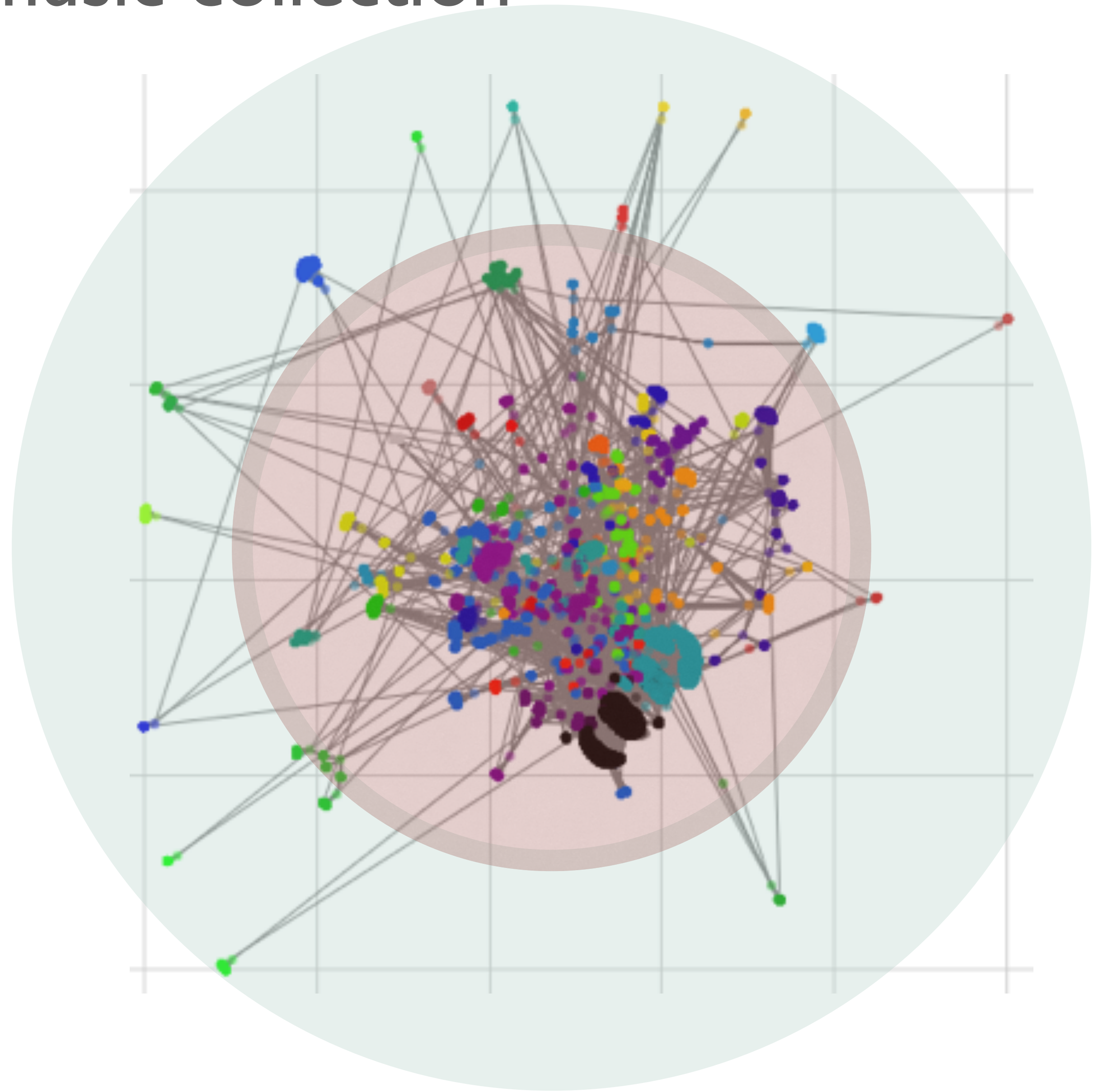
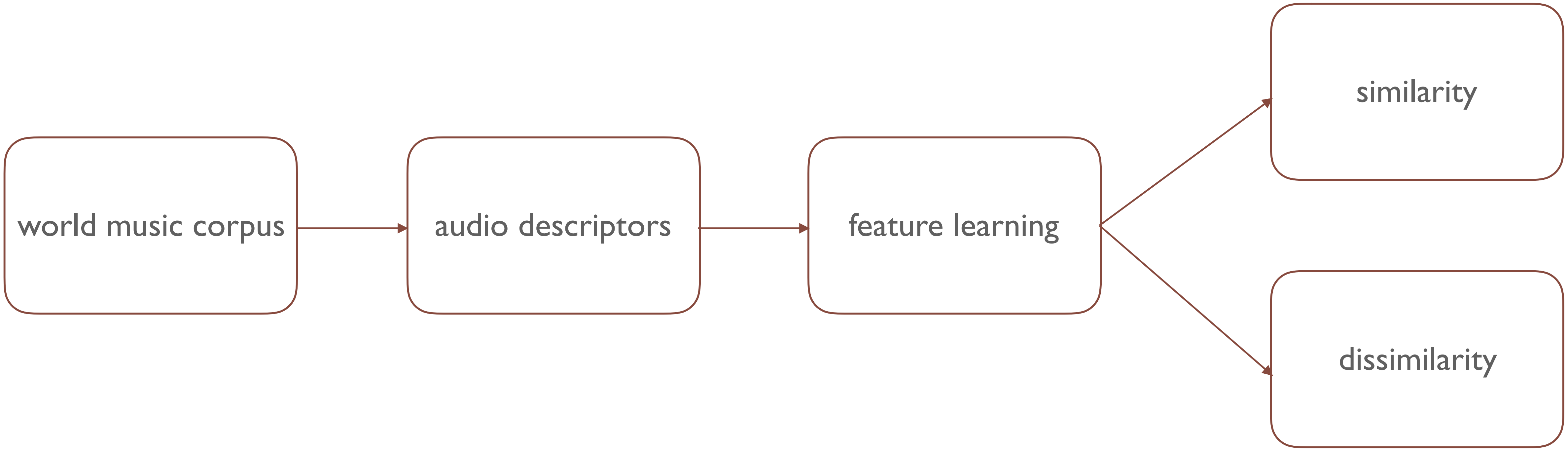
LEARNING A FEATURE SPACE FOR SIMILARITY IN WORLD MUSIC

