

Adapting Music Similarity Models to Geographic User Groups

Daniel Wolff, Tillman Weyde, Andrew MacFarlane

Music Informatics Group
Department of Computer Science
daniel.wolff.1@soi.city.ac.uk



**CITY UNIVERSITY
LONDON**

Music Similarity Modelling

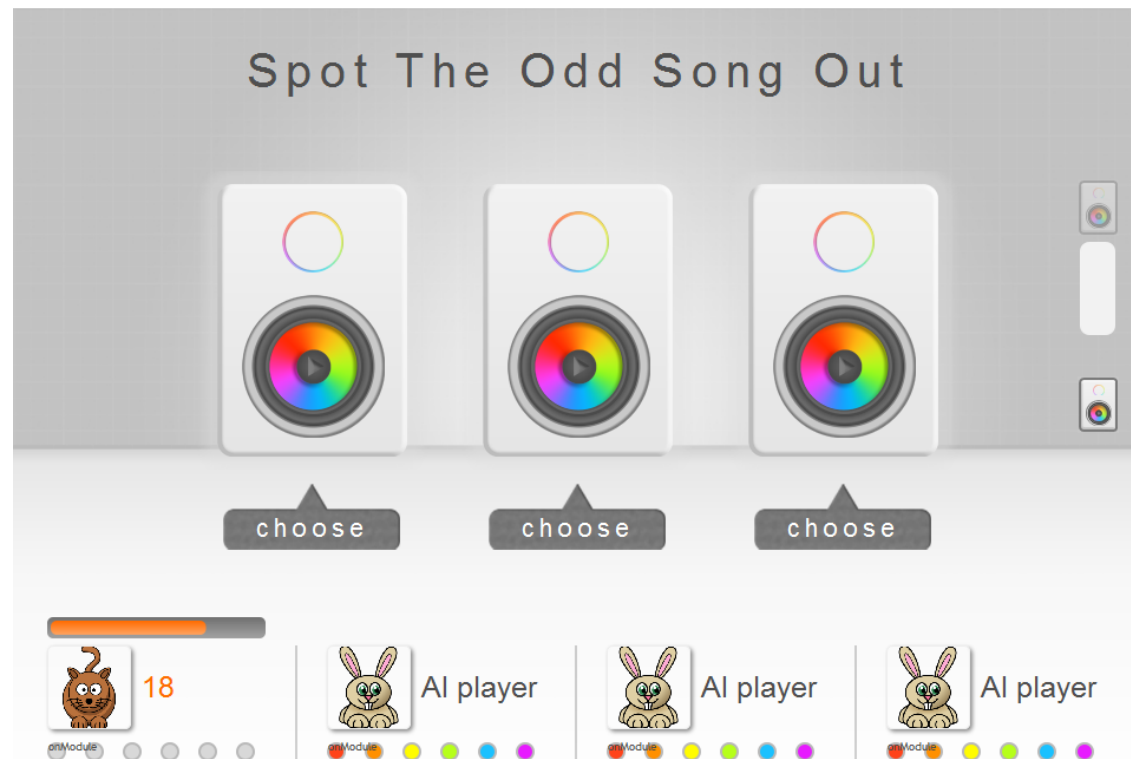


- Computer models of human music similarity judgements
- Computer systems applications:
 - Music exploration
 - Music recommendation
 - Music indexing and retrieval
- Research applications
 - Music psychology
 - Geographic models => ethnomusicology

