



# MIREX 2019 Evaluation Results

# J. Stephen Downie, Yun Hao and Task Captains

# University of Illinois at Urbana-Champaign

Email: jdownie@illinois.edu

# I ILLINOIS

# Train-Test Classification Tasks

SubID	Participants	Composer	Latin	Mixed	Mood	Accuracy			
						K-Mood by K	K-Mood by A	K-Genre by K	K-Genre by A
		(2011) 0.79	(2011) 0.8231	(2014) 0.8355	(2017) 0.6983	(2017) 0.6534	(2017) 0.6534	(2014) 0.8355	(2017) 0.6774
AS*	Sainath Adapa, Manoj Sukhavasi	-	-	0.6800	-	0.6033	-	-	-
KN1	Taejun Kim, Juhan Nam	0.3142	0.4388	0.6069	0.6033	0.5692	0.5616	0.5882	0.5861
KN2	Taejun Kim, Juhan Nam	0.4242	0.5956	0.6773	0.6667	0.6193	0.6131	0.6241	0.6183
LJ1	Jingru Li	-	-	0.1183	0.2217	-	-	-	-
LJ3	Jingru Li	-	-	0.1173	0.2250	-	-	-	-
LJ4	Jingru Li	-	-	-	0.3000	-	-	-	-

**Task Captain: Yun Hao (IMIRSEL)**

# Audio Melody Extraction

SubID ▾	Participants	Overall Accuracy					
		ADC04 Dataset	MIREX05 Dataset	INDIAN08 Dataset	MIREX09 -5dB Dataset	MIREX09 +5dB Dataset	Orcheset15 Dataset
		(2014) 0.863	(2014) 0.7484	(2016) 0.8444	(2019) 0.8278	(2019) 0.7159	(2019) 0.8712
AH1	An-Qi Huang, Shenyuan Li, Zhi Dong	0.6052	0.6065	0.6639	0.7991	0.6844	0.8598
BH1	Jiří Balhar, Jan Hajič	0.7079	0.6786	0.6468	0.606	0.5297	0.6688
BH2	Jiří Balhar, Jan Hajič	0.6983	0.6691	0.6627	0.6414	0.5982	0.6714
HH4	Huang Hsiang-Yu	0.3314	0.3265	0.4879	0.5156	0.4754	0.5289
HLD1	An-Qi Huang, Shenyuan Li, Zhi Dong	0.5233	0.5368	0.6419	0.8278	0.7159	0.8712
KN4	Juhan Nam, Sangeun Kum	0.6786	0.6107	0.7507	0.7059	0.7931	0.8066
ZC1	Zhengyu Cao	0.664	0.5491	0.6988	0.7304	0.5808	0.2118

## **Task Captain: Anqi Huang**

# Audio Key Detection

SubID ▾	Participants	Weighted Key Score				
		MIREX 2015	GiantSteps	Billboard 2012	Isophonics	RobbieWilliams
		(2019) 89.19	(2017) 87.38	(2019) 90.41	(2017) 86.30	(2019) 84.76
CG1	David Castells-Rufas, Adria Galin	76.67	59.74	49.86	49.66	64.97
CN1	Chris Cannam, Katy Noland	88.07	50.53	67.4	64.99	62.28
CN2	Chris Cannam, Katy Noland	89.19	57.85	73.29	74.05	73.23
JXC1	Junyan Jiang, Guangyu Xia, David Carlton	86.53	65.61	85.48	79.68	84.76
JXC2	Junyan Jiang, Guangyu Xia, David Carlton	88.45	75.81	90.41	79.66	82.06
NN1	Nestor Napoles Lopez	88.65	67.91	78.42	75.7	73.86

**Task Captain: Johan Pauwels**

# Audio Fingerprinting

SubID ▾	Participants	Size of Database (MB)	Top-1 Hit Rate (%)
JY1	Jeong-soo Yu	610.05	87.53
LKZL1		714.35	86.3
LKZL2	Xiao Lu,	1248.88	88.76
LKZL4	Lester Kong,	3169.11	90.69
LKZL5	Ethan Zhao,	5840.09	87.28
LKZL6	Simon Lui	3933.24	90.76

**Task Captain: Chung-Che Wang**

# Audio Chord Estimation

SubID ▾	Participants								
	i Chien, Song Rong Lee, Yeh Ssuhung, Tzu-Chun Yeh, Jyh-Shing Roger Jang		Chris Cannam, Matthias Mauch						
SubID ▾	Root	MajMin	MajMinBass	Sevenths	SeventhsBass	Avg. Hamming	Combined Measure	Avg. Under-Segmentation	Avg. Over-Segmentation
<b>Isophonics2009</b>									
	(2017) 87.38	(2017) 86.8	(2018) 84.44	(2015) 76.04	(2015) 74.43		(2018) 90.33	(2018) 89.03	(2018) 93.48
CLSYJ1	79.89	76.5	74.33	67.79	65.82		82.61	81.56	85.11
CM1	78.66	75.51	72.58	54.78	52.36		85.87	87.22	85.98
<b>Billboard 2012</b>									
	(2017) 86.30	(2017) 86.02	(2017) 83.12	(2018) 71.26	(2018) 70.11		(2017) 88.38	(2018) 90.21	(2017) 90.97
CLSYJ1	76.32	73.38	71.6	63.59	61.92		82.05	81.74	83.77
CM1	74.23	72.31	70.25	55.44	53.48		83.6	85.33	83.31
<b>Billboard 2013</b>									
	(2017) 80.6	(2017) 78.37	(2018) 76.74	(2018) 64.22	(2018) 63.2		(2018) 85.55	(2018) 87.7	(2018) 90.28
CLSYJ1	73.69	69.47	67.66	57.41	55.78		78.8	78.67	82.52
CM1	71.23	67.36	65.28	49.07	47.25		81.5	83.13	82.53
<b>Jay Chou</b>									
	(2018) 84.77	(2018) 84.49	(2018) 81.75	(2015) 64.43	(2015) 61.39		(2018) 93.16	(2018) 92.26	(2018) 94.16
CLSYJ1	65.44	62.49	55.27	46.83	40.6		78.95	80.78	77.55
CM1	72.82	72.15	65.55	54.46	49.05		86.58	86.94	86.84
<b>Robbie Williams</b>									
	(2017) 89.38	(2017) 88.12	(2017) 85.06	(2017) 83.38	(2017) 80.68		(2018) 91.03	(2018) 90.85	(2018) 93.73
CLSYJ1	79.89	75.45	73.73	68.57	66.91		83.6	83.72	84.32
CM1	81.94	78.29	76.09	57.97	55.94		87.95	89	87.39
<b>RWC-Popular</b>									
	(2017) 87.20	(2018) 86.71	(2018) 84.93	(2018) 73.78	(2018) 72.17		(2018) 91.15	(2018) 90.66	(2018) 94.16
CLSYJ1	70.33	66.77	63.42	56.09	52.97		79.26	79.17	80.14
CM1	79.12	77.95	74.3	63.33	59.89		88.62	88.11	89.67
<b>USPOP2002 Chords</b>									
	(2018) 85.66	(2018) 84.89	(2018) 82.95	(2018) 75.08	(2018) 73.35		(2018) 89.57	(2018) 90.37	(2018) 92
CLSYJ1	76.23	72.3	68.75	62.27	58.92		82.13	82.12	83.24
CM1	78.41	76.24	72.74	59.84	56.67		85.81	86.81	86.02

**Task Captain: Johan Pauwels**

# Automatic Lyrics-to-Audio Alignment

Participants				
SubID ▾				
ED1	Emir Demirel			
GYL1	Chitralekha Gupta, Haizhou Li, Emre Yilmaz			
SDE1	Daniel Stoller, Simon Durand, Sebastian Ewert			
SDE2	Daniel Stoller, Simon Durand, Sebastian Ewert			
SubID ▾	Avg. absolute error	Median absolute error	Correct segments (%)	Correct onsets with tolerance (%)
<b>Hansen Clean Recordings</b>				
ED1	0.28	0.04	0.74	0.93
GYL1	0.13	0.03	0.79	0.97
SDE2	0.21	0.08	0.47	0.94
<b>Hansen Mixed Recordings</b>				
ED1	3.47	0.26	0.49	0.69
GYL1	0.1	0.04	0.78	0.97
SDE2	0.39	0.09	0.44	0.88
<b>Mauch Mixed Recordings</b>				
ED1	11.65	8.46	NA	0.39
GYL1	0.19	0.1	NA	0.91
SDE1	0.34	0.11	NA	0.88
SDE2	0.26	0.11	NA	0.87
<b>Gracenote Mixed Recordings</b>				
ED1	0.23	0.06	NA	0.85
GYL1	0.09	0.04	NA	0.96
<b>Jamendo Lyrics</b>				
ED1	7.29	4.78	NA	0.57
GYL1	0.22	0.05	NA	0.94
SDE1	0.82	0.1	NA	0.85
SDE2	0.38	0.1	NA	0.87

Task Captains: Georgi Dzhambazov Daniel Stoller

## Multi-F0 Estimation

SubID ▾	Participants	MIREX Dataset		Su Dataset
		Accuracy	Accuracy	
		(2014) 0.723	(2017) 0.51	
AR2	Anton Runov	<b>0.69</b>	0.406	
CB1	Chris Cannam, Emmanouil Benetos	0.498	0.234	
CB2	Chris Cannam, Emmanouil Benetos	0.42	0.221	
HH2	Huang Hsiang-Yu	0.141	0.053	
SBJ1	Peter Steiner, Peter Birkholz, Azarakhsh Jalalvand	0.632	0.385	
SBJ2	Peter Steiner, Peter Birkholz, Azarakhsh Jalalvand	0.618	0.419	
SBJ3	Peter Steiner, Peter Birkholz, Azarakhsh Jalalvand	0.626	0.439	
SBJ4	Peter Steiner, Peter Birkholz, Azarakhsh Jalalvand	0.611	<b>0.455</b>	
Task Captain: Yun Hao (IMIRSEL)				

## Multi-F0 Note Tracking – Piano-Only Dataset

SubID ▾	Participants	Avg. F-measure		Avg. Overlap Onset-Offset
		Onset-Offset	Onset Only	
<b>MIREX Dataset</b>				
		(2017) 0.6201	(2016) 0.82	(2017) 0.86
AR2	Anton Runov	0.2339	0.6468	0.806
BK1	Sebastian Böck, Rainer Kelz	<b>0.4816</b>	0.6464	<b>0.824</b>
CB1	Chris Cannam, Emmanouil Benetos	0.238	0.6681	0.813
CB2	Chris Cannam, Emmanouil Benetos	0.175	0.497	0.797
KN3	Taegyun Kwon, Juhan Nam, Dasaem Jung	0.3431	0.6009	0.769
KNJ1	Taegyun Kwon, Juhan Nam, Dasaem Jung	0.3941	0.6653	0.809
KY1	Changhyun Kim, Sangeon Yong	0.4476	0.6731	0.803
KY2	Changhyun Kim, Sangeon Yong	0.4068	0.6713	0.819
YK1	Sangeon Yong, Changhyun Kim	0.4031	0.7023	0.804
YK2	Sangeon Yong, Changhyun Kim	0.3976	<b>0.7032</b>	0.819
<b>Su Dataset</b>				
		(2017) 0.182	(2019) 0.6297	(2017) 0.883
AR2	Anton Runov	0.1009	0.3768	0.583
BK1	Sebastian Böck, Rainer Kelz	0.2005	0.5446	<b>0.72</b>
CB1	Chris Cannam, Emmanouil Benetos	0.0892	0.3686	0.585
CB2	Chris Cannam, Emmanouil Benetos	0.0744	0.2522	0.576
KN3	Taegyun Kwon, Juhan Nam, Dasaem Jung	<b>0.2006</b>	0.5501	0.602
KNJ1	Taegyun Kwon, Juhan Nam, Dasaem Jung	0.0743	0.6038	0.615
KY1	Changhyun Kim, Sangeon Yong	0.1602	0.6141	0.624
KY2	Changhyun Kim, Sangeon Yong	0.1324	0.6245	0.588
YK1	Sangeon Yong, Changhyun Kim	0.1553	<b>0.6297</b>	0.553
YK2	Sangeon Yong, Changhyun Kim	0.1387	0.6235	0.584
Task Captain: Yun Hao (IMIRSEL)				

## Multi-F0 Note Tracking – Mixed Dataset

SubID ▾	Participants	Avg. F-measure	Avg. F-measure	Avg. Overlap
		Onset-Offset	Onset Only	Onset-Offset
<b>MIREX Dataset</b>				
		(2014) 0.5817	(2014) 0.821	(2017) 0.903
AR2	Anton Runov	0.406	0.624	<b>0.881</b>
CB1	Chris Cannam,	0.305	0.503	0.865
CB2	Emmanouil Benetos	0.206	0.374	0.862
KY1	Changhyun Kim,	<b>0.438</b>	<b>0.674</b>	0.845
KY2	Sangeon Yong	0.415	0.664	0.848
YK1	Sangeon Yong,	0.399	0.639	0.844
YK2	Changhyun Kim	0.392	0.642	0.846
<b>Su Dataset</b>				
		(2017) 0.1511	(2017) 0.3334	(2017) 0.841
AR2	Anton Runov	0.0856	0.2742	<b>0.808</b>
CB1	Chris Cannam,	0.0614	0.228	0.731
CB2	Emmanouil Benetos	0.0491	0.1653	0.803
KY1	Changhyun Kim,	0.1262	0.3979	0.676
KY2	Sangeon Yong	0.1175	0.4047	0.676
YK1	Sangeon Yong,	<b>0.1317</b>	<b>0.4422</b>	0.683
YK2	Changhyun Kim	0.1235	0.4412	0.679
Task Captain: Yun Hao (IMIRSEL)				

## Cover Song Identification

SubID ▾	Participants	Total Number of Covers in Top 10	
		Mixed Dataset	Sapp's Mazuraka
		(2019) 2368	(2018) 5179
CZWZ1	Zimo Cheng, Junfan Zhao, Guanbo Wang, Wei-Qiang Zhang	75	1609
JS1	Jin Soo Seo	-	5146
JS2		-	5041
YXCY1	Zhesong Yu, Xiaoshuo Xu, Xiaoou Chen, Deshun Yang	2368	5143
Task Captains: Yun Hao (IMIRSEL)			

## Patterns for Prediction – Part. 1

SubID ▾	Participants	Explicit Task					
		symMono (Modulo12Pitch)			symPoly (Pitch)		
		Mean	Median	STD	Mean	Median	STD
TD1	Timothy de Reuse	0.596	0.612	0.292	0.594	0.598	0.238
	Task Captains	0.583	0.608	0.195	0.506	0.508	0.176
Task Captains: Iris Ren, Berit Janssen, Tom Collins							

## Patterns for Prediction – Part. 2

SubID ▾	Participants	Implicit Task				
		Data	Oberservations	Accuracy	Mean Probability	Variance
EP1	Jeff Ens, Philippe Pasquier	mono	500	1	1	0
EP1	Jeff Ens, Philippe Pasquier	poly	500	0.998	0.992	0.002
YB2	Adrien Ycart, Emmanouil Benetos	poly	499	0.703	0.527	0.004
YB5	Adrien Ycart, Emmanouil Benetos	poly	499	0.731	0.554	0.012
Task Captains: Iris Ren, Berit Janssen, Tom Collins						

## Audio Beat Tracking

SubID ▾	Participants	F-measure		
		SMC Dataset	MAZ Dataset	MCK Dataset
		(2015) 58.1935	(2015) 74.2809	(2017) 63.9475
BD1	Sebastian Böck, Matthew Davies			
CD1	Chris Cannam, Matthew Davies			
CD2	Chris Cannam, Matthew Davies			
GBK1	Aggelos Gkiokas, Grigoris Bastas, Vassilis Katsouras			
SB1	Sebastian Böck			
Task Captain: Aggelos Gkiokas				

## Drum Transcription

|
<th rowspan
|  |

## Patterns for Prediction

SubID ▾ Participants	Explicit Task								Implicit Task					
	symMono (Modulo12Pitch)			symPoly (Pitch)			nr_obs	accuracