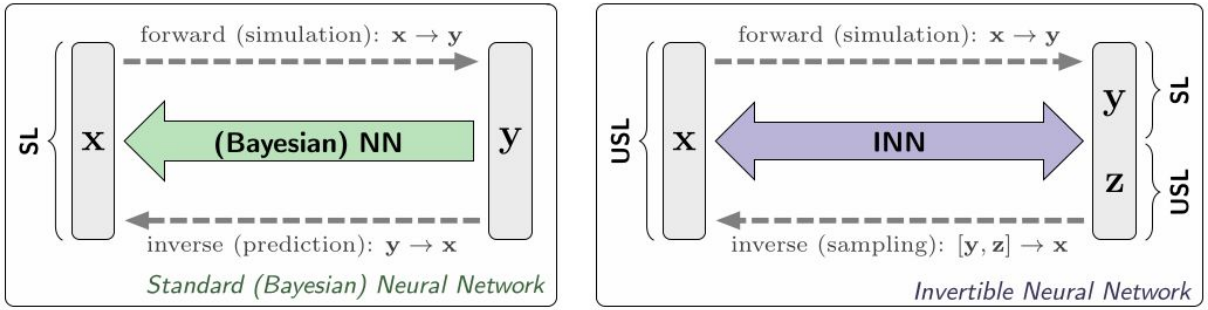
- Introduced in (Deco 1995)
- Rediscovered in (Baird 2005)
- This work is based on (Ardizzone 2019)

Properties

- The mapping from inputs to outputs is bijective, i.e. its inverse exists,
- both forward and inverse mapping are efficiently computable

Rainer Kelz and Gerhard Widmer, “Towards Interpretable Polyphonic Transcription with Invertible Neural Networks” (ISMIR, 2019), preprint: https://arxiv.org/pdf/1909.01622.pdf

Towards Interpretable Polyphonic Transcription with Invertible Neural Networks

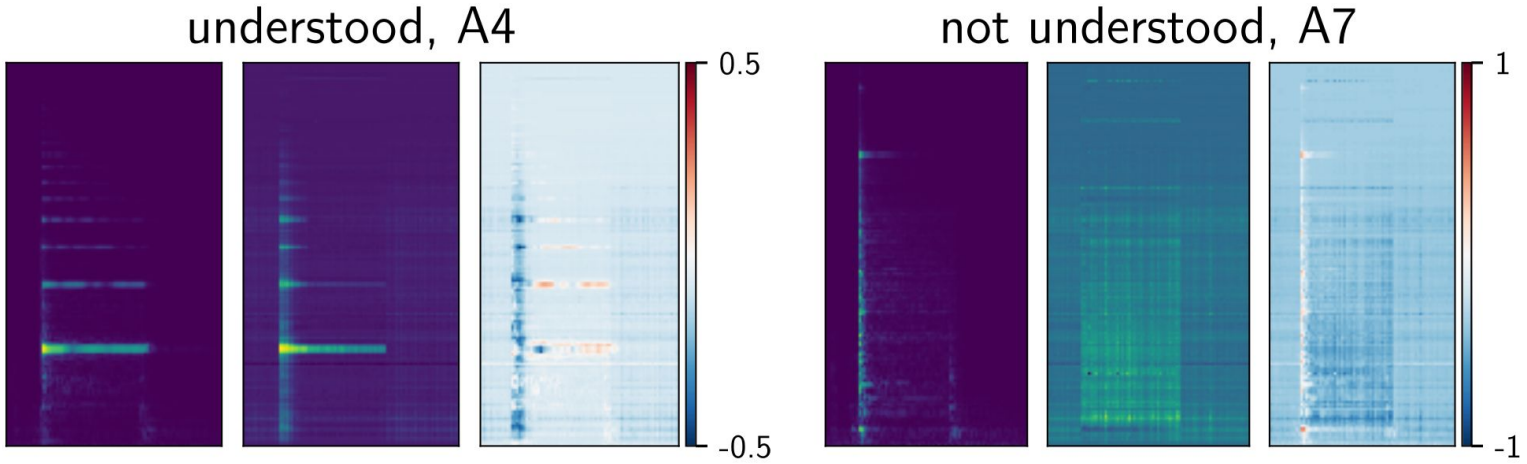


Perfect scenario: given an input x

- All semantic information (note phases, velocities, instrument) ➜ ŷ vector
- All nuisance information (other acoustic variability, such as microphone characteristics, room reverberation or actual noise) ➜ ẑ
- ẑ is distributed as N (0, I)

Rainer Kelz and Gerhard Widmer, “Towards Interpretable Polyphonic Transcription with Invertible Neural Networks” (ISMIR, 2019), preprint: https://arxiv.org/pdf/1909.01622.pdf

Towards Interpretable Polyphonic Transcription with Invertible Neural Networks



Rainer Kelz and Gerhard Widmer, “Towards Interpretable Polyphonic Transcription with Invertible Neural Networks” (ISMIR, 2019), preprint: https://arxiv.org/pdf/1909.01622.pdf

Gradient-Based Explanations for Multi Instrument Classifiers

Task: Multi Instrument Classification (https://github.com/cosmir/openmic-2018)

Gradient Based Attribution Maps:

- Assign each input feature and importance value
- Shows positive & negative contribution in a heatmap

Several methods

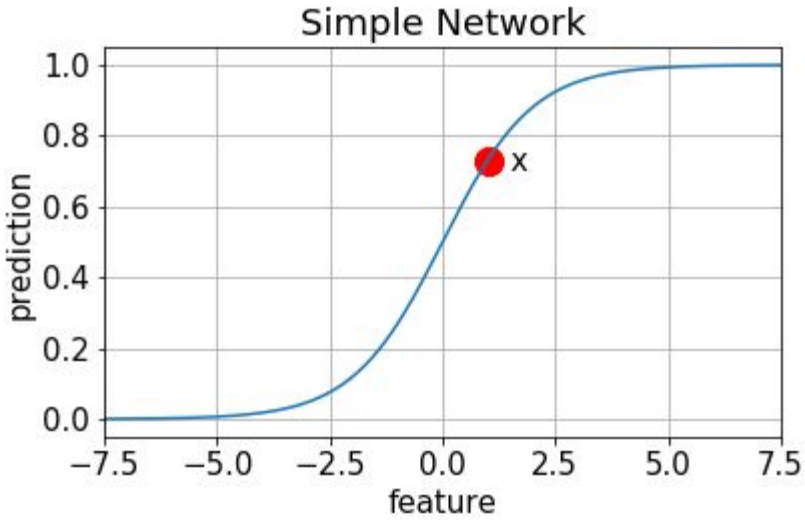
- Occlusion Maps (Zeiler 2014)
- Vanilla Gradients (Simonyan 2013)
- SmoothGrad (Smilkov 2017)
- Integrated Gradients (Sundararajan 2017)
- GradCAM (Selvaraju 2016)

Paul Primus, “Gradient-Based Explanations for Audio Classifiers” (Master thesis, 2019)

Vanilla Gradients (Simonyan 2013)

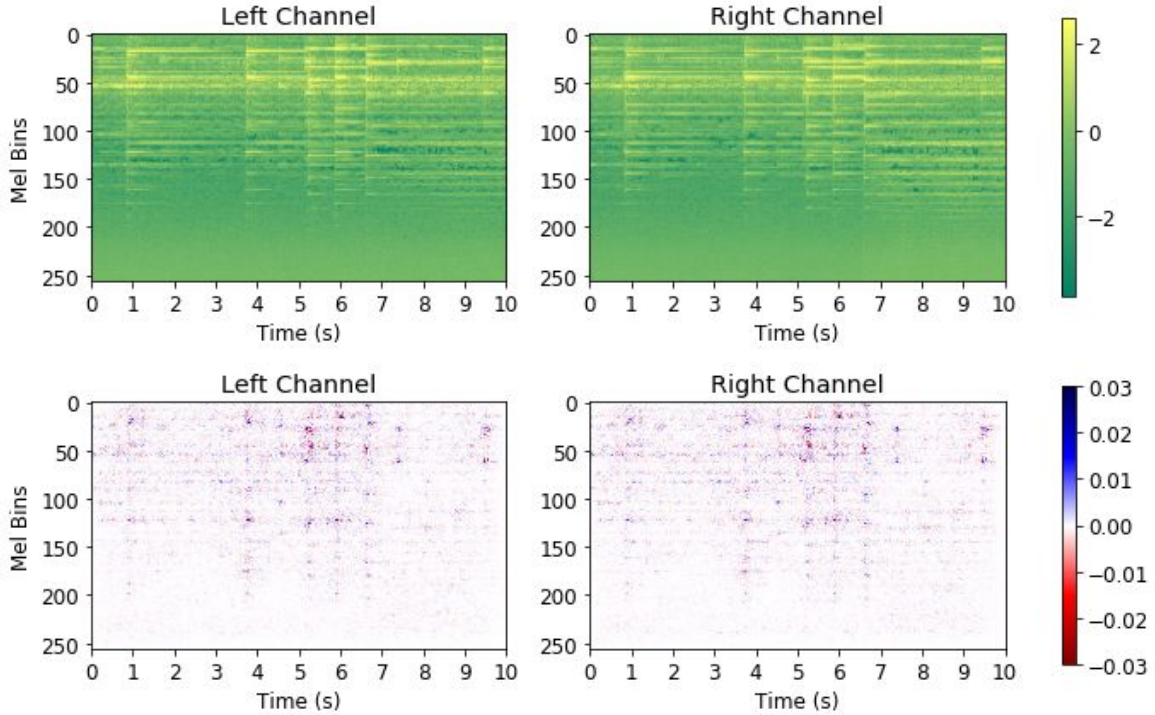
Gradient of Output wrt. Input:

$$M_C(x) = \frac{\partial f_c(x)}{\partial x}$$



Paul Primus, “Gradient-Based Explanations for Audio Classifiers” (Master thesis, 2019)

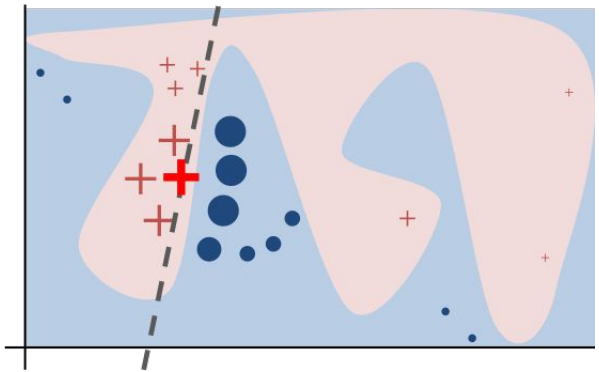
Vanilla Gradients - Explanation for Piano



Paul Primus, “Gradient-Based Explanations for Audio Classifiers” (Master thesis, 2019)

Local Interpretable Model-Agnostic Explanations (LIME)

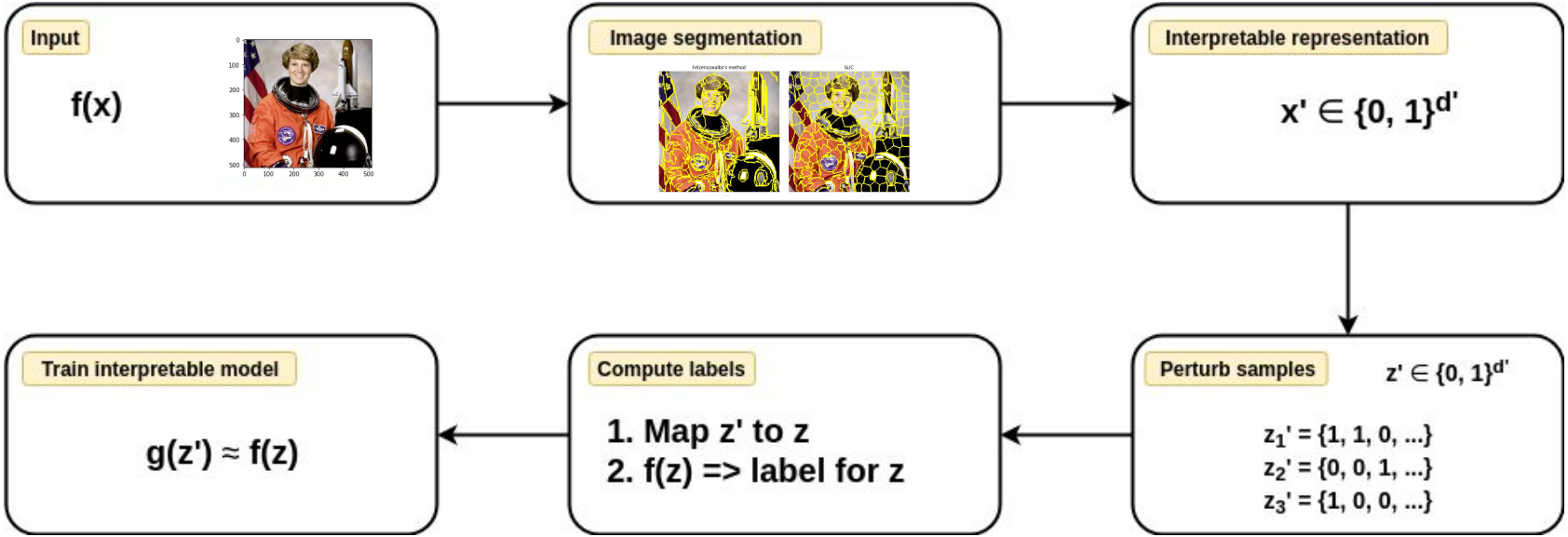
- Used to explain individual predictions of black box machine learning models
- LIME trains local surrogate models to explain individual predictions