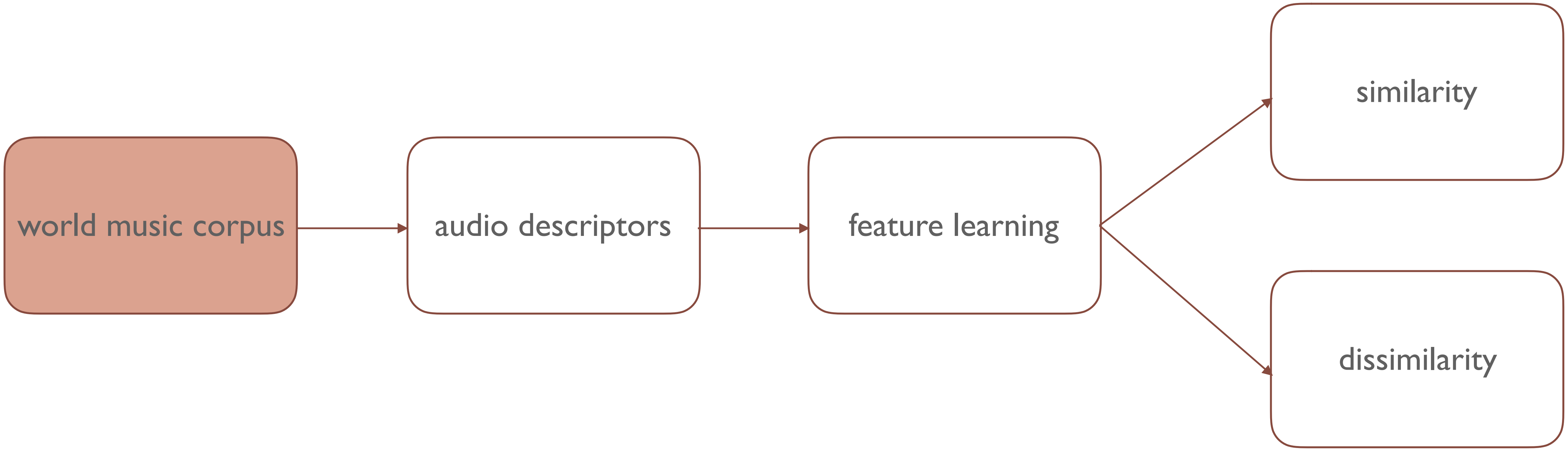
Maria Panteli, Emmanouil Benetos, Simon Dixon
Centre for Digital Music
Queen Mary University of London

- the goal
  - study similarity in a large world music collection



- related research
  - music universals
    - [Lomax, AAAS 1968, Brown & Jordania, Psychology of Music 2011, Savage et al., PNAS 2015]
  - world music features & classification
    - [Gomez et al., ISMIR 2009, Kruspe et al., AES 2011, Zhou et al., ICDM 2014]
  - corpus analysis
    - [Serra et al., Nature 2012, Mauch et al., Royal Society 2015, Moelants et al., ISMIR 2009]
  - compmusic Computational models for the discovery of the World's Music · Digital Music Lab An AHRC Digital Transformations Project · Telemeta : collaborative multimedia asset management system telemeta


