



MIREX 2018 Evaluation Results

J. Stephen Downie, Yun Hao and Task Captains

University of Illinois at Urbana-Champaign

Email: jdownie@illinois.edu



Train-Test Classification Tasks

SubID	Participants	Accuracy			
		Mood	K-Mood by K	K-Mood by A	K-Genre by A
		(2017) 0.6983	(2017) 0.6534	(2017) 0.6534	
GS1	Guangxiao Song, Shenyi Ding, Zhijie Wang	0.5533	0.5303	0.5685	
WB1	Wenhao Bian	0.6117	-	-	

Task Captain: Yun Hao (IMIRSEL)

Audio Fingerprinting

SubID	Participants	Size of Database (MB)	Top-1 Hit Rate (%)
			(2015) 91.88
PK3	Harikrishnan Potty, Srikanth Konjeti	554.4	90.43
PK4	Harikrishnan Potty, Srikanth Konjeti	743.13	90.27
PK6	Harikrishnan Potty, Srikanth Konjeti	554.4	90.43
PK7	Harikrishnan Potty, Srikanth Konjeti	556.37	91.62
PK8	Harikrishnan Potty, Srikanth Konjeti	556.29	91.3
TY1	Te-Lun Yang	4110.43	91.41
TY2	Te-Lun Yang	4152.14	90.78
TY3	Te-Lun Yang	4110.43	91.41
ZL1	Yuanzhong Zheng, Huaping Liu	927.48	86.1
ZL3	Yuanzhong Zheng, Huaping Liu	1291.19	91.73

Task Captain: Chung-Che Wang

Audio Melody Extraction

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	(2018) 0.6437	(2015) 0.869	(2016) 0.5928
KN1_Dense	Sanguen Kum, Juhan Nam	0.7034	0.6388	0.7605	0.6843	0.5646	0.7777	0.237
KN1_Res	Sanguen Kum, Juhan Nam	0.6916	0.6285	0.7666	0.7588	0.6352	0.8296	0.2086
KN3	Sanguen Kum, Juhan Nam	0.6772	0.6462	0.758	0.7648	0.6437	0.8309	0.2252
LS1	Li Su	0.6586	0.5046	0.6894	0.7219	0.5933	0.7756	0.123

Task Captain: Derek Wu

Audio Chord Estimation – Part. 1

SubID	Participants
CM1	Chris Cannam, Matthias Mauch
FK2	Florian Krebs, Filip Korzeniowski, Sebastian Böck
JLCX1, JLCX2	Junyan Jiang, Wei Li, Ke Chen, Guangyu Xia
SG1	Franz Strasser, Stefan Gaser

SubID	Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation	
		(2017) 87.38	(2017) 86.8	(2018) 84.44	(2015) 76.04	(2015) 74.43	(2018) 90.33	(2018) 89.03	(2018) 93.48
CM1	78.66	75.51	72.58	54.78	52.36	85.87	87.22	85.98	
FK2	87.38	86.8	83.43	75.55	72.6	89.29	87.24	92.35	
JLCX1	86.75	86.25	84.44	75.87	74.39	90.33	88.36	93.38	
JLCK2	86.51	86.05	84.23	75.64	74.17	90.12	87.94	93.48	
SG1	82.03	78.67	76.84	69.2	67.55	82.34	89.03	77.87	
		(2017) 86.30	(2017) 86.02	(2017) 83.12	(2018) 71.26	(2018) 70.11	(2017) 88.38	(2018) 90.21	(2017) 90.97
CM1	74.23	72.31	70.25	55.44	53.48	83.6	85.33	83.31	
FK2	86.3	86.01	83.12	61.12	58.75	88.38	86.78	90.96	
JLCX1	83.09	82.79	81.48	70.88	69.72	87.72	86.65	89.85	
JLCK2	83.6	83.31	82.01	71.26	70.11	88.35	86.94	90.9	
SG1	82.11	79.28	78.04	66.57	65.53	81.6	90.21	76.22	
		(2017) 80.6	(2017) 78.37	(2018) 76.74	(2018) 64.22	(2018) 63.2	(2018) 85.55	(2018) 87.7	(2018) 90.28
CM1	71.23	67.36	65.28	49.07	47.25	81.5	83.13	82.53	
FK2	80.6	78.37	75.85	55.81	53.57	83.59	83.01	87.54	
JLCX1	78.76	77.3	76.14	63.69	62.71	84.86	83.12	89.21	
JLCK2	79.38	77.92	76.74	64.22	63.2	85.55	83.53	90.28	
SG1	77.92	72.41	71.04	58.26	57.1	77.95	87.7	73.46	