What Similarity



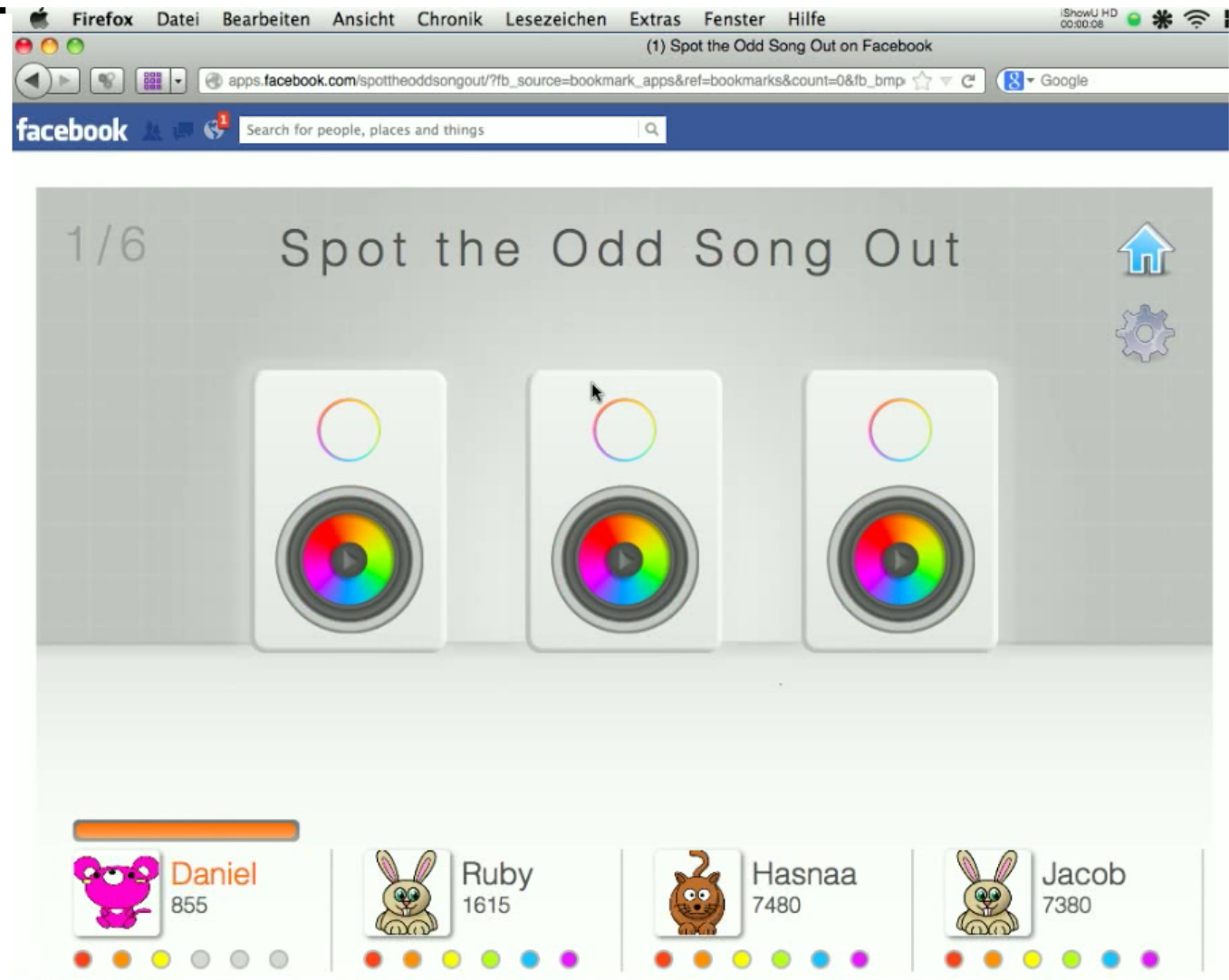
mi.soi.city.ac.uk/camir/game/



Spot The Odd Song Out

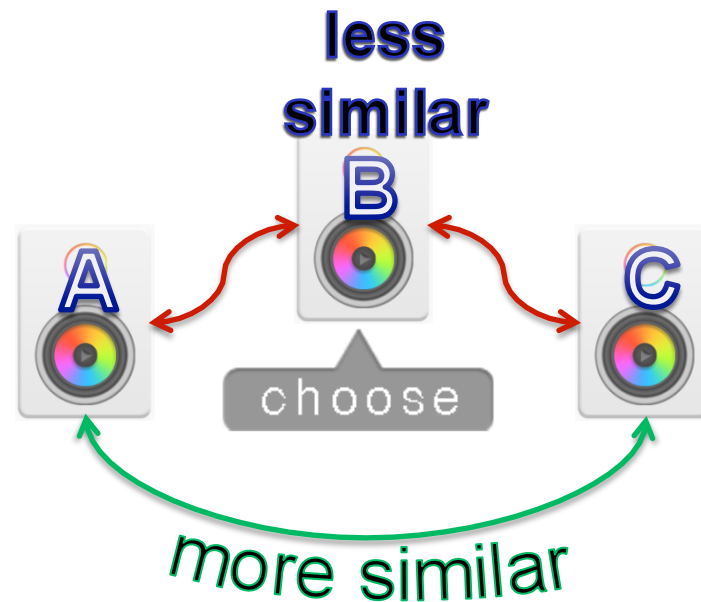


CITY UNIVERSITY
LONDON



Relative Similarity Data

- 1 vote for “Odd Song Out” **B** implies 2 constraints
 $\text{sim}(A, C) > \text{sim}(A, \mathbf{B})$ AND $\text{sim}(A, C) > \text{sim}(\mathbf{B}, C)$
- Compared to absolute similarity:
 - Fast input
 - No calibration needed
 - Fewer algorithms available
 - Fewer data available



Modelling for User Groups

- Learn specific models
 - Better similarity prediction for specific group?
- Applications
 - Personalisation
 - Group recommendation
 - Analyse model differences
 - Isolate attributes influencing similarity votes



Grouping by Geography

- First experiments with user groups
- Most frequent attribute:
 - Input location
- 4 European countries with most data
- “Macroculture” pop music
 - Regional “microcultures” [Slobin 92]

12/17/13



CITY UNIVERSITY
LONDON

Europe



Four (Different) Sets

Oktoberfest



	DE	FR	SE	UK
# unique constraints	459	463	309	411
# clips	151	151	123	151

Audio Features

- Million Song Dataset
 - Extractor: The Echo Nest “Analyse” API
- Key, mode, time signature, tempo
- Chroma & timbre features averaged per clip
- Statistics like [Slaney 2007]
 - Loudness, energy, beat, segmentation,

Similarity Model

- Free parameter: Mahalanobis Matrix W
 - W transforms comparison space

