Universitat Pompeu Fabra Barcelona

MTG Music Technology Group

Human-centered machine learning for music emotion recognition

Juan Sebastián Gómez-Cañón

Supervisors:

Dr. Emilia Gómez
Joint Research Centre,
European Commission, Seville, Spain

Dr. Estefanía Cano
Songquito UG,
Erlangen, Germany

Dr. Perfecto Herrera
Music Technology Group,
Universitat Pompeu Fabra, Barcelona, Spain

Committee members:

Dr. Blair Kaneshiro
Center for Computer Research in Music and Acoustics,
University of Stanford, USA

Dr. Isabelle Hupont
Joint Research Centre,
European Commission, Seville, Spain

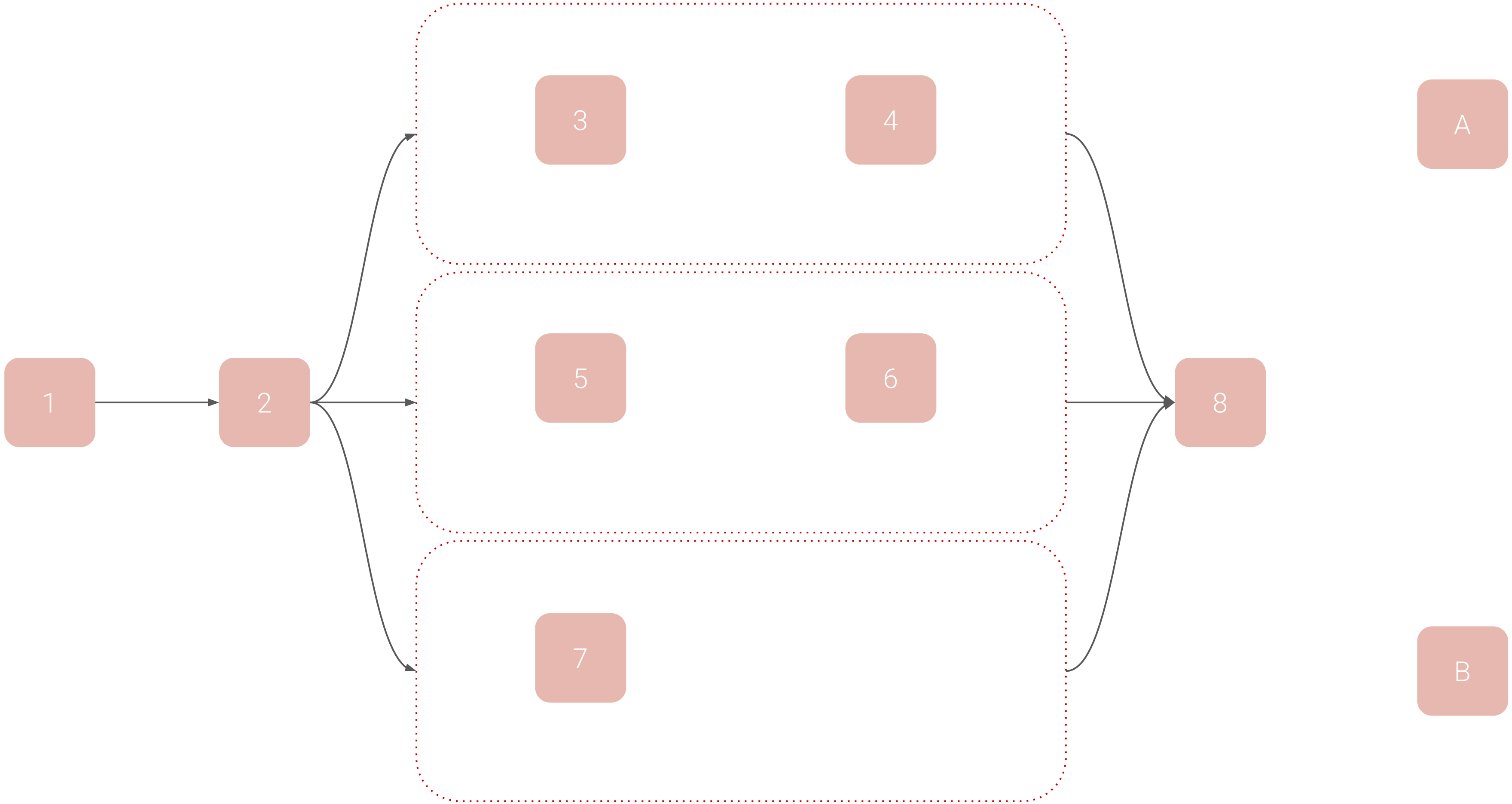
Dr. Arthur Flexer
Institute of Computational Perception,
Johannes Kepler University Linz, Austria

TROMPA

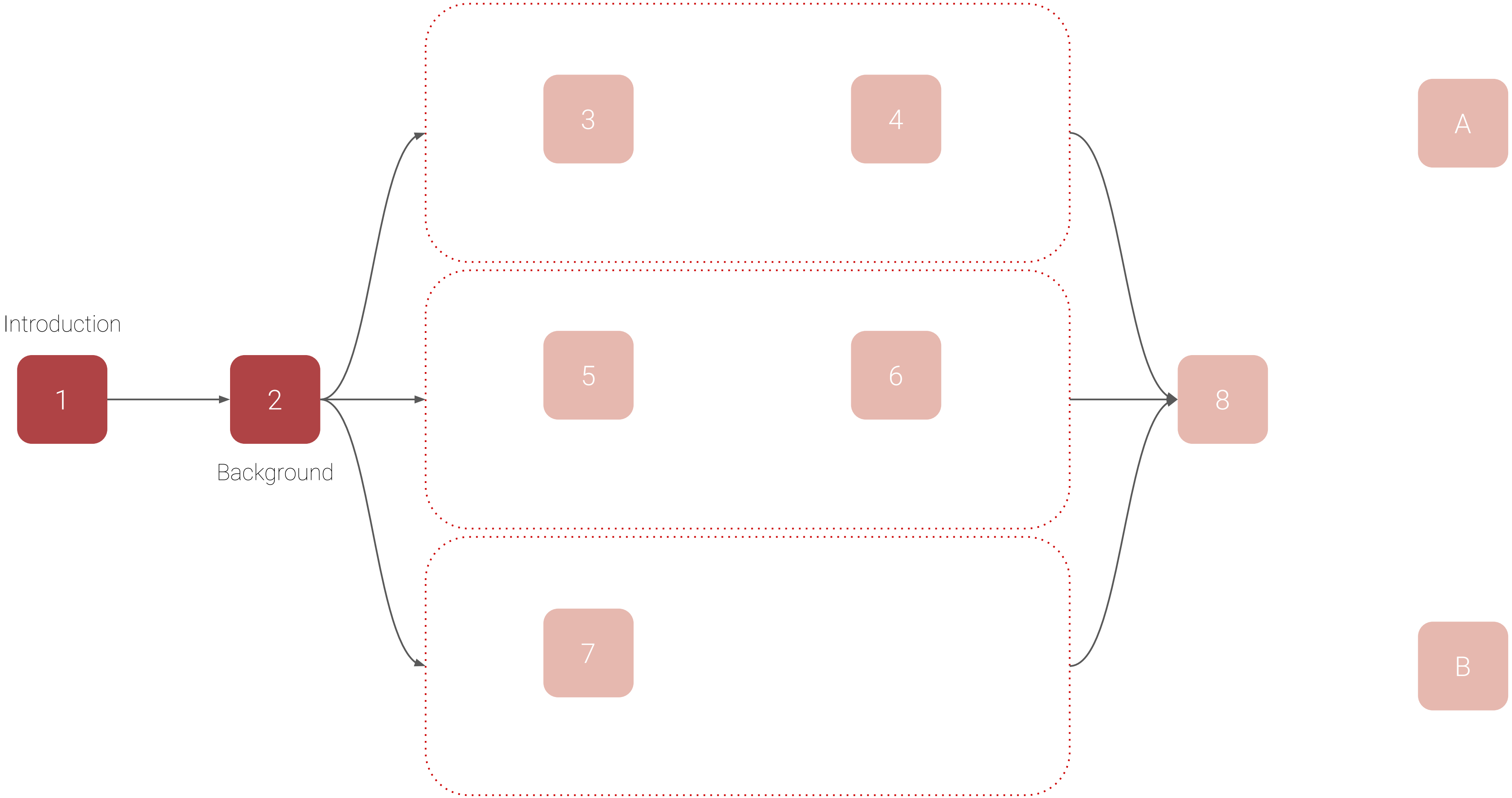
MUSICAL AI_

Ph.D. defense - Barcelona, October 26, 2022

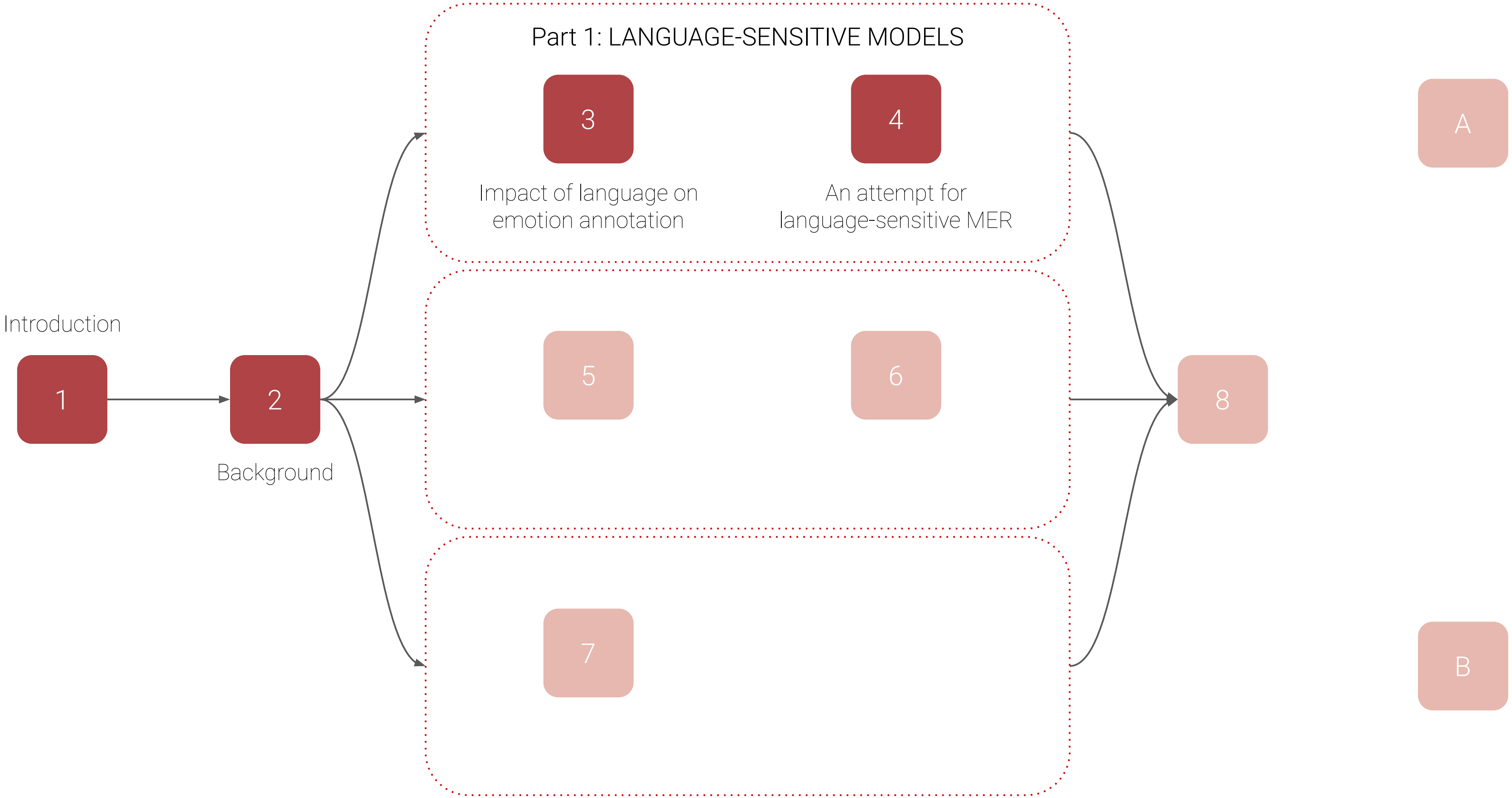
Outline



Outline

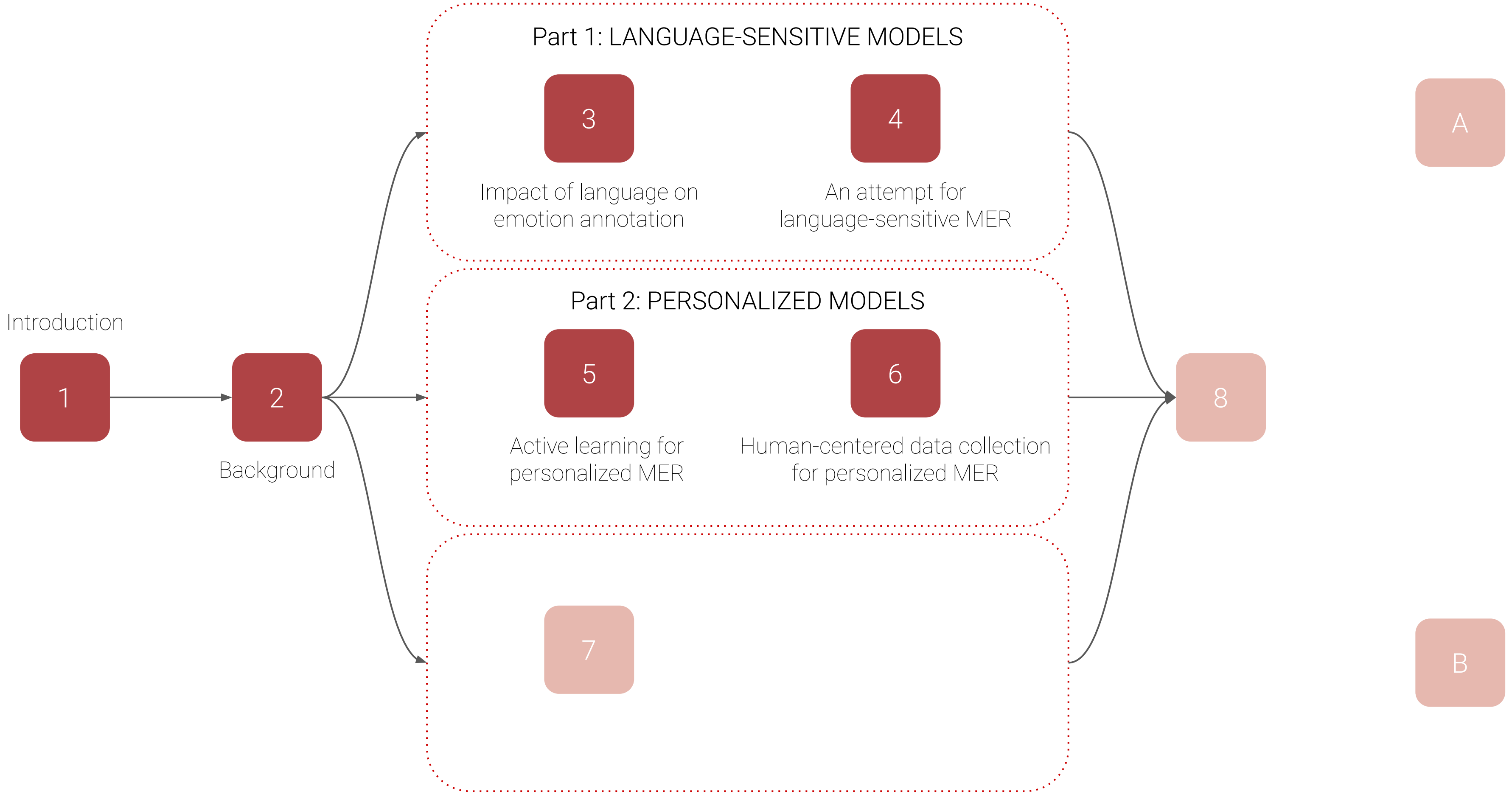


Outline

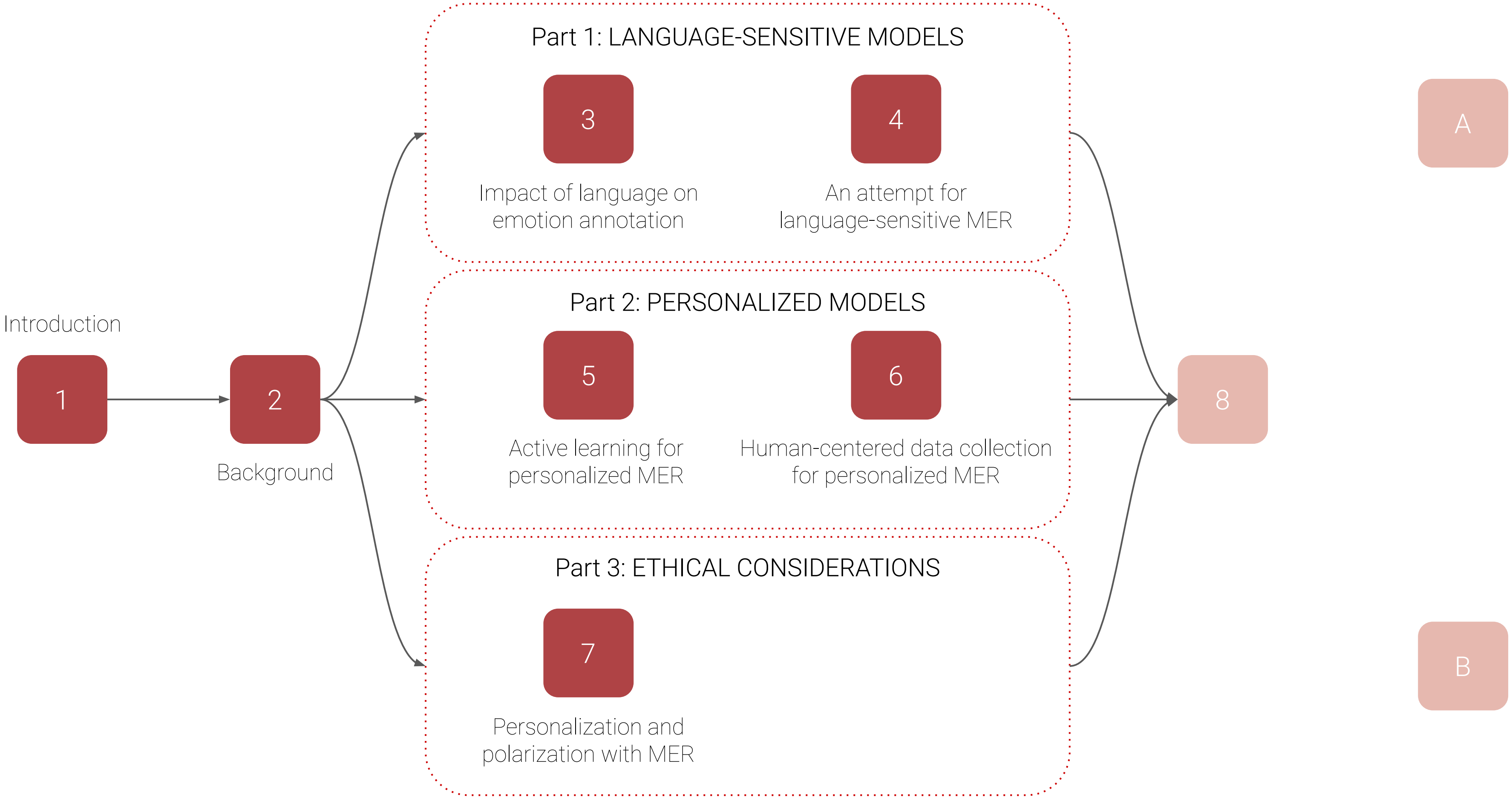


Part 1: LANGUAGE-SENSITIVE MODELS
3
Impact of language on emotion annotation
4
An attempt for language-sensitive MER
A
Introduction
1
2
Background
5
6
8
7
B

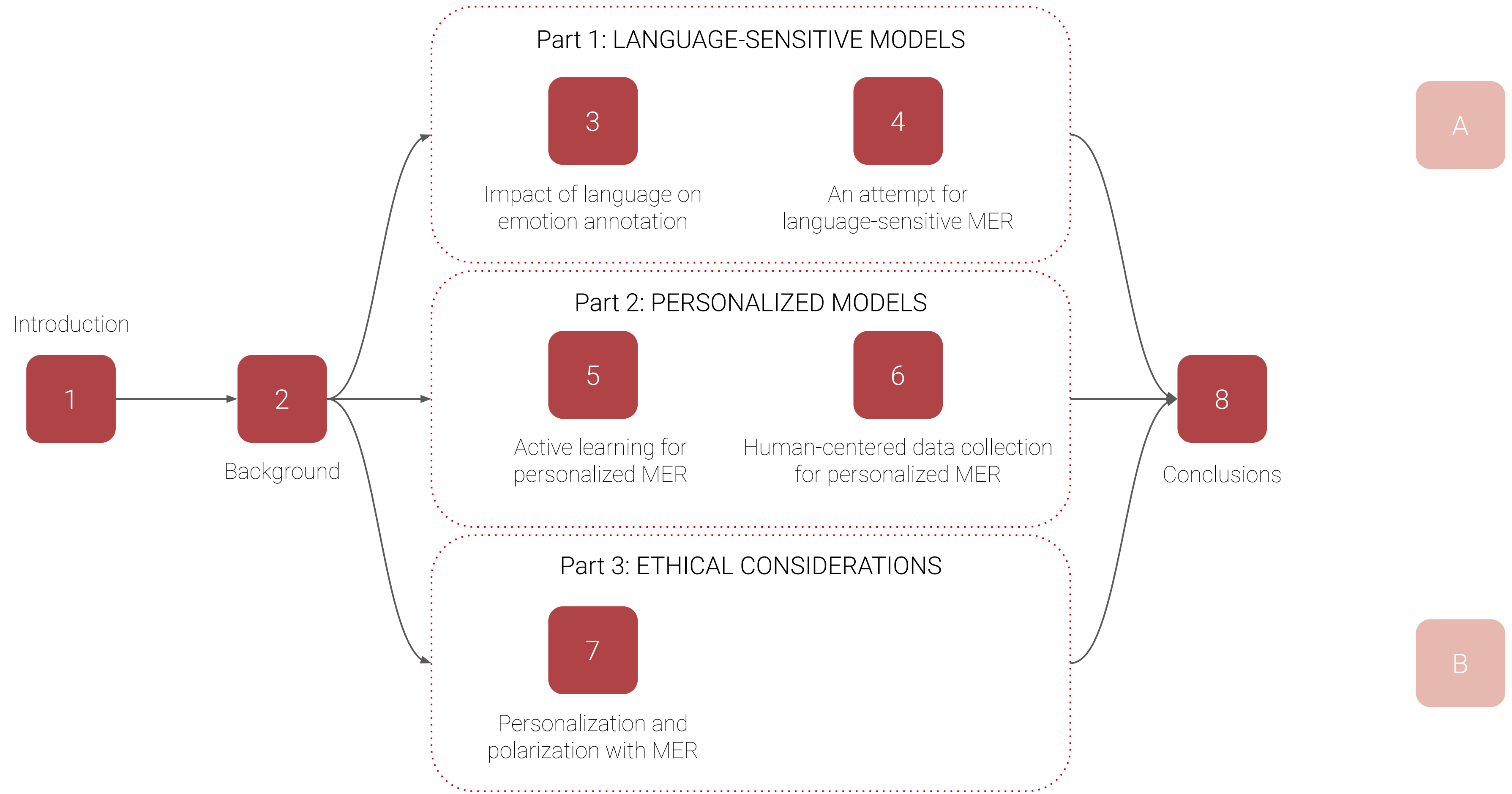
Outline


Part 1: LANGUAGE-SENSITIVE MODELS
3
Impact of language on emotion annotation
4
An attempt for language-sensitive MER
Part 2: PERSONALIZED MODELS
5
Active learning for personalized MER
6
Human-centered data collection for personalized MER
Introduction
1
2
Background
7
8
A
B

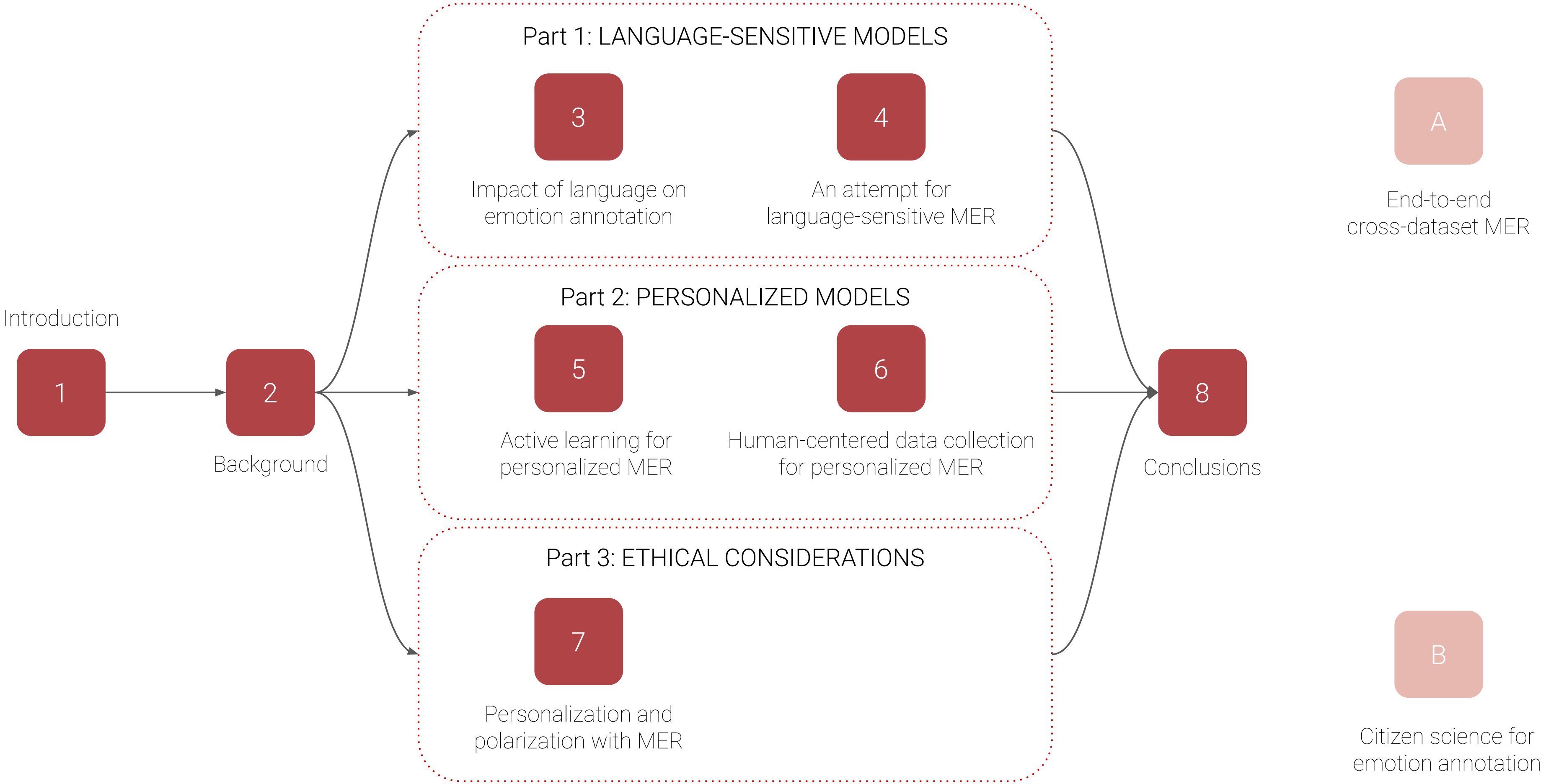
Outline


Part 1: LANGUAGE-SENSITIVE MODELS
3
Impact of language on emotion annotation
4
An attempt for language-sensitive MER
Part 2: PERSONALIZED MODELS
5
Active learning for personalized MER
6
Human-centered data collection for personalized MER
Part 3: ETHICAL CONSIDERATIONS
7
Personalization and polarization with MER
Introduction
1
2
Background
8
A
B

Outline


Introduction
1
2
Background
Part 1: LANGUAGE-SENSITIVE MODELS
3
Impact of language on emotion annotation
4
An attempt for language-sensitive MER
Part 2: PERSONALIZED MODELS
5
Active learning for personalized MER
6
Human-centered data collection for personalized MER
Part 3: ETHICAL CONSIDERATIONS
7
Personalization and polarization with MER
8
Conclusions
A
B

Outline



Part 1: LANGUAGE-SENSITIVE MODELS
3
Impact of language on emotion annotation
4
An attempt for language-sensitive MER
A
End-to-end cross-dataset MER
Part 2: PERSONALIZED MODELS
Introduction
1
2
Background
5
Active learning for personalized MER
6
Human-centered data collection for personalized MER
8
Conclusions
Part 3: ETHICAL CONSIDERATIONS
7
Personalization and polarization with MER
B
Citizen science for emotion annotation

Why music and emotions?