Task Captain: Johan Pauwels

Audio Chord Estimation – Part. 2

SubID	Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Combined Hamming Measure	Avg. Under-Segmentation	Avg. Over-Segmentation	
		(2018) 84.77	(2018) 84.49	(2018) 81.75	(2015) 64.43	(2015) 61.39	(2018) 93.16	(2018) 92.26	(2018) 94.16
CM1	72.82	72.15	65.55	54.46	49.05	86.58	86.94	86.84	
FK2	81.42	80.48	69.66	51.77	43.18	89.08	87.67	90.87	
JLCX1	84.41	84.12	81.42	63.14	61.08	92.59	91.82	93.46	
JLCK2	84.77	84.49	81.75	63.38	61.31	93.16	92.26	94.16	
SG1	80.59	75.65	72.28	56.43	53.8	82.51	91.49	75.3	
		(2017) 89.38	(2017) 88.12	(2017) 85.06	(2017) 83.38	(2017) 80.68	(2018) 91.03	(2018) 90.85	(2018) 93.73
CM1	81.94	78.29	76.09	57.97	55.94	87.95	89	87.39	
FK2	89.38	88.12	85.06	83.38	80.68	90.96	89.52	92.84	
JLCX1	87.14	85.33	83.31	75.91	74.1	90.95	89.3	93.23	
JLCK2	87.25	85.45	83.4	76.01	74.17	91.03	89.04	93.73	
SG1	84.87	80.28	78.58	73.78	72.16	85.24	90.85	80.75	
		(2017) 87.20	(2018) 86.71	(2018) 84.93	(2018) 73.78	(2018) 72.17	(2018) 91.15	(2018) 90.66	(2018) 94.16
CM1	79.12	77.95	74.3	63.33	59.89	88.62	88.11	89.67	
FK2	87.2	86.58	80.92	58.04	52.78	89.94	87.94	92.37	
JLCX1	87.11	86.71	84.93	73.78	72.17	91.15	88.77	93.99	
JLCK2	86.9	86.53	84.76	73.51	71.88	91.07	88.48	94.16	
SG1	82.22	78.06	76.06	63.65	61.86	82.62	90.66	76.44	
		(2018) 85.66	(2018) 84.89	(2018) 82.95	(2018) 75.08	(2018) 73.35	(2018) 89.57	(2018) 90.37	(2018) 92
CM1	78.41	76.24	72.74	59.84	56.67	85.81	86.81	86.02	
FK2	82.34	80.72	76.04	64.5	60.3	86.59	85.49	89.17	
JLCX1	85.66	84.89	82.95	75.08	73.35	89.57	88.22	91.88	
JLCK2	85.34	84.67	82.72	74.86	73.12	89.35	87.72	92	
SG1	81.67	77.38	75.16	65.69	63.68	83.46	90.37	78.59	

Task Captain: Johan Pauwels

Multi-F0 Estimation

SubID	Participants	MIREX Dataset	Su Dataset
		Accuracy	Accuracy
		(2014) 0.723	(2017) 0.51
CB1	Chris Cannam, Emmanouil Benetos	0.498	0.236
CB2	Chris Cannam, Emmanouil Benetos	0.42	0.224

Task Captain: Yun Hao (IMIRSEL)

Multi-F0 Note Tracking – Mixed Dataset

SubID	Participants	Avg. F-measure Onset-Offset	Avg. F-measure Onset Only	Avg. Overlap Onset-Offset
		(2014) 0.5817	(2014) 0.821	(2017) 0.903
CB1	Chris Cannam, Emmanouil Benetos	0.3047	0.5029	0.865
CB2	Chris Cannam, Emmanouil Benetos	0.2064	0.3742	0.862
		(2017) 0.1511	(2017) 0.3334	(2017) 0.841
CB1	Chris Cannam, Emmanouil Benetos	0.0614	0.2280	0.731
CB2	Chris Cannam, Emmanouil Benetos	0.0491	0.1653	0.803

Task Captain: Yun Hao (IMIRSEL)

