

# MIREX 2016 Evaluation Results

J. Stephen Downie & IMIRSEL  
University of Illinois at Urbana-Champaign  
Email: jdownie@illinois.edu

## Classical Composer Identification

SubID	Participants	▼ Accuracy (2011) 0.79
LS1	Thomas Lidy, Alexander Schindler	0.6760
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.6180

## Latin Classification - MIREX08 Dataset

SubID	Participants	▼ Accuracy (2011) 0.8231
LS1	Thomas Lidy, Alexander Schindler	0.6988
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.6278

## Mood Classification - MIREX08 Dataset

SubID	Participants	▼ Accuracy (2011) 0.695
LS1	Thomas Lidy, Alexander Schindler	0.6333
LS2	Thomas Lidy, Alexander Schindler	0.6250
LS3	Thomas Lidy, Alexander Schindler	0.6033
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.5700

## Genre Classification (Mixed) - MIREX08 Dataset

SubID	Participants	▼ Accuracy (2014) 0.8355
WWWJ1	Ming-Ju Wu, JC Wu, Jyh-Shing Roger Jang	0.7627
LS1	Thomas Lidy, Alexander Schindler	0.7314
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.7044

## KPOP Genre Classification (Annotated by Korean Annotators)

SubID	Participants	▼ Accuracy (2014) 0.6558
LS1	Thomas Lidy, Alexander Schindler	0.6436
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.5713

## KPOP Genre Classification (Annotated by American Annotators)

SubID	Participants	▼ Accuracy (2014) 0.6325
LS1	Thomas Lidy, Alexander Schindler	0.6235
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.5824

## KPOP Mood Classification (Annotated by Korean Annotators)

SubID	Participants	▼ Accuracy (2014) 0.6235
LS1	Thomas Lidy, Alexander Schindler	0.6075
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.5971

## KPOP Mood Classification (Annotated by American Annotators)

SubID	Participants	▼ Accuracy (2014) 0.6423
LS1	Thomas Lidy, Alexander Schindler	0.6298
FFF1	Juliano Henrique Foleiss, Tiago Fernandes Tavares	0.5901

## Multi-F0 Estimation

SubID	Participants	Accuracy	
		▼ MIREX Dataset (2014) 0.723	Su Dataset (2015) 0.39
MM1	Matija Marolt	0.537	0.310
CB1	Chris Cannam, Emmanouil Benetos	0.486	0.234
DT1	Daylin Troxel	0.44	0.016
CB2	Chris Cannam, Emmanouil Benetos	0.42	0.221

## Multi-F0 Note Tracking - Mixed Dataset

SubID	Participants	MIREX Dataset		Su Dataset	
		▼ Avg. F-measure Onset Only	Avg. Overlap	▼ Avg. F-measure Onset Only	Avg. Overlap
		(2014) 0.821	(2012) 0.893	(2015) 0.319	(2015) 0.837
DT1	Daylin Troxel	0.712	0.852	0.002	0.089
MM1	Matija Marolt	0.618	0.877	0.315	0.719
CB1	Chris Cannam, Emmanouil Benetos	0.503	0.865	0.228	0.731
CB2	Chris Cannam, Emmanouil Benetos	0.373	0.862	0.165	0.803

## Multi-F0 Note Tracking - Piano Only

SubID	Participants	MIREX Dataset		Su Dataset	
		▼ Avg. F-measure Onset Only	Avg. Overlap	▼ Avg. F-measure Onset Only	Avg. Overlap
		(2016) 0.82	(2009) 0.844	(2015) 0.500	(2015) 0.862
DT1	Daylin Troxel	0.82	0.796	0.004	0.000
MM1	Matija Marolt	0.754	0.813	0.470	0.819
CB1	Chris Cannam, Emmanouil Benetos	0.667	0.813	0.369	0.838
CB2	Chris Cannam, Emmanouil Benetos	0.497	0.797	0.252	0.774
KB1 (with bugs)	Rainer Kelz, Sebastian Böck	0.485	0.118	0.063	0.000