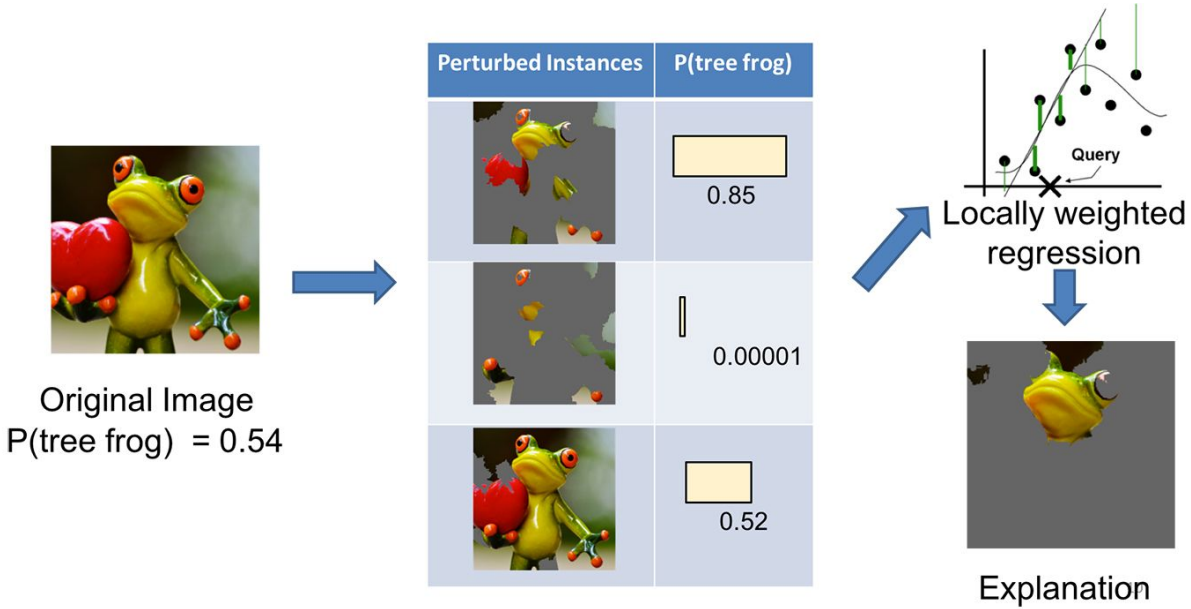
Source: https://github.com/marcotcr/lime

Tulio Ribeiro, Marco; Singh, Sameer; Guestrin, Carlos: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 2016, 13-17.

Local Interpretable Model-Agnostic Explanations (LIME)



Local Interpretable Model-Agnostic Explanations (LIME)



Source: https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime

Desired Explanations in Music Information Processing

- Visual: highlight parts in the spectrogram
  - Using LIME on spectrograms? (Related Work: SoundLIME (Mishra 2017))
- Listenable
  - Reconstruct enhanced spectrograms
  - Silence irrelevant time frames
- Textual / semantic
  - “This song is classified as classical, because it contains piano”
  - “This song has a … score because it has high value for … and a low value for …”

Perceptually Explainable Emotion Recognition in Music

Johnny Cash - Hurt

Nine Inch Nails - Hurt

rafnaj1 3 years ago

Both versions are good for different reasons. I like this one because it sounds a lot more angry and hopeless. Johnny's sound like that of a tired man who has lived a long, hard life. Of a man who is trying to make up for the things he has done, even if he knows he'll continue to hurt those he loves.

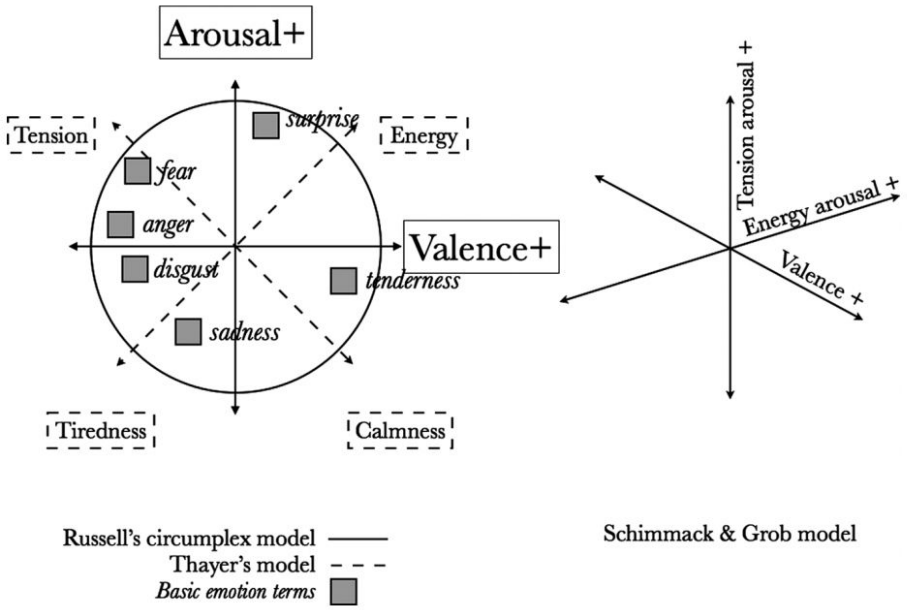
6K REPLY

What MIR can and cannot do

- Tracking, classification, tagging, transcription

- Distinguish playing styles of two versions of the same song
- Provide explanations for predictions
- Understand how certain aspects of a piece affect the listener

Emotion models



In our context: perceived emotions in music

Discrete/Dimensional models

Emotion recognition and emotion experience

=> lots of vague definitions

=> noisy data

T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," Psychology of Music, vol. 39, no. 1, pp. 18–49, Jan. 2011.