world music corpus
audio descriptors
feature learning
similarity
dissimilarity



world music corpus

audio descriptors

feature learning

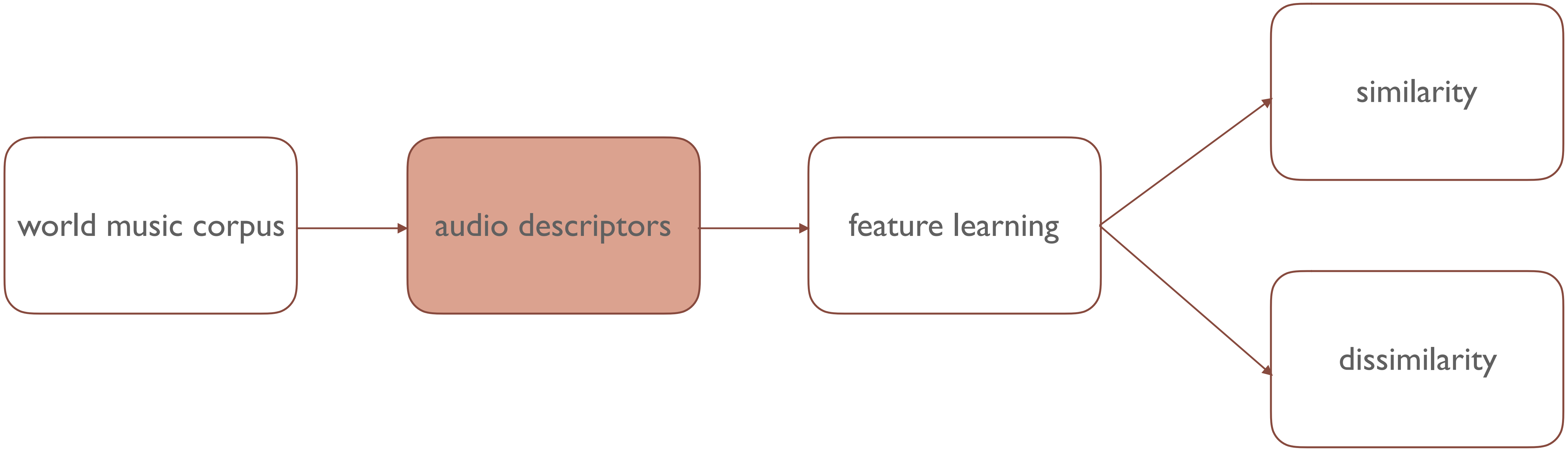
similarity

dissimilarity

- what is world music?
  - “basically, all the music of the world” [Bohlman, Oxford University Press 2002]

Smithsonian Folkways Recordings

- folk and traditional, from as many countries as possible
- 30-second audio previews
- balanced dataset: 31 countries x 70 recs = **2170 recs**

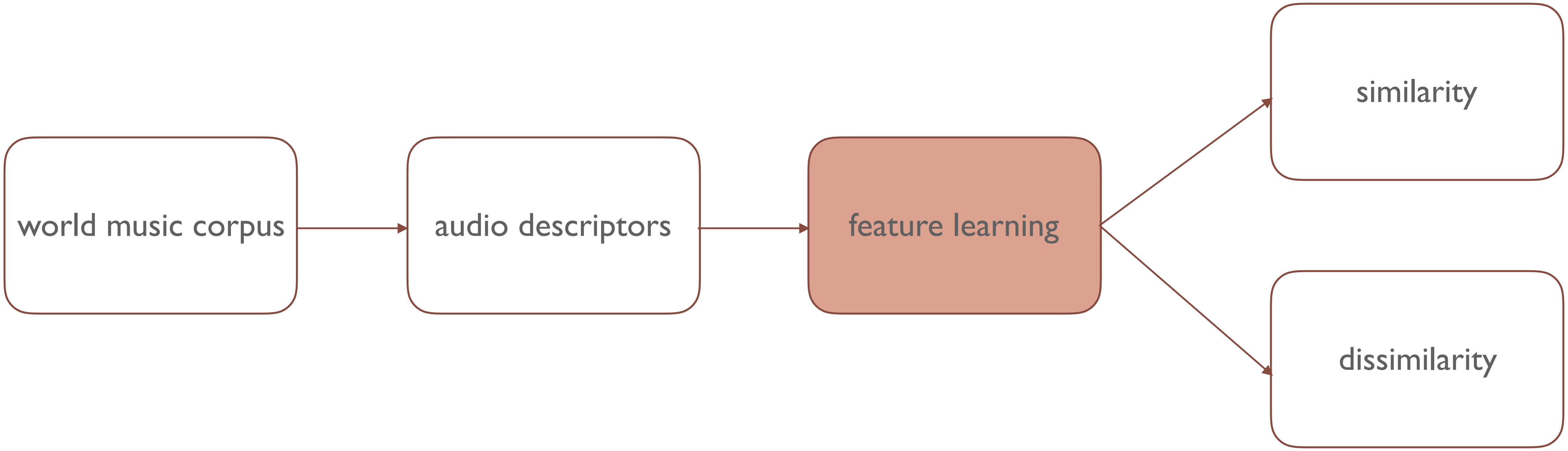


world music corpus
audio descriptors
feature learning
similarity
dissimilarity

- which descriptors?
  - “style can be recognized by characteristic uses of form, **texture, harmony, melody, and rhythm**” [Sadie et al., Oxford University Press 2001]

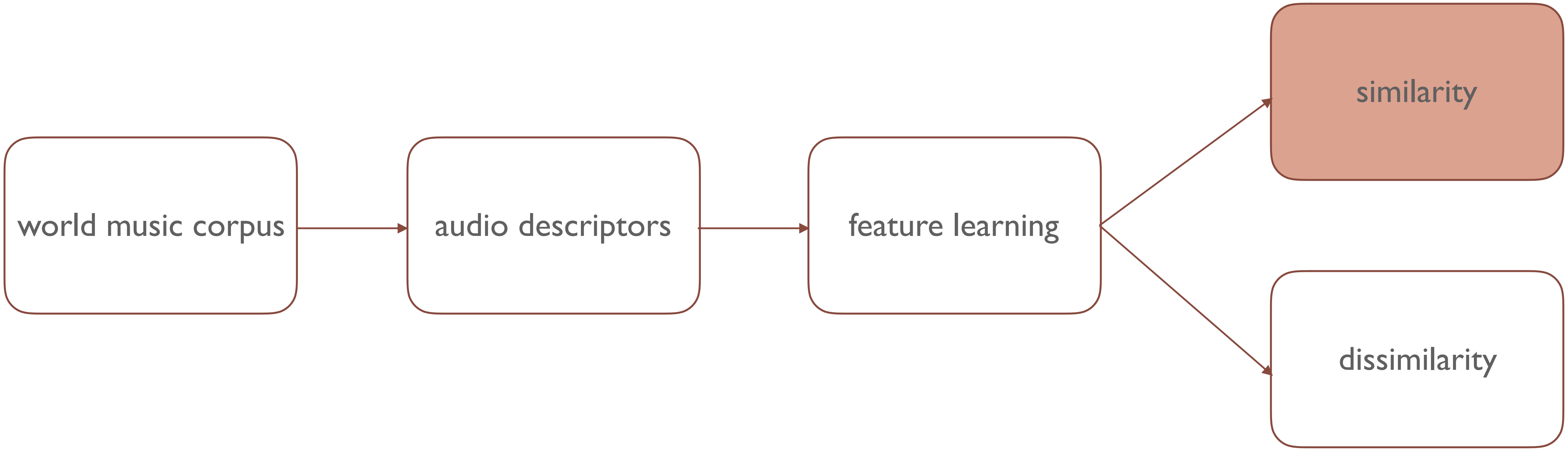
  - timbre: MFCCs [Aucouturier et al., IEEE-Multimedia 2005]
  - rhythm: scale transform [Holzapfel & Stylianou, IEEE-ASLP 2011]
  - melody: pitch bihistogram [Van Balen et al., ISMIR 2014]
  - harmony: average chroma [Bartsch & Wakefield, IEEE-Multimedia, 2005]