## Onset Detection

SubID	Participants	▼ Avg. F-measure (2016) 0.8732
SB3	Jan Schlüter, Sebastian Böck	0.8537
SB4	Sebastian Böck	0.852
CC3	ChunTa Chen	0.8407
SB5	Sebastian Böck	0.8393
SB7	Sebastian Böck	0.8335
SB6	Sebastian Böck	0.8247
LSY1	Che-Yuan Liang, Li Su, Yi-Hsuan Yang	0.8154
LSY2	Che-Yuan Liang, Li Su, Yi-Hsuan Yang	0.7345
CC2	Chris Cannam	0.6763
CS1	Chris Cannam, Dan Stowell	0.3659
CZ2	Chaogang Zhang	0.3648
CZ1	Chaogang Zhang	0.2057
YF1	Yin Feng	

## Cover Song Identification

SubID	Participants	Total Number of Covers in Top 10	
		▼ Mixed (2016) 1918	Sapp's Mazuraka (2016) 5147
CL1	Ning Chen, Hai-Dong Xiao	1371	4919
CYWW1	Chuan-Yau Chan, Ming-Chi Yen, Ju-Chiang Wang, Hsin-Min Wang	103	5020
SYBK1	Diego Silva, Chin-Chia Michael Yeh, Gustavo Enrique de Almeida Prado Alves Batista, Eamonn Keogh	6	1
YWJ2	Ming-Chi Yen, Hsin-Min Wang, Jyh-Shing Roger Jang	3	2
YWJ3	Ming-Chi Yen, Hsin-Min Wang, Jyh-Shing Roger Jang		

## Structure Segmentation

SubID	Participants	▼ MIREX '09 F-measure (2012) 0.6528	MIREX '10 RWC/Quaero SB@3sec (2015) 0.7934	MIREX '10 RWC/RWC F-measure (2012) 0.6883	Salami F-measure (2012) 0.5809
		MC1	Matthias Mauch, Chris Cannam	0.6063	0.4915
ON4	Oriol Nieto	0.5261	0.5282	0.5701	0.5271
CC1	Chris Cannam	0.5054	0.6039	0.5136	0.4491
ON1	Oriol Nieto	0.4992	0.6242	0.5173	0.4632
ON2	Oriol Nieto	0.4966	0.6044	0.5255	0.4708
ON3	Oriol Nieto	0.4896	0.6625	0.5261	0.4608
ON5	Oriol Nieto	0.4881	0.6100	0.5343	0.4719

## Singing Voice Separation

SubID	Participants	▼ Singing Voice (GNSDR, dB) (2015) 6.8236	Music (GNSDR, dB) (2016) 11.1878
		LCP2	Yi Luo, Zhuo Chen, John R. Hershey, Jonathan Le Roux, Daniel P. W. Ellis
LCP1	Yi Luo, Zhuo Chen, John R. Hershey, Jonathan Le Roux, Daniel P. W. Ellis	6.0726	10.9256
MC3	Marius Miron, Pritish Chandna	5.492	9.8049
MC2	Marius Miron, Pritish Chandna	5.2891	9.6678
HC1	Yi-Chun Huang, Tai-Shih Chi	4.6309	7.818
RSGP1	Gerard Roma, Emad M. Grais, Andrew J. R. Simpson, Mark D. Plumbley	3.2589	8.7664
GD1	Georgi Dzhambazov	-2.281	0.3954

## Audio Key Detection

SubID	Participants	Weighted Key Score	
		▼ MIREX 2015 Dataset (2013) 0.8683	GiantSteps Dataset
CN1	Chris Cannam, Katy Noland	0.817	0.4697
BD1	Gilberto Bernardes, Matthew Davies	0.7613	0.6230
FJH2	Angel Faraldo, Sergi Jordà, Perfecto Herrera	0.7491	0.4465
FJH3	Angel Faraldo, Sergi Jordà, Perfecto Herrera		(2016) 0.6826

## Real-time Audio to Score Alignment (a.k.a. Score Following) Results

SubID	Participants	▼ Total Precision (2016) 0.9743
VCR1	Francisco Jose Rodriguez-Serrano, Pedro Vera-Candeas, Julio Jose Carabias-Orti	0.9484
VCR2	Francisco Jose Rodriguez-Serrano, Pedro Vera-Candeas, Julio Jose Carabias-Orti	

