On the Evaluation of Rhythmic and Melodic Descriptors for Music Similarity

Maria Panteli, Simon Dixon
{m.panteli, s.e.dixon}@qmul.ac.uk
Centre for Digital Music, Queen Mary University of London

Aim: Assess invariance of rhythmic and melodic descriptors to transformations of timbre, recording quality, tempo, pitch and the effects of style and polyphonic structure.

Motivation: To investigate similarity in large world music collections we require features which are robust to audio transformations and agnostic to culture-specific music characteristics.

Contributions: A dataset of synthesised rhythms and melodies and a strategy for feature invariance evaluation.

Dataset

- 3000 synthesised rhythms
- 3000 synthesised melodies
- (30 rhythms/melodies x 100 transformations)

Table: Styles for each melody and rhythm*.
* (M) is monotimbral/monophonic and (P) is polytimbral/polyphonic.

Rhythm	No.	Melody	No.
Afro-American (M) [1]	5	Dutch Folk (M) [2]	5
North-Indian (M) [1]	5	Classical (M)	5
African (M) [1]	5	Byzantine (M)	5
Classical (M)	5	Pop (M) [3]	5
EDM (P)	5	Classical (P)	5
Latin-Brazilian (P)	5	Pop (P) [3]	5

Table: Transformations for each melody and rhythm.

Transformations	Values
Timbre (Timb)	25 distinct timbres (similar freq. range and instrument)
Global Tempo (GTemp)	25 deviations in [-20, 20] percent from orig. tempo
Rec. Quality (RecQ) [4]	25 degradations including reverb, compression, noise
Key Transpos. (KeyT) - melody only	25 transpostions in [-10, 10] semitones from orig. key
Local Tempo (LTemp) - rhythm only	25 deviations in [-20, 20] percent from orig. tempo

Features

- low-level descriptors
- for content description and similarity estimation

Table: Rhythmic and melodic descriptors.

Rhythm	Melody
Scale Transform (ST) [5]	Pitch Bihistogram (PB) [8]
Onset Patterns (OP) [6]	Intervalgram (IG) [9]
Fluctuation Patterns (FP) [7]	2D Fourier Magnitudes (FTM) [10]

Results

Evaluation method:

- classification experiments: predict one of the 30 rhythms or melodies using 5-fold cross validation.
- retrieval experiments: query one of the 30 rhythms or melodies and assess the recall rate of its 99 transformations.

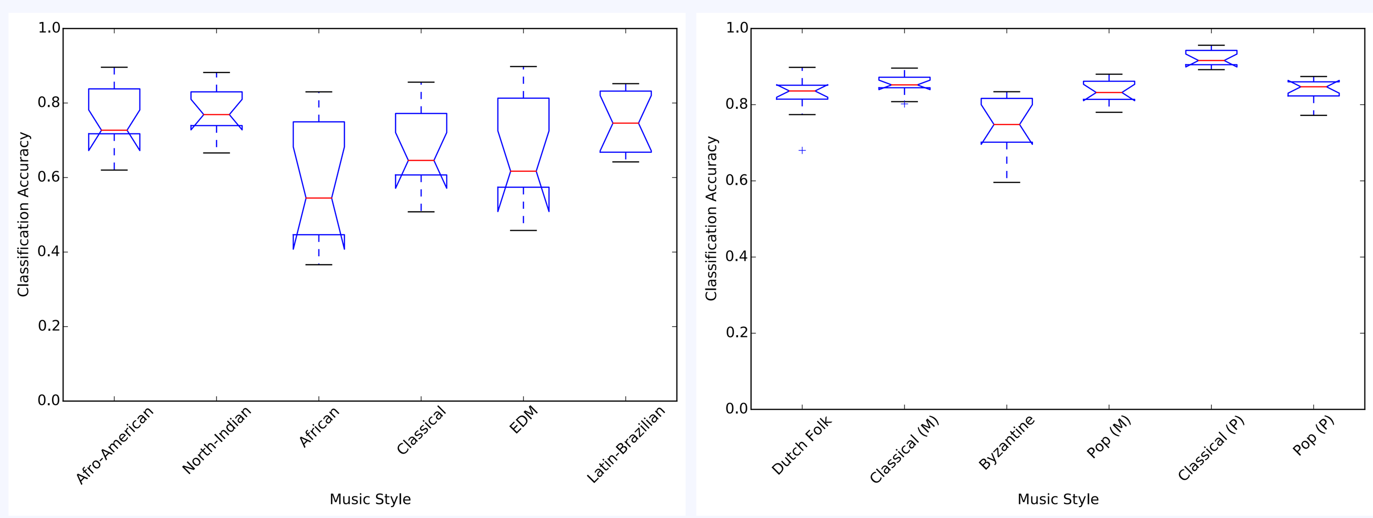
Table: Overall accuracy for classification* and retrieval experiments.
*classifiers: KNN-K Nearest Neighbors, LDA-Linear Discriminant analysis, NB-Naive Bayes, SVM-Support Vector Machines.

	Rhythm			Melody		
Metric	ST	OP	FP	PB	IG	FTM
Classification						
KNN	**0.86**	0.71	0.68	**0.88**	0.83	0.86
LDA	**0.82**	0.66	0.59	**0.83**	0.82	0.82
NB	**0.80**	0.62	0.58	**0.84**	0.76	0.81
SVM	**0.87**	0.66	0.59	0.86	0.86	**0.87**
Retrieval						
Euclidean	**0.65**	0.47	0.42	**0.80**	0.56	0.67
Cosine	**0.66**	0.47	0.42	**0.80**	0.55	0.68
Correlation	**0.66**	0.47	0.42	**0.80**	0.54	0.67
Mahalanobis	**0.61**	0.48	0.40	**0.81**	0.60	0.72

Table: Classification accuracy per transformation.

Metric	Feature	Timb	GTemp	RecQ	LTemp
Classification					
KNN	ST	**0.98**	**0.90**	**0.93**	0.62
KNN	OP	0.97	0.20	0.92	**0.75**
KNN	FP	0.91	0.18	0.92	0.71
KNN	PB	0.97	**0.99**	**0.78**	0.76
KNN	IG	0.95	**0.99**	0.62	0.77
KNN	FTM	**0.98**	0.96	0.71	**0.79**

Figure: Classification accuracy per style (left: rhythm, right: melody).



Conclusions: Scale transform most invariant for rhythm similarity and pitch bihistogram most invariant for melody. Descriptors are not completely invariant to music style characteristics (low accuracy for African rhythms and Byzantine melodies).

References

[1] Thul and Toussaint, BRIDGES 2008.
[2] Van Kranenburg et al., Meertens Institute 2014.
[3] Bittner et al., ISMIR 2014.
[4] Mauch and Ewert, ISMIR 2013.
[5] Holzapfel and Stylianou, ICASSP 2011.
[6] Esparza et al., JNMR 2014.
[7] Pampalk et al., ISMIR 2005.
[8] Van Balen et al., ISMIR 2014.
[9] Walters et al., CMMR 2012.
[10] Bertin-Mahieux and Ellis, ISMIR 2012.