- why these descriptors?
  - low-level representations more likely to be robust to the diversity of world music
  - state-of-the-art performance in related music similarity tasks
  - invariance - check out the poster today! [Panteli & Dixon, ISMIR 2016]

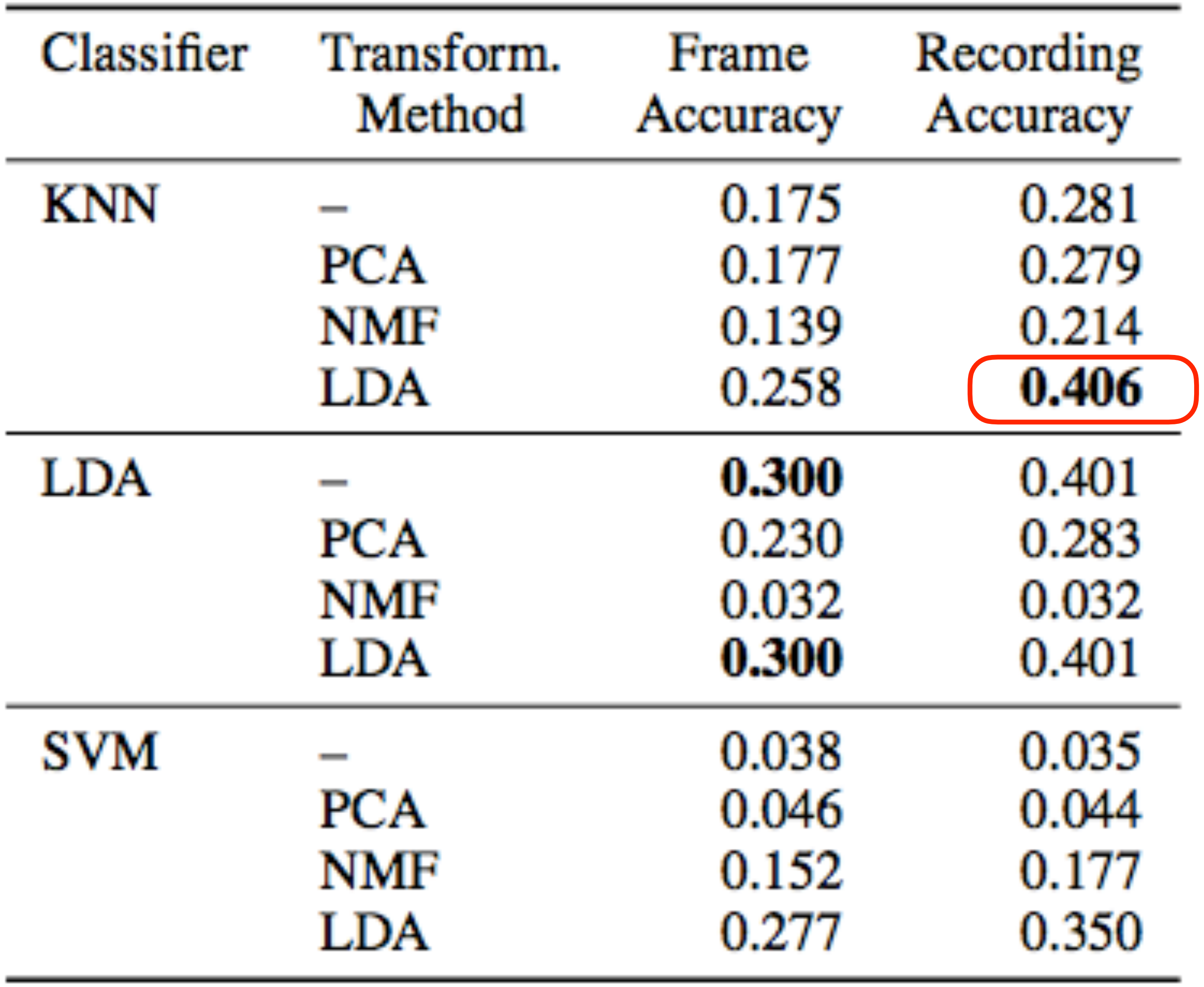

world music corpus
audio descriptors
feature learning
similarity
dissimilarity

- feature vector for 8-second frames
  - rhythm, melody, timbre, harmony

- from ~1000 dimensions to ~30 using
  - Principal Component Analysis (PCA)
  - Non-negative Matrix Factorization (NMF)
  - Linear Discriminant Analysis (LDA)


world music corpus
audio descriptors
feature learning
similarity
dissimilarity

- similarity via *country* classification



Classifier	Transform. Method	Frame Accuracy	Recording Accuracy
KNN	–	0.175	0.281
	PCA	0.177	0.279
	NMF	0.139	0.214
	LDA	0.258	**0.406**
LDA	–	**0.300**	0.401
	PCA	0.230	0.283
	NMF	0.032	0.032
	LDA	**0.300**	0.401
SVM	–	0.038	0.035
	PCA	0.046	0.044
	NMF	0.152	0.177
	LDA	0.277	0.350