Multi-F0 Note Tracking – Piano-Only Dataset

SubID	Participants	Avg. F-measure Onset-Offset	Avg. F-measure Onset Only	Avg. Overlap Onset-Offset
		(2017) 0.6201	(2016) 0.82	(2017) 0.86
CB1	Chris Cannam, Emmanouil Benetos	0.2216	0.7002	0.812
CB2	Chris Cannam, Emmanouil Benetos	0.2085	0.5642	0.797
KB1	Rainer Kelz, Sebastian Böck	0.5205	0.6869	0.827
		(2017) 0.182	(2018) 0.6090	(2017) 0.883
CB1	Chris Cannam, Emmanouil Benetos	0.0892	0.3686	0.838
CB2	Chris Cannam, Emmanouil Benetos	0.0744	0.2522	0.774
KB1	Rainer Kelz, Sebastian Böck	0.2136	0.6090	0.837

Task Captain: Yun Hao (IMIRSEL)

Audio Beat Tracking

SubID	Participants	F-measure		
		SMC Dataset	MAZ Dataset	MCK Dataset
		(2015) 58.1935	(2015) 74.2809	(2017) 63.9475
BK1	Sebastian Böck, Florian Krebs	56.8363	57.5132	63.6107
BK2	Sebastian Böck, Filip Korzeniowski	52.8343	73.8882	62.5296
CD1	Chris Cannam, Simon Dixon	30.3524	42	52.2974
CD2	Chris Cannam, Matthew Davies	33.6605	49.6137	52.7871
SB1	Sebastian Böck	52.1352	58.6525	63.9464
SB2	Sebastian Böck	49.8633	52.2371	63.8436

Task Captain: Aggelos Gkiokas

Automatic Lyrics-to-Audio Alignment – Part. 1

Subtask 1: A cappella Mandarin Chinese pop songs

SubID	Participants	Clean Recordings			Mixed Recordings		
		Mean ASE	Mean PCS	Mean PCSTW	Mean ASE	Mean PCS	Mean PCSTW
CW1	Chung-Che Wang	0.12	0.82	0.94	3.35	0.364	0.54
GSLW1	Chitralkha Gupta,						
GSLW2	Bidisha Sharma,						
GSLW3	Haizhou Li,						

Task Captains: Rong Gong, Georgi Dzhambazov

Automatic Lyrics-to-Audio Alignment – Part. 2

Subtask 2: Mixed English Pop Songs

SubID	Participants	Hansen Clean Recordings			Hansen Mixed Recordings		
		Mean ASE	Mean PCS	Mean PCSTW	Mean ASE	Mean PCS	Mean PCSTW
CW2	Chung-Che Wang	2.33	0.51	0.64	3.57	0.38	0.57
CW3	Chung-Che Wang	0.47	0.62	0.81	2.07	0.45	0.70
GSLW1	Chitralkha Gupta,						
GSLW2	Bidisha Sharma,						
GSLW3	Haizhou Li,						

SubID	Participants	Mauch Mixed Recordings			Gracenote Mixed Recordings		
		Mean ASE	Mean PCS	Mean PCSTW	Mean ASE	Mean PCS	Mean PCSTW
CW2	Chung-Che Wang	6.92	0.25	0.37	0.70	0.46	0.61
CW3	Chung-Che Wang	4.13	0.35	0.52	2.33	0.51	0.64
GSLW1	Chitralkha Gupta,						
GSLW2	Bidisha Sharma,						
GSLW3	Haizhou Li,						

Task Captains: Rong Gong, Georgi Dzhambazov

Audio Onset Detection

SubID	Participants	Avg. F-measure
		(2017) 0.8737
AR3	Axel Roebel	0.8604
AR4	Axel Roebel	0.8572
CC1	Chris Cannam	0.7404
CS1	Chris Cannam, Dan Stowell	0.6769
KPK1	Ricardo Koscky, Flavio Pereira,	0.7982
KPK2	Karin Komati	0.8024
SB4	Jan Schlüter, Sebastian Böck	0.8737
SB5	Sebastian Böck	0.8534
SB6	Sebastian Böck	0.8346
SB7	Sebastian Böck	0.8404

Task Captain: Sebastian Böck

Cover Song Identification

SubID	Participants	Total Number of Covers in Top 10	
		Mixed Dataset	Sapp's Mazuraka
		(2017) 2111	(2018) 5179
CL1	Ning Chen, Mingyu Li	2060	5179
LCLL1	Juheon Lee, Sungkyun Chang, Donmoon Lee, Kyogu Lee	2104	5072

Task Captains: Yun Hao (IMIRSEL), Chris Tralie

Patterns for Prediction (NEW)

Explicit task – Algorithms take an excerpt of music as input (the prime), and output a predicted continuation of the excerpt

Implicit task - Algorithms take a prime and one or more continuations as input, and output the likelihood that each continuation is the genuine extension of the prime.