- Complex ➞ well researched
- Main reason ➞ understandable
- Categorization of music collections

Introduction

Why music and emotions?

- Complex ➞ well researched
- Main reason ➞ understandable
- Categorization of music collections

What is music emotion recognition (MER)?

- Emotionally relevant features of music
- **Perceived** or **induced** emotions
- Supervised machine learning (Yang, 2011)
- “Ground truth” ➞ subjective

Reference:
Yang & Chen. **Music emotion recognition.** CRC Press, 2011.

Musical communication of emotions

Musician

Listener

Intention

↑

Felt emotion?

Musical communication of emotions

Musician

Expression of emotion

Intention

↑

Felt emotion?

Listener

Musical communication of emotions

Musician

Expression of emotion

→ **Measurable sounds** →

Listener

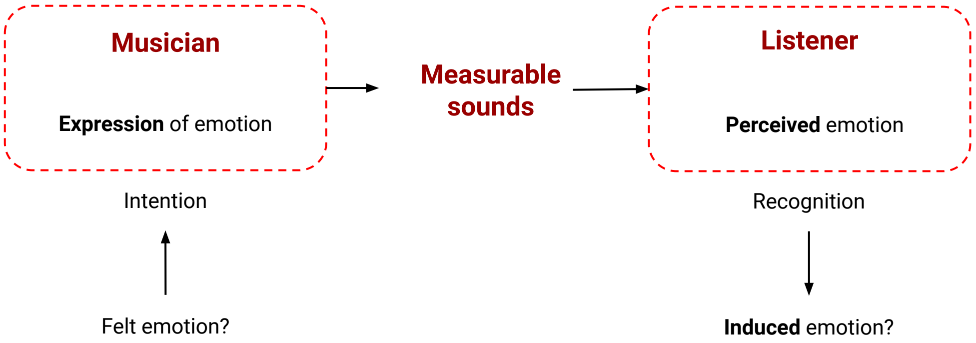
Perceived emotion

Intention

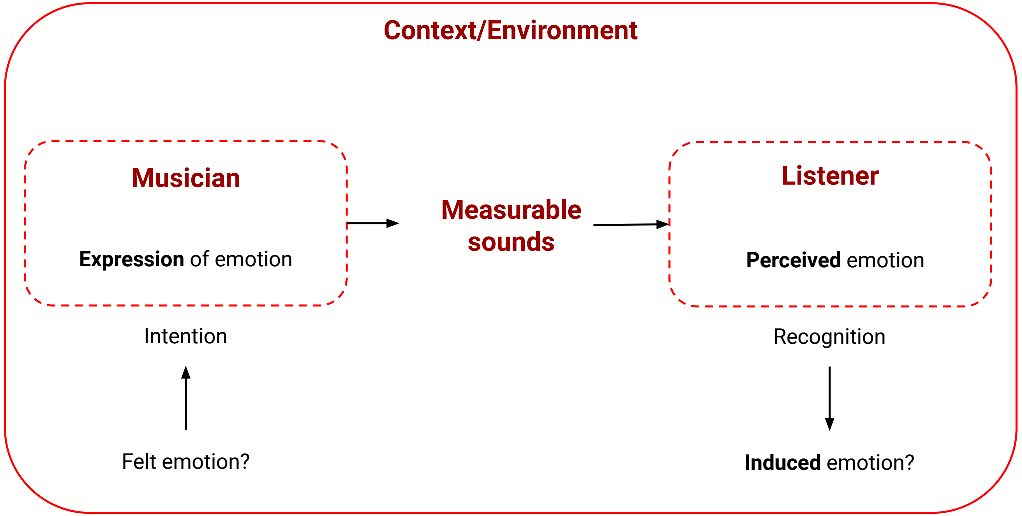
↑

Felt emotion?

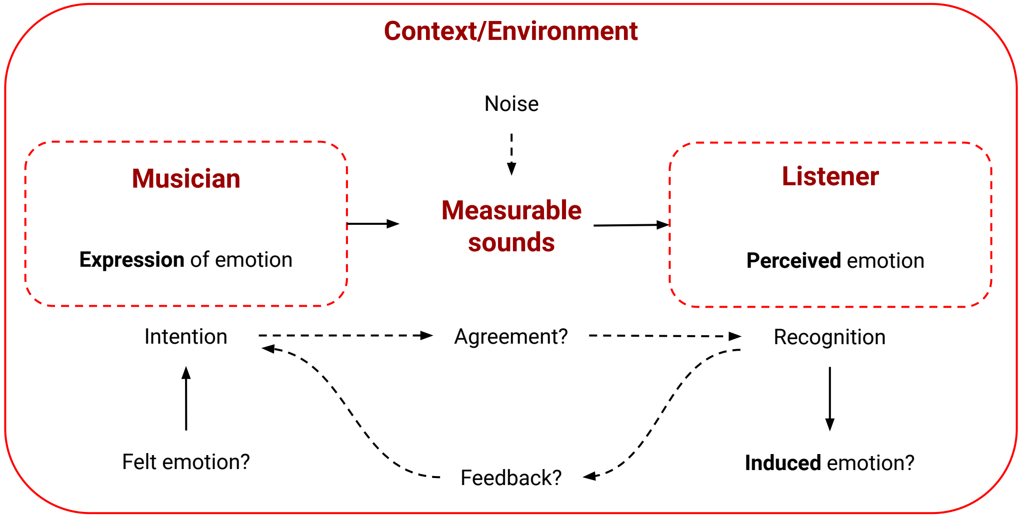
Musical communication of emotions

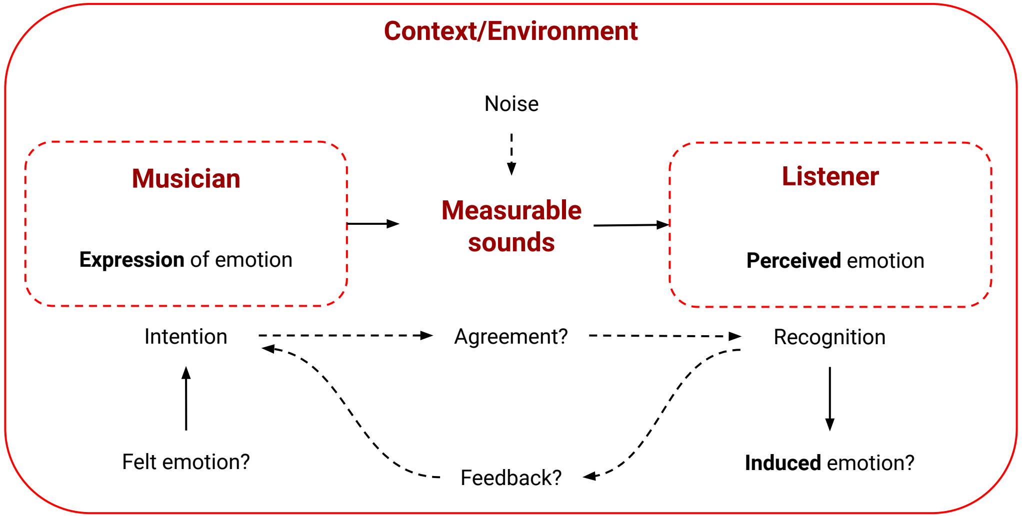


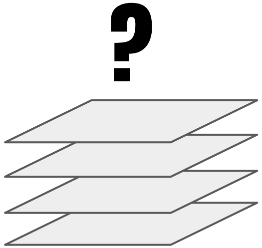
Musical communication of emotions



Musical communication of emotions

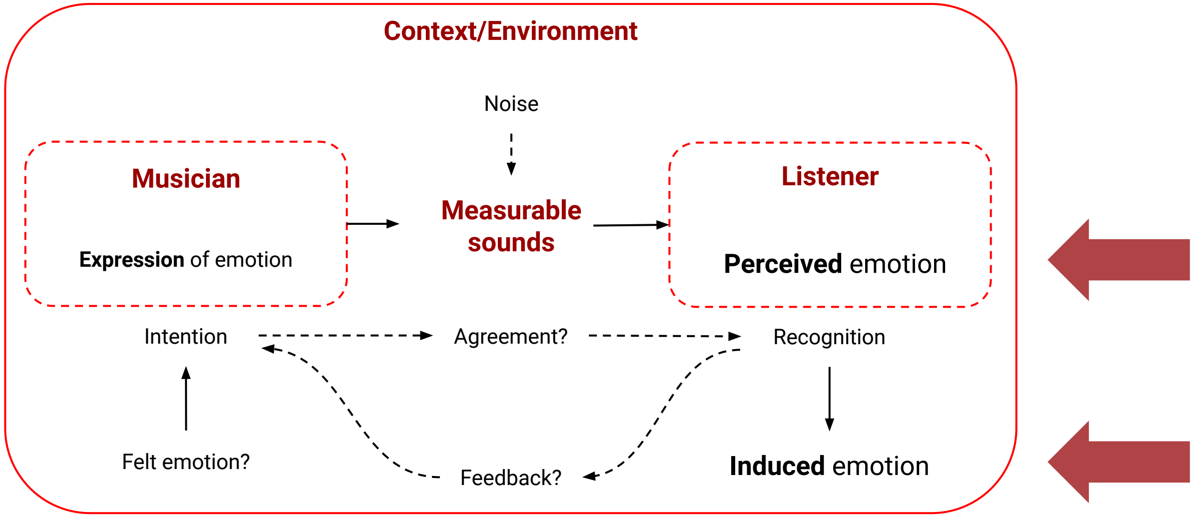






“Ground truth”

Felt emotion ⇋ Expressed emotion ⇋ Perceived emotion ⇋ Induced emotion



Context/Environment
Noise
Musician
Expression of emotion
Measurable sounds
Listener
Perceived emotion
Intention
Agreement?
Recognition
Felt emotion?
Feedback?
Induced emotion

Motivation

Why is music emotion recognition relevant?

- Perceived emotions:
  - Mood/emotion search
  - Indexing and categorization
- Induced emotions:
  - Mood regulation
  - Learning and well-being

Research questions for this dissertation

- For whom?
  - Subjectivity ➞ MIR
- What for?
  - Affective multimedia recommendation

Research questions for this dissertation

- For whom?
  - Subjectivity ➞ MIR
- What for?
  - Affective multimedia recommendation
- Individual judgment ➞ possible?

Why is human-centric ML reasonable for MER?

- For whom?
  - Person at the center ➞ PERSONALIZATION
- Role of individuality (Yang et al., 2007):
  - Group-based MER
  - Personalized MER

Reference:
Yang, Y.-H., et al. **Music emotion recognition: the role of individuality**. Proceedings of the International Workshop on Human-Centered Multimedia, pp. 13-22, 2007.

Context and goals

Why is human-centric ML reasonable for MER?

- What for?
  - Impact of ML on humans ➞ CONTEXT
- Age of context-aware music systems (Herrera, 2018):
  - Physiology
  - Cultural background
  - Activities while listening
  - Environmental context
  - Temporal context
  - Other data

Reference:
Herrera. **MIRages: an account of music audio extractors, semantic description and content-awareness, in the three ages of MIR**. PhD thesis, Universitat Pompeu Fabra, 2018.

Why is human-centric ML reasonable for MER?

- Age of context-aware music systems (Herrera, 2018)
  - More context!
- Role of individuality (Yang et al., 2007):
  - More personalization!
- Hypothesis of this dissertation
  - More effort on human-centric approaches!

2 Scientific Background and State of the Art

Publication:

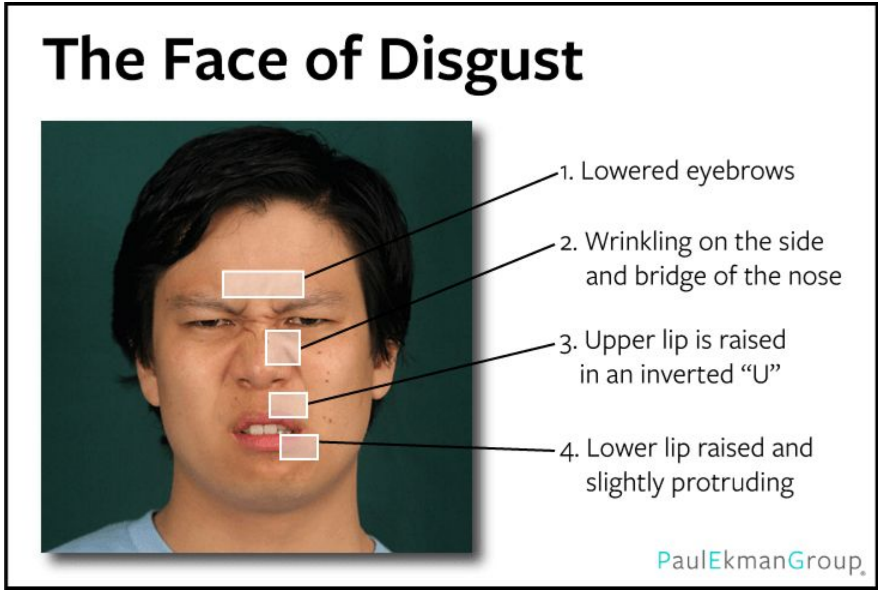
Gómez-Cañón, Cano, Eerola, Herrera, Hu, Yang & Gómez. ***Music Emotion Recognition: toward new, robust standards in personalized and context-sensitive applications***. IEEE Signal Processing Magazine, 38(6), 2021.

https://github.com/juansgomez87/datasets_emotion

Lack of a unifying theory of emotions

Disgust (before):

- Universal

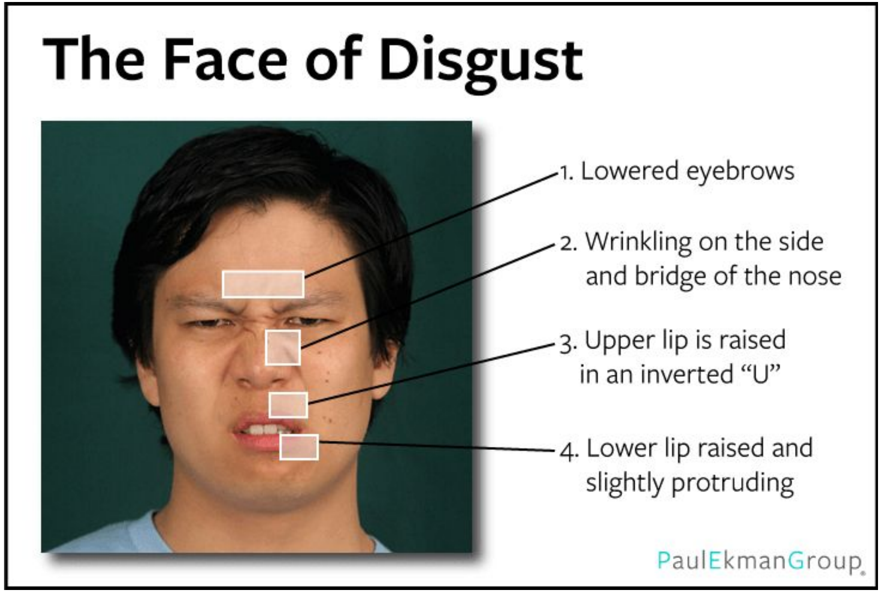


Reference:
Ekman, P. **An argument for basic emotions**. Cognition and Emotion, 6(3), pp. 169-200, 1992.

Lack of a unifying theory of emotions

Disgust (now):

- “Universal” (?)
- Risks?

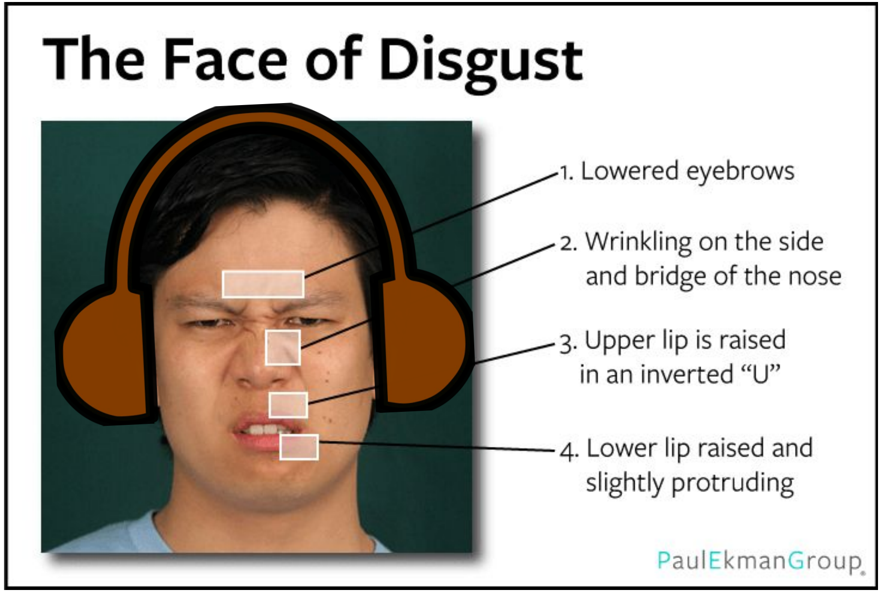


Barrett, L.F. **How emotions are made: the secret life of the brain**. Houghton Mifflin Harcourt, 2017.
Hupont, I., et al. **The landscape of facial processing applications in the context of the EU AI Act and the development of trustworthy systems**, Scientific Reports, 12, 2022.

Lack of a unifying theory of emotions

Disgust (now):

- "Universal" (?)
- Risks?



Music:

- Description?
- Induced?

Barrett, L.F. **How emotions are made: the secret life of the brain**. Houghton Mifflin Harcourt, 2017.
Hupont, I., et al. **The landscape of facial processing applications in the context of the EU AI Act and the development of trustworthy systems**, Scientific Reports, 12, 2022.

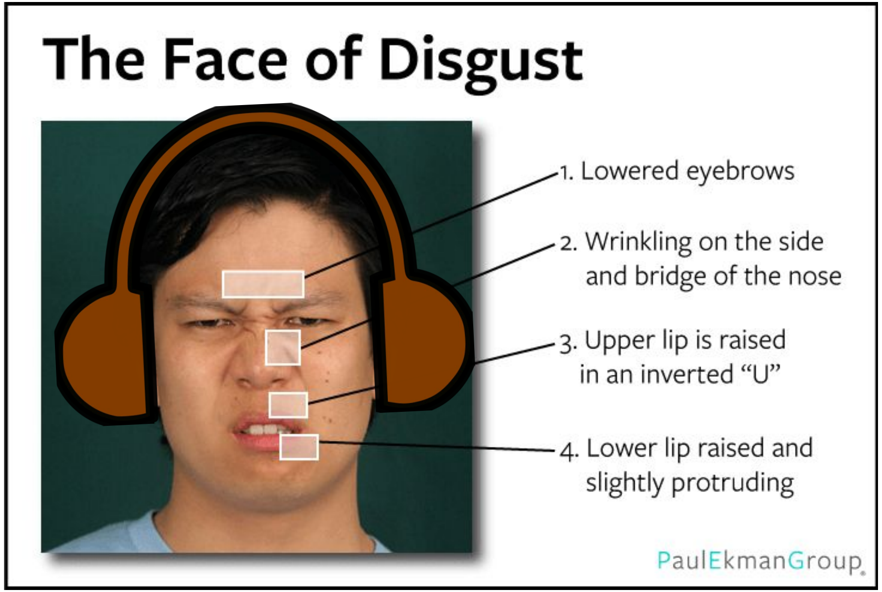
Lack of a unifying theory of emotions

Disgust (now):

- “Universal” (?)
- Risks?

Music:

- Description?
- Induced?
- Subtle differences



Barrett, L.F. **How emotions are made: the secret life of the brain**. Houghton Mifflin Harcourt, 2017.
Hupont, I., et al. **The landscape of facial processing applications in the context of the EU AI Act and the development of trustworthy systems**, Scientific Reports, 12, 2022.

Definitions

Emotions, moods, feelings…

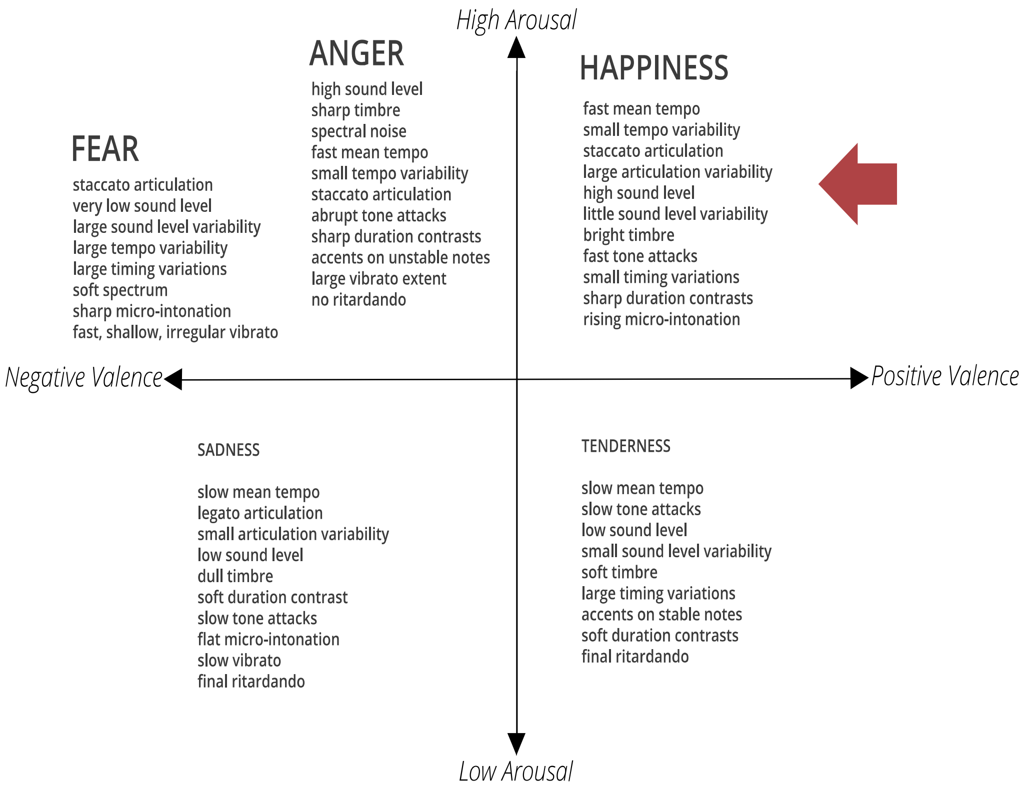
Emotions, moods, feelings…

- Brief and intense reactions
- Synchronicity
- Focus on an object

Reference:
Juslin. **Musical emotions explained.** Oxford University Press, 2019.

Taxonomies of emotion

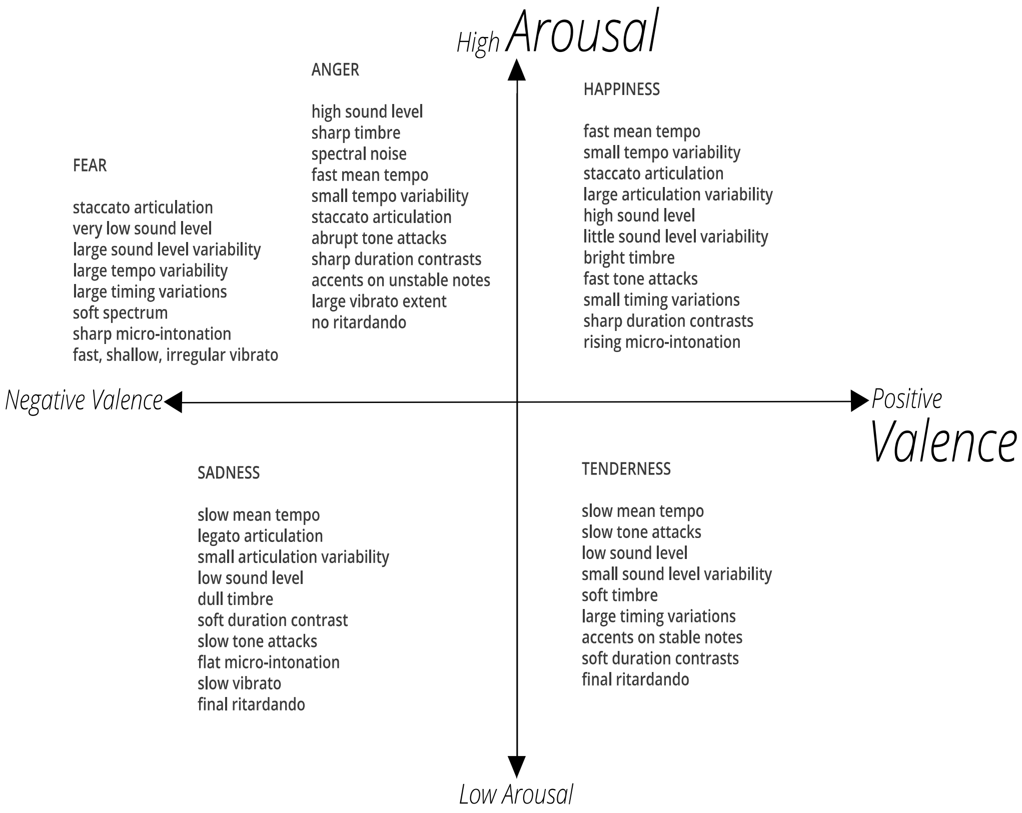
- **Categorical or discrete**
- Dimensional or continuous



Reference:
Juslin. **Musical emotions explained.** Oxford University Press, 2019.