Table: Classification accuracies for the predicted frame labels and the predicted recording labels based on a vote count (31 classes x 70 instances, baseline at 0.03)

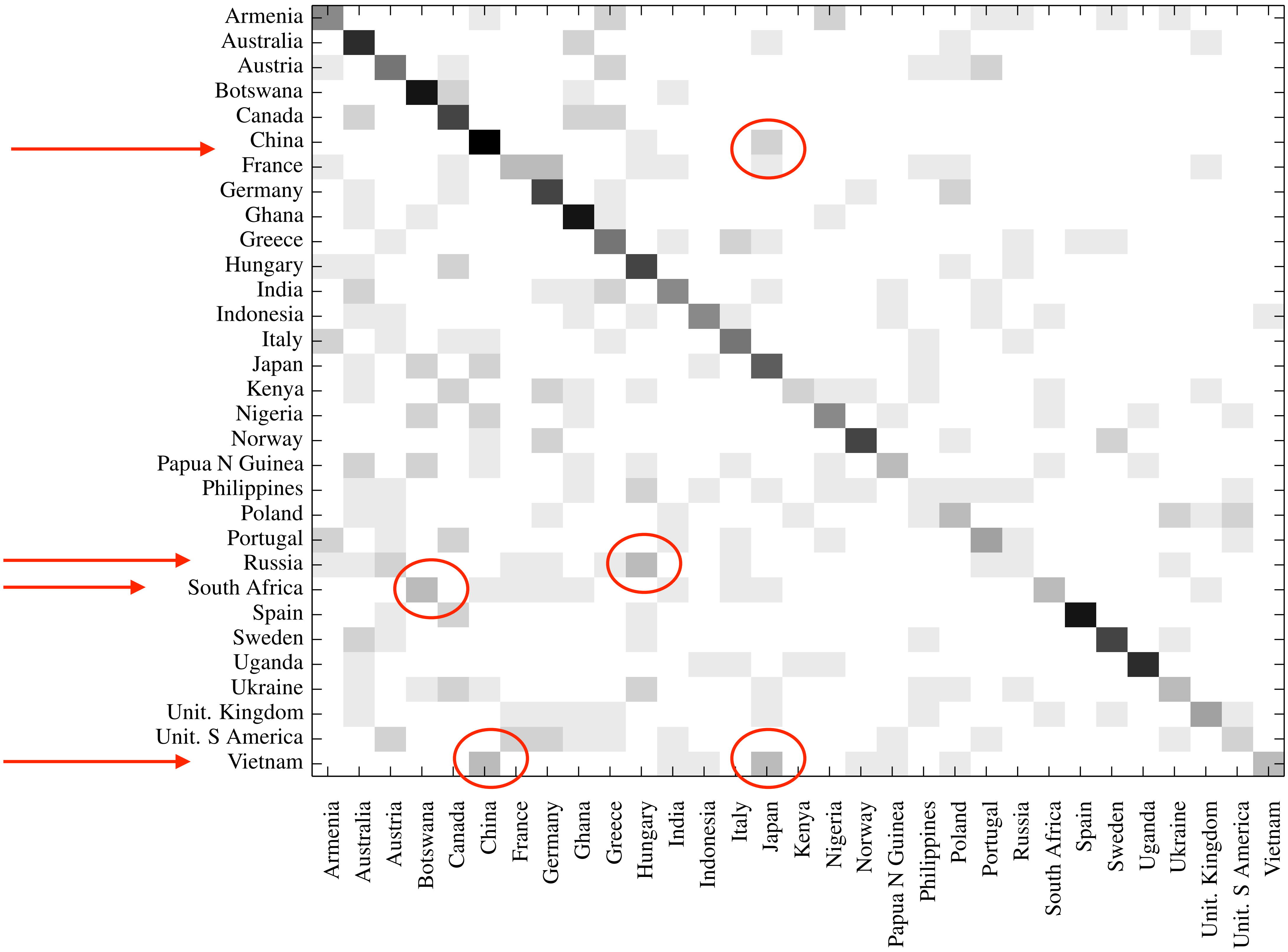
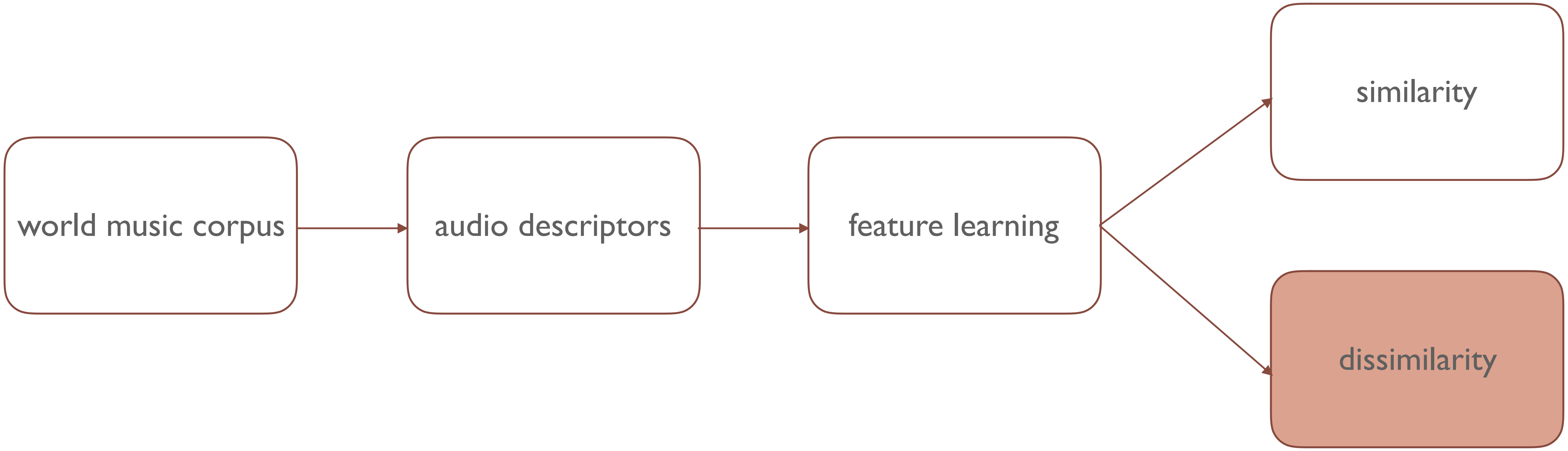