## Audio Beat Tracking

SubID	Participants	F-measure		
		▼ SMC Dataset (2015) 58.1935	MAZ Dataset (2015) 74.2809	MCK Dataset (2015) 63.8961
BK1	Sebastian Böck, Florian Krebs	56.8313	57.5286	63.6093
BK3	Sebastian Böck, Filip Korzeniowski	52.8343	73.8911	62.5299
BK2	Sebastian Böck, Florian Krebs	52.3142	55.1651	62.7315
SB9	Sebastian Böck	52.0970	58.9357	63.8961
SB8	Sebastian Böck	49.8366	52.2792	63.8436
JZ1	Jose R. Zapata	36.8192	50.6300	52.6104
JZ2	Jose R. Zapata	36.4073	47.8740	53.2314
CD3	Chris Cannam, Matthew Davies	33.6630	49.6229	52.8758
CD2	Chris Cannam, Simon Dixon	30.3400	41.5218	52.6579

## Set List Identification

SubID	Participants	▼ Avg. ED (2015) 11.50	Avg. sBD (2015) 131.147	Avg. eBD (2015) 160.695
		YWYW1	Ming-Chi Yen, Ju-Chiang Wang, Yi-Hsuan Yang, Hsin-Min Wang	11.35
YWJ1	Ming-Chi Yen, Hsin-Min Wang, Jyh-Shing Roger Jang	10.35	131.147	160.695

## Query-by-Singing/Humming Results

SubID	Participants	▼ Hidden Jang Dataset (MRR) (2016) 0.9537	Jang Dataset (MRR) (2016) 0.9734	ThinkIt Dataset (MRR) (2014) 0.9689	IOACAS Dataset (MRR) (2015) 0.8232	Subtask2 Dataset (Simple Count) (2016) 9.7162
		AT5	ACRCloud Team	0.9537	0.9734	0.9646
AT4	ACRCloud Team	0.9537	0.9734	0.9689	-	9.7162
BS2	Bartłomiej Stasiak	0.8699	0.9178	0.8464	0.6754	9.6824
BS1	Bartłomiej Stasiak	0.8655	0.9160	0.8012	0.6407	9.6647



# MIREX 2016 Evaluation Results

J. Stephen Downie & IMIRSEL  
University of Illinois at Urbana-Champaign  
email: jdownie@illinois.edu

## Audio Chord Estimation (MIREX 2009)

SubID	Participants	▼ Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation
		(2016) 86.09	(2016) 85.53	(2016) 82.24	(2015) 76.04	(2015) 74.43	(2016) 87.76	(2016) 87.97	(2015) 91.57
FK2	Filip Korzeniowski	86.09	85.53	82.24	74.42	71.54	87.76	85.79	90.73
KO1	Maksim Khadkevich, Maurizio Omologo	82.93	82.19	79.61	76.04	73.43	87.69	85.66	91.24
FK4	Filip Korzeniowski	82.28	80.93	78.03	70.91	68.26	85.62	82.4	90.89
DK3	Junqi Deng, Yu-Kwong Kwok	80.03	77.55	74.79	68.4	65.88	85.81	82.5	91.53
DK1	Junqi Deng, Yu-Kwong Kwok	79.21	76.19	74	66.02	64.15	85.71	82.62	91.23
CM1	Chris Cannam, Matthias Mauch	78.56	75.41	72.48	54.67	52.26	85.9	87.17	86.09
DK2	Junqi Deng, Yu-Kwong Kwok	77.84	74.49	71.93	61.61	59.47	85.82	82.72	91.28
DK4	Junqi Deng, Yu-Kwong Kwok	76.05	72.96	71.41	62.77	61.44	78.19	87.97	72.43