Taxonomies of emotion

- Categorical or discrete
- **Dimensional or continuous**

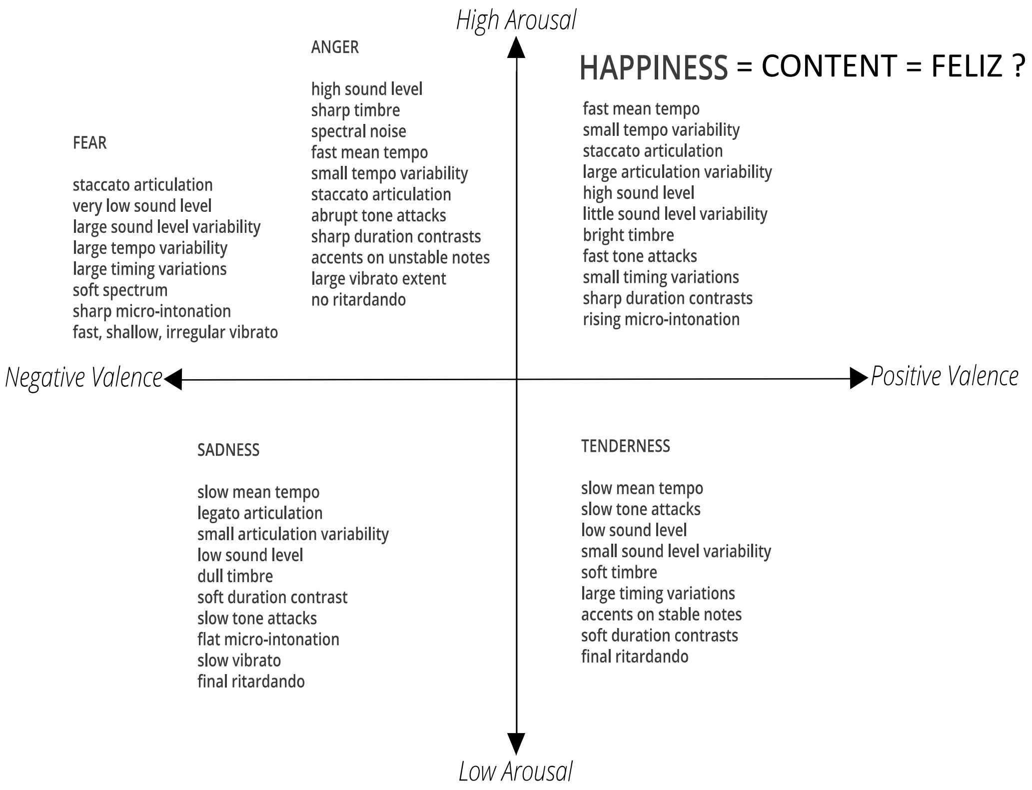


Reference:
Juslin. **Musical emotions explained.** Oxford University Press, 2019.

Definitions

Taxonomies of emotion

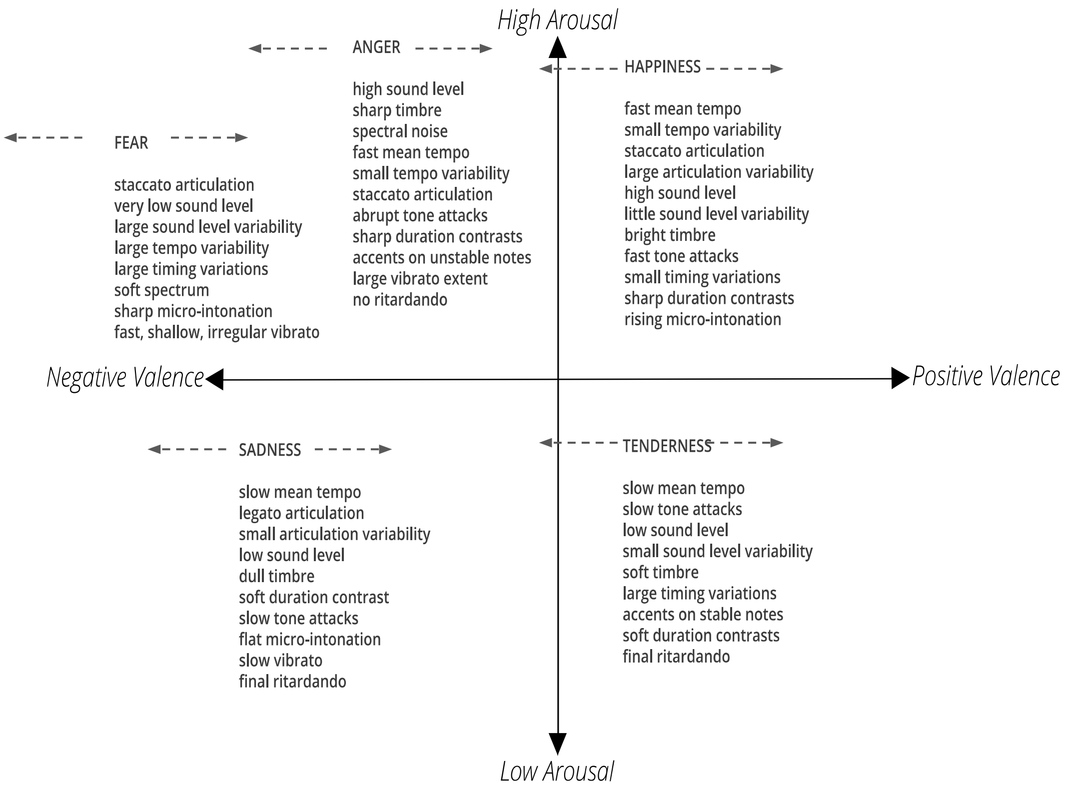
- Categorical or discrete
- Dimensional or continuous
- **Ambiguity** ⇋ Granularity



Reference:
Juslin. **Musical emotions explained.** Oxford University Press, 2019.

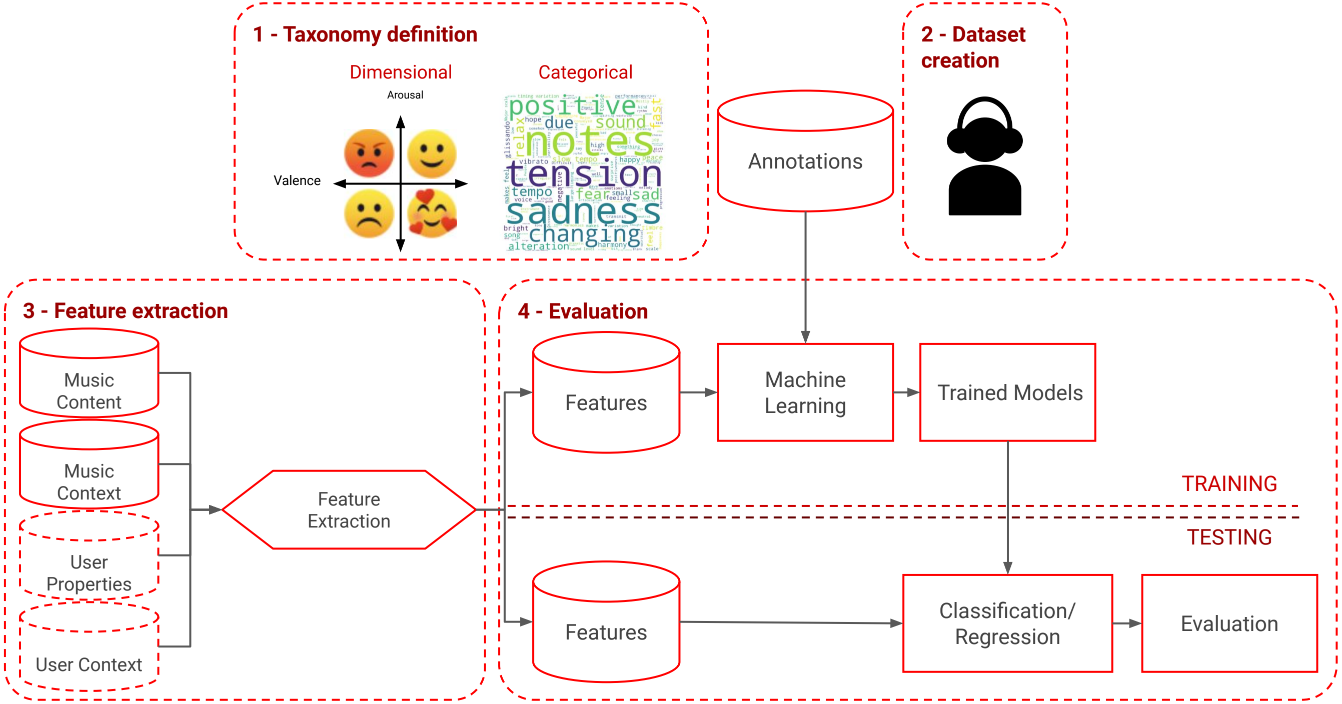
Taxonomies of emotion

- Categorical or discrete
- Dimensional or continuous
- Ambiguity ⇋ **Granularity**



Reference:
Juslin. **Musical emotions explained.** Oxford University Press, 2019.

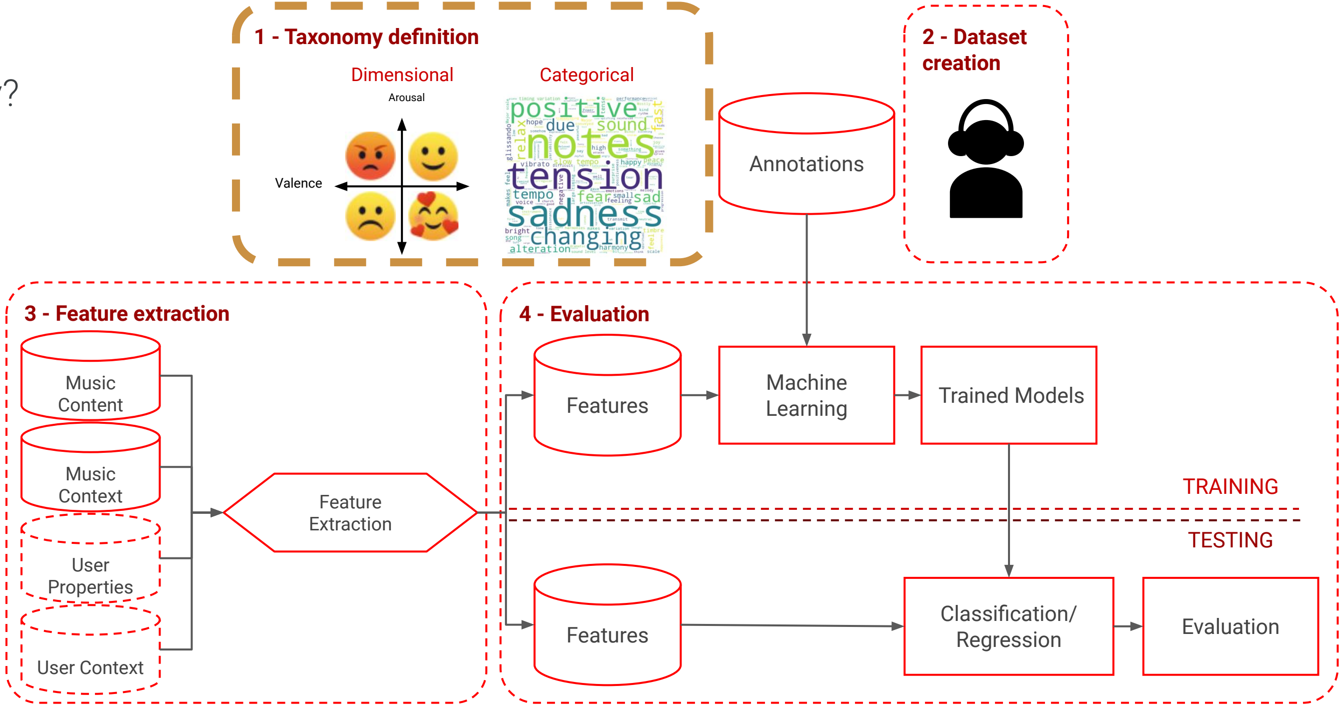
Traditional music emotion recognition systems



Traditional music emotion recognition systems

Step 1:

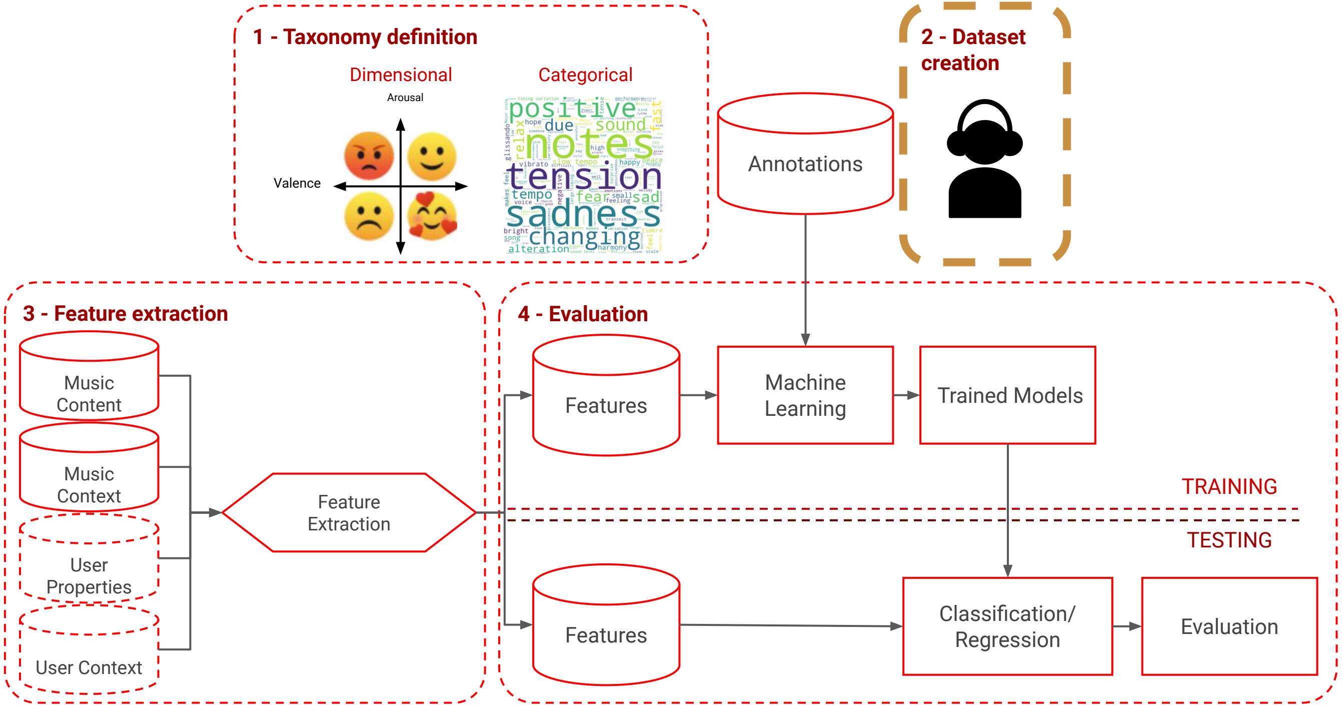
- Categories? Granularity?
- Time variation?
- Excerpt length?



Traditional music emotion recognition systems

Step 2:

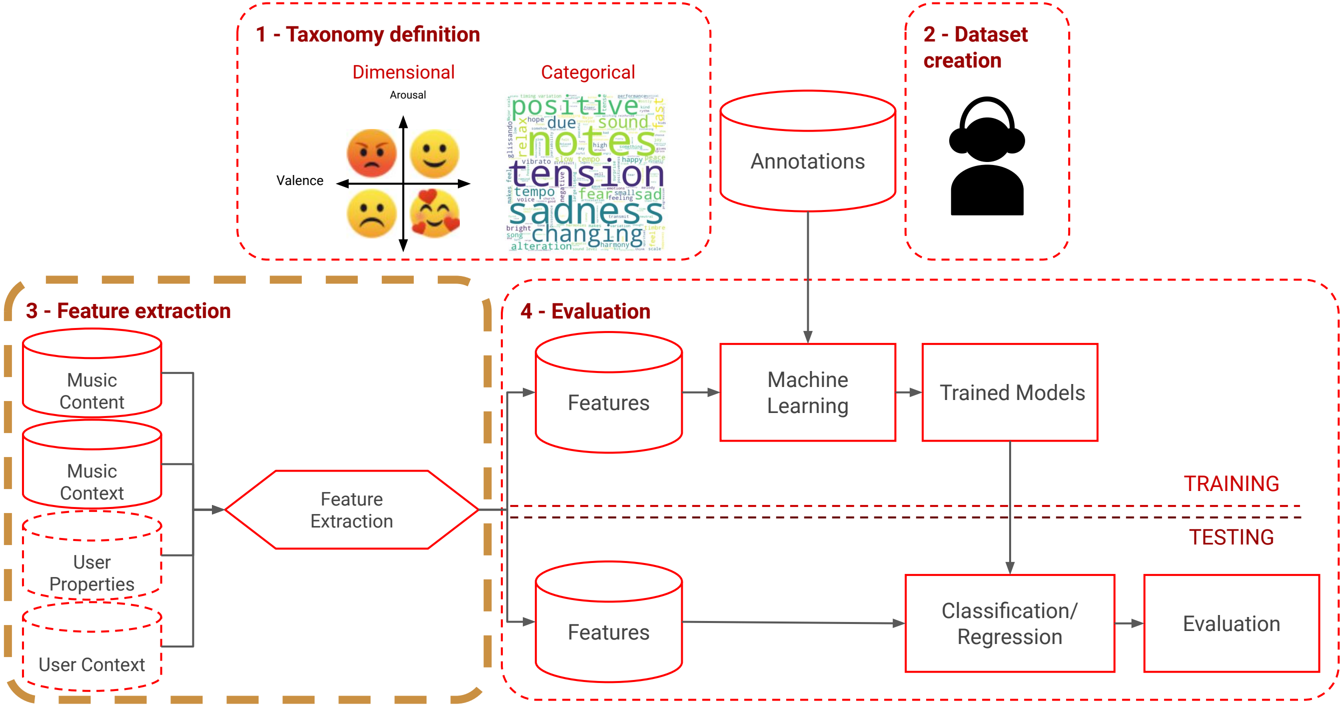
- Listening tests
- Annotation
- Cognitive load



Traditional music emotion recognition systems

Step 3:

- Music content
- Music context
- User properties
- User context

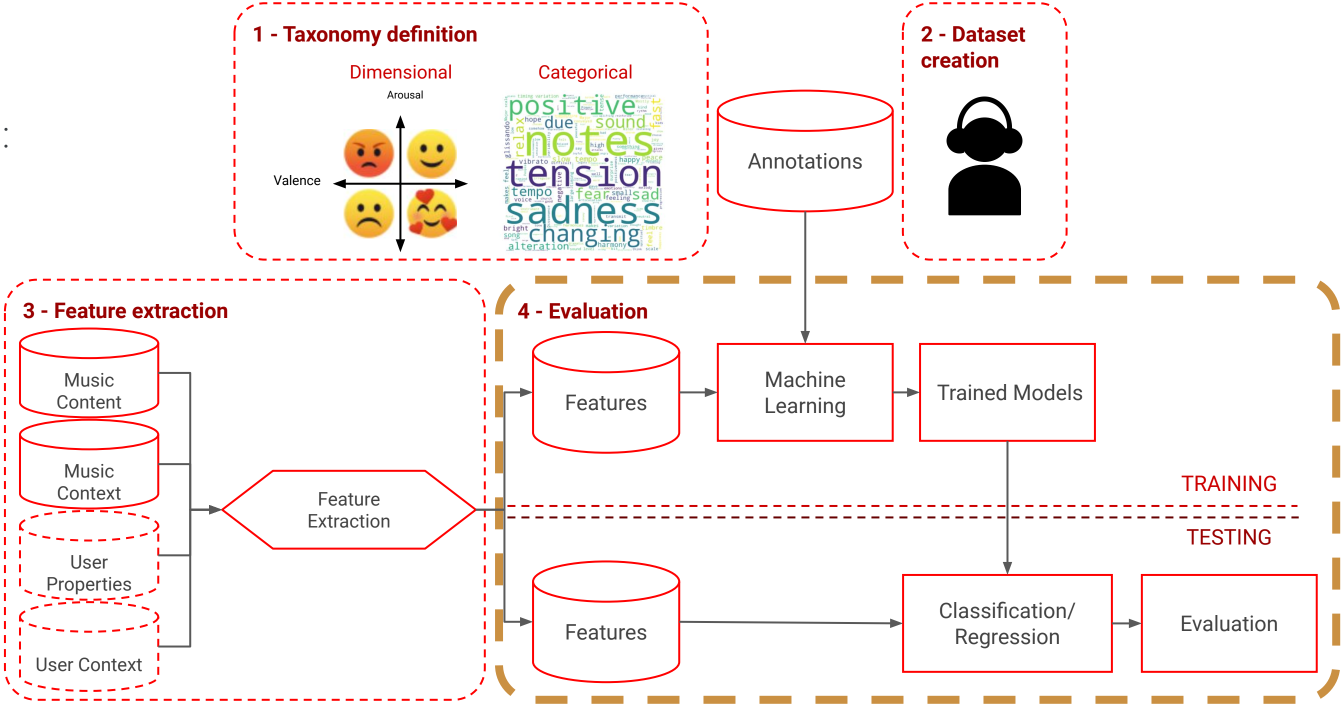


Reference:
Schedl, M., Flexer, A., Urbano, J. **The neglected user in music information retrieval research.** Journal of Intelligent Information Systems, 41, pp. 523-539, 2013.

Traditional music emotion recognition systems

Step 4:

- Assemble + split
- Taxonomy → approach:
  - D: Regression
  - C: Classification
- Test metrics



Problematic demo

For whom?

What for?

Problematic demo

For whom?

What for?

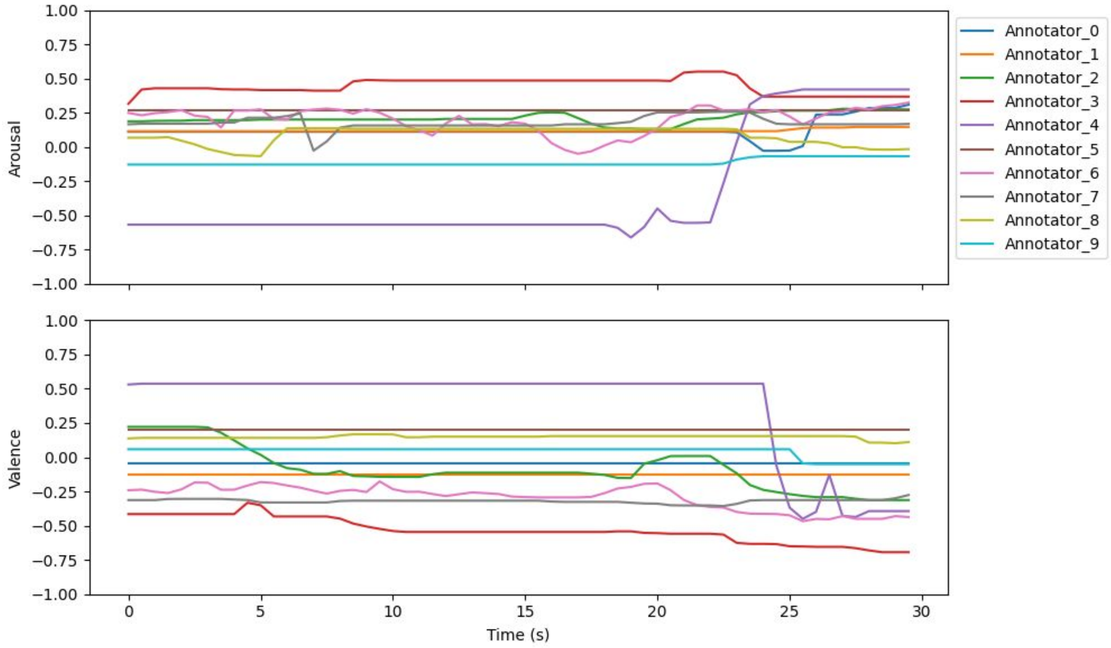
Reference:
Aljanaki, A., Yang, Y.-H, Soleymani, M. **Developing a benchmark for emotional analysis of music.** PLoS One, 12(3), pp. 1-22, 2017.

Problematic demo

For whom?

What for?

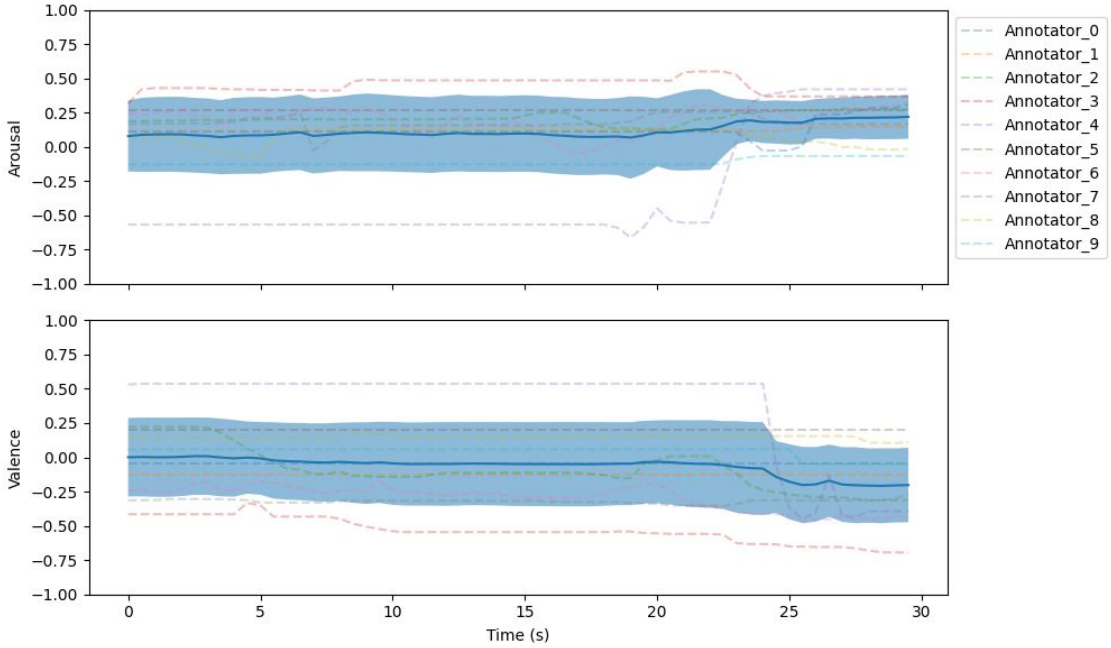
Arousal, Valence ⊂ [-1, 1]



Problematic demo

For whom?