$$d_W(x, y) = \sqrt{(x - y)^T W (x - y)}$$

- With feature vectors x, y
- Iff W is positive definite, d_W qualifies as metric
 - Can be slightly violated without causing problems

Relative ITML

- Information-Theoretic Metric Learning [Davis et al. 2007]
- Optimises a Mahalanobis distance
 - Input: feature difference of clips (A,B)
 - Targets: absolute similarity data $d_W(A,B)$
 - Optional regularisation by template matrix
- Extended to relative data: RITML
 - Iteratively generate absolute targets and update W extending [Zheng et al. 2007]

Results on Country Sets

- Train and test per country
- 10-fold cross-validation
- Test-set results:
 - Baseline: Euclidean metric
 - RITML slightly outperforms baseline
 - But below expectation ~ **66%** [Wolff & Weyde 2013]

	DE	FR	SE	UK	AVERAGE
RITML	64.35	62.71	61.75	63.78	63.15
Euclidean	60.79	62.09	58.11	62.65	60.91

Transfer Learning: Template



- Transfer knowledge from other datasets
 - Via pre-computed Mahalanobis matrix W_0
- Compute four W_0 via RITML:
 - On combined 3-country datasets
 - (FR+SE+UK for DE dataset)
 - Using matrix with best generalisation (to 3-country datasets) as W_0
- Use W_0 as starting point for RITML on one-country data
 - Leverage commonalities in the datasets

Transfer Learning Benchmarks



CITY UNIVERSITY
LONDON

- W_0 -Direct:
 - Test W_0 directly on one-country test-sets

- JOINT:
 - Combine all (DEFRSEUK) training data
 - Learn W directly via RITML
 - Evaluate per country