Figure: Confusion matrix for best classifier in the country classification task.


world music corpus
audio descriptors
feature learning
similarity
dissimilarity

- outliers: “recordings that stand out in the collection”
  - detection via Mahalanobis distance [Aggarwal & Yu, ACM SIGMOD 2001]

country: Greece,
instrument: bells

country: Vietnam,
instrument: dav dav

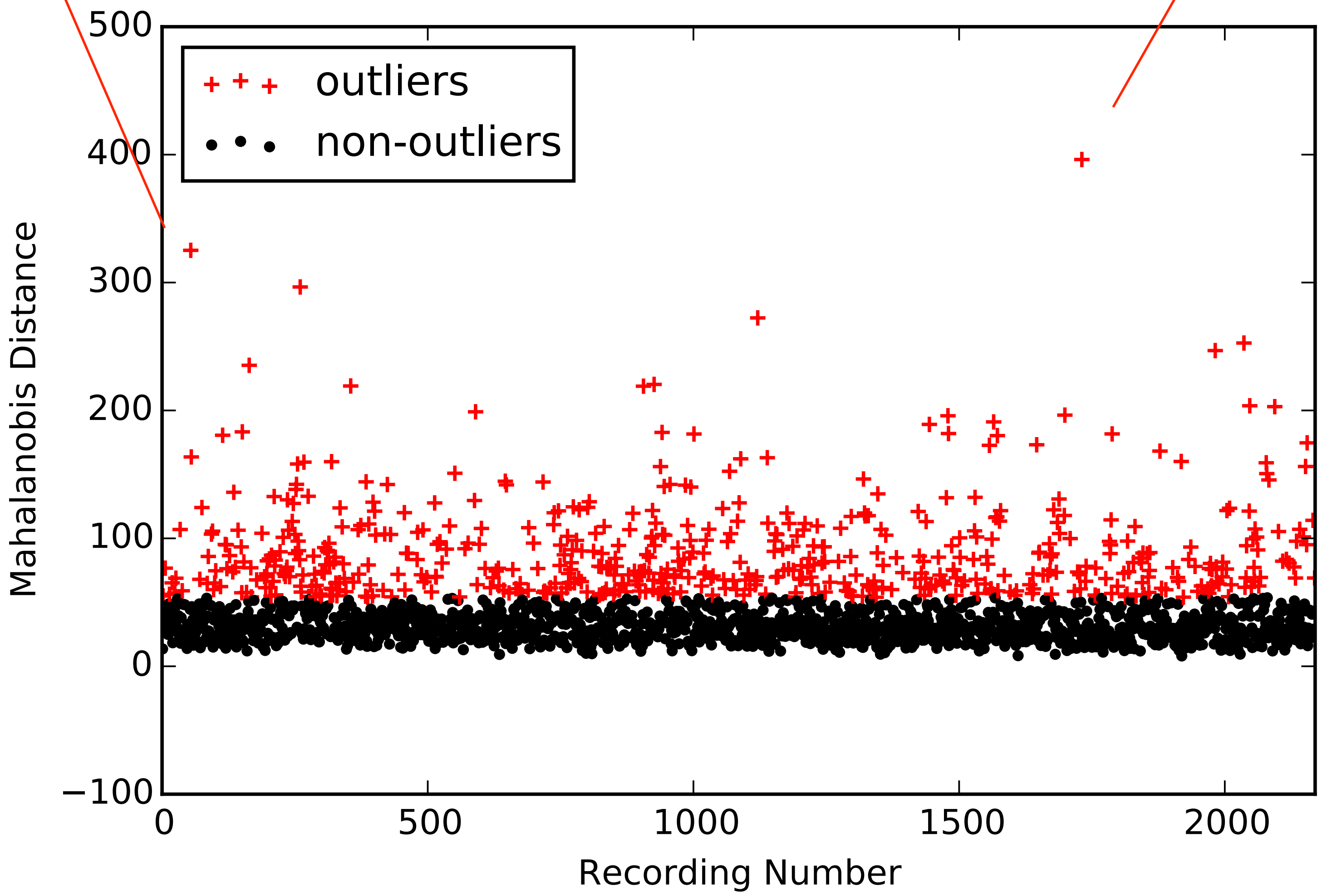


Figure: Mahalanobis distances and outliers at the 99.5% quantile of chi-square distribution.