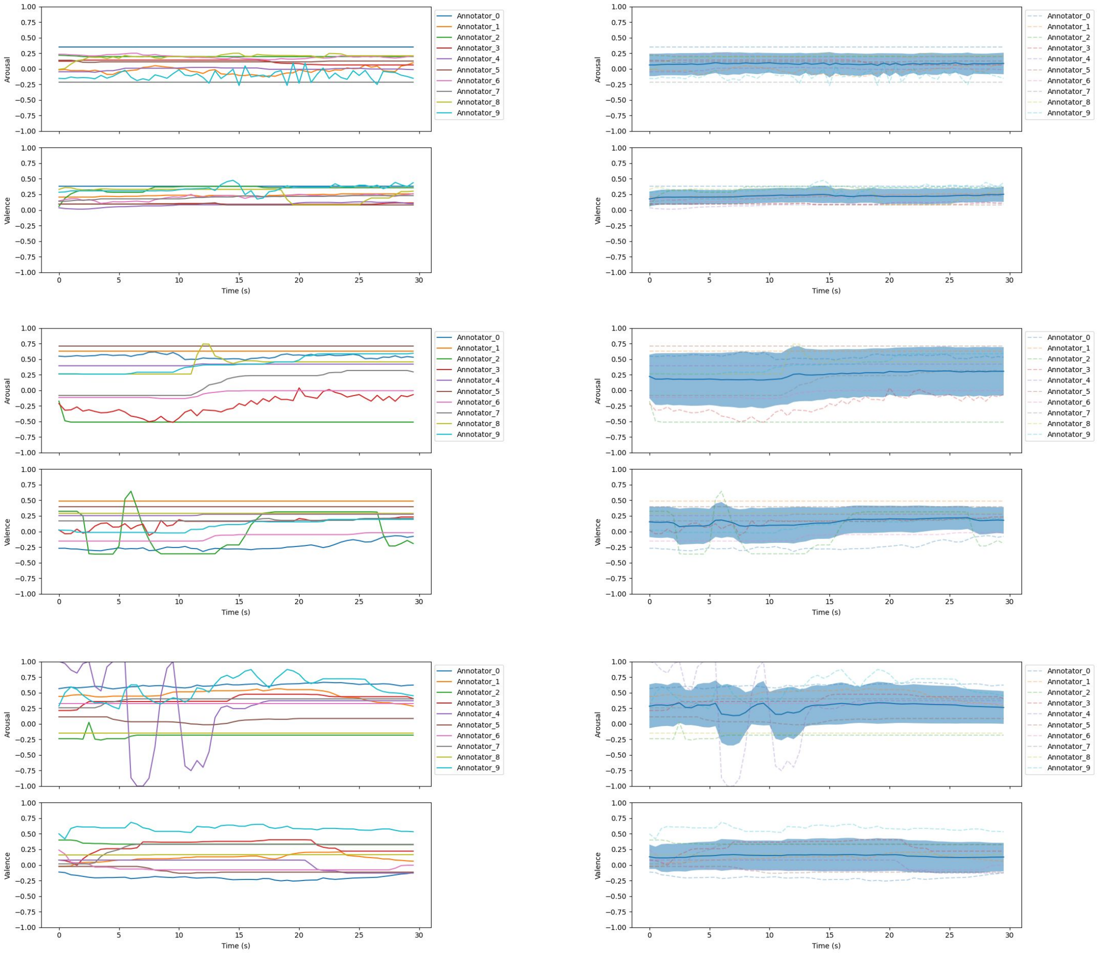
What for?



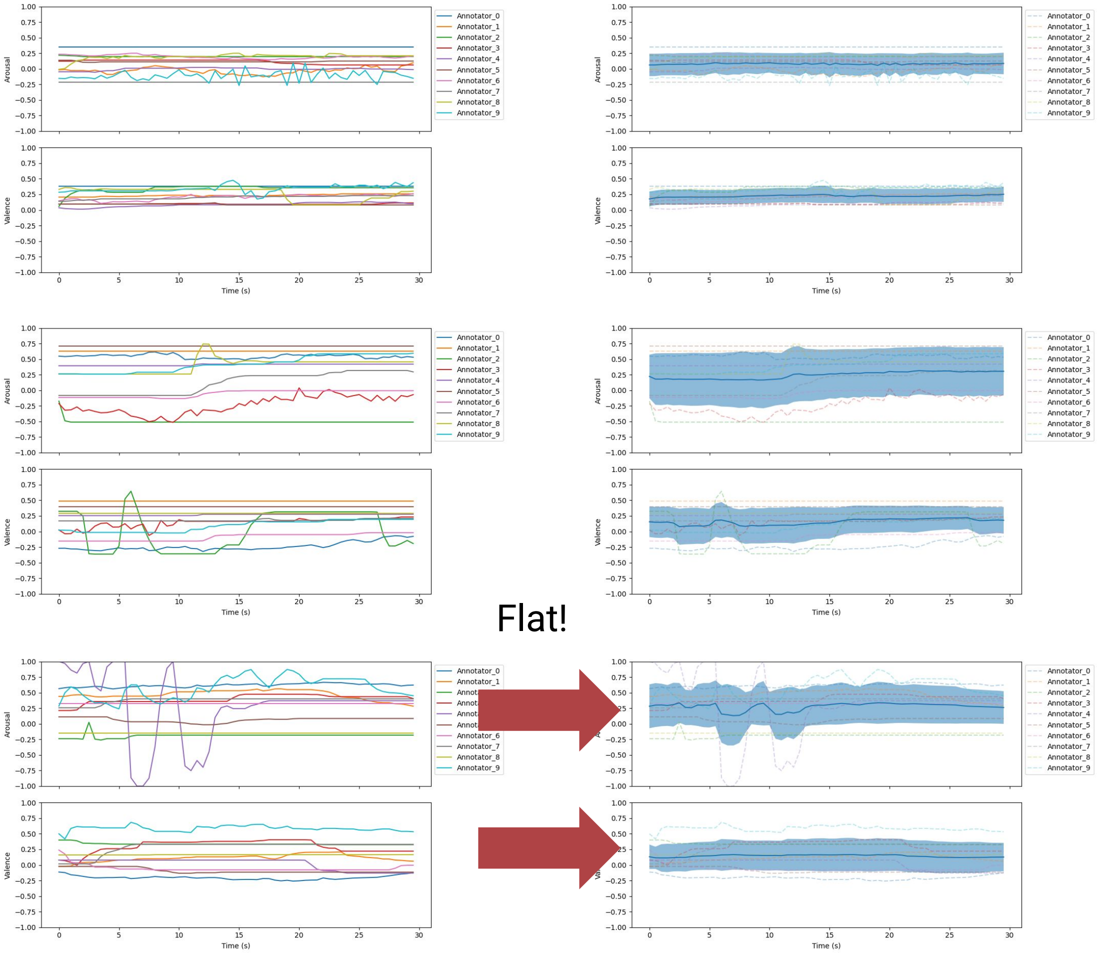
Arousal μ=0.12, $\boldsymbol{\sigma}$= 0.25   Valence μ=-0.06, $\boldsymbol{\sigma}$=0.29

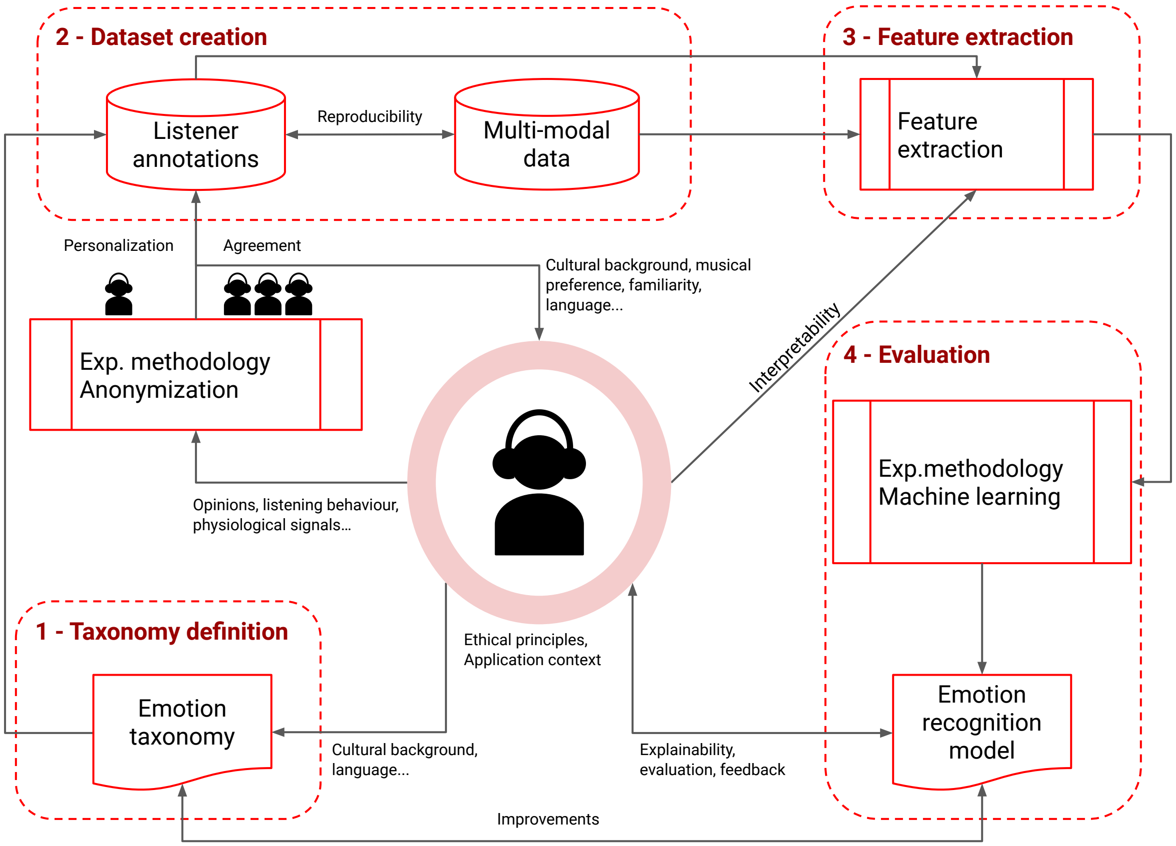
Problematic demo

A system that is 100% accurate is merely predicting the mean of an aggregated annotation!



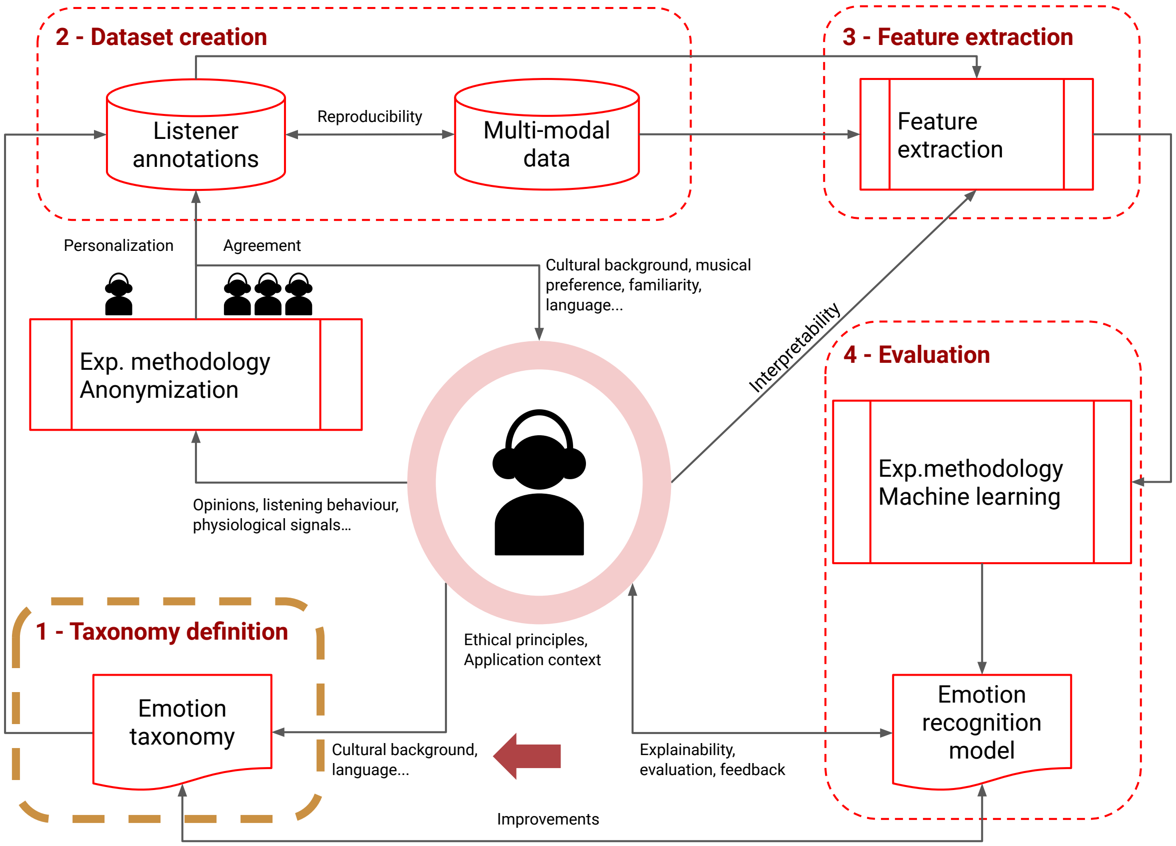
A system that is 100% accurate is merely predicting the mean of an aggregated annotation!



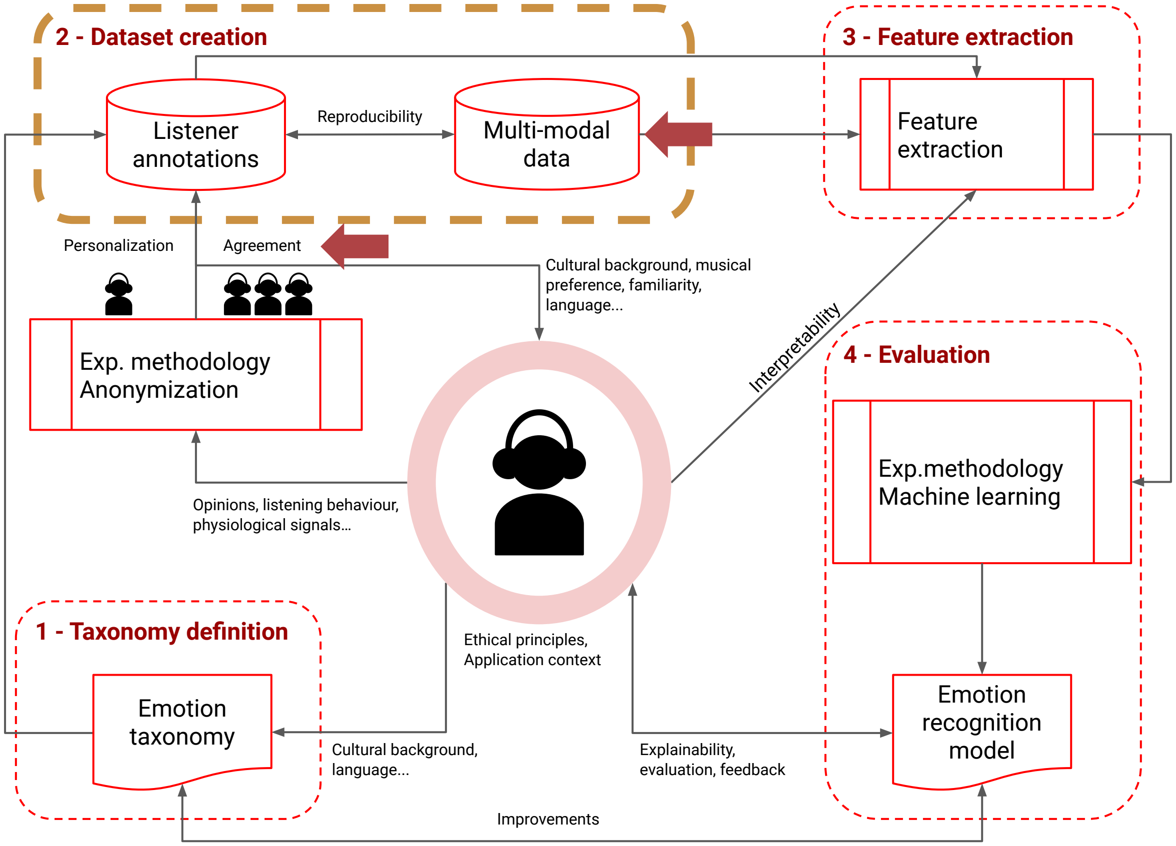


2 - Dataset creation
Listener annotations
Reproducibility
Multi-modal data
3 - Feature extraction
Feature extraction
Personalization
Agreement
Cultural background, musical preference, familiarity, language...
Interpretability
Exp. methodology Anonymization
4 - Evaluation
Exp.methodology Machine learning
Opinions, listening behaviour, physiological signals...
1 - Taxonomy definition
Ethical principles, Application context
Emotion taxonomy
Emotion recognition model
Cultural background, language...
Explainability, evaluation, feedback
Improvements

Human-centered music emotion recognition systems

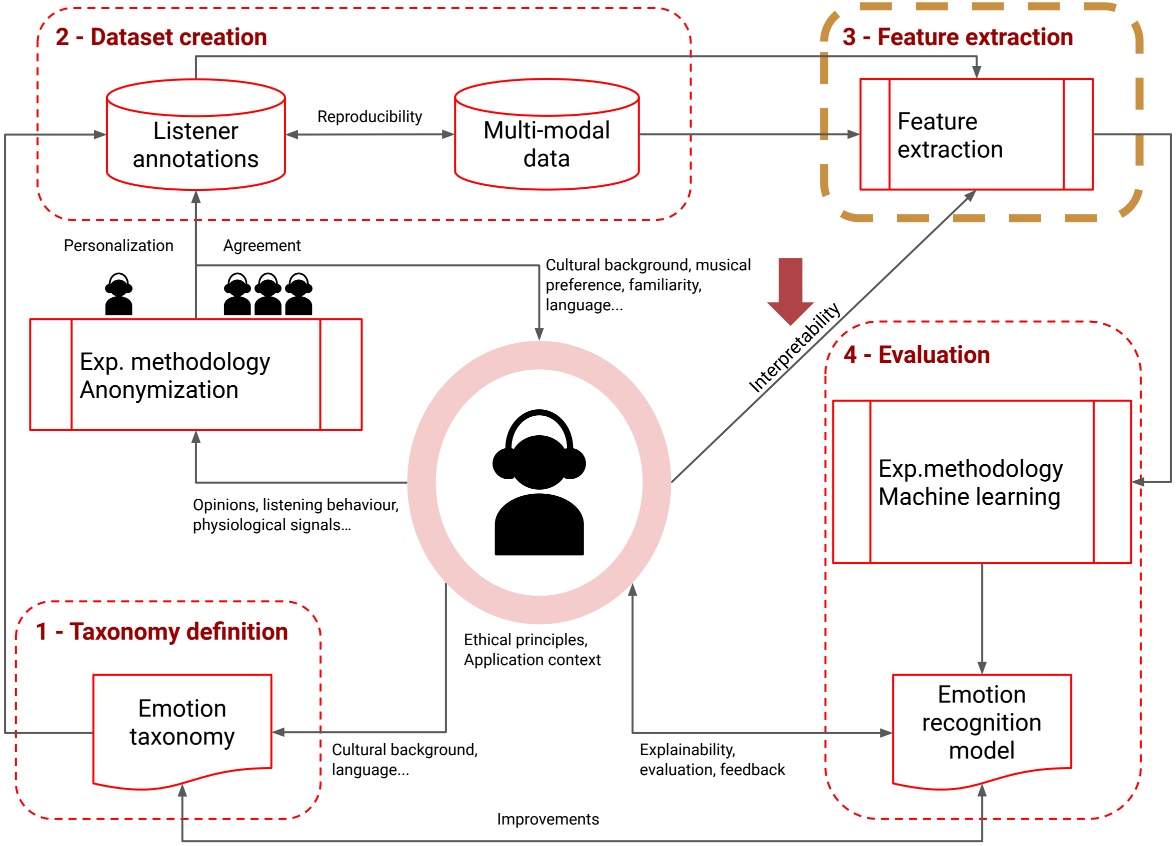


Human-centered music emotion recognition systems

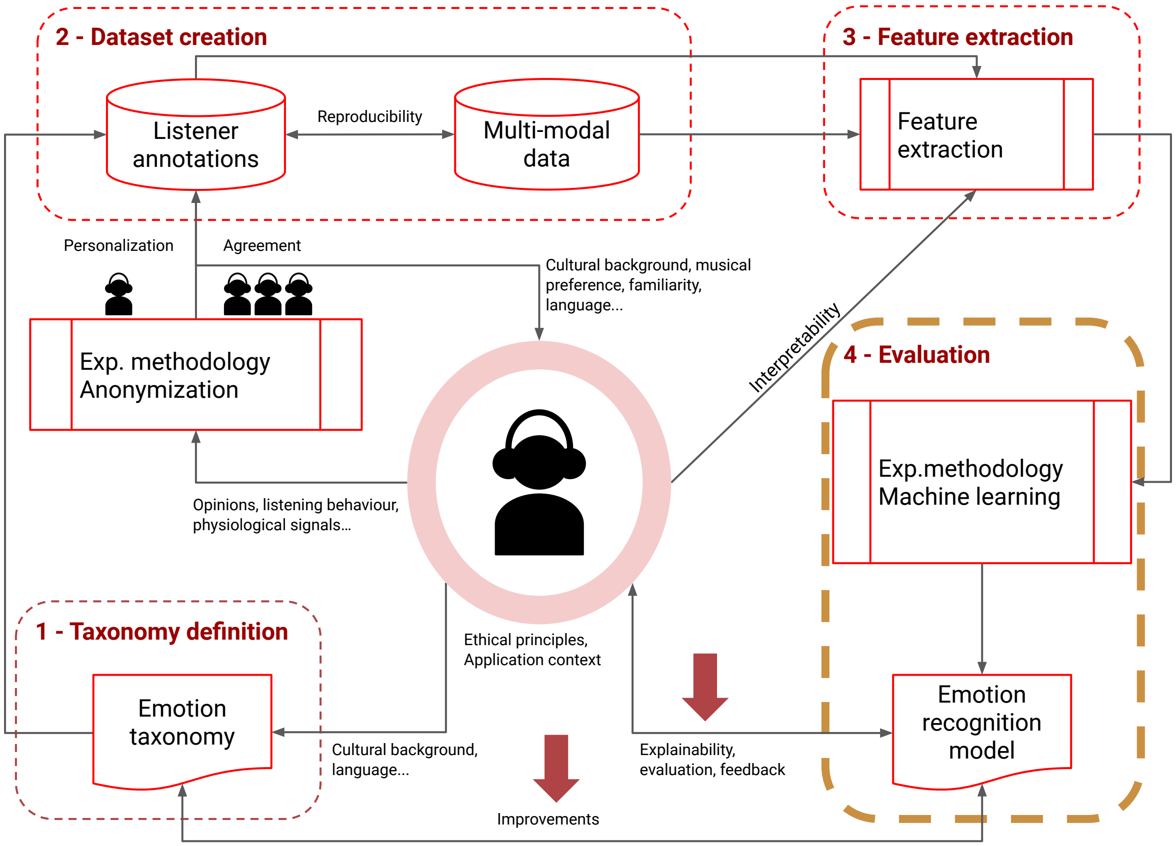


Reference: Kaneshiro, B. et al. **Naturalistic music EEG datasets.** Stanford University, 2016-2021.

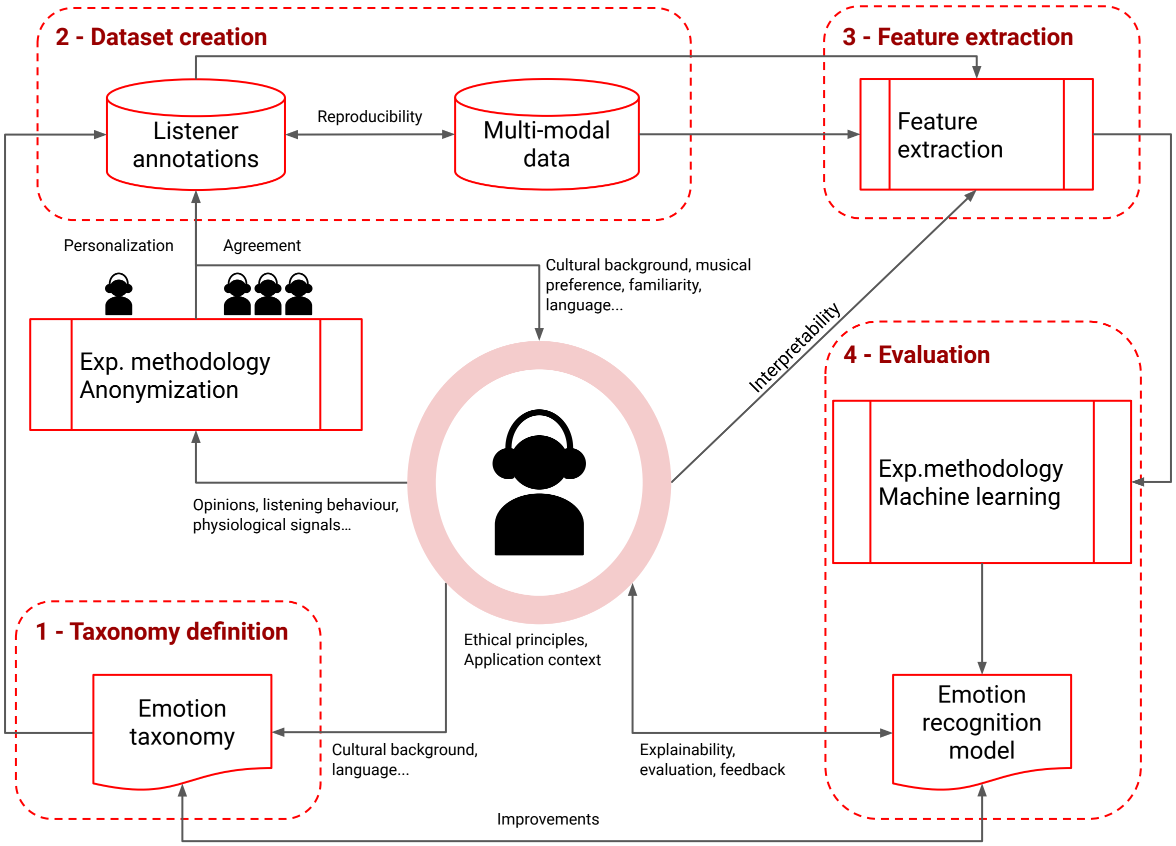
Human-centered music emotion recognition systems



Human-centered music emotion recognition systems



Our contributions...