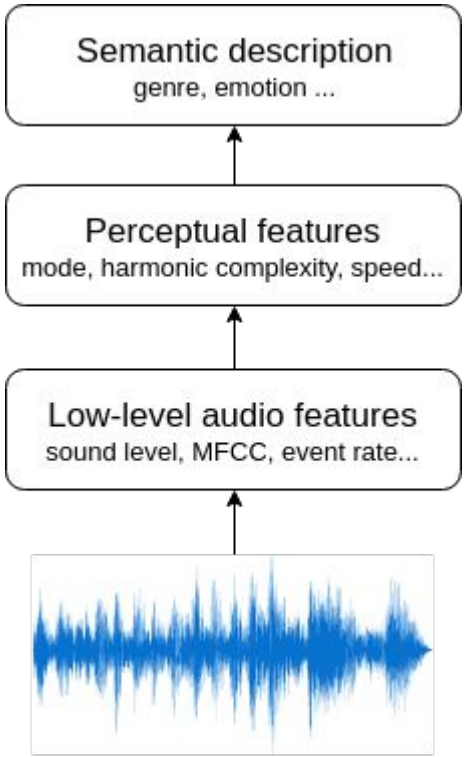
Emotion models

Clusters	Mood Adjectives
Cluster 1	passionate, rousing, confident, boisterous, rowdy
Cluster 2	rollicking, cheerful, fun, sweet, amiable/good natured
Cluster 3	literate, poignant, wistful, bittersweet, autumnal, brooding
Cluster 4	humorous, silly, campy, quirky, whimsical, witty, wry
Cluster 5	aggressive, fiery, tense/anxious, intense, volatile, visceral

Y. E. Kim et al., “MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW,” p. 12, 2010.

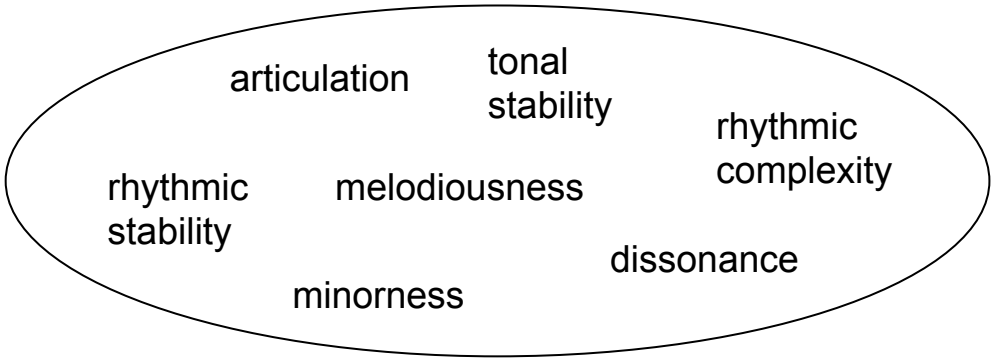
Perceptual features

What do we really hear when we listen to music?



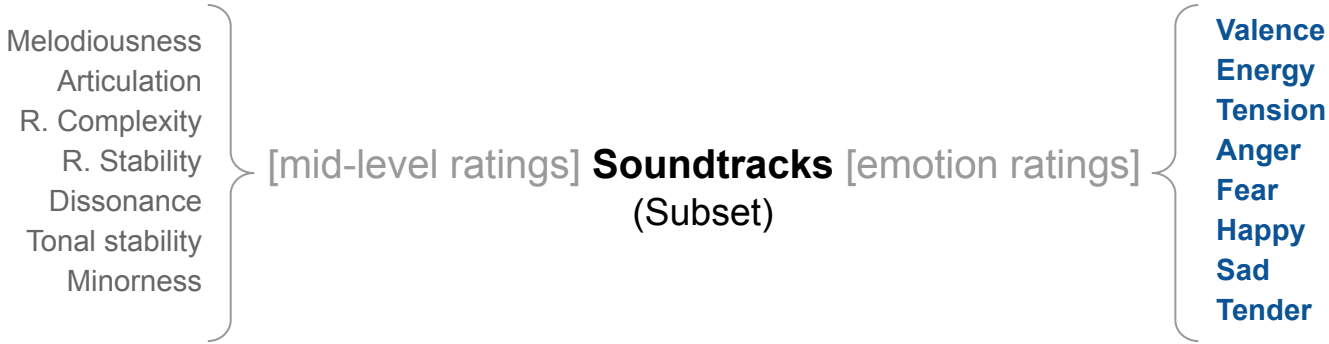
Mid level perceptual features dataset

- 5000 audio snippets (~15 sec each)
- Each snippet is annotated with ratings (1-10) for 7 mid-level features



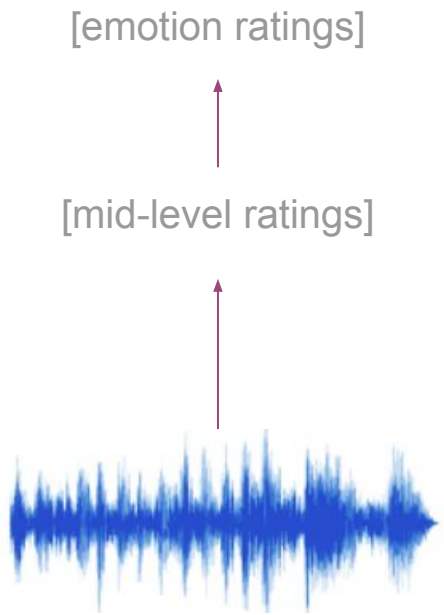
[1]A. Aljanaki and M. Soleymani, “A DATA-DRIVEN APPROACH TO MID-LEVEL PERCEPTUAL MUSICAL FEATURE MODELING,” p. 7, 2018.

Mid level perceptual features dataset



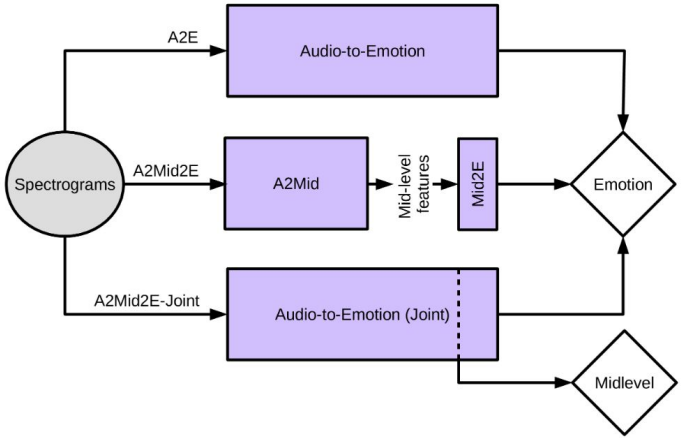
Mid level perceptual features dataset

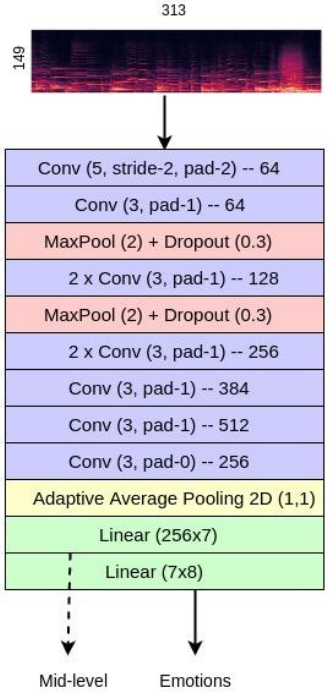
Basic idea - using mid-level features to explain (instance-based) predictions of emotions.