## Audio Chord Estimation (Billboard 2012)

SubID	Participants	▼ Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation
		(2016) 85.64	(2016) 85.38	(2016) 82.55	(2016) 60.7	(2016) 58.53	(2016) 87.62	(2016) 87.72	(2016) 90.13
FK2	Filip Korzeniowski	85.64	85.38	82.55	60.7	58.38	87.62	86.09	90.13
FK4	Filip Korzeniowski	79.23	78.62	76.2	56.53	54.51	85.09	81.98	89.94
KO1	Maksim Khadkevich, Maurizio Omologo	77.45	75.58	73.51	57.68	55.82	84.16	82.8	87.44
DK3	Junqi Deng, Yu-Kwong Kwok	75.92	74.75	72.69	53.42	51.67	83.39	79.97	88.92
DK1	Junqi Deng, Yu-Kwong Kwok	75.28	73.57	71.87	59.98	58.53	83.35	80.26	88.52
CM1	Chris Cannam, Matthias Mauch	74.15	72.22	70.21	55.35	53.4	83.64	85.31	83.39
DK2	Junqi Deng, Yu-Kwong Kwok	73.77	71.69	69.86	58.66	57	83.57	80.4	88.7
DK4	Junqi Deng, Yu-Kwong Kwok	72.59	70.85	69.78	56.29	55.36	76.13	87.72	70.05

## Audio Chord Estimation (Billboard 2013)

SubID	Participants	▼ Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation
		(2016) 80.07	(2016) 77.89	(2016) 75.42	(2016) 55.41	(2016) 53.29	(2016) 82.94	(2016) 86.31	(2015) 88.74
FK2	Filip Korzeniowski	80.07	77.89	75.42	55.41	53.22	82.94	82.43	86.8
KO1	Maksim Khadkevich, Maurizio Omologo	75.36	71.39	69.43	53.57	51.78	81.63	79.61	87.75
FK4	Filip Korzeniowski	74.66	71.85	69.44	51.93	49.8	80.61	77.19	88.7
DK3	Junqi Deng, Yu-Kwong Kwok	72.39	68.53	66.55	48.99	47.28	80.76	77.26	88.3
DK1	Junqi Deng, Yu-Kwong Kwok	72.06	68.69	67.26	54.54	53.29	80.82	77.58	88.06
CM1	Chris Cannam, Matthias Mauch	71.16	67.28	65.2	48.99	47.17	81.54	83.11	82.63
DK2	Junqi Deng, Yu-Kwong Kwok	70.18	66.54	64.66	52.97	51.41	80.85	77.68	88.02
DK4	Junqi Deng, Yu-Kwong Kwok	69.56	65.83	64.78	51.81	50.93	74.55	86.31	69.18

## Audio Chord Estimation (JayChou)

SubID	Participants	▼ Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation
		(2016) 79.51	(2016) 78.66	(2015) 73.1	(2015) 64.43	(2015) 61.39	(2015) 88.46	(2016) 87.87	(2016) 92.08
FK2	Filip Korzeniowski	79.51	78.66	68.15	50.69	42.34	86.81	85.43	88.56
KO1	Maksim Khadkevich, Maurizio Omologo	78.73	77.69	66.87	54.16	44.55	88.46	87.12	90.11
FK4	Filip Korzeniowski	76.13	75.44	64.36	49.69	40.74	84.55	81.22	88.95
DK3	Junqi Deng, Yu-Kwong Kwok	75.01	74.75	63.56	49.27	40.24	86.76	82.54	92.08
DK1	Junqi Deng, Yu-Kwong Kwok	74.7	73.87	70.33	54.98	52.25	86.76	82.78	91.79
CM1	Chris Cannam, Matthias Mauch	72.75	72.08	65.48	54.39	48.98	86.6	86.89	86.91
DK2	Junqi Deng, Yu-Kwong Kwok	72.19	72.55	69.1	54.09	51.46	87.09	83.35	91.75
DK4	Junqi Deng, Yu-Kwong Kwok	71.51	69.03	65.93	50.07	47.45	78.11	87.87	70.56

## Audio Chord Estimation (RobbieWilliams) - NEW

SubID	Participants	▼ Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation
FK2	Filip Korzeniowski	88.53	87.23	84.19	82.57	79.88	90.04	88.62	91.88
KO1	Maksim Khadkevich, Maurizio Omologo	83.55	80.33	78.16	73.54	71.39	88.04	85.39	91.68
FK4	Filip Korzeniowski	83.37	80.96	78.42	77.04	74.76	87.22	84.5	91.02
CM1	Chris Cannam, Matthias Mauch	81.9	78.25	76.05	57.92	55.9	87.96	88.96	87.45
DK3	Junqi Deng, Yu-Kwong Kwok	81.85	78.56	76.16	74.71	72.55	86.98	82.95	92.34
DK1	Junqi Deng, Yu-Kwong Kwok	81.5	77.77	76.1	68.88	67.34	87.03	83.22	92.11
DK2	Junqi Deng, Yu-Kwong Kwok	79.01	75.97	73.57	65.26	62.98	87.2	83.4	92.23
DK4	Junqi Deng, Yu-Kwong Kwok	78.92	75.15	73.66	66.72	65.34	81.82	88.44	76.88