Our contributions...

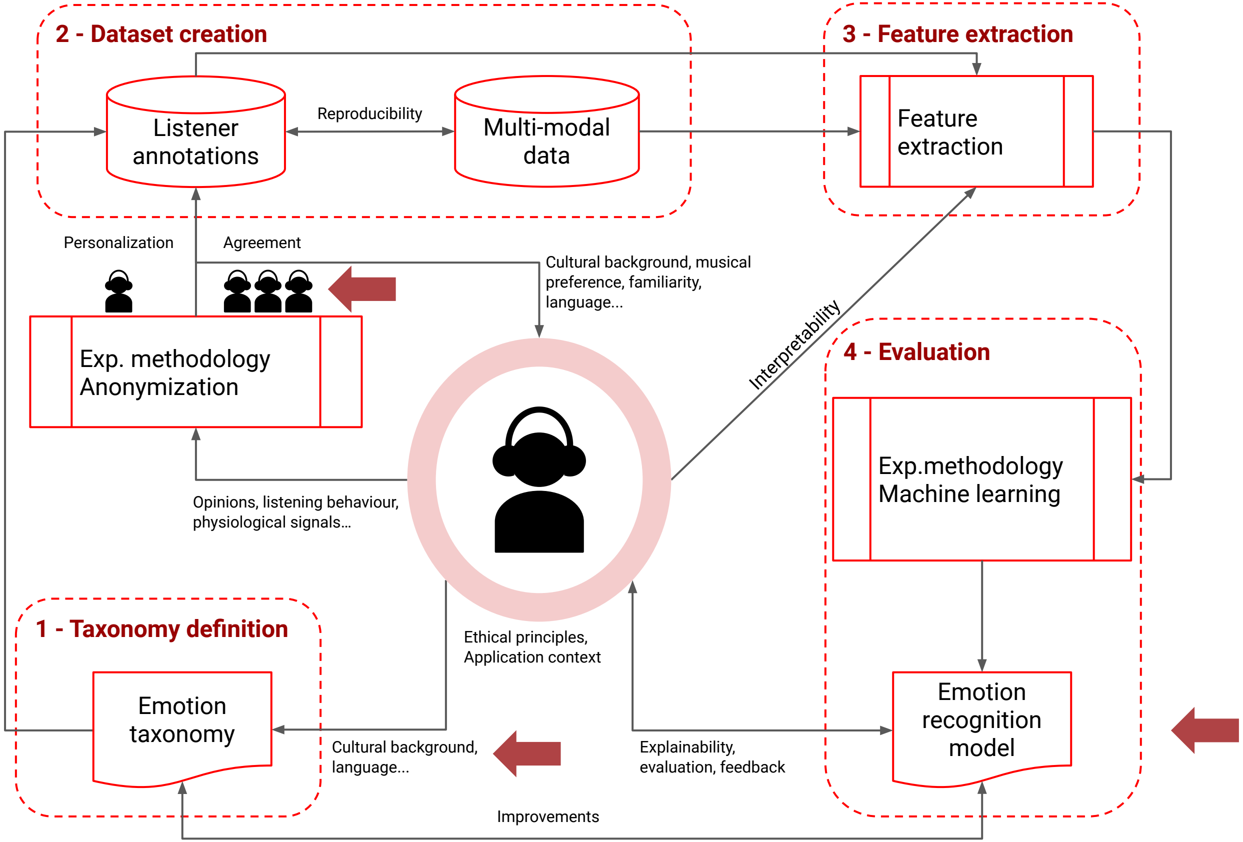
3

4

5

6

7



Our contributions...

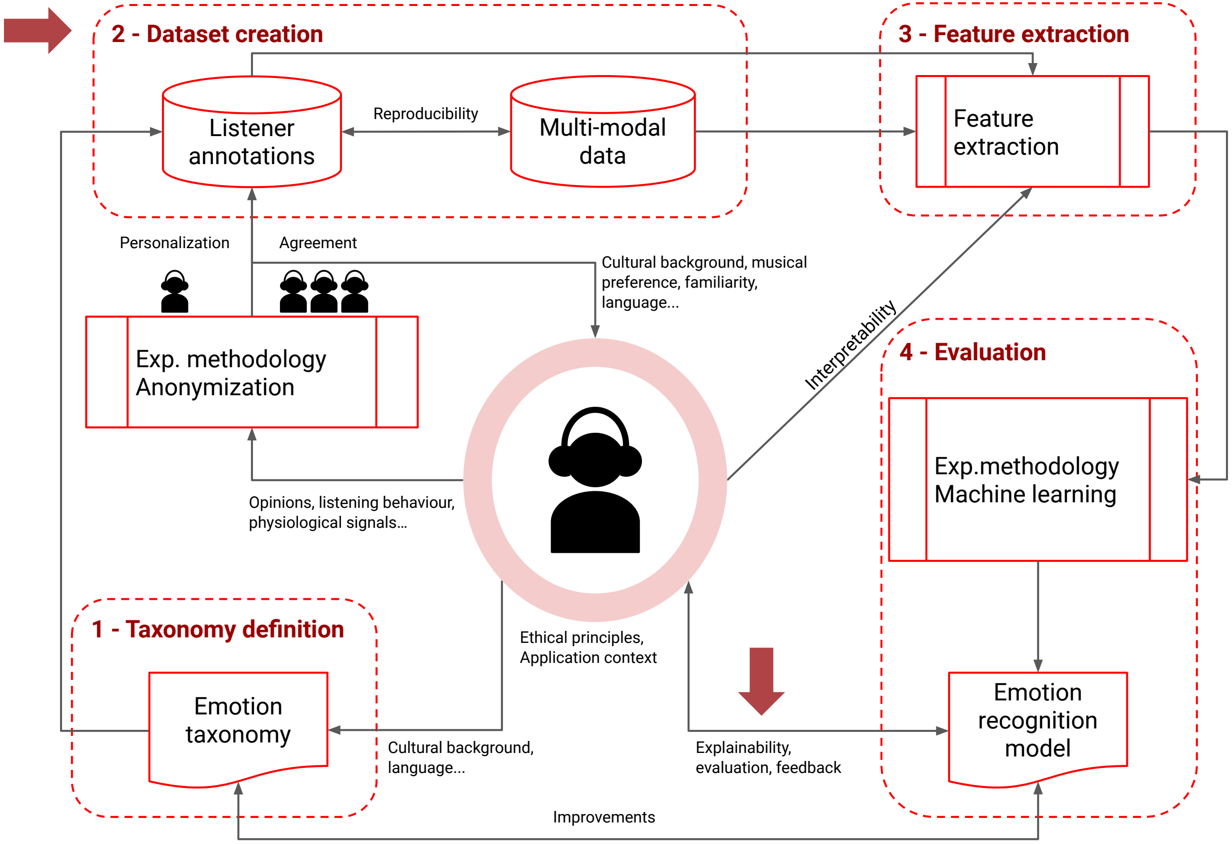
3

4

5

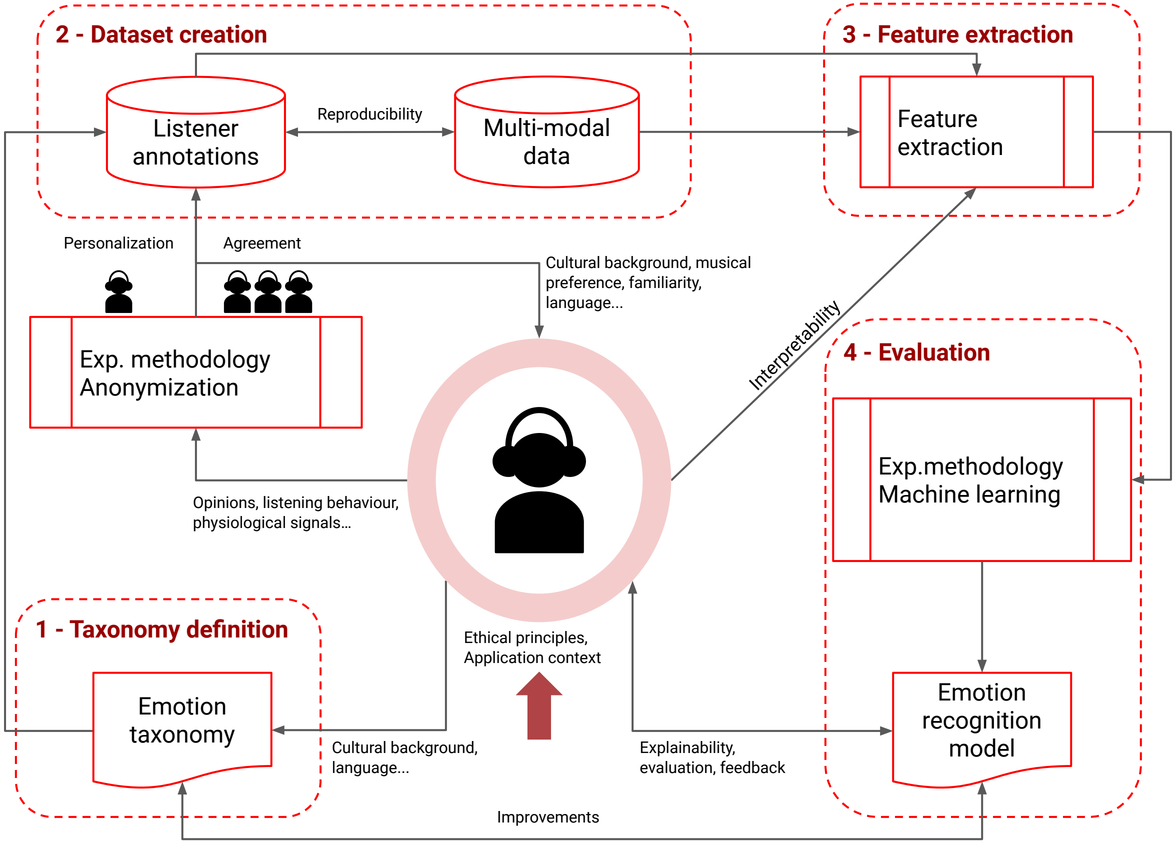
6

7



Our contributions...

3
4
5
6
7



3 Impact of language on emotion annotation

Publication:

Gómez-Cañón, Cano, Herrera & Gómez. ***Joyful for you and tender for us: the influence of individual characteristics and language on emotion labeling and classification***. Proceedings of ISMIR 2020, pp. 853-860.

https://github.com/juansgomez87/agreement-emotion

Annotation analysis and group-based MER

Motivation

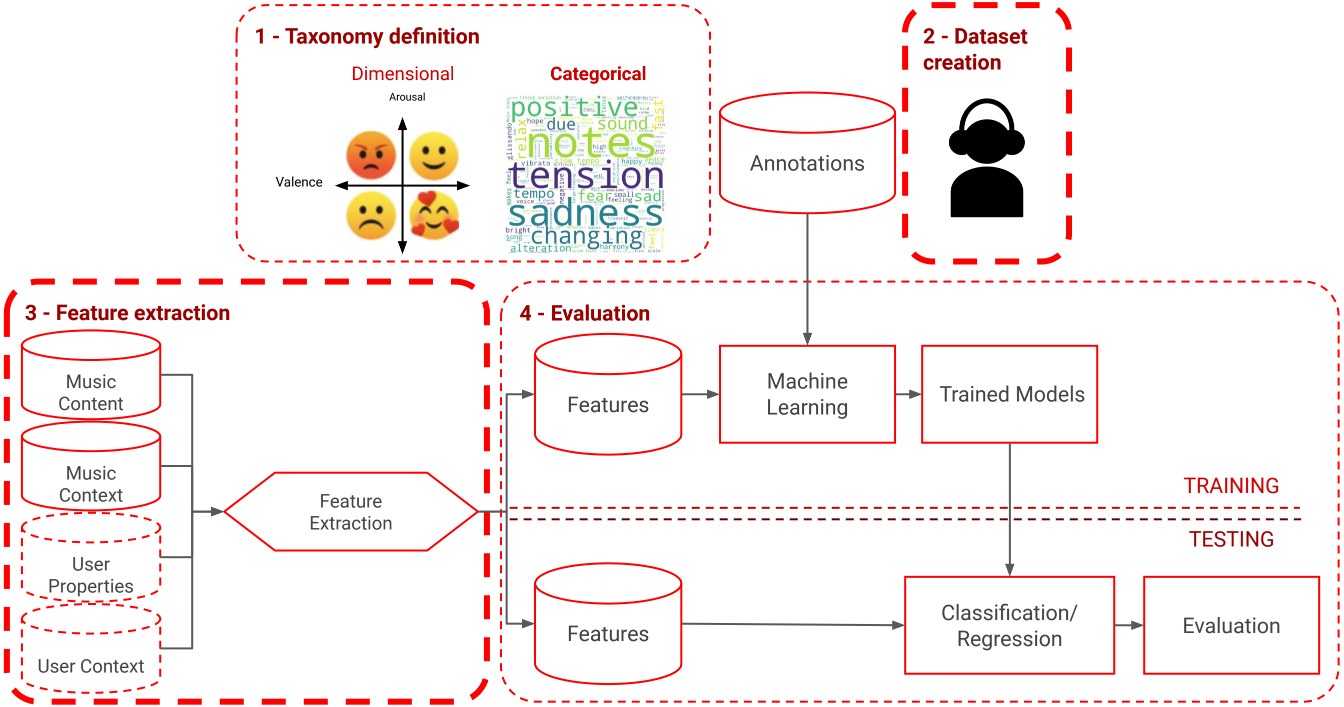
- Subjectivity is a complex issue (Schedl et al, 2018) ➞ Eroica symphony
- Pop and rock music - agreement by language?
- Group-based MER

Research questions

- Differences in annotation?
- Can we improve MER?

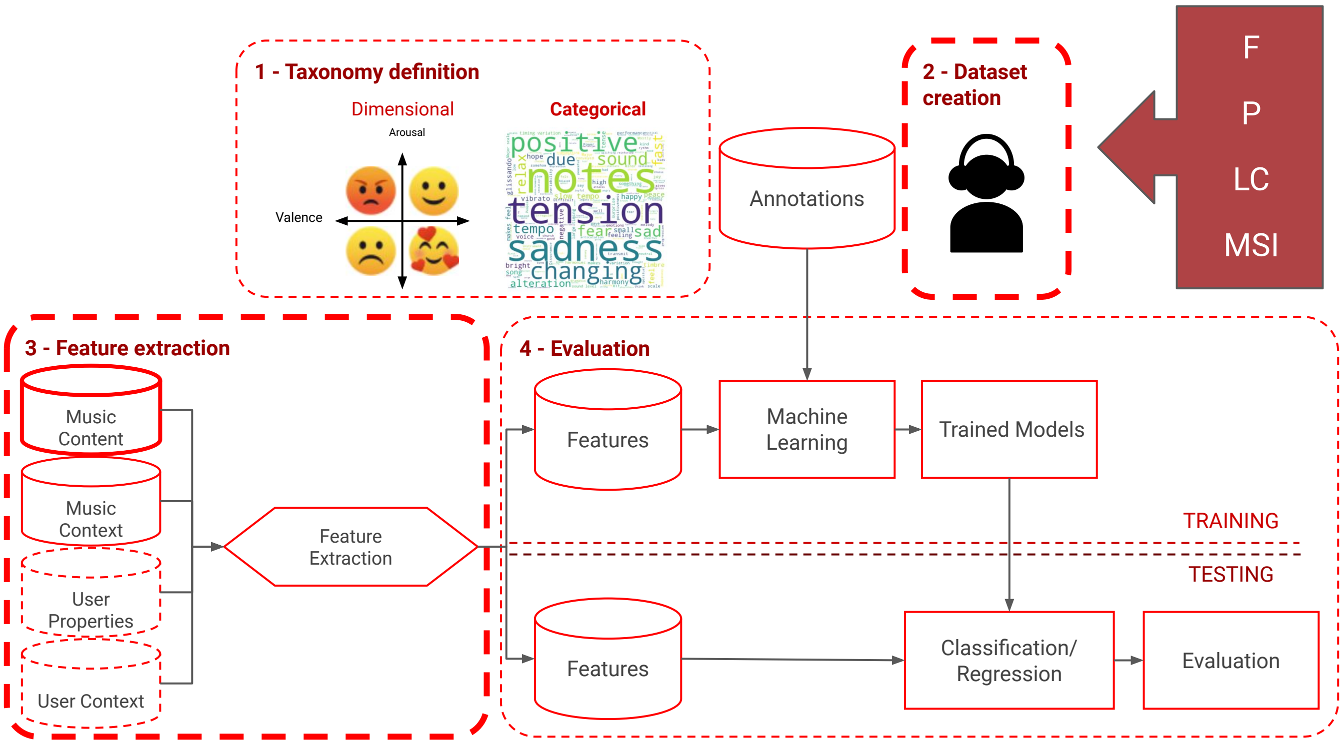
Reference:
Schedl, M. et al. **On the interrelation between listener characteristics and the perception of emotions in classical orchestra music.** IEEE Transactions on Affective Computing, 9(4), pp. 507-525, 2018.

Annotation analysis and group-based MER



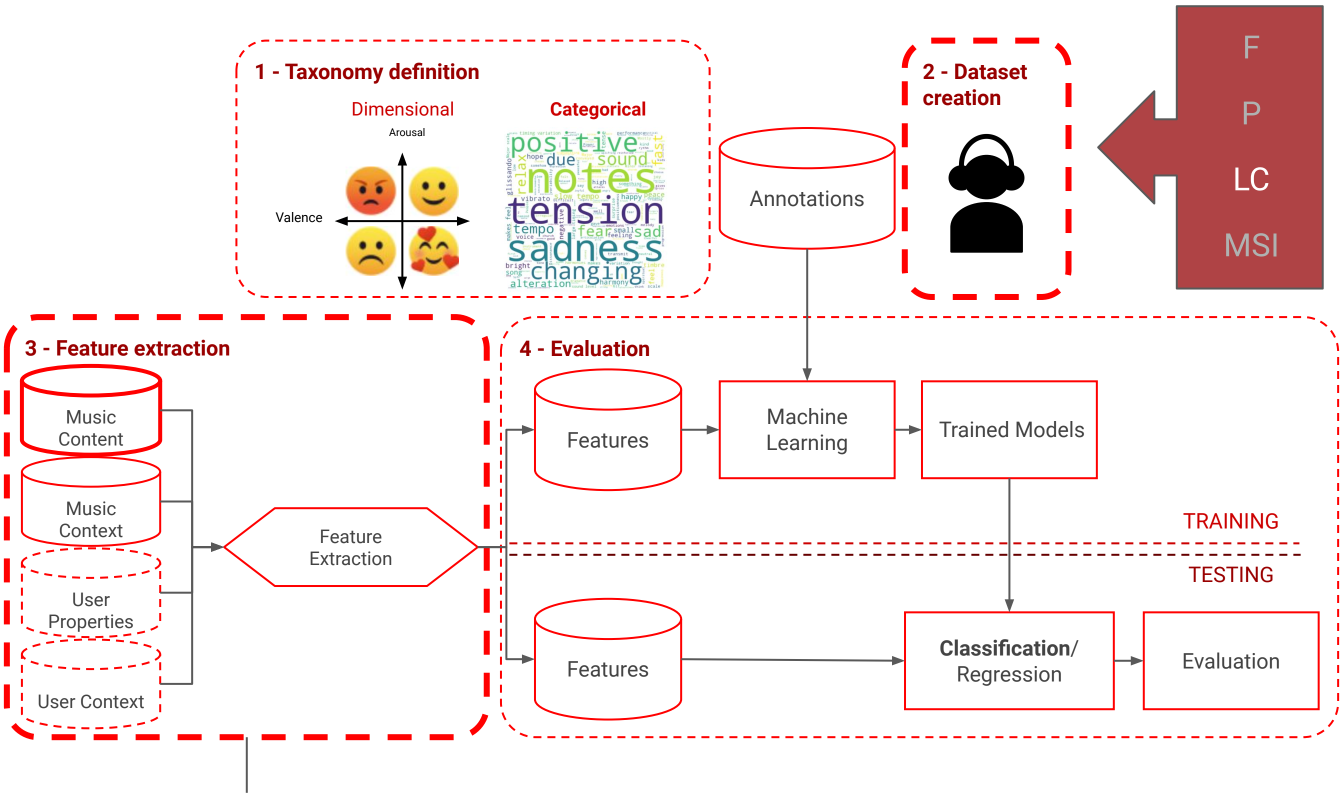
- Annotations:
  - 22 excerpts
  - 11 categories
  - 4 languages (English, Spanish, German, Mandarin)

Annotation analysis and group-based MER



- Annotations:
  - 22 excerpts
  - 11 categories
  - 4 languages
- Inter-rater reliability:
  - 30k+ annotations
  - Krippendorff's α
  - Groups

Annotation analysis and group-based MER



Fixed

- Annotations:
  - 22 excerpts
  - 11 categories
  - 4 languages
- Inter-rater reliability:
  - 30k+ annotations
  - Krippendorff's α
  - Groups
- Evaluation:
  - Clusterability
  - Group and classify
  - Support vector machines

Overall low inter-rater agreement

- $0.05 < \boldsymbol{\alpha} < 0.58$

Research questions

- Differences in annotation?
- Can we improve MER?

Research questions

- Differences in annotation?
- Can we improve MER?

Overall low inter-rater agreement

- $0.05 < \boldsymbol{\alpha} < 0.58$

Significant differences

- Emotional annotations vary across languages (Jackson et al., 2019)
- Group-based annotations are more similar amongst them

Reference:
Jackson, J.C. et al. **Emotion semantics show both cultural variation and universal structure.** Science, 1522, pp. 1517-1522, 2019.

Results

Research questions

- Differences in annotation?
- Can we improve MER?

Overall low inter-rater agreement

- 0.05 < α < 0.58

Significant differences

- Emotional annotations vary across languages (Jackson et al., 2019)
- Group-based annotations are more similar amongst them

Multi-label and group-based classification

- Up to 18 percentage points improvement in F1-scores
- Group-based models < general models
  - Except lyrics comprehension!

Reference:
Jackson, J.C. et al. **Emotion semantics show both cultural variation and universal structure.** Science, 1522, pp. 1517-1522, 2019.

4 An attempt for language-sensitive MER

Publications:

Gómez-Cañón, Cano, Herrera & Gómez. ***Transfer learning from speech to music: towards language-sensitive emotion recognition models***. Proceedings of EUSIPCO 2020, pp. 136-140.

https://github.com/juansgomez87/quad-pred

Gómez-Cañón, Cano, Pandrea, Herrera & Gómez. ***Language-sensitive music emotion recognition models: are we really there yet?*** Proceedings of ICASSP 2021, pp. 576-580.

https://github.com/juansgomez87/lang-sens-mer

Motivation

- Speech as source of data (Coutinho & Schuller, 2017)
- If language is important, can we use it somehow?

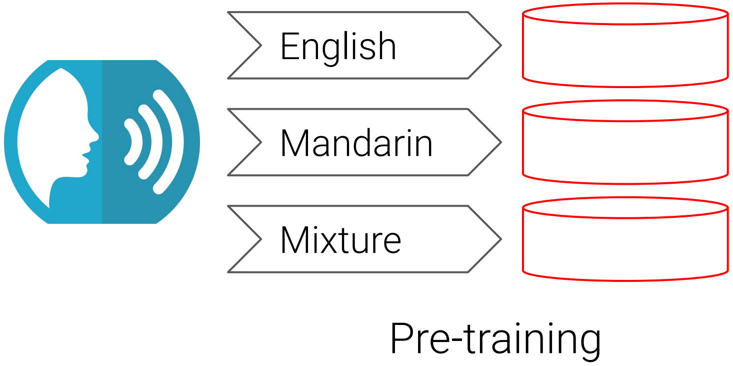
Research question

- Transfer learning to create language-sensitive models?

Reference:
Coutinho, E., & Schuller, B. **Shared acoustic codes underlie emotional communication in music and speech - evidence from deep learning.** PLoS One, 12(6), 2017.

