

# THE 2007 MUSIC INFORMATION RETRIEVAL EVALUATION EXCHANGE (MIREX 2007) RESULTS OVERVIEW

The IMIRSEL Group led by J. Stephen Downie  
Graduate School of Library and Information Science  
University of Illinois at Urbana-Champaign

## MIREX Task Results

Audio Genre Classification		
Rank	Participant	Avg. Raw Accuracy
1	IMIRSEL (svm)	68.29%
2	Lidy, Rauber, Pertusa & Iñesta	66.71%
3	Mandel & Ellis	66.60%
4	Mandel & Ellis (spec)	65.50%
5	Tzanetakis, G.	65.34%
6	Guaus & Herrera	62.89%
7	IMIRSEL (knn)	54.87%

Audio Classical Composer Identification		
Rank	Participant	Avg. Raw Accuracy
1	IMIRSEL (svm)	53.72%
2	Mandel & Ellis (spec)	52.02%
3	IMIRSEL (knn)	48.38%
4	Mandel & Ellis	47.84%
5	Lidy, Rauber, Pertusa & Iñesta	47.26%
6	Tzanetakis, G.	44.59%
7	Lee, K.	19.70%

Multi F0 Estimation		
Rank	Participant	Accuracy
1	Ryynänen & Klapuri (1)	0.605
2	Yeh, C.	0.589
3	Zhou & Reiss	0.582
4	Pertusa & Iñesta	0.580
5	Vincent, Bertin & Badeau (2)	0.543
6	Cao, Li, Liu & Yan (1)	0.510
7	Raczyński, Ono & Sagayama	0.484
8	Vincent, Bertin & Badeau (1)	0.466
9	Poliner & Ellis (1)	0.444
10	Leveau, P.	0.394
11	Cao, Li, Liu & Yan (2)	0.359
12	Egashira, Kameoka & Sagayama (2)	0.336
13	Egashira, Kameoka & Sagayama (1)	0.327
14	Cont, A. (2)	0.311
15	Cont, A. (1)	0.277
16	Emiya, Badeau & David (1)	0.145

Audio Onset Detection		
Rank	Participant	Avg. F-Measure
1	Zhou & Reiss	0.808
2	Lee, Shiu & Kuo (0.3)	0.802
3	Lee, Shiu & Kuo (0.4)	0.801
4	Lee, Shiu & Kuo (0.2)	0.800
5	Lee, Shiu & Kuo (lp)	0.796
	Röbel, A. (1)	0.796
	Röbel, A. (3)	0.796
8	Röbel, A. (2)	0.793
	Röbel, A. (4)	0.793
10	Stowell & Plumbley (cd)	0.784
11	Stowell & Plumbley (som)	0.770
12	Stowell & Plumbley (pow)	0.769
13	Stowell & Plumbley (rcd)	0.762
14	Stowell & Plumbley (wpd)	0.753
15	Lacoste, A.	0.743
16	Stowell & Plumbley (mkl)	0.717
17	Stowell & Plumbley (pd)	0.565

Audio Mood Classification		
Rank	Participant	Avg. Raw Accuracy
1	Tzanetakis, G.	61.50%
2	Laurier, C.	60.50%
3	Lidy, Rauber, Pertusa & Iñesta	59.67%
4	Mandel & Ellis	57.83%
5	Mandel & Ellis (spec)	55.83%
	IMIRSEL (svm)	55.83%
7	Lee, K. (1)	49.83%
8	IMIRSEL (knn)	47.17%
9	Lee, K. (2)	25.67%

Audio Artist Identification		
Rank	Participant	Avg. Raw Accuracy
1	IMIRSEL (svm)	48.14%
2	Mandel & Ellis (spec)	47.16%
3	Mandel & Ellis	40.46%
4	Lidy, Rauber, Pertusa & Iñesta	38.76%
5	Tzanetakis, G.	36.70%
6	IMIRSEL (knn)	35.29%
7	Lee, K.	9.71%

Multi F0 Note Tracking		
Rank	Participant	Avg. F-Measure
1	Ryynänen & Klapuri (2)	0.614
2	Vincent, Bertin & Badeau (4)	0.527
3	Poliner & Ellis (2)	0.485
4	Vincent, Bertin & Badeau (3)	0.453
5	Pertusa & Iñesta (2)	0.408
6	Egashira, Kameoka & Sagayama (4)	0.268
7	Egashira, Kameoka & Sagayama (3)	0.246
8	Pertusa & Iñesta (3)	0.219
9	Emiya, Badeau & David (2)	0.202
10	Cont, A. (4)	0.093
11	Cont, A. (3)	0.087