Results of Transfer Learning



CITY UNIVERSITY
LONDON

- All results on one-country test-sets

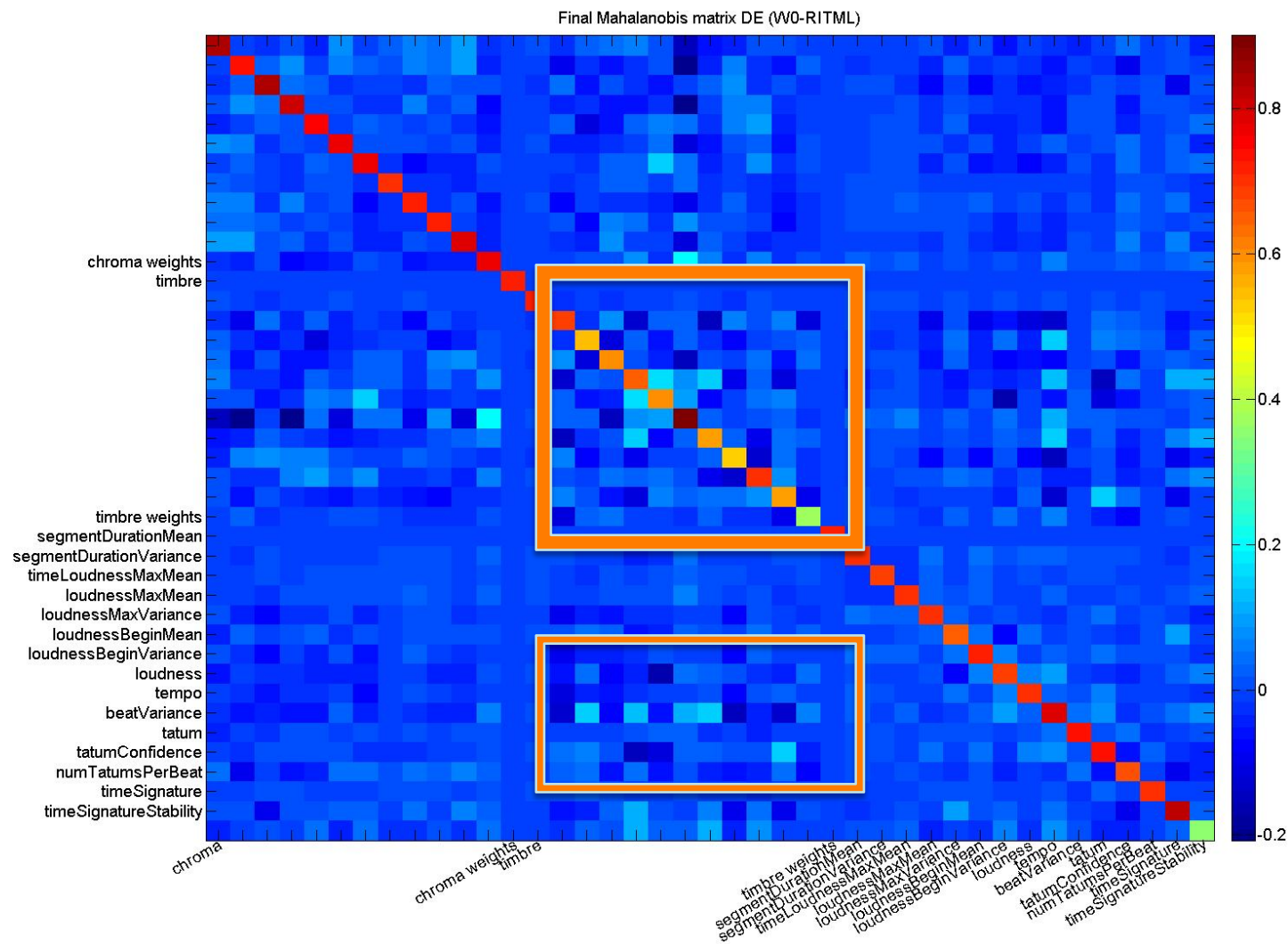
	DE	FR	SE	UK	AVERAGE	TL
W_0 -RITML	69.28	64.34	64.40	70.36	67.09	
W_0 -Direct	67.61	65.97	64.81	69.02	66.85	
JOINT	67.80	67.39	64.05	70.46	67.43	
RITML	64.35	62.71	61.75	63.78	63.15	
Euclidean	60.79	62.09	58.11	62.65	60.91	

- Transfer learning
 - Outperforms RITML/ Euclidean (stat. significant)
 - Similar to JOINT model
- W_0 -RITML only successful for DE and UK