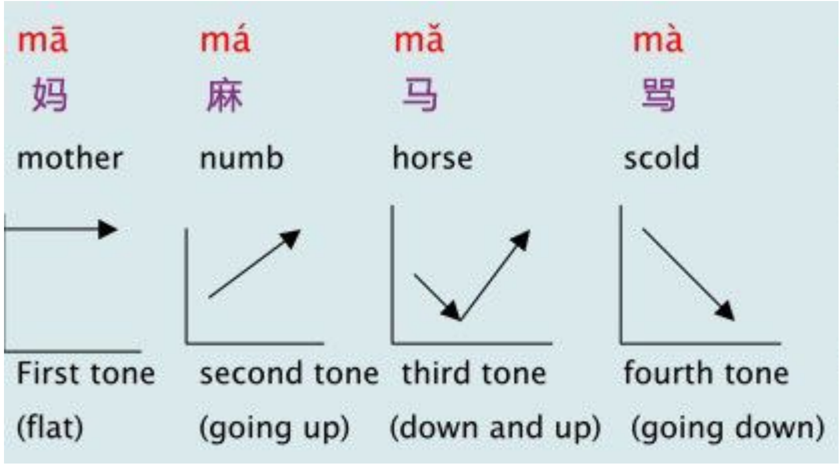
Transfer learning from speech to music

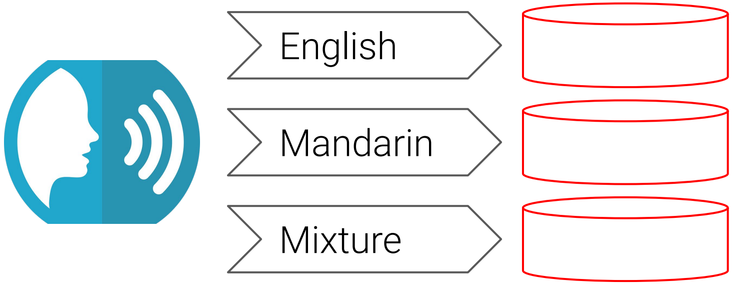
- Pretraining with speech:
  - English
  - Mandarin
  - Mixture 50/50



Transfer learning from speech to music

- Pretraining with speech:
  - English
  - Mandarin **(tonal language)**
  - Mixture 50/50

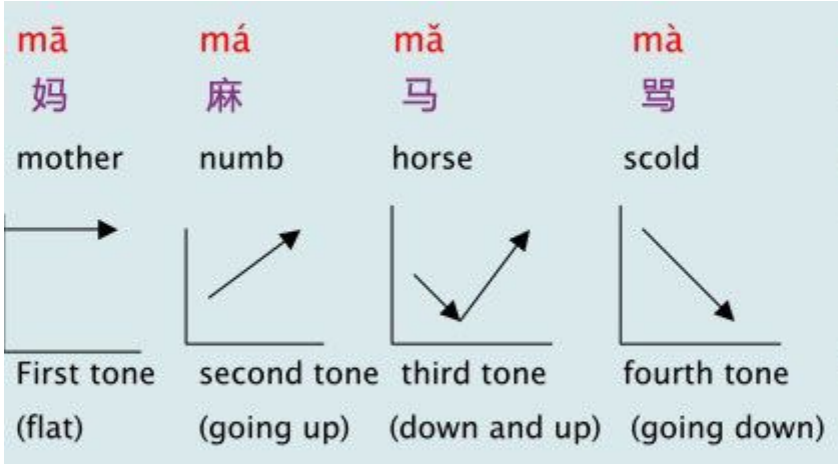


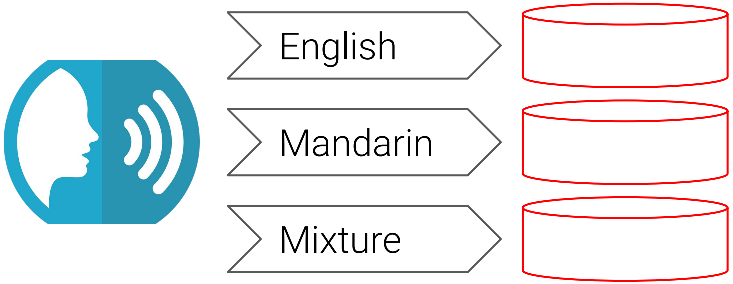


Pre-training

Transfer learning from speech to music

- Pretraining with speech:
  - English
  - Mandarin **(tonal language)**
  - Mixture 50/50



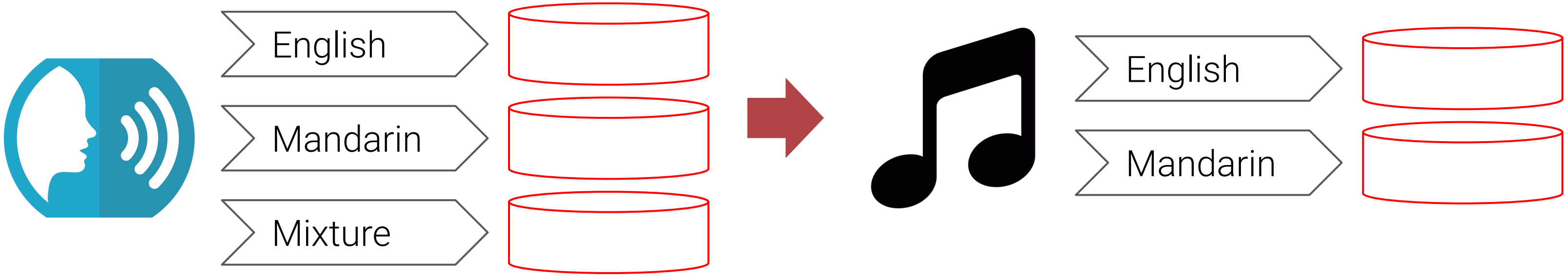


Pre-training

Reference:
Purves, D. **Music as biology: the tones we like and why.** Harvard University Press, 2017.

Transfer learning from speech to music

- Pretraining with speech (U, S):
  - English
  - Mandarin
  - Mixture 50/50
- Fine-tune on music (TL):
  - English
  - Mandarin

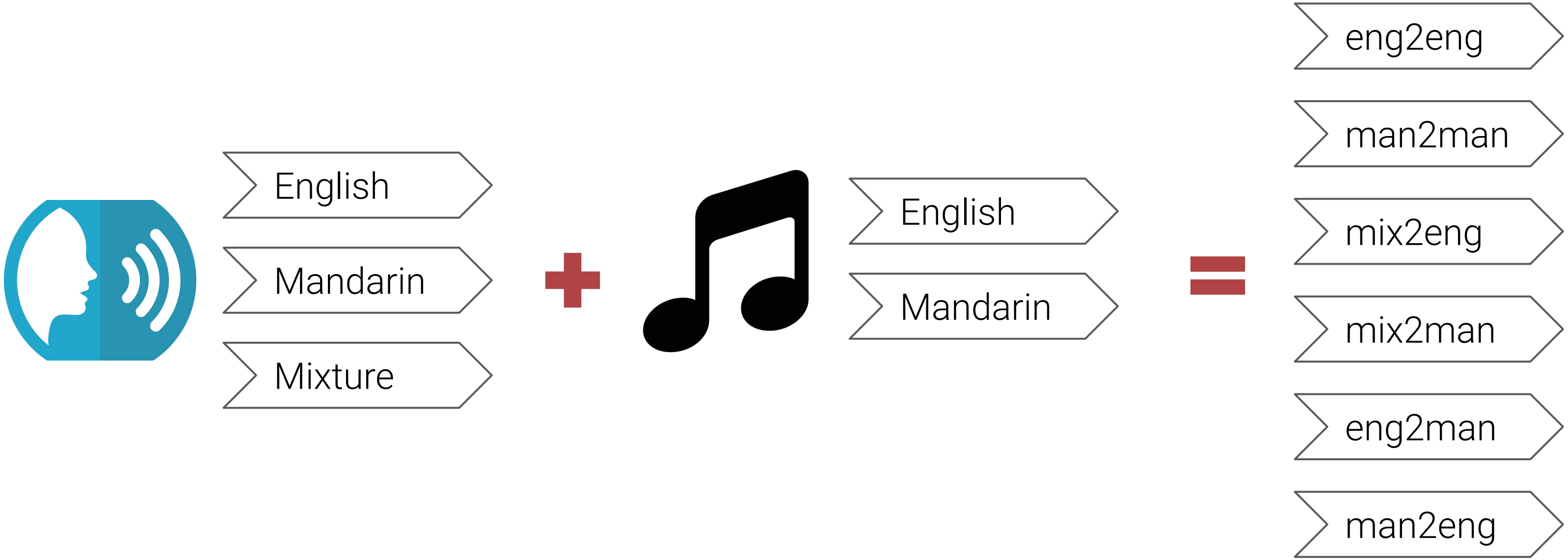


Pre-training

Transfer Learning

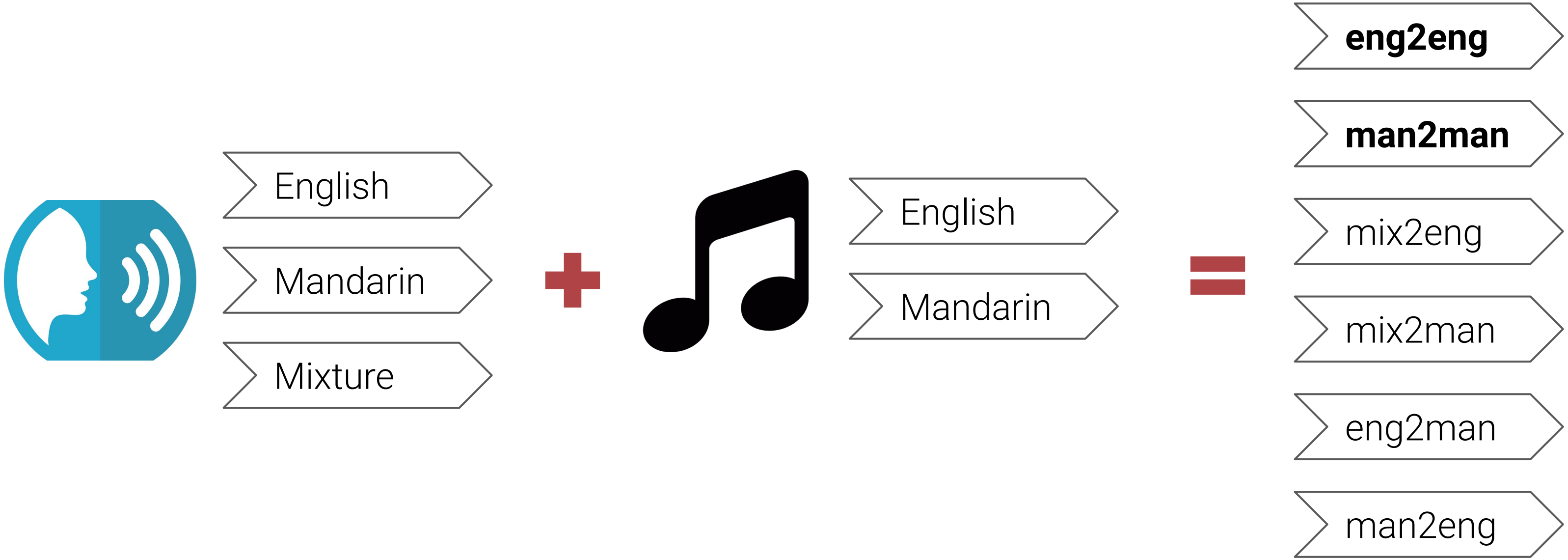
Transfer learning from speech to music

- Two general settings:
  - Intra-linguistic
  - Cross-linguistic



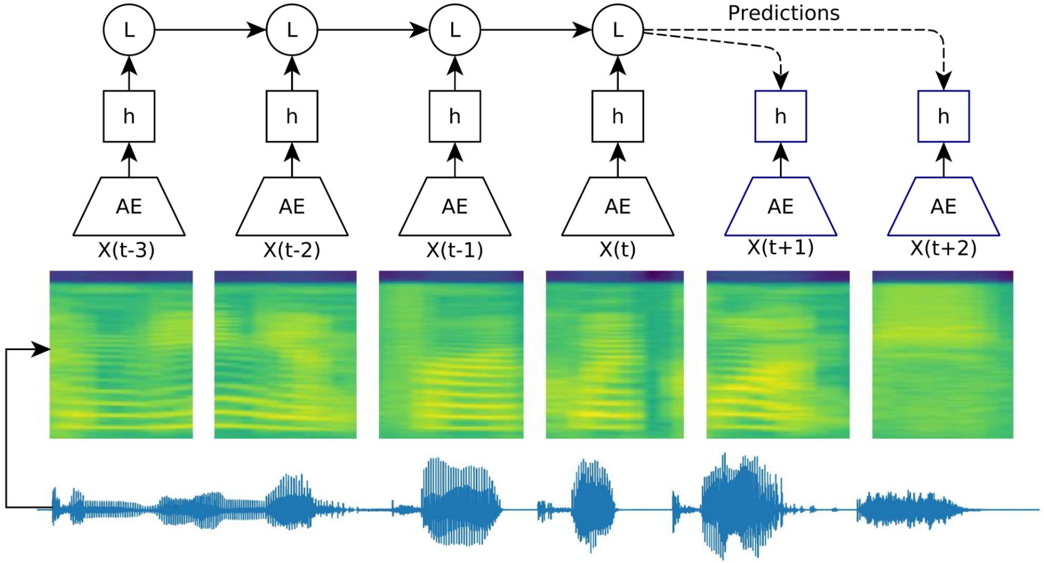
Transfer learning from speech to music

- Two general settings:
  - **Intra-linguistic**
  - Cross-linguistic



Un-, self-supervised learning

- Sparse Convolutional Autoencoder
  - SCAE Feature Extractor
  - SCAE Full
- Contrastive Predictive Coding
  - CPC Feature Extractor
  - CPC Full
- Multi-task learning



Results

No meaningful features are transferred

- Eng2eng and man2man do not outperform other models

Diversity in pre-training data

- Mix2eng and mix2man improve performance

Research questions

- Transfer learning to create language-sensitive models?

5 Active learning for personalized MER

Publication:

Gómez-Cañón, Cano, Yang, Herrera & Gómez. ***Let's agree to disagree: consensus entropy active learning for personalized music emotion recognition***. Proceedings of ISMIR 2021, pp. 237-245.

https://github.com/juansgomez87/consensus-entropy

Motivation

- Language-sensitive models 🤔

Motivation

- Language-sensitive models 🤔
- Group-based MER ⇒ personalized MER
  - Use of active learning (Su & Fung, 2012)

Reference:
Su, D. & Fung, P. **Personalized music emotion classification via active learning.** ACM workshop on MIR with user-centered and multi-modal strategies, 2012.

Personalization and music emotion recognition

Motivation

- Language-sensitive models 🤔
- Group-based MER ⇒ personalized MER
  - Use of active learning (Su & Fung, 2012)

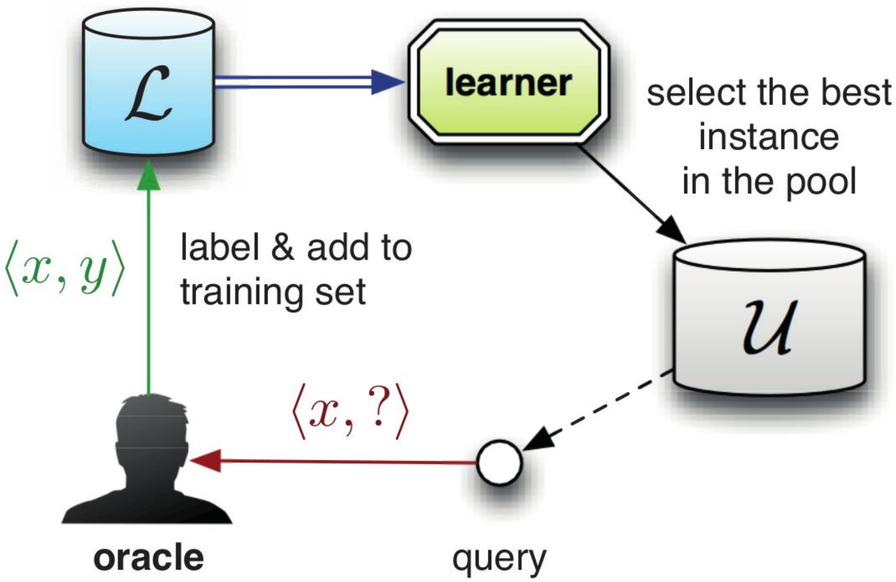
Research questions

- Can agreement be used as input?
- Which ML algorithms can be personalized?

Reference:
Su, D. & Fung, P. **Personalized music emotion classification via active learning.** ACM workshop on MIR with user-centered and multi-modal strategies, 2012.

Active learning

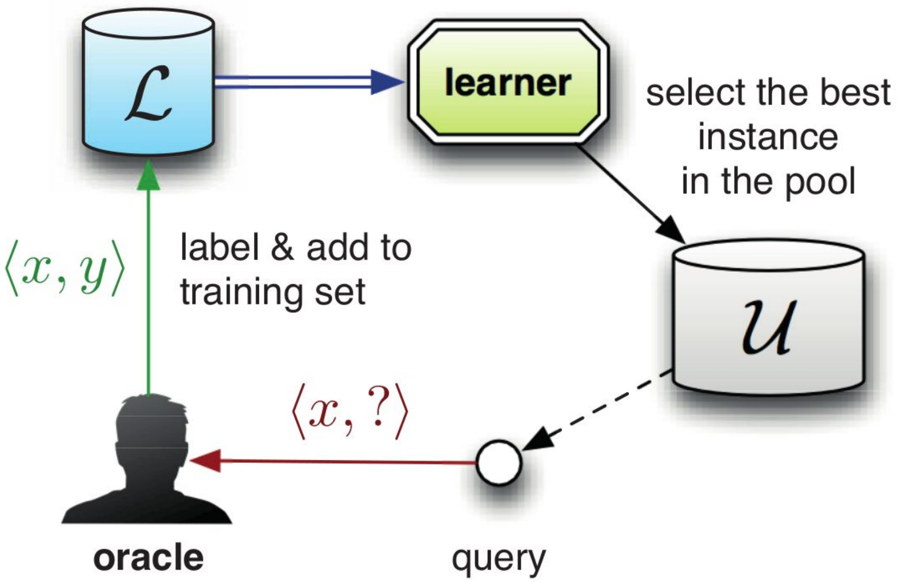
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling



Settles, B. **Active learning.** Morgan and Claypool publishers, 2012.
Aggarwal et al., **Active learning: a survey**. Data classification algorithms and applications, CRC Press, 2014.
Zhang et al., **Learning from crowdsourced labeled data: a survey**. Artificial Intelligence review, 2016.

Active learning

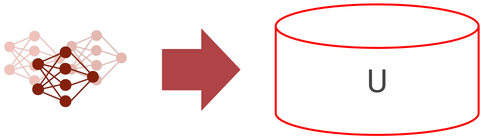
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples



Settles, B. **Active learning.** Morgan and Claypool publishers, 2012.
Aggarwal et al., **Active learning: a survey**. Data classification algorithms and applications, CRC Press, 2014.
Zhang et al., **Learning from crowdsourced labeled data: a survey**. Artificial Intelligence review, 2016.

Active learning

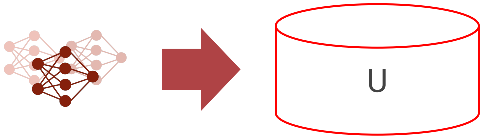
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - **Classifiers (MC)**
  - Humans (HC)
  - Hybrid (MIX)
  - Random baseline



	Output probability	Entropy
Song 1		
Song 2		
Song 3		
⋮		
Song n		

Active learning

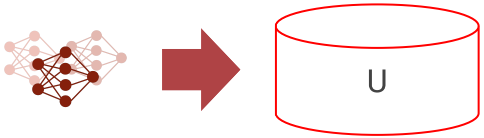
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - **Classifiers (MC)**
  - Humans (HC)
  - Hybrid (MIX)
  - Random baseline



	Output probability	Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2		
Song 3		
⋮		
Song n		

Active learning

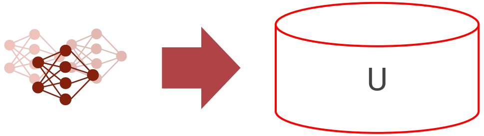
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - **Classifiers (MC)**
  - Humans (HC)
  - Hybrid (MIX)
  - Random baseline



	Output probability	Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2	{Q1 : 0.25, Q2 : 0.25, Q3 : 0.25, Q4 : 0.25}	1.39
Song 3		
⋮		
Song n		

Active learning

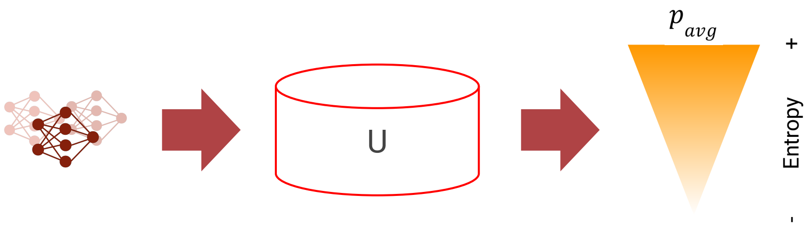
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - **Classifiers (MC)**
  - Humans (HC)
  - Hybrid (MIX)
  - Random baseline



	Output probability	Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2	{Q1 : 0.25, Q2 : 0.25, Q3 : 0.25, Q4 : 0.25}	1.39
Song 3	{Q1 : 0.0, Q2 : 0.5, Q3 : 0.0, Q4 : 0.5}	0.69
⋮		⋮
Song n		0.13

Active learning

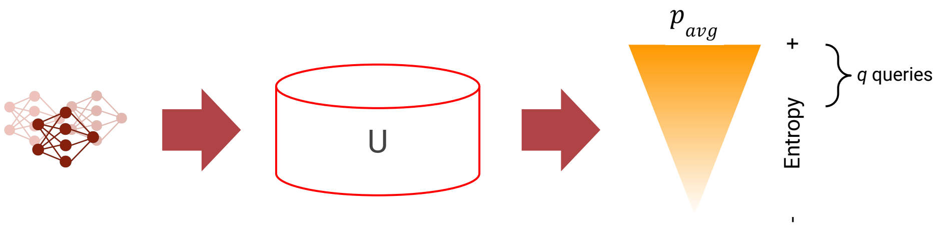
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - **Classifiers (MC)**
  - Humans (HC)
  - Hybrid (MIX)
  - Random baseline



	Output probability	Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2	{Q1 : 0.25, Q2 : 0.25, Q3 : 0.25, Q4 : 0.25}	1.39
Song 3	{Q1 : 0.0, Q2 : 0.5, Q3 : 0.0, Q4 : 0.5}	0.69
⋮		⋮
Song n		0.13

Active learning

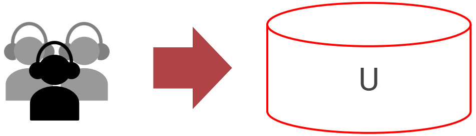
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - **Classifiers (MC)**
  - Humans (HC)
  - Hybrid (MIX)
  - Random baseline



	Output probability	Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2	{Q1 : 0.25, Q2 : 0.25, Q3 : 0.25, Q4 : 0.25}	1.39
Song 3	{Q1 : 0.0, Q2 : 0.5, Q3 : 0.0, Q4 : 0.5}	0.69
⋮		⋮
Song n		0.13

Active learning

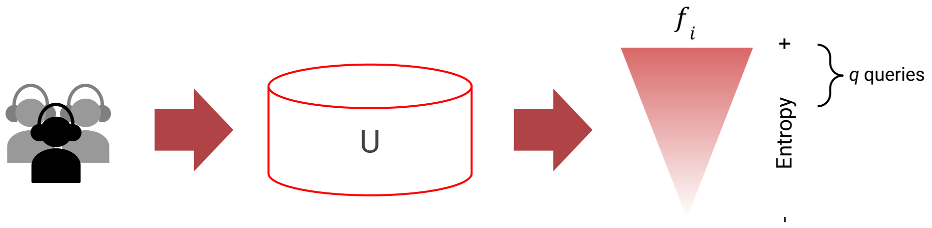
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - Classifiers (MC)
  - **Humans (HC)**
    - **Assumption: Disagreement ➪ Boundaries**
  - Hybrid (MIX)
  - Random baseline



		Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2	{Q1 : 0.25, Q2 : 0.25, Q3 : 0.25, Q4 : 0.25}	1.39
Song 3	{Q1 : 0.0, Q2 : 0.5, Q3 : 0.0, Q4 : 0.5}	0.69
⋮		⋮
Song n		0.13

Active learning

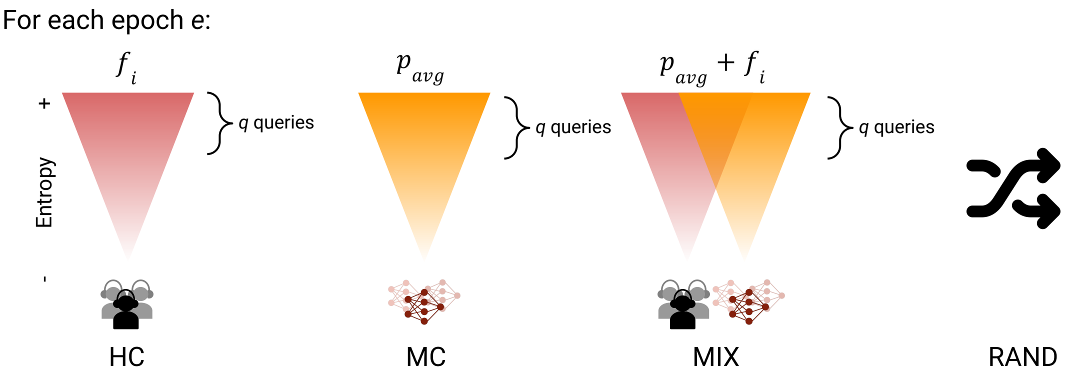
- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - Classifiers (MC)
  - **Humans (HC)**
    - **Assumption: Disagreement ➪ Boundaries**
  - Hybrid (MIX)
  - Random baseline



	Relative frequency	Entropy
Song 1	{Q1 : 1.0, Q2 : 0.0, Q3 : 0.0, Q4 : 0.0}	0
Song 2	{Q1 : 0.25, Q2 : 0.25, Q3 : 0.25, Q4 : 0.25}	1.39
Song 3	{Q1 : 0.0, Q2 : 0.5, Q3 : 0.0, Q4 : 0.5}	0.69
⋮		⋮
Song n		0.13

Active learning

- Consensus entropy (1994):
  - Query-by-committee
  - Uncertainty sampling
- Informative samples:
  - Classifiers (MC)
  - Humans (HC)
  - **Hybrid (MIX)**
  - **Random baseline**



Algorithms

- Probability outputs:
  - Gaussian naive bayes (GNB)
  - Logistic regression (SGD)
  - Extreme gradient boosting (XGB)
  - Short-chunk convolutional neural network (CNN)

Algorithms

- Probability outputs:
  - Gaussian naive bayes (GNB)
  - Logistic regression (SGD)
  - Extreme gradient boosting (XGB)
  - Short-chunk convolutional neural network (CNN) - (Won et al., 2020)

Reference:
Won, M., et al **Evaluation of cnn-based automatic music tagging models.** Proceedings of the 17th Sound and Music Computing conference, 2020.

Algorithms

- Probability outputs:
  - Gaussian naive bayes (GNB) x 5
  - Logistic regression (SGD) x 5
  - Extreme gradient boosting (XGB) x 5
  - Short-chunk convolutional neural network (CNN) x 5
- 20 models per user:
  - Pre-trained to have different outputs

Algorithms

- Probability outputs:
  - Gaussian naive bayes (GNB) x 5
  - Logistic regression (SGD) x 5
  - Extreme gradient boosting (XGB) x 5
  - Short-chunk convolutional neural network (CNN) x 5
- 20 models per user:
  - Pre-trained to have different outputs
- 46 users:
  - More than 150 annotations

Simple agreement can be beneficial

- HC outperforms SOTA ~ 15 percentage points!

Research questions

- Agreement as input?
- Which ML algorithms?

Simple agreement can be beneficial

- HC outperforms SOTA ~ 15 percentage points!
  - SOTA is not great 🙃 mean F-score: 0.35

Research questions

- Agreement as input?
- Which ML algorithms?

Results

Research questions

- Agreement as input?
- Which ML algorithms?

Simple agreement can be beneficial

- HC outperforms SOTA ~ 15 percentage points!
  - SOTA is not great 🙃 mean F-score: 0.35

Not all algorithms work

- GNB is **naive**
- SGD shows **no differences** across methods
- XGB have the **highest amount** of personalized models (HC > MC, HC ≠ RAND)
- CNN has **best performance (HC > ALL)**
  - 40 percentage points for some cases

Results

Simple agreement can be beneficial

- HC outperforms SOTA ~ 15 percentage points!
  - SOTA is not great 🙃 mean F-score: 0.35
  - **Test with new data…**

Not all algorithms work

- GNB is naive
- SGD shows no differences across methods
- **XGB** have the highest amount of personalized models (HC > MC, HC ≠ RAND)
- **CNN** has best performance (HC > ALL)
  - 40 percentage points for some cases

Research questions

- Agreement as input?
- Which ML algorithms?