Training a model to test cost of explainability

- VGG-like model
- Joint-model: Losses summed





Shreyan Chowdhury, Andreu Vall, Verena Haunschmid, Gerhard Widmer, *"Towards Explainable Music Emotion Recognition: The Route via Mid-level Features"* (under review, ISMIR 2019)

Training a model to test cost of explainability

	Valence	Energy	Tension	Anger	Fear	Happy	Sad	Tender	Avg.
Mid2E (Aljanaki)	0.88	0.79	0.84	0.65	0.82	0.81	0.73	0.72	0.78
Mid2E (Ours)	0.88	0.80	0.84	0.65	0.82	0.81	0.74	0.73	0.79
A2E	0.81	0.79	0.84	0.82	0.81	0.66	0.60	0.75	0.76
A2Mid2E	0.79	0.74	0.78	0.72	0.77	0.64	0.58	0.67	0.71
A2Mid2E-Joint	0.82	0.78	0.82	0.76	0.79	0.65	0.64	0.72	0.75
$CoE_{A2Mid2E}$	0.02	0.05	0.06	0.10	0.03	0.02	0.02	0.08	0.05
$CoE_{A2Mid2E\text{-}Joint}$	-0.02	0.01	0.02	0.06	0.02	0.01	-0.04	0.03	0.01

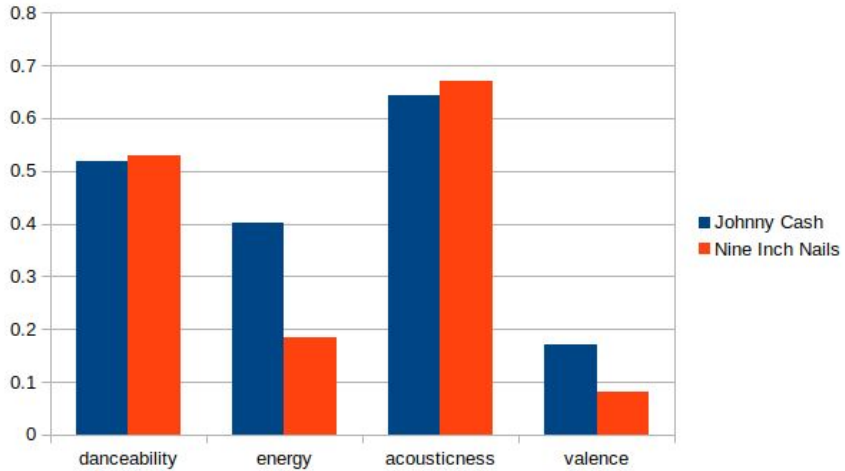
Multitask-learning works even better - 0.77

Shreyan Chowdhury, Andreu Vall, Verena Haunschmid, Gerhard Widmer, *"Towards Explainable Music Emotion Recognition: The Route via Mid-level Features"* (under review, ISMIR 2019)