SubID	Participants	Explicit Task				Implicit Task	
		F1_Pitch	F1-IOI	F1-Pairs	Recall-IOI	Accuracy (Monophonic)	Accuracy (Polyphonic)
EN1	Eric Nichols	0.66	0.87	0.69	0.87	-	-
FC1	Florian Colombo	0.8	0.91	0.82	0.92	0.87*	0.92*
MM	Task Captains	0.77	0.92	0.79	0.87	-	-

*Accuracy greater than 0.54 constitutes significant performance above chance

Task Captains: Iris Ren, Berit Janssen, Tom Collins

Audio Key Detection

SubID	Participants	Weighted Key Score				
		MIREX 2015 (2018)	GiantSteps (2017)	Billboard 2012 (2018)	Isophonics (2017)	RobbieWilliams (2018)
		88.78	87.38	88.22	86.30	81.38
CG1	David Castells-Rufas, Adria Galin	88.78	28.21	63.01	61.86	54.66
CN1	Chris Cannam, Katy Noland	88.07	50.53	67.4	64.99	62.28
FK1	Filip Korzeniowski	82.5	74.62	88.22	82.3	81.38
GC1	Adria Galin, David Castells-Rufas	76.58	30.03	69.66	68.05	62.7
NA1	Nestor Napoles Lopez, Claire Arthur	84.06	59.42	79.59	75.72	77.2
OM1	James Owers,	82.52	58.11	71.58	62.92	63.86
OM2	Andrew McLeod	73.43	49.72	55.41	51.15	51.32
OM3		49.64	48.58	39.04	41.13	57.99

Task Captain: Johan Pauwels

Audio Downbeat Estimation

SubID	Participants	F-measure						
		Ballroom	Beatles	Carnatic	Turkish	Cretan	Hjdb	Gtzan
		(2016) 0.908	(2016) 0.872	(2015) 0.474	(2014) 0.775	(2014) 0.854	(2016) 0.970*	(2018) 0.663
BK3	Sebastian Böck, Florian Krebs	0.907	0.860	0.369	0.538	0.635	0.970	0.639
JCXYL1	Bijue Jia, Dongdong Chen, Huanmin Xu, Xue Yang, Jiancheng Lv	0.100	0.024	0.030	0.039	0.021	0.018	0.018
KB2	Florian Krebs, Sebastian Böck	0.874	0.604	0.355	0.338	0.268	0.883	0.663
WCL1	Jian Wang, Dongdong Chen, Jiancheng Lv	0.105	0.098	0.081	0.084	0.173	0.119	0.098

Task Captains: Mickaël Zehren, Paolo Bientinesi

Audio Tempo Estimation

SubID	Participants	P-Score (8% tolerance)	P-Score (4% tolerance)
		MIREX06 Dataset	
CD3	Chris Cannam, Matthew Davies	0.7386	
GK1	Aggelos Gkiokas, Vassilis Katsouras	0.8290	
HS1	Hendrik Schreiber	0.7939	
HS2	Hendrik Schreiber	0.8789	
HS3	Hendrik Schreiber	0.8716	
OL1	Olivier Lartillot	0.8141	
OL2	Olivier Lartillot	0.8056	
SB3	Sebastian Böck	0.8977	
ZL4	Zeyu Li	0.6064	
GiantSteps			
CD3	Chris Cannam, Matthew Davies	0.3962	
GK1	Aggelos Gkiokas, Vassilis Katsouras	0.8758	
HS1	Hendrik Schreiber	0.9604	
HS2	Hendrik Schreiber	0.9648	
HS3	Hendrik Schreiber	0.9598	
OL1	Olivier Lartillot		
OL2	Olivier Lartillot		
SB3	Sebastian Böck	0.8457	
ZL4	Zeyu Li	0.9507	