6 Human-centered data collection for personalized MER

Publication:

Gómez-Cañón, Gutiérrez-Páez, Porcaro, Porter, Cano, Herrera, Gkiokas, Santos, Hernández-Leo, Karreman & Gómez. ***TROMPA-MER: an open dataset for personalized music emotion recognition***. Journal of Intelligent Information Systems, 2022.

https://github.com/juansgomez87/vis-mtg-mer

Human-centered data collection

Motivation

- Consensus entropy works! 😁
  - Imbalanced classes
- Citizen science
- Datasheet to a MER dataset (Gebru et al., 2021)
  - Explain disagreement to researchers!

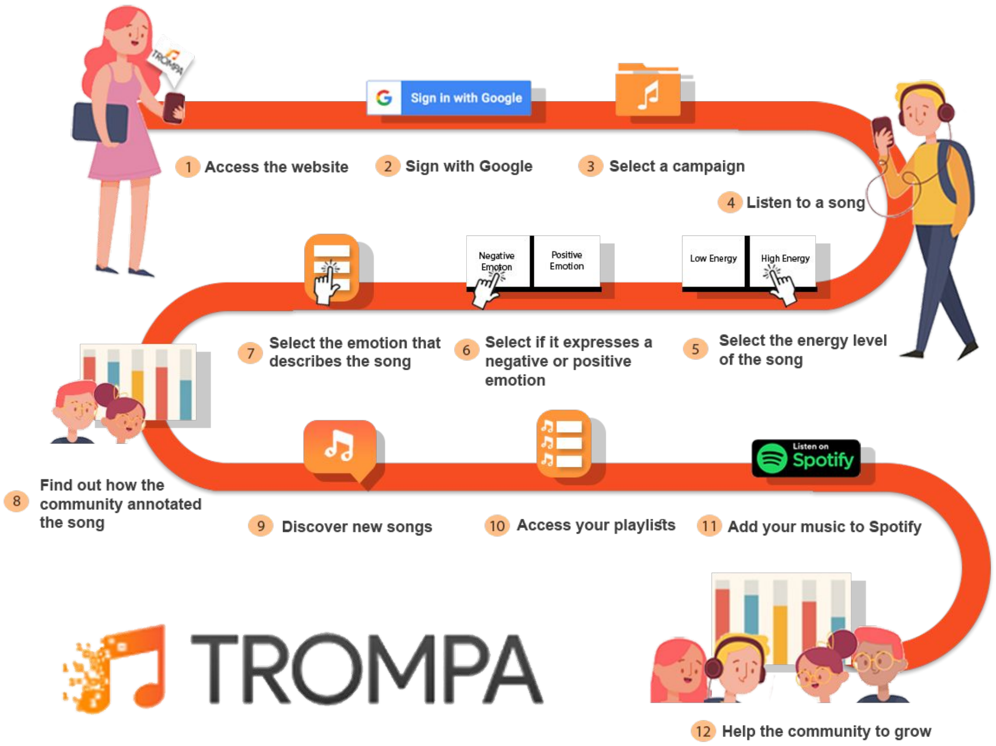
Research questions

- Differences of emotion judgments (perceived and induced)?
- Can consensus entropy generalize?

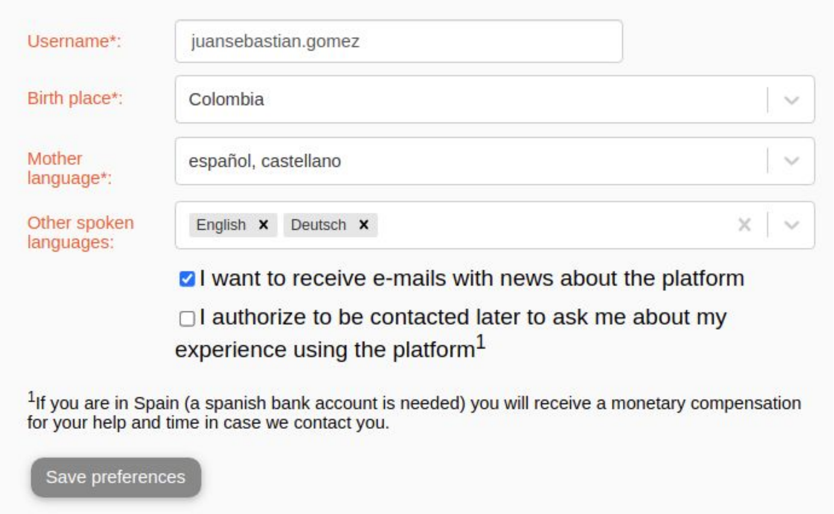
Reference:
Gebru, T. et al. **Datasheets for datasets.** Communications of the ACM, 64(12), pp. 86-92, 2021.

Music enthusiasts platform

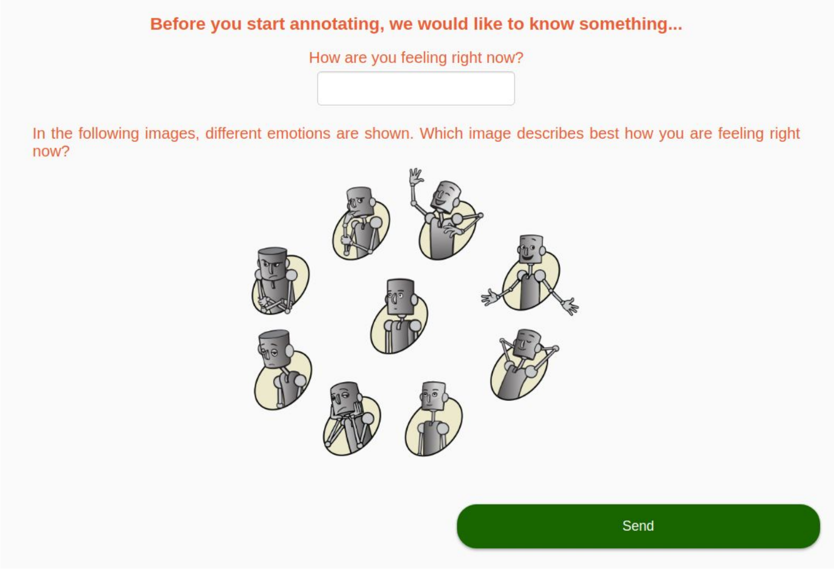
- TROMPA EU project
- Musical training
- Diverse annotations
  - Perceived, induced, free-text, native language..
- Reasoning behind annotations
- Explicit feedback
  - Personalization
  - Music recommendation



Human-centered data collection



Individual differences
(English, Spanish, Italian, Dutch, Mandarin)

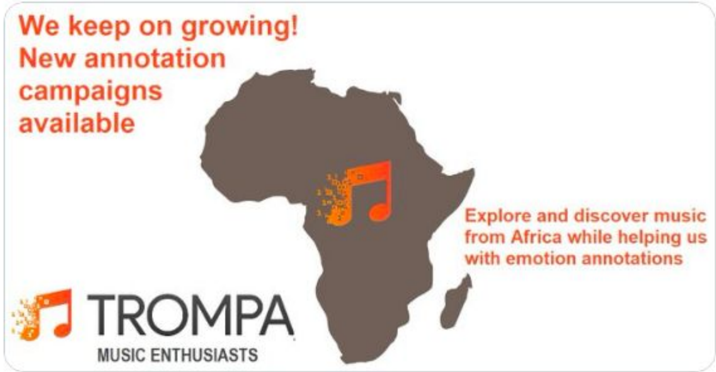


Free-text and forced-choice Pick a Mood

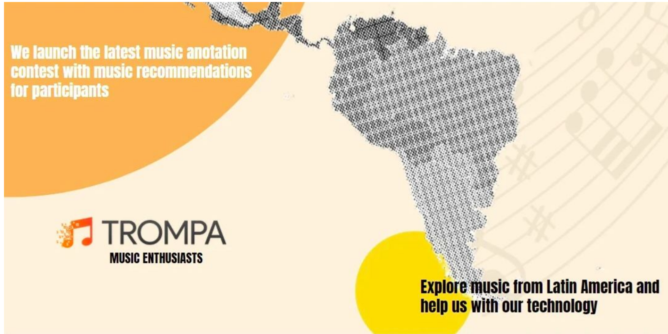
Human-centered data collection

TROMPA Music
@TrompaMusic

Join our second online music annotation contest and win new prizes! from October 14-20th, use our app and annotate the #emotion your perceive in #music from #Africa ilde.upf.edu/trompa/rc/news #music #citizenscience



Music from West Africa - 1 week
23 participants - 655 annotations



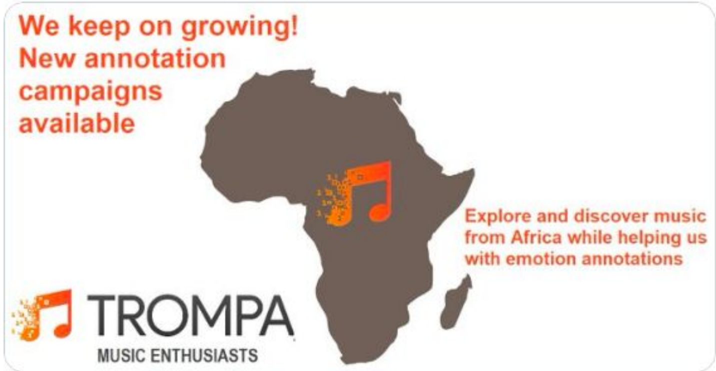
Music from Latin America - 4 weeks
26 participants - 183 annotations

Music from the Middle East - ongoing...

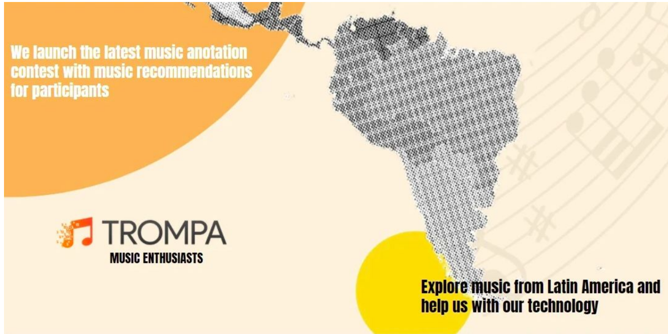
Human-centered data collection

TROMPA Music
@TrompaMusic

Join our second online music annotation contest and win new prizes! from October 14-20th, use our app and annotate the #emotion your perceive in #music from #Africa ilde.upf.edu/trompa/rc/news #music #citizenscience



Music from West Africa - 1 week
23 participants - 655 annotations



Music from Latin America - 4 weeks
26 participants - 183 annotations

Music from the Middle East - ongoing…

Incentivization strategies…

Appendix B:
Gutiérrez-Páez et al., **Emotion annotation of music: a citizen science approach.** Proceedings of the Collaboration Technologies and Social Computing conference, pp. 51-66, 2021.

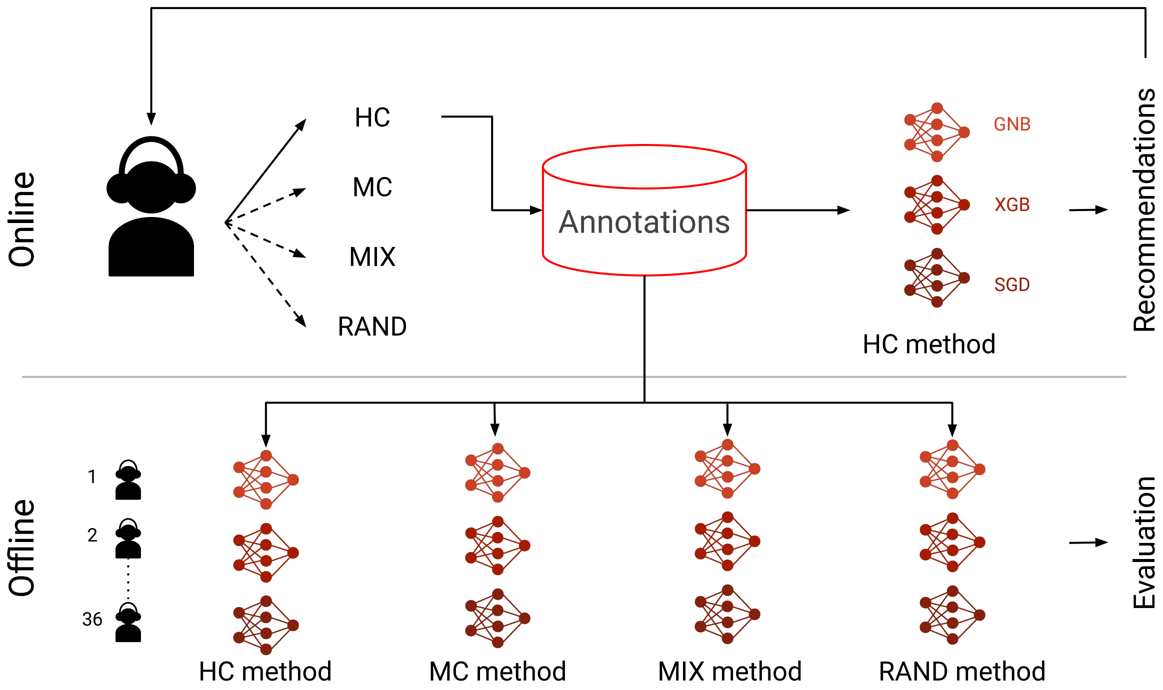
Human-centered data collection

Online personalization

- Random assignment
- Personalized recommendations
- CNN not viable
    - GNB, SGD, XGB

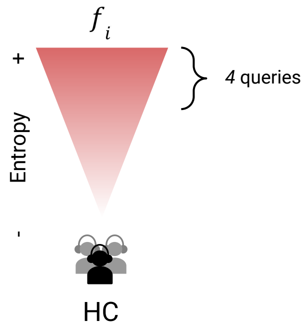
Offline evaluation

- 36 users, over 80 annotations
- All methods
- Evaluate *q* and *e*



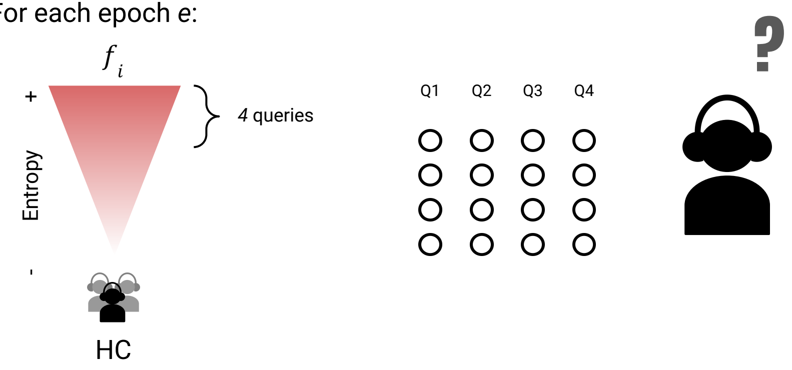
Imbalanced classification

For each epoch *e*:



Imbalanced classification

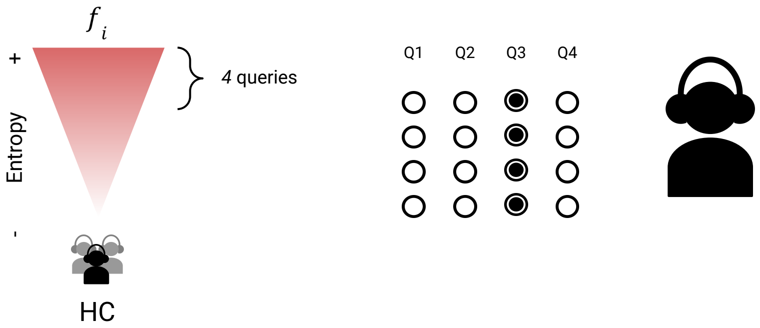
For each epoch *e*:



Imbalanced classification

- Bias in models and listeners

For each epoch *e*:

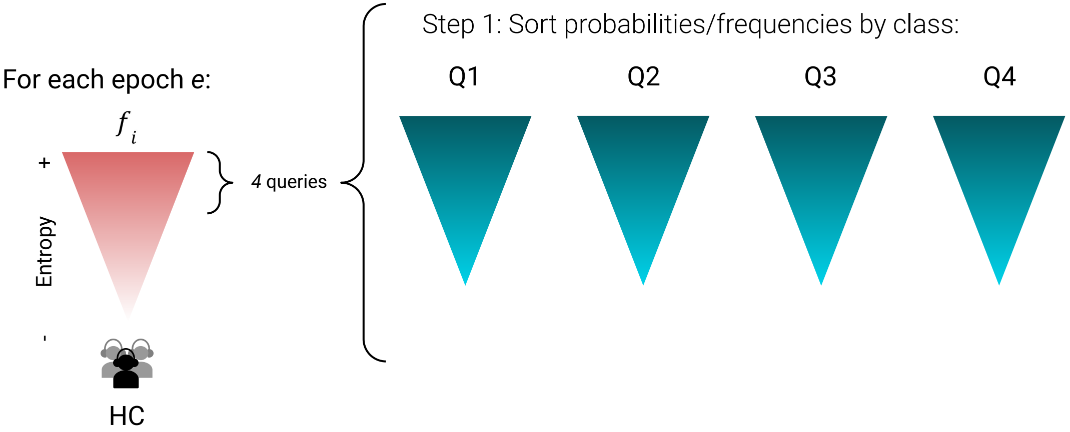


Human-centered data collection

Imbalanced classification

- Bias in models and listeners

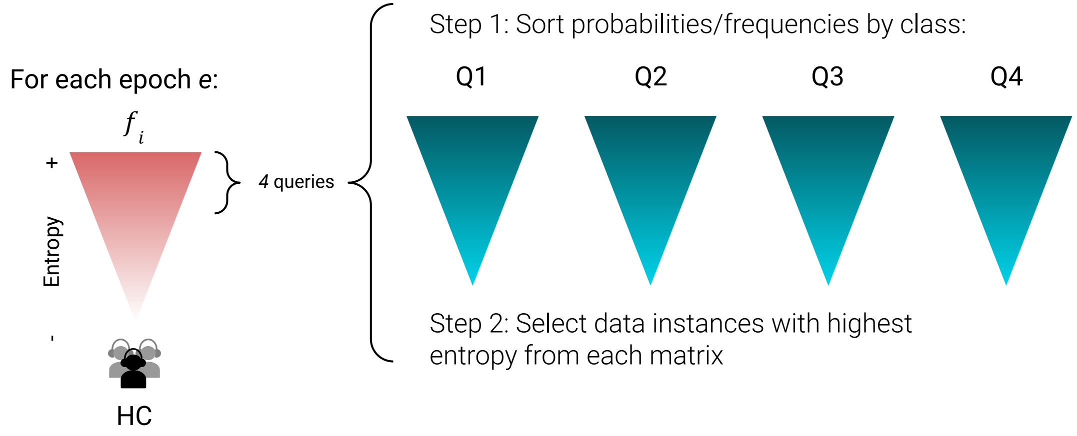
Sort by class!



Imbalanced classification

- Bias in models and listeners

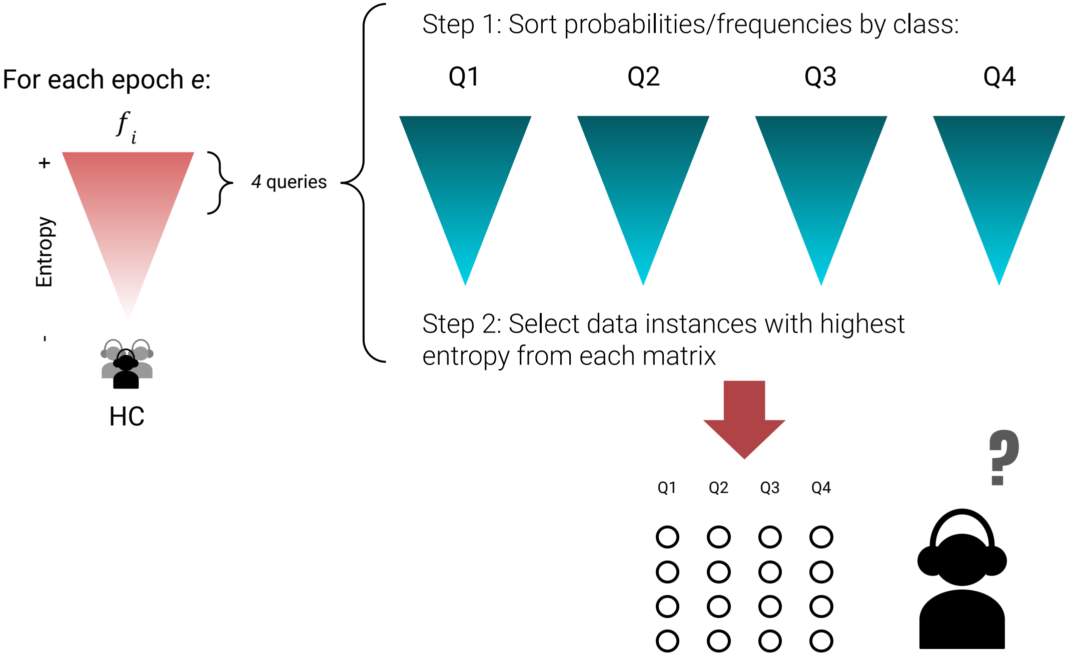
Sort by class!



Imbalanced classification

- Bias in models and listeners

Sort by class!



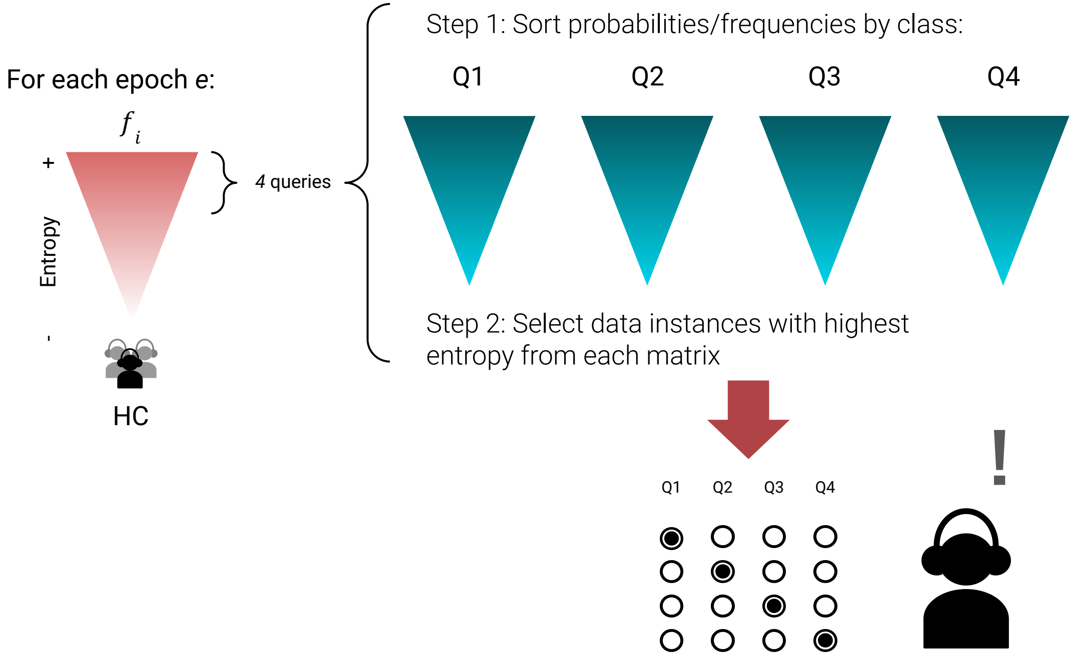
Human-centered data collection

Imbalanced classification

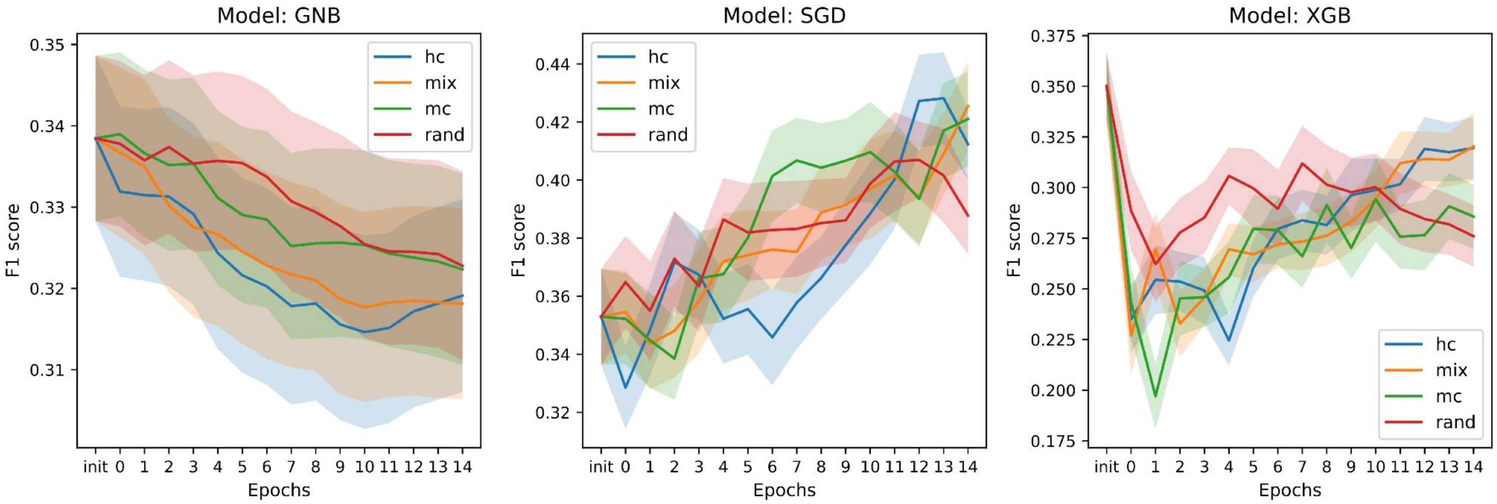
- Bias in models and listeners

Sort by class!