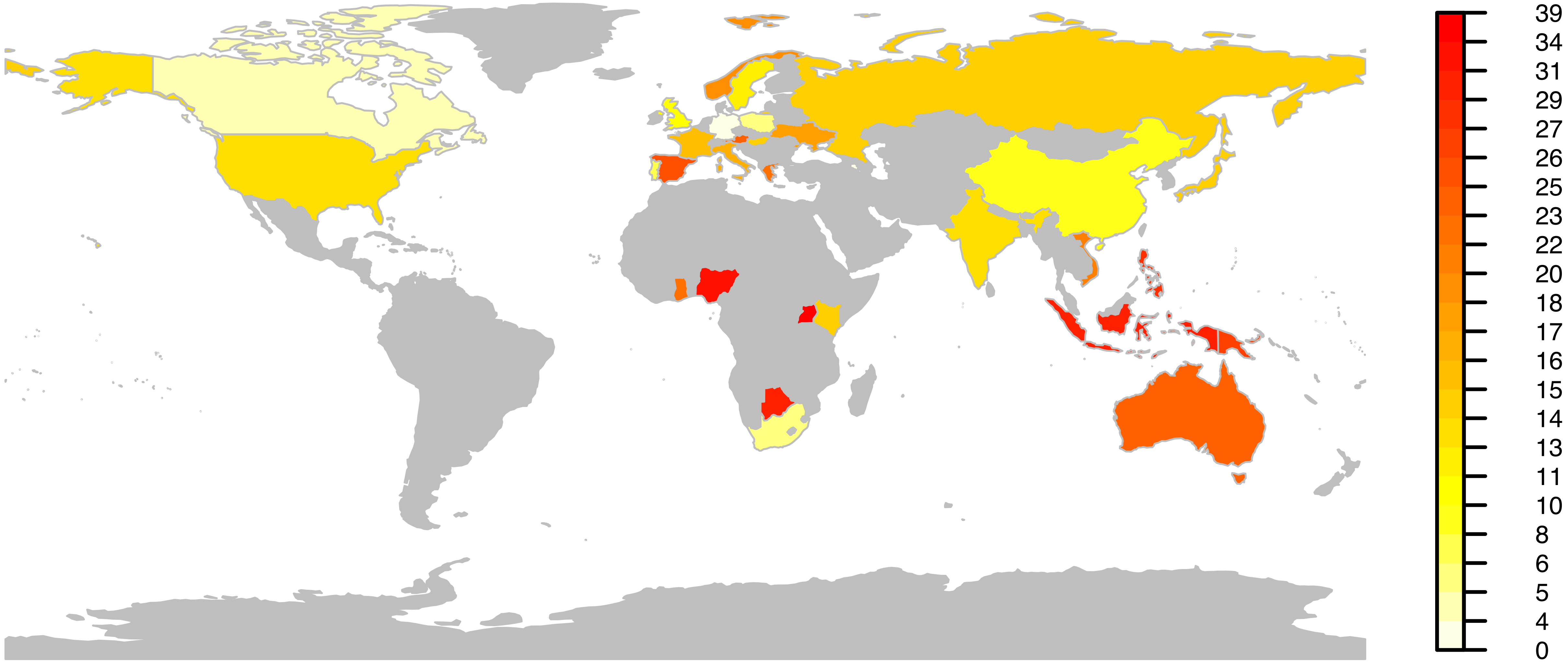
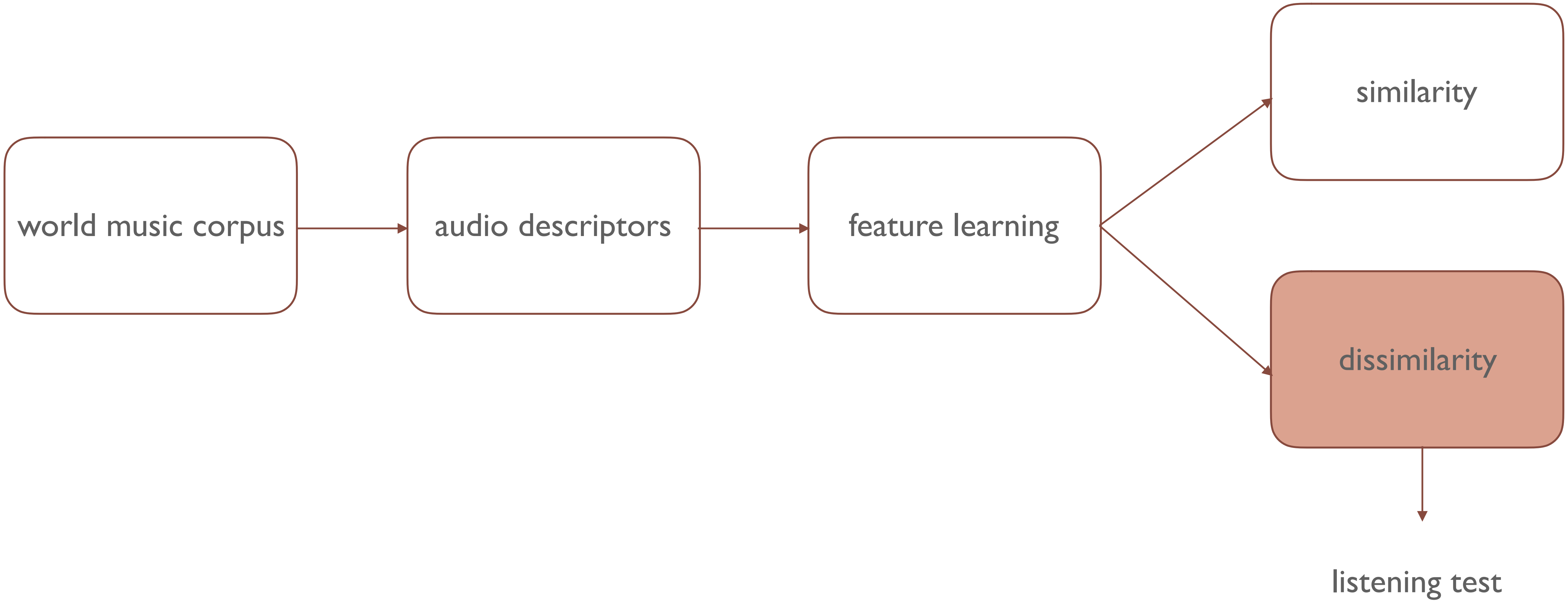
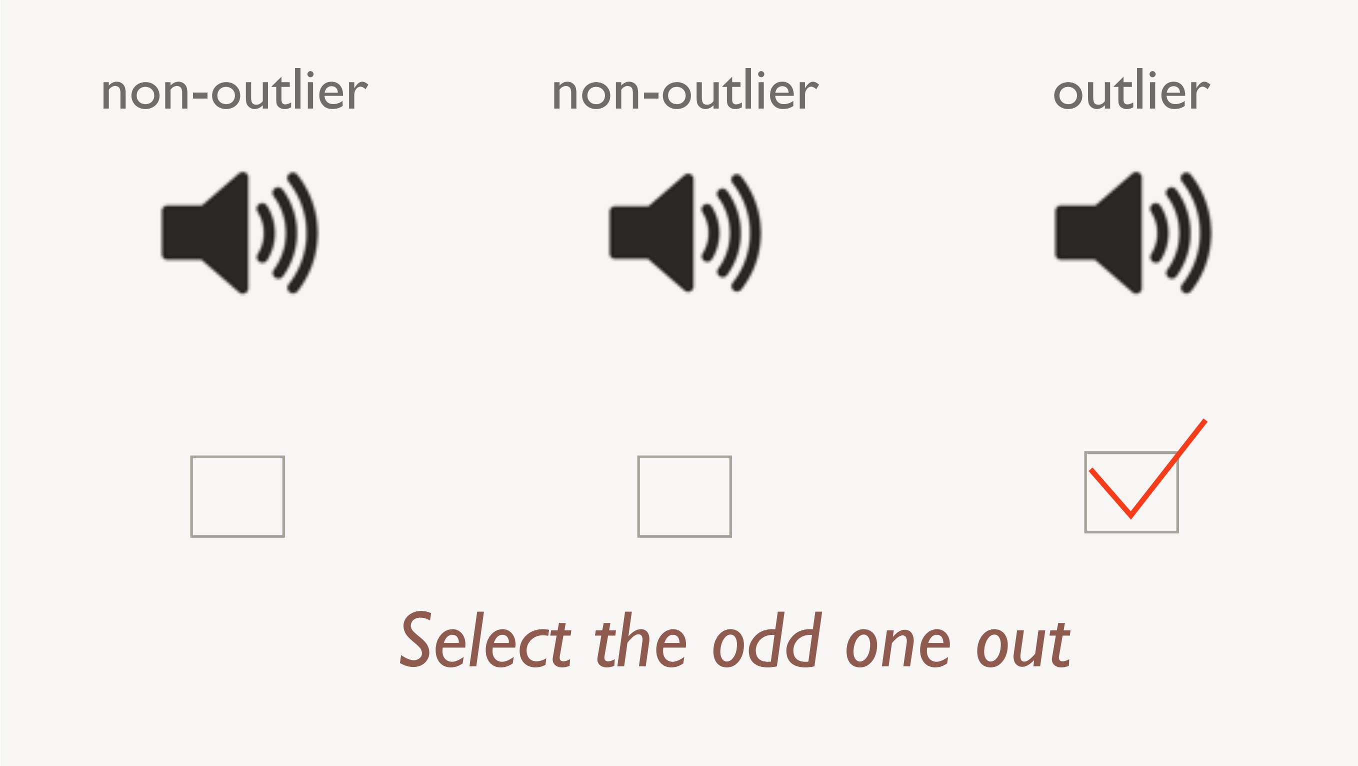


Figure: Number of outliers for each country in our collection (grey areas denote missing data). Botswana had the most outliers (39 out of 70) and Germany the least (0 out of 70).


world music corpus
audio descriptors
feature learning
similarity
dissimilarity
listening test



Select the odd one out

- listening test: odd one out [Wolff & Weyde, ISMIR 2011]
  - 60 outliers x 10 triads each
  - **53% agreement** (random baseline 33%)

- some remarks
    - *representative* dataset
    - *reliable* descriptors
    - *advanced* embeddings
    - *quantitative vs qualitative* evaluation

- conclusion
  - feature learning for world music content description
  - classification confusion between countries of cultural proximity
  - geographical regions with most and least outliers
    - moderate agreement in perceptual evaluation of outliers
  - …world music analysis is interesting for MIR research

Thank you!

- **demo**: http://tinyurl.com/hk6kq6k

Indonesia
JavaneseIOsing,Osing,World
Outlier(MD):False(39)