## Audio Downbeat Estimation Results

SubID	Participants	F-measure							
		▼ Ballroom	Beatles	Carnatic	Turkish	Cretan	Hjdb	Rwc_classical	Gtzan
		(2016) 0.908*	(2016) 0.872*	(2015) 0.474	(2014) 0.775*	(2014) 0.854*	(2016) 0.970*	(2016) 0.599*	(2016) 0.647
BK4	Sebastian Böck, Florian Krebs	0.908*	0.865*	0.369*	0.537*	0.635*	0.970*	0.599*	0.638
KB1	Florian Krebs, Sebastian Böck	0.898*	0.803	0.269	0.352	0.433	0.690	0.436	0.630
KB2	Florian Krebs, Sebastian Böck	0.86*	0.818*	0.330*	0.336*	0.443*	0.851*	0.428*	0.647
DBDR1	Simon Durand, Juan Bello, Bertrand David, Gael Richard	0.838*	0.849	0.201	0.306	0.426	0.578	0.527*	0.615
DBDR2	Simon Durand, Juan Bello, Bertrand David, Gael Richard	0.783	0.872*	0.231	0.415	0.418	0.629	0.532*	0.619
DSR1	Matthew Davies, Adam Stark, Andrew Robertson	0.463	0.665	0.184	0.317	0.265	0.208	0.251	0.397
CD4	Matthew Davies, Chris Cannam	0.412	0.604	0.186	0.218	0.250	0.334	0.174	0.460

\*) Rows marked by an asterisk should be taken with care as in those cases overlapping test and training sets were used. This could lead to overestimated metrics.

## Audio Melody Extraction Results

SubID	Participants	Overall Accuracy						
		ADC04 Dataset ▼	MIREX05 Dataset	INDIAN08 Dataset	MIREX09 0dB Dataset	MIREX09 -5dB Dataset	MIREX09 +5dB Dataset	Orcheset15 Dataset
		(2014) 0.863	(2014) 0.7484	(2016) 0.8444	(2016) 0.7669	(2016) 0.6116	(2015) 0.869	(2016) 0.5928
IY1	Yukara Ikemiya	0.7187	0.6738	0.8444	0.6799	0.5310	0.7935	0.3204
WFJY1	Chung-Che Wang	0.7035	0.5732	0.8011	0.7539	0.6105	0.8243	0.1851
WFJY3	Chung-Che Wang	0.6989	0.5706	0.7988	0.7543	0.6116	0.8243	0.1863
BG2	Juan J, Emilia Gómez	0.697	0.6370	0.7637	0.5847	0.4631	0.6672	0.5310
BG1	Juan J, Emilia Gómez	0.6886	0.5814	0.7323	0.4621	0.3660	0.5398	0.5928
WFJY2	Chung-Che Wang	0.6677	0.5769	0.8160	0.7669	0.6077	0.8466	0.1574
KON1	Sangeun Kum	0.6236	0.5774	0.8312	0.7005	0.5761	0.7962	0.2143
FJ1	Zhe-Cheng Fan	0.6034	0.5863	0.6819	0.7653	0.6075	0.8638	0.1411
FJ2	Zhe-Cheng Fan	0.5563	0.4748	0.6027	0.6507	0.4570	0.8020	0.0990

Special Thanks to:  
The Andrew W. Mellon Foundation, the National Science Foundation (Grant No. NSF IIS-0327371 and No. NSF IIS-0328471), task captains, IMIRSEL members (K. Choi, J. Downie, Y. Hao, P. Organisciak, B. Owen), the content providers, Evalutron graders, the MIR community, and the ISMIR 2016 organizing committee.