Spotify features

Johnny Cash - Hurt

Nine Inch Nails - Hurt

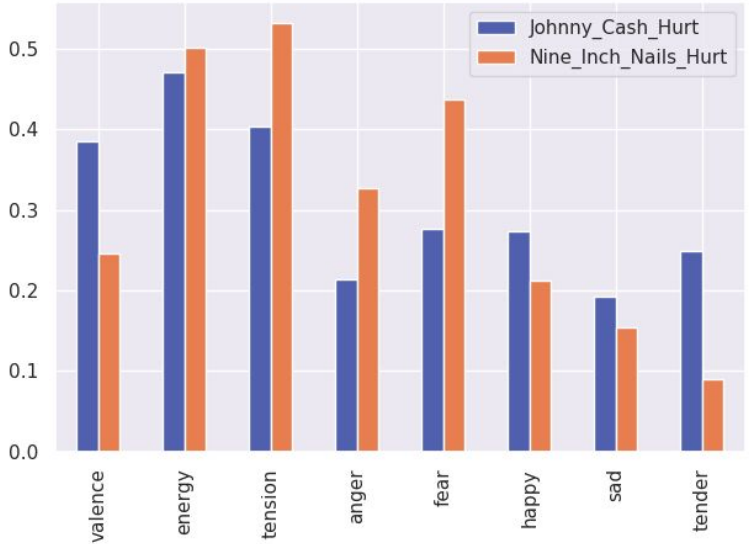


https://developer.spotify.com/console/get-audio-features-track

Emotions

Johnny Cash - Hurt

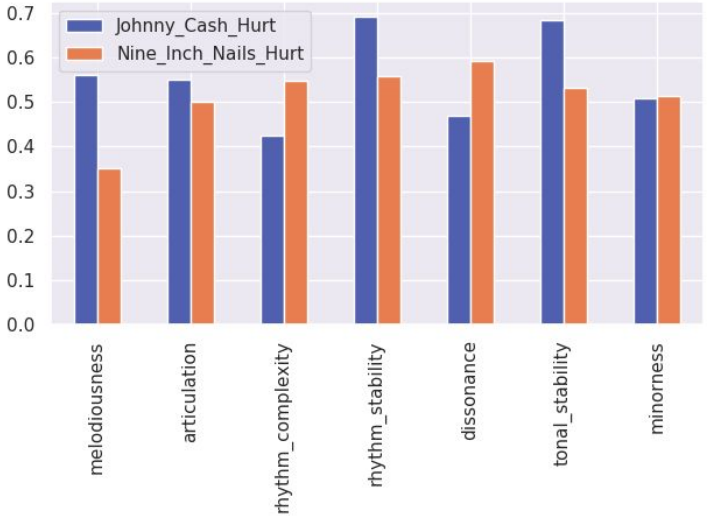
Nine Inch Nails - Hurt



Mid level features

Johnny Cash - Hurt

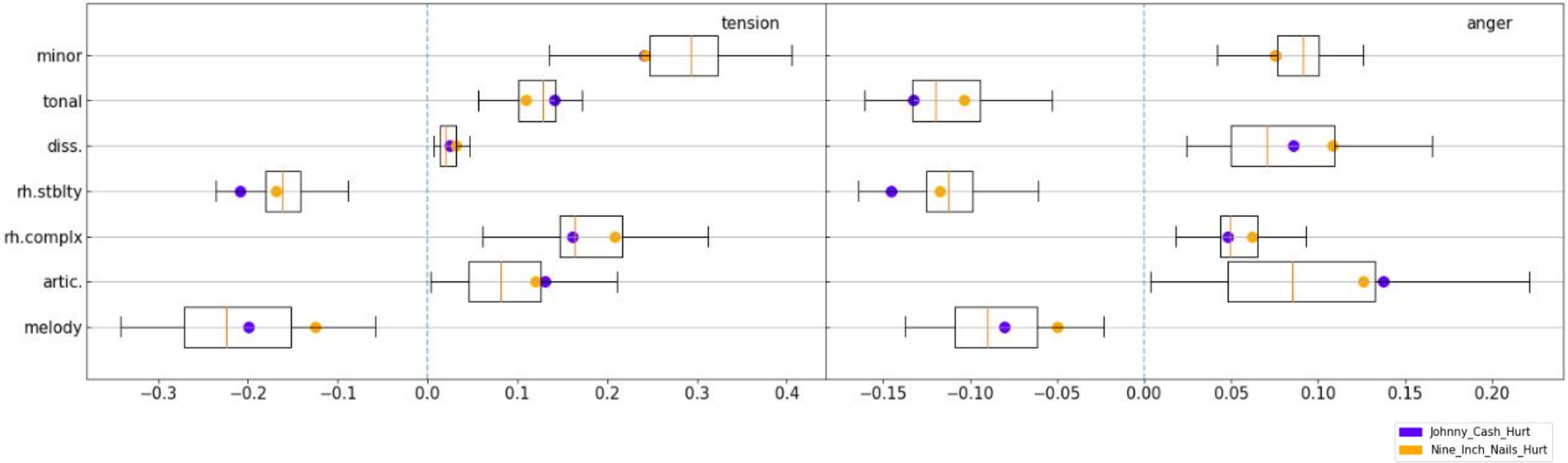
Nine Inch Nails - Hurt



Making sense of intermediate representations - effects plots

What musical characteristics make the Nine Inch Nails version sound "angrier"?

Linear effects



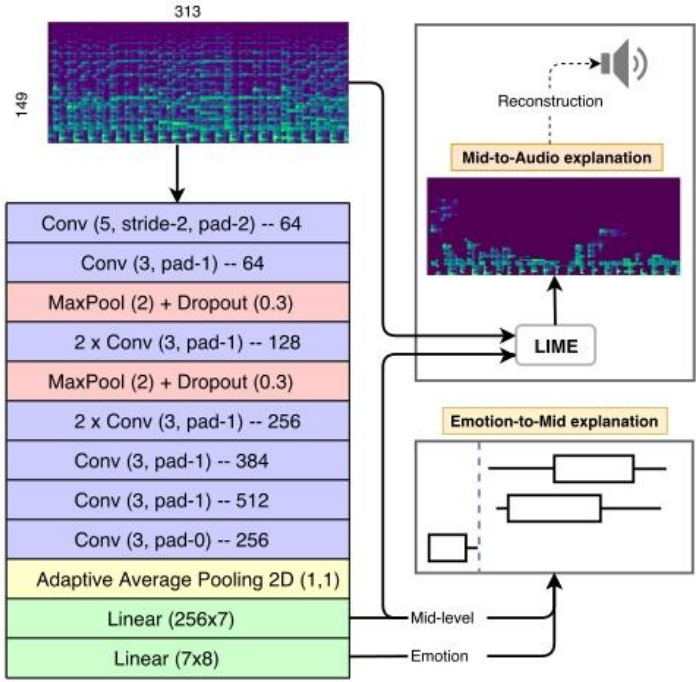
$$\text{effect}_j^{(i)} = w_j x_j^{(i)}$$

Shreyan Chowdhury, Andreu Vall, Verena Haunschmid, Gerhard Widmer, *"Towards Explainable Music Emotion Recognition: The Route via Mid-level Features"* (under review, ISMIR 2019)

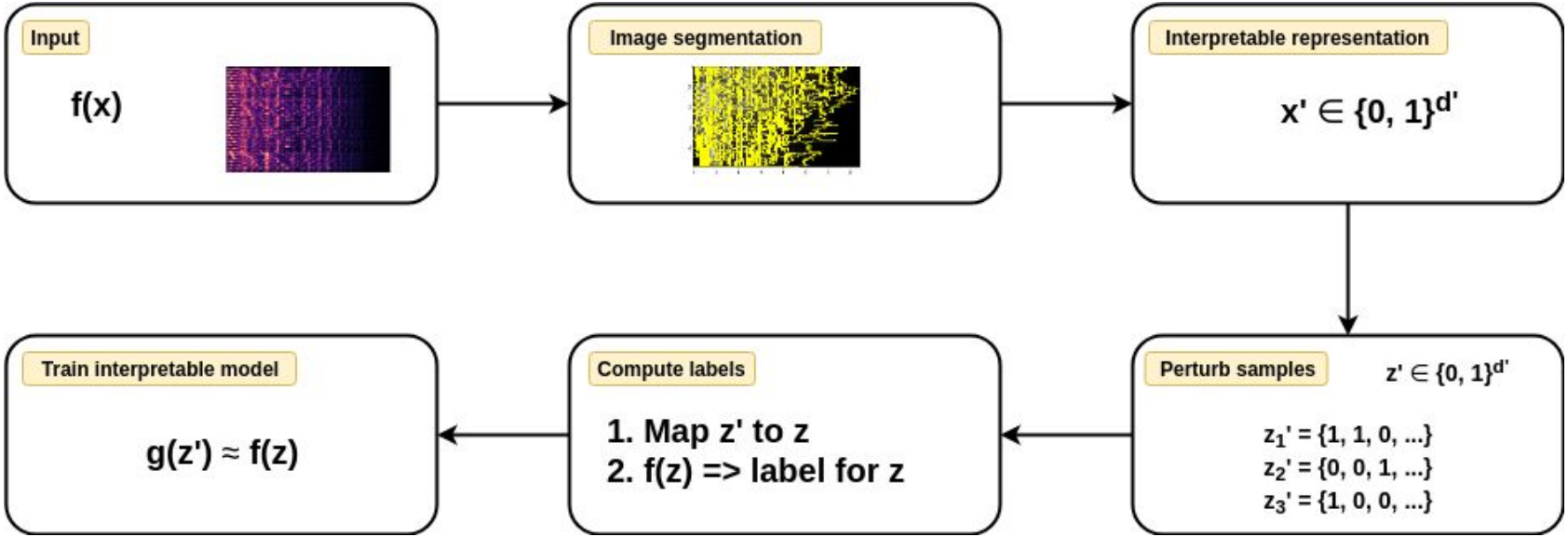
Explaining intermediate representations

Which part of the audio/spectrogram makes the model predict high values for "articulation"?