Query by Singing/Humming (Jang's Collection)		
Rank	Participant	MRR
1	Wu & Li (2)	0.925
2	Wu & Li (1)	0.909
3	Jang, Lee & Hsu (2)	0.872
4	Jang, Lee & Hsu (1)	0.704
5	Lemström & Mikkilä	0.576
6	Gómez, Abad-Mota & Ruckhaus	0.477
7	Ferraro, Hanna, Allali & Robine	0.355
8	Uitdenbogerd, A. (1)	0.240
9	Uitdenbogerd, A. (3)	0.110
10	Uitdenbogerd, A. (2)	0.093

Audio Cover Song Identification		
Rank	Participant	Avg. Precision
1	Serrà & Gómez	0.521
2	Ellis & Cotton	0.330
3	Bello, J.	0.267
4	Jensen, Ellis, Christensen & Jensen	0.238
5	Lee, K. (1)	0.130
6	Lee, K. (2)	0.086
7	Kim & Perelstein	0.061
8	IMIRSEL	0.017

Audio Music Similarity		
Rank	Participant	Avg. Fine Score
1	Pohle & Schnitzer	0.568
2	Tzanetakis, G.	0.554
3	Barrington, Turnbull, Torres & Lanskrict	0.541
4	Bastuck, C. (1)	0.539
5	Lidy, Rauber, Pertusa & Iñesta (1)	0.519
6	Mandel & Ellis	0.512
7	Lidy, Rauber, Pertusa & Iñesta (2)	0.491
8	Bastuck, C. (2)	0.446
9	Bastuck, C. (3)	0.439
10	Bosteels & Kerre (1)	0.412
11	Paradzinets & Chen	0.377
12	Bosteels & Kerre (2)*	0.178

\* The results for Bosteels & Kerre (2) were interpolated from partial data due to a runtime error.

Query by Singing/Humming (ThinkIT Collection)		
Rank	Participant	MRR
1	Wu & Li 2 (Wu pitches)	0.937
2	Wu & Li 1 (Wu notes)	0.917
3	Wu & Li 2 (Jang pitches)	0.883
4	Gómez, Abad-Mota & Ruckhaus (Wu notes)	0.715
5	Lemström & Mikkilä (Wu notes)	0.618
6	Jang, Lee & Hsu 2 (Jang pitches)	0.536
7	Ferraro, Hanna, Allali & Robine (Wu notes)	0.452
8	Jang, Lee & Hsu 2 (Wu pitches)	0.379
9	Jang, Lee & Hsu 1 (Jang pitches)	0.345
10	Jang, Lee & Hsu 1 (Wu pitches)	0.305

## MIREX 2007 Challenges

- Non-compliance with input/output formats
- Robustness issues:
  - Tolerance for silence, graceful failing, scalability issues (dealing with larger data sets), parallelizability, writing out results incrementally (useful for failures)
- Library dependency issues
- Pre-compiled MEX files not running
- JAVA object serialization/de-serialization issues
- Low quality and trustworthiness of the metadata for music
- General shortage of new test sets
- Effectively managing results output: From standard out to Wiki

Symbolic Melodic Similarity			
Rank	Participant	ADR	Fine
1	Gómez, Abad-Mota & Ruckhaus (1)	0.712	0.586
2	Gómez, Abad-Mota & Ruckhaus (2)	0.739	0.581
3	Ferraro, Hanna, Allali & Robine	0.730	0.540
4	Uitdenbogerd, A. (3)	0.706	0.532
5	Uitdenbogerd, A. (1)	0.666	0.532
6	Uitdenbogerd, A. (2)	0.698	0.528
7	Pinto, A. (1)	0.031	0.292
8	Pinto, A. (2)	0.024	0.281

Special Thanks to: The Andrew W. Mellon Foundation, the National Science Foundation (Grant No. NSF IIS-0327371 and No. NSF IIS-0328471), the content providers and the MIR community, all of the IMIRSEL group - M. Bay, A. Ehmann, A. Gruzdz, X. Hu, M. C. Jones, J. H. Lee, M. McCrory, K. West, X. Xiang and all the volunteers who evaluated the MIREX results, and the ISMIR 2007 organizing committee