W₀-RITML Matrix DE



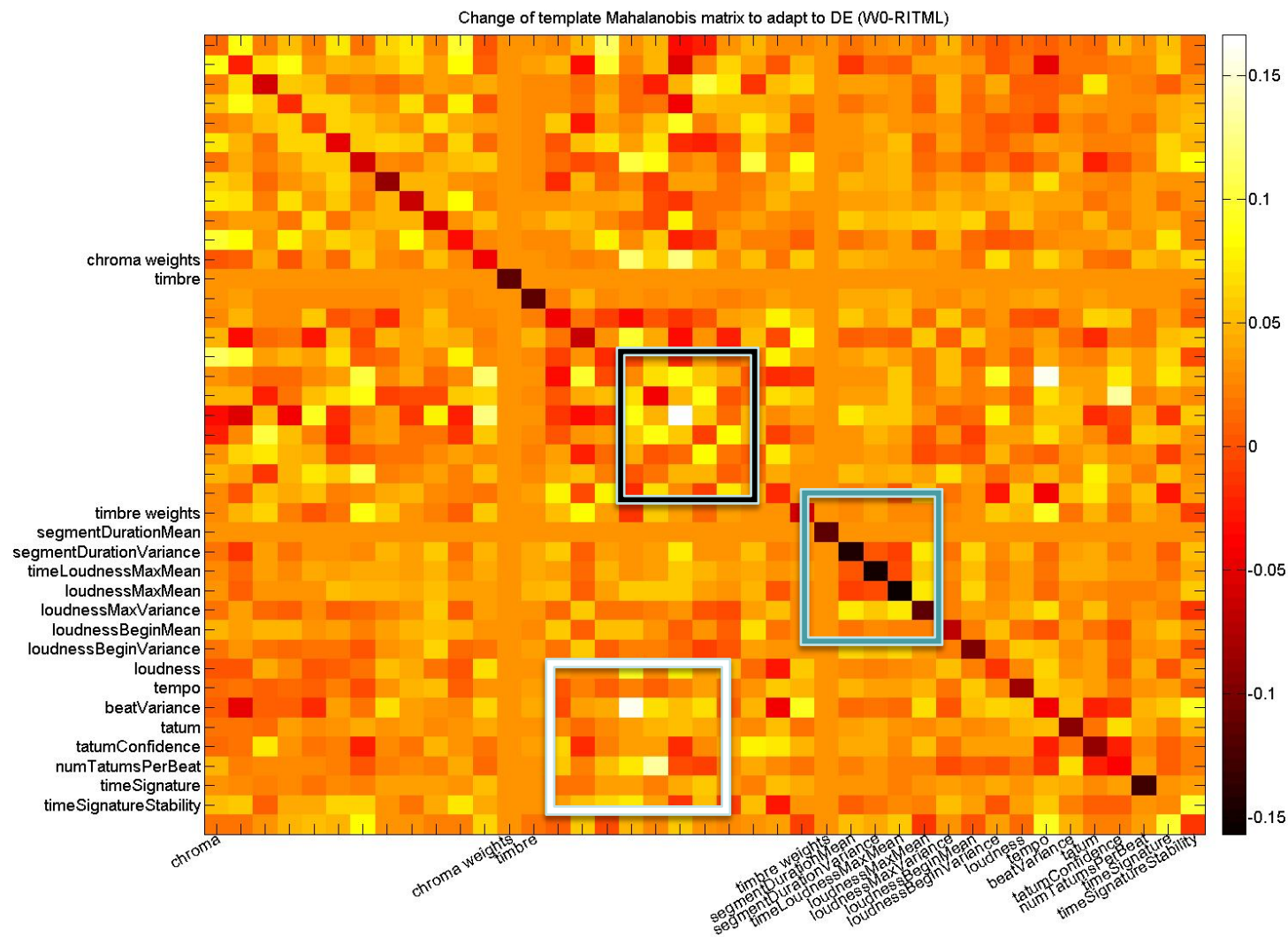
CITY UNIVERSITY
LONDON



Difference to W_0



CITY UNIVERSITY
LONDON



Conclusion

- Introduced transfer learning via W_0 for relative similarity
 - Supports reuse of learning results without need for original data
 - Effects of fine-tuning on W show differences in similarity data

 - Used datasets too variable / small
 - to train robust models per country
 - to analyse model differences
- ⇒ More data needed to find useful country-specific models (if they exist)
- ⇒ Future goal: automatic selection of attributes