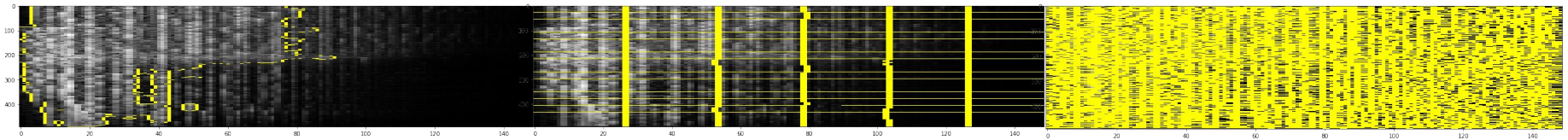
Two-level Explanations



Two-level Explanations

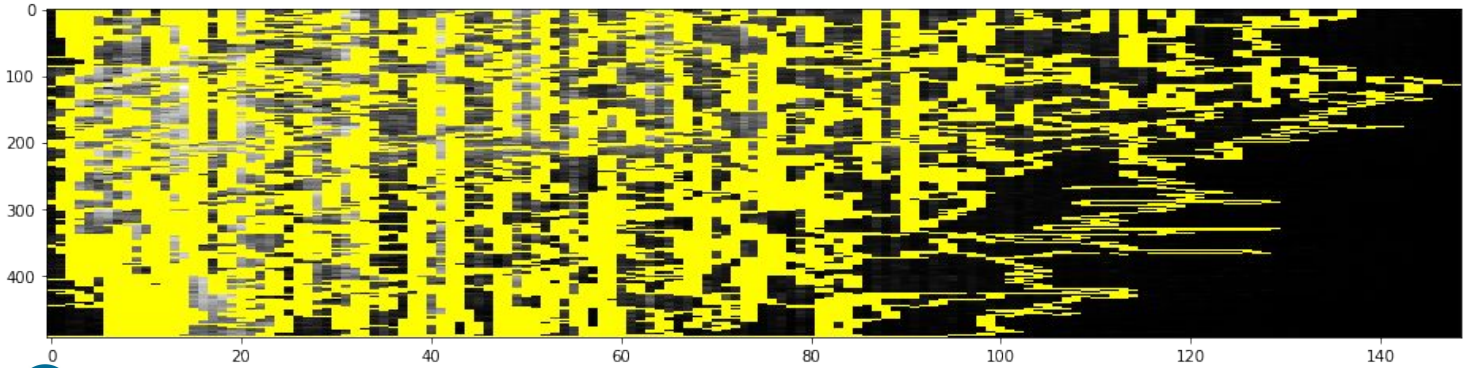


Spectrogram segmentation



Chan Vese SLIC Watershed

Felzenszwalb

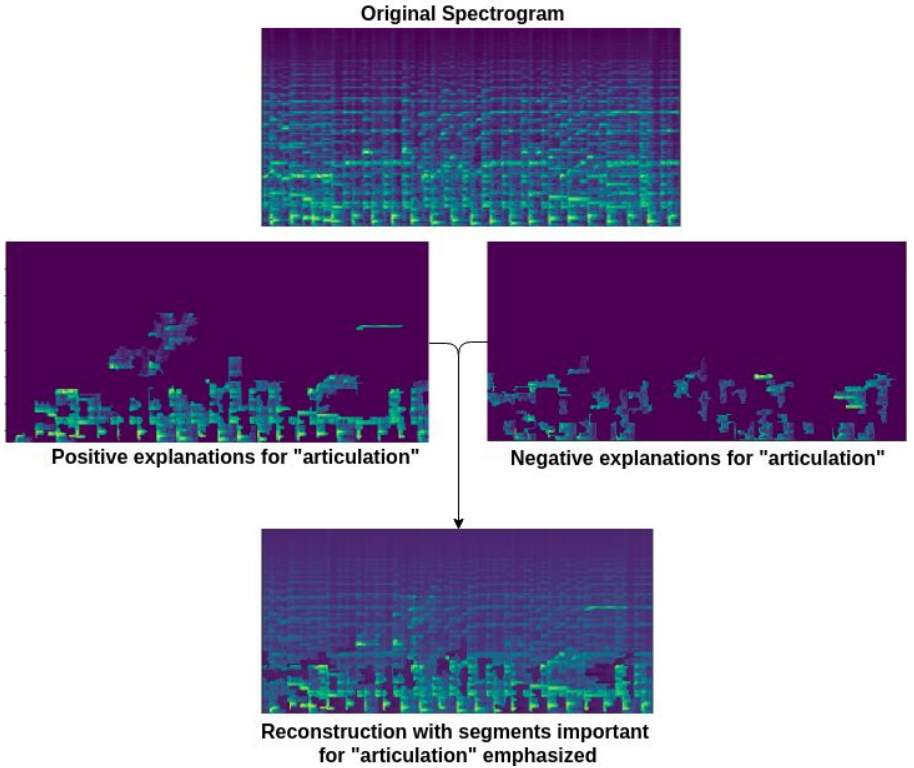


Extending LIME

- Instability when repeating sampling process
  - Experiments conducted with 50k perturbed samples
- How many important image segments?
  - Thresholding the p-value : weight ratio
  - For our experiments: t = 10e-6
  - Resulting in 30 to 60 image segments (out of ~300)
- Image segmentation
  - Felzenszwalb (scale=25, min_size=40)

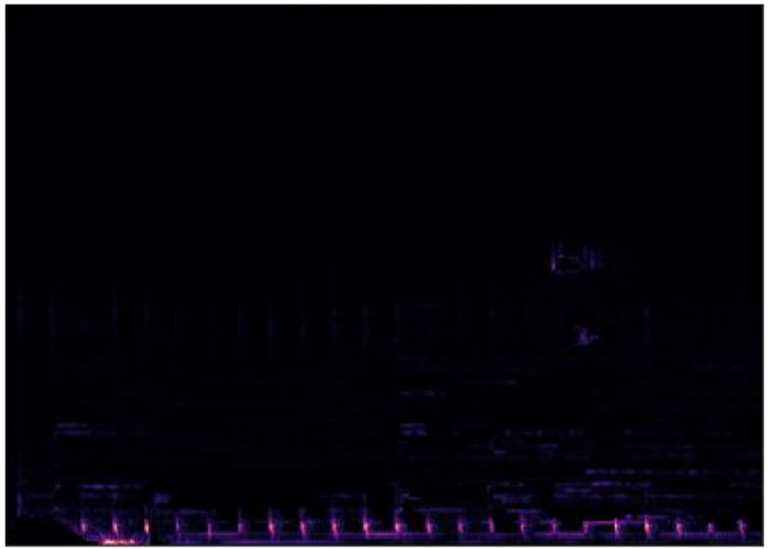
Audio-to-Mid explanations

- Input: original spectrogram
- LIME + Post-Processing:
  - "Positive" explanations
  - "Negative" explanations
- Experimental: Audio reconstruction