- Diminish likelihood of imbalance
- q = k x num_class



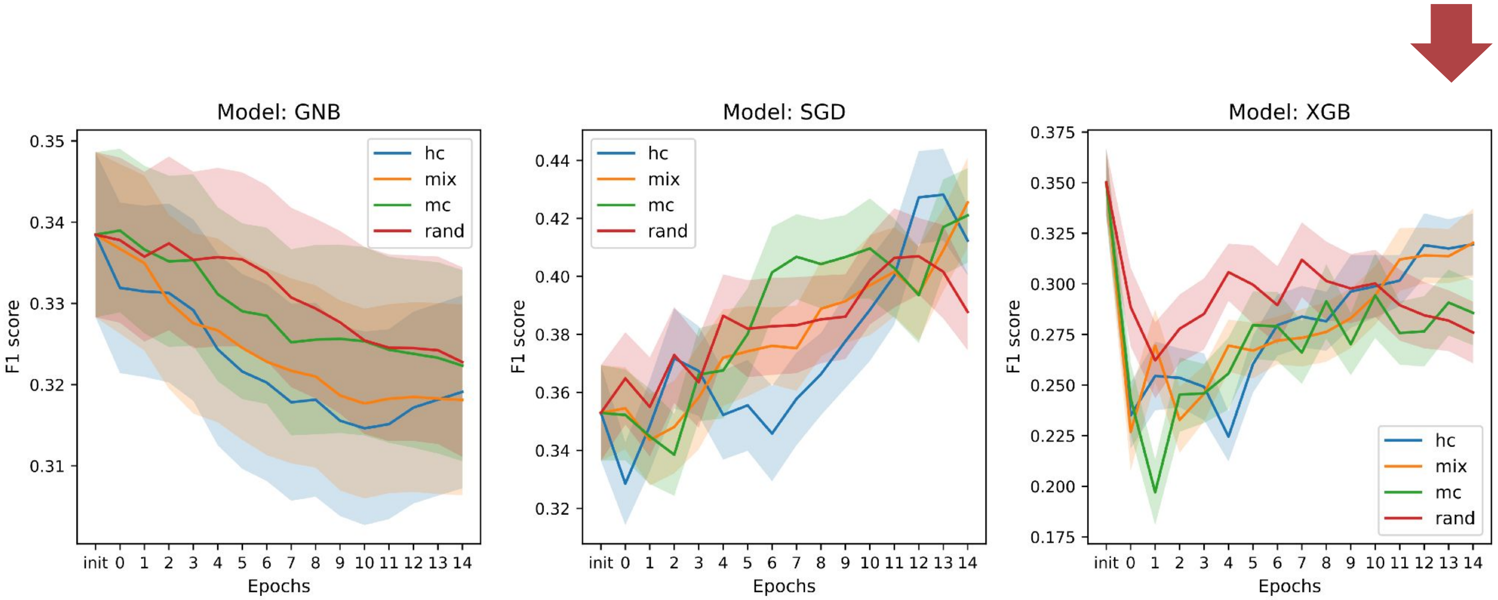
Results



2160 trained classifiers: 36 users × 3 classifiers × 5 models per pre-training split × 4 consensus entropy methods

d.f. = 179, statistical significance $p < 0.0125$ with Bonferroni correction

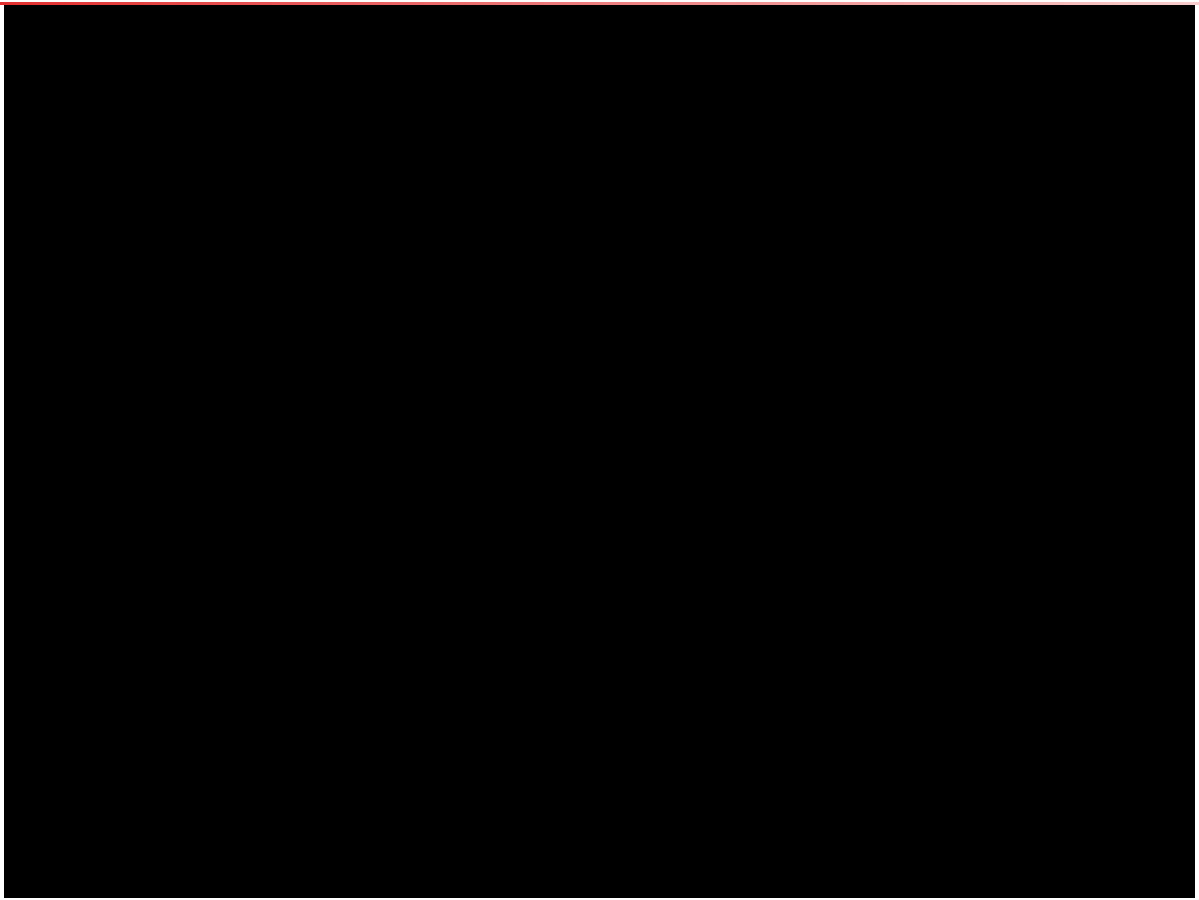
Results



2160 trained classifiers: 36 users × 3 classifiers × 5 models per pre-training split × 4 consensus entropy methods

d.f. = 179, statistical significance $p < 0.0125$ with Bonferroni correction

Human-centered data collection



Datasheet: https://trompa-mtg.upf.edu/vis-mtg-mer/

Results

Research questions

- Difference in judgment?
- Can HC generalize?

Generalized lack of agreement

- Illusion of universality ➞ averaging
- Listeners get confused
- Enriched dataset ➞ greater response diversity
  - Broad AV ➞ **perceived emotion**
  - Specific, free-text, native ➞ **induced emotion**

Personalization

- HC and MIX methods are **significantly better** than RAND
- **Embracing subjectivity**

7 Personalization and polarization with MER

Publications & research stay: Durham University

Gómez-Cañón, Cano, Herrera & Gómez. ***Personalized musically induced emotions of not-so-popular Colombian music***. Human centered AI Workshop at NeurIPS 2021, pp. 1-5.

Gómez-Cañón, Lennie, Eerola, Aragón, Cano, Herrera & Gómez. ***Polarization through Colombian not-so-popular music and algorithms: appraisal guided musically induced emotions***. Music and Science (under review), 2022.

Motivation

- Consensus entropy works! 😁

Motivation

- Consensus entropy works! 😱

Personalization and polarization

Motivation

- Consensus entropy works! 😱
- Negative emotions?
- User profiling?
- Induced emotions - goal-directed mechanisms (Lennie & Eerola, 2022)
    - Test psychological theories

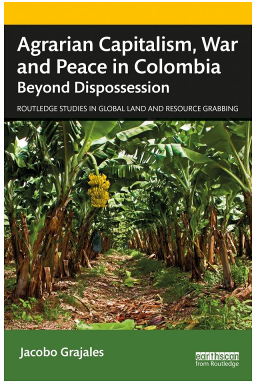
Research questions

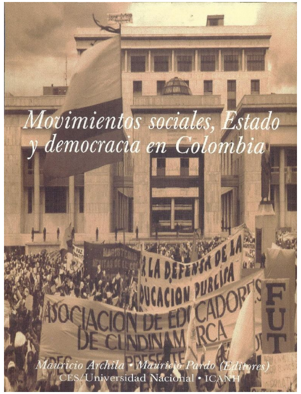
- Do political values affect musically induced emotions?
- Reveal sensitive information from a listener?

Reference:
Lennie, T. and Eerola, T. **The CODA model: a review and skeptical extension of the constructionist model of emotional episodes induced by music.** Frontiers in Psychology, 13, 2022.

Historical context:

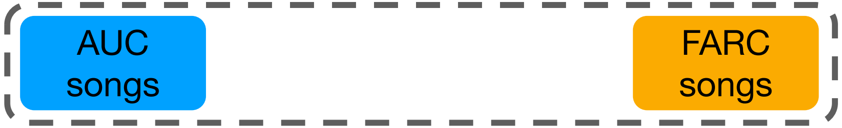
- More than 420.000 violent deaths
- 11 million Colombians displaced
- Illegal armies left-wing FARC and right-wing AUC
- Political identities:
    - Not-so-popular music (?)
    - Functionalities are different!





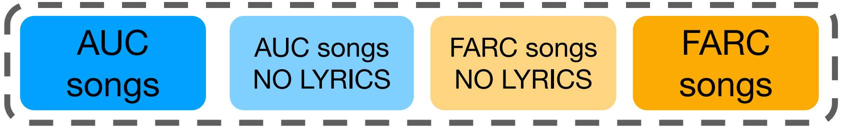
Musical styles:

- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos



Musical styles:

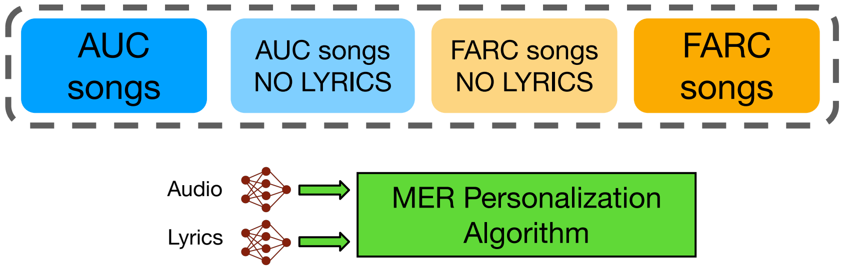
- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos
- Strong political content in lyrics
  - Source separation



Personalization and polarization

Musical styles:

- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos
- Strong political content in lyrics
  - Source separation



Personalization:

- Consensus entropy MC
- Topic modeling
  - Word frequency + Logistic regression

Personalization and polarization

Musical styles:

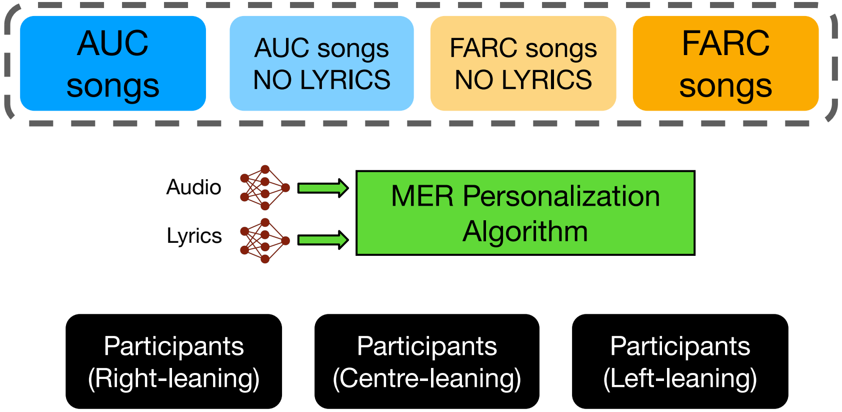
- FARC-songs
    - Canción social or vallenato
- AUC-songs
    - Corridos
- Strong political content in lyrics
    - Source separation

Personalization:

- Consensus entropy MC
- Topic modeling

Hypothesis (pilot):

- Political stance

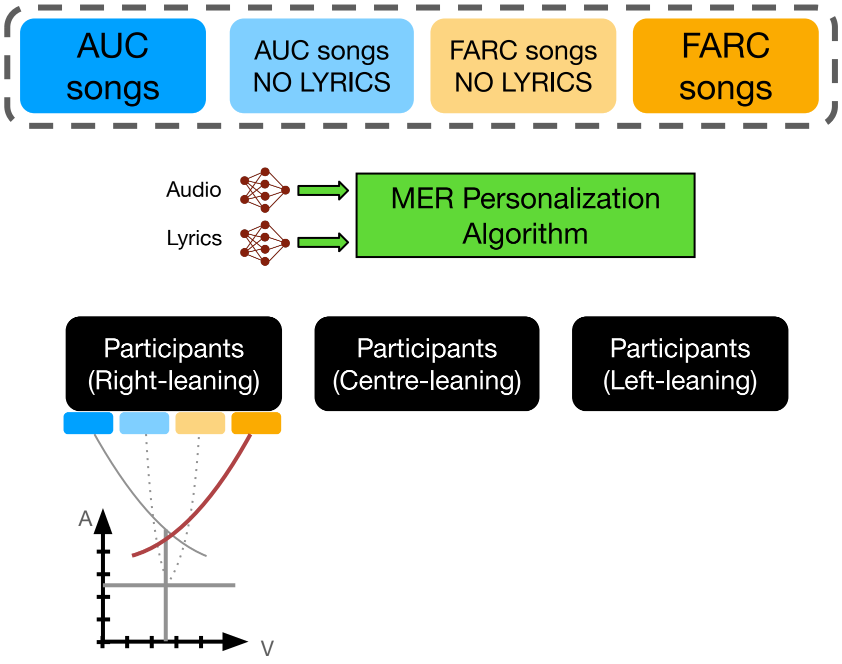


Personalization and polarization

Musical styles:
- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos
- Strong political content in lyrics
  - Source separation

Personalization:
- Consensus entropy MC
- Topic modeling

Hypothesis (pilot):
- Political stance
- Intrinsic goals



Personalization and polarization

Musical styles:

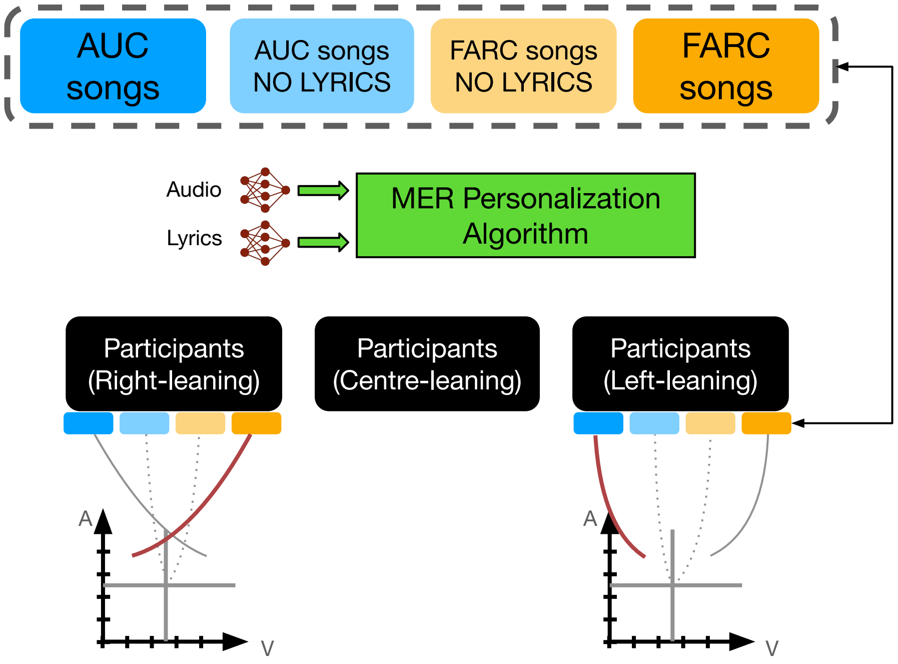
- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos
- Strong political content in lyrics
  - Source separation

Personalization:

- Consensus entropy MC
- Topic modeling

Hypothesis (pilot):

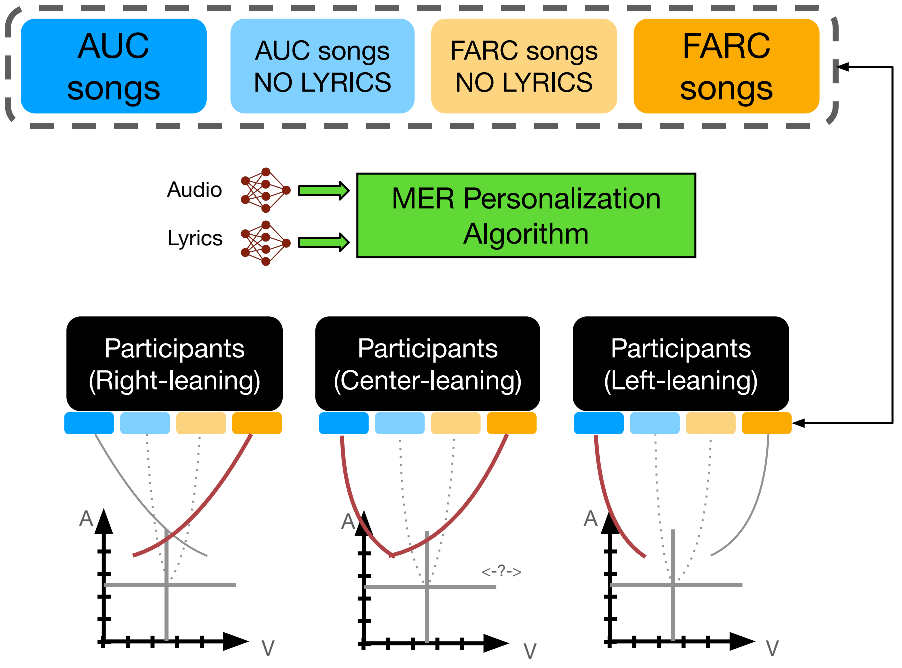
- Political stance
- Intrinsic goals



Personalization and polarization

Musical styles:

- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos
- Strong political content in lyrics
  - Source separation



Personalization:

- Consensus entropy MC
- Topic modeling

Hypothesis (pilot):

- Political stance
- Intrinsic goals

Personalization and polarization

Musical styles:

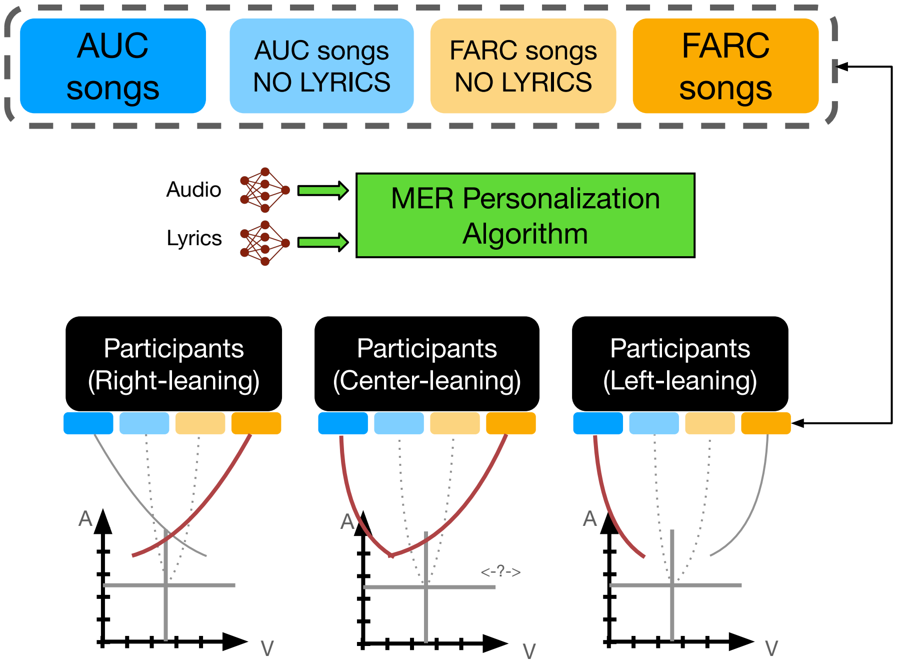
- FARC-songs
  - Canción social or vallenato
- AUC-songs
  - Corridos
- Strong political content in lyrics
  - Source separation

Personalization:

- Consensus entropy MC
- Topic modeling

Hypothesis (pilot):

- Political stance
- Intrinsic goals
- Methodology:
  - RWA, SDO, Colombian-specific
  - N = 52, during elections…



Results

Research questions

- Political ⇋ emotions?
- Bias in MER system?

No significant difference in annotation

- Questionable groupings!
- Goals do influence the appraisal of emotions

Personalization can reflect political stance!

- 10 out of 27 personalized models
- With vs. without lyrics
- “Toying with emotions and personalization” (European Commission, 2022)

Interdisciplinary research

- MER research
- Music cognition research

Reference:
European Commission. **Behavioral study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation.** 2022.

Results

Research questions

- Political ⇋ emotions?
- Bias in MER system?

No significant difference in annotation

- Questionable groupings!
- Goals do influence the appraisal of emotions

Personalization can reflect political stance!

- 10 out of 27 personalized models
- With vs. without lyrics
- “Toying with emotions and personalization” (European Commission, 2022)

Interdisciplinary research

- MER research
- Music cognition research

→ What for?

Reference:
European Commission. **Behavioral study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation.** 2022.

8 Conclusions

Summary

Impact of language on MER

- Overall low agreement
  - Ground truth ➞ judgment
- Un- or self-supervised pre-training is too coarse

Impact of language on MER

- Overall low agreement
  - Ground truth ➞ judgment
- Un- or self-supervised pre-training is too coarse
  - Language-sensitive MER is a 🦄

Summary

Impact of language on MER

- Overall low agreement
  - Ground truth ➞ judgment
- Un- or self-supervised pre-training is too coarse
  - Language-sensitive MER is a 🦄

Personalization for MER

- Human consensus entropy works!
  - Not all models are effectively personalized
- Citizen science of music and emotions from the Global South

Summary

Impact of language on MER

- Overall low agreement
  - Ground truth ➞ judgment
- Un- or self-supervised pre-training is too coarse
  - Language-sensitive MER is a 🦄

Personalization for MER

- Human consensus entropy works!
  - Not all models are effectively personalized
- Citizen science of music and emotions from the Global South
  - Low agreement ➞ rich response diversity ➞ opportunity to learn better!

Summary

Impact of language on MER

- Overall low agreement
  - Ground truth ➞ judgment
- Un- or self-supervised pre-training is too coarse
  - Language-sensitive MER is a 🦄

Personalization for MER

- Human consensus entropy works!
  - Not all models are effectively personalized
- Citizen science of music and emotions from the Global South
  - Low agreement ➞ rich response diversity ➞ opportunity to learn better!

Ethical considerations

- Misuse/manipulation/profiling:
  - Unfair digital asymmetries ➞ “colonial value and power paradigms”
  - Regulation over emotion recognition

“And which are the harmonies
expressive of sorrow?
You are musical and can tell me.

The harmonies which you mean are the
mixed or tenor
Lydian, and the full-toned or bass
Lydian and such alike.”

Plato, Republic Book III (307 BCE)

Final remarks

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary

Final remarks

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary

Western
Educated
Industrialized
Rich
Democratic

Final remarks

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary
  - Decolonize our own minds in the GS (Smit et al., 2022)

Western
Educated
Industrialized
Rich
Democratic

Reference:
Smit., E. A., et al. **Emotional responses in Papua New Guinea show negligible evidence for a universal effect of major versus minor music.** PLOS One, 2022.

Final remarks

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary
  - **Decolonize our own minds in the GS** (Smit et al., 2022)

Problematic categorization

- "Skill to understand our environment"
- Representations vs. experiences
- "Expert" judgments (Kahneman et al., 2022)
  - Bias and noise
  - Overestimate agreement, underestimate noise

Reference:
Kahneman, D., et al. **Noise: a flaw in human judgment.** Little, Brown and Company, 2022.

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary
  - **Decolonize our own minds in the GS** (Smit et al., 2022)

Problematic categorization

- "Skill to understand our environment"
- Representations vs. experiences
- "Expert" judgments (Kahneman et al., 2022)
  - Bias and noise
  - Overestimate agreement, underestimate noise ➡ Skepticism!

Reference:
Kahneman, D., et al. **Noise: a flaw in human judgment.** Little, Brown and Company, 2022.

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary
  - **Decolonize our own minds in the GS** (Smit et al., 2022)

Problematic categorization

- "Skill to understand our environment"
- Representations vs. experiences
- "Expert" judgments (Kahneman et al., 2022)
  - Bias and noise
  - Overestimate agreement, underestimate noise

More human-centric efforts in AI

- Thought on personalization
  - Use?
  - Feelings, perceptions, engagement?
  - Enjoyment?

Final remarks

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary
    - **Decolonize our own minds in the GS** (Smit et al., 2022)

Problematic categorization

- "Skill to understand our environment"
- Representations vs. experiences
- "Expert" judgments (Kahneman et al., 2022)
    - Bias and noise
    - Overestimate agreement, underestimate noise

More human-centric efforts in AI

- Thought on personalization ➡ **What for?**

Final remarks

WEIRD history, research, researchers, and listeners

- How music is used, taught, learned?
- Magnificently diverse and interdisciplinary
  - **Decolonize our own minds in the GS** (Smit et al., 2022)

Problematic categorization

- "Skill to understand our environment"
- Representations vs. experiences
- "Expert" judgments (Kahneman et al., 2022)
  - Bias and noise
  - Overestimate agreement, underestimate noise

More human-centric efforts in AI

- Thought on personalization
- Just beginning…

→ **What for?**

Universitat Pompeu Fabra Barcelona
MTG Music Technology Group

Human-centered machine learning for music emotion recognition

Juan Sebastián Gómez-Cañón

Supervisors:

Dr. Emilia Gómez
Joint Research Centre,
European Commission, Seville, Spain

Dr. Estefanía Cano
Songquito UG,
Erlangen, Germany

Dr. Perfecto Herrera
Music Technology Group,
Universitat Pompeu Fabra, Barcelona, Spain

Committee members:

Dr. Blair Kaneshiro
Center for Computer Research in Music and Acoustics,
University of Stanford, USA

Dr. Isabelle Hupont
Joint Research Centre,
European Commission, Seville, Spain

Dr. Arthur Flexer
Institute of Computational Perception,
Johannes Kepler University Linz, Austria

TROMPA MUSICAL AI_

Ph.D. defense - Barcelona, October 26, 2022

Universitat Pompeu Fabra Barcelona
MTG Music Technology Group

Human-centered machine learning for music emotion recognition

Juan Sebastián Gómez-Cañón

Supervisors:

Dr. Emilia Gómez
Joint Research Centre,
European Commission, Seville, Spain

Dr. Estefanía Cano
Songquito UG,
Erlangen, Germany

Dr. Perfecto Herrera
Music Technology Group,
Universitat Pompeu Fabra, Barcelona, Spain

Committee members:

Dr. Blair Kaneshiro
Center for Computer Research in Music and Acoustics,
University of Stanford, USA

Dr. Isabelle Hupont
Joint Research Centre,
European Commission, Seville, Spain

Dr. Arthur Flexer
Institute of Computational Perception,
Johannes Kepler University Linz, Austria

TROMPA
MUSICAL AI_

Ph.D. defense - Barcelona, October 26, 2022