Task Captain: Aggelos Gkiokas

Drum Transcription

SubID	Participants	3-class Datasets								8-class Datasets	
		mean fm	sum fm	KD mean fm	KD sum fm	SD mean fm	SD sum fm	HH mean fm	HH sum fm	mean fm	sum fm
		(2017) 0.71	(2018) 0.74	(2017) 0.82	(2018) 0.79	(2017) 0.7	(2018) 0.69	(2018) 0.54	(2018) 0.54	(2018) 0.62	(2018) 0.68
CS2	Carl Southall	0.63	0.7	0.78	0.78	0.59	0.59	0.5	0.5	0.47	0.54
CS4	Carl Southall	0.62	0.68	0.77	0.77	0.58	0.58	0.47	0.47	0.5	0.56
JAR1	Céline Jacques, Achille Aknin, Axel Roebel Jacques	0.69	0.72	0.78	0.78	0.65	0.65	0.53	0.53	-	-
JAR2	Céline Jacques, Achille Aknin, Axel Roebel Jacques	0.68	0.73	0.78	0.78	0.65	0.65	0.53	0.53	-	-
JAR3	Céline Jacques, Achille Aknin, Axel Roebel Jacques	0.69	0.73	0.79	0.79	0.65	0.65	0.54	0.54	-	-
JAR5	Céline Jacques, Achille Aknin, Axel Roebel Jacques	0.67	0.71	0.77	0.77	0.63	0.63	0.53	0.53	-	-
JS1	Julien Schroeter	0.61	0.62	0.75	0.75	0.59	0.59	0.43	0.43	0.55	0.57
JS2	Julien Schroeter	0.62	0.64	0.78	0.78	0.63	0.63	0.4	0.4	0.58	0.61
JS3	Julien Schroeter	0.61	0.62	0.76	0.76	0.63	0.63	0.36	0.36	0.57	0.62
JS4	Julien Schroeter	0.57	0.58	0.73	0.73	0.61	0.61	0.32	0.32	0.54	0.59
RV1	Richard Vogl	0.69	0.74	0.78	0.78	0.69	0.69	0.53	0.53	-	-
RV2	Richard Vogl	0.65	0.72	0.75	0.75	0.65	0.65	0.5	0.5	-	-
RV3	Richard Vogl	-	-	-	-	-	-	-	-	0.62	0.68

Task Captains: Richard Vogl, Carl Southall, Chih-Wei Wu

Music and/or Speech Detection

SubID	Participants	Task 1: Music Detection				Task 2: Speech Detection				Task 3: Music and Speech Detection			
		Dataset 1		Dataset 2		Dataset 1		Dataset 2		Dataset 1		Dataset 2	
		Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure	MUSIC_F	SPEECH_F	MUSIC_F	SPEECH_F
DD1	David Doukhan	0.6860	0.5424	0.9257	0.9334	0.8770	0.9186	0.9617	0.9583	-	-	-	-
JHKK1	Byeong-Yong Jang, Woon-Haeng Heo, Jung-Hyun Kim, Oh-Wook Kwon	0.7798	0.7123	0.9415	0.9487	-	-	-	-	-	-	-	-
JHKK2	Byeong-Yong Jang, Woon-Haeng Heo, Jung-Hyun Kim, Oh-Wook Kwon	0.8005	0.7415	0.9153	0.9309	-	-	-	-	-	-	-	-
JHKK3	Byeong-Yong Jang, Woon-Haeng Heo, Jung-Hyun Kim, Oh-Wook Kwon	-	-	-	-	0.8307	0.8795	0.8575	0.8305	-	-	-	-
LN1 (GAFMFSF)	Minsuk Choi, Jongpil Lee, Juhan Nam	0.6251	0.5022	0.7814	0.8053	0.6908	0.7472	0.8636	0.8314	0.4936	0.9683	0.7855	0.8455
LN1 (GAFMF)	Minsuk Choi, Jongpil Lee, Juhan Nam	-	-	0.7751	0.7936	-	-	0.8754	0.8483	-	-	0.7592	0.7888
LN1 (GAFSF)	Minsuk Choi, Jongpil Lee, Juhan Nam	-	-	0.7996	0.8247	-	-	0.8597	0.8256	-	-	0.7968	0.8227
MM1	Matija Marolt	0.6135	0.3899	0.9150	0.9228	0.8626	0.9115	0.9367	0.9326	0.3899	0.8795	0.9228	0.9326
MM2	Matija Marolt	0.6807	0.5478	0.9032	0.9158	0.8619	0.909	0.9226	0.914	0.5478	0.8945	0.9158	0.914
MM3	Matija Marolt	0.6075	0.3124	0.8725	0.8791	0.8508	0.9086	0.8973	0.8973	0.3124	0.8383	0.8791	0.8973
MMG1	Blai Meléndez-Catalán, Emilio Molina, Emilia Gómez	0.9049	0.8996	0.9025	0.9223	-	-	-	-	-	-	-	-
MMG2	Blai Meléndez-Catalán, Emilio Molina, Emilia Gómez	0.8506	0.8211	0.9490	0.9574	-	-	-	-	-	-	-	-

Task Captains: Blai Meléndez-Catalán, Emilio Molina, David Doukhan, Jan Schlüter