Visualizing some examples

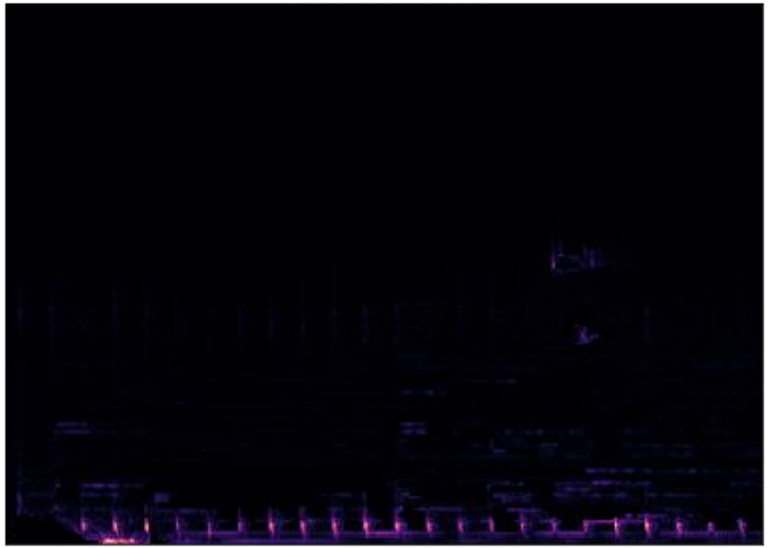
articulation POS



articulation NEG

Experimental: Listening to explanations

articulation POS



articulation NEG

More examples

Original

Articulation

Melodiousness

Demonstration

- https://shreyanc.github.io/ICML_example.html

Acknowledgements: Shreyan Chowdhury

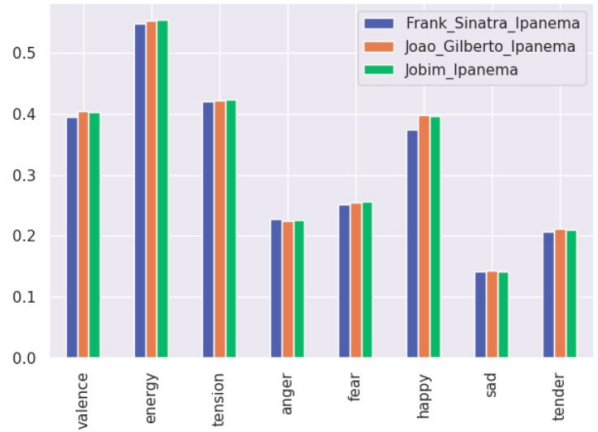
Summary

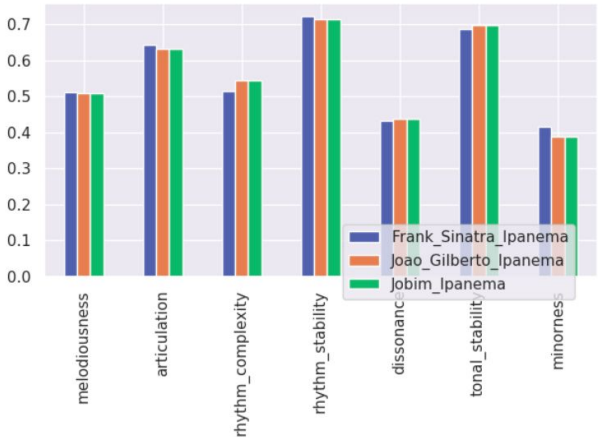
- Explanations in terms of audio
  - LIME detected meaningful patterns in the spectrogram ⇒ leading to listenable explanations
  - Provide a basis for targeted modifications of the audio
    - E.g. enhance or weaken certain musical qualities to control perceived emotional qualities
    - Some features are easier to enhance than others

- LIME not useful out of the box

Future Work

- Replace image segmentation by NMF / Source Separation
- Evaluate other interpretability methods
- Improve sensitivity to subtle differences





Presented Work

Verena Haunschmid, Shreyan Chowdhury, Gerhard Widmer (2019). "Two-Level Explanations for Music Emotion Recognition". Machine Learning for Music Discovery Workshop @ ICML.

Shreyan Chowdhury, Andreu Vall, Verena Haunschmid, Gerhard Widmer (2019, to appear). "Towards Explainable Music Emotion Recognition: The Route via Mid-level Features". International Society of Music Information Retrieval.

Rainer Kelz, Gerhard Widmer (2019, to appear). “Towards Interpretable Polyphonic Transcription with Invertible Neural Networks”. International Society of Music Information Retrieval.

Paul Primus (2019), “Gradient-Based Explanations for Audio Classifiers”. Master Thesis.

K. Koutini, H. Eghbal-zadeh, M. Dorfer, and G. Widmer, “The Receptive Field as a Regularizer in Deep Convolutional Neural Networks for Acoustic Scene Classification,” arXiv:1907.01803 [cs, eess, stat], Jul. 2019.

References (1)

Tim Miller (2017). “Explanation in artificial intelligence: Insights from the social sciences.” arXiv Preprint arXiv:1706.07269.

Been Kim, Rajiv Khanna, and Oluwasanmi O. Koyejo (2016). “Examples are not enough, learn to criticize! Criticism for interpretability.” Advances in Neural Information Processing Systems.

Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin (2016). "Why Should I Trust You? Explaining the Predictions of Any Classifier."

Tuomas Eerola and Jonna Katariina Vuoskoski (2011). "A comparison of the discrete and dimensional models of emotion in music". *Psychology of Music, 39(1), 18-49*

Anna Aljanaki and Mohammed Soleymani (2018). "A Data-driven Approach to Mid-level Perceptual Musical Feature Modeling."

References (2)

Christoph Molnar (2019). “Interpretable Machine Learning. A Guide for Making Black Boxes Explainable”.

Saumitra Mishra, Bob L. Sturm and Simon Dixon (2017), “Local Interpretable Model-Agnostic Explanations for Music Content Analysis”, International Society of Music Information Retrieval.

JOHANNES KEPLER
UNIVERSITY LINZ

Thank you!

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Str. 69
4040 Linz, Austria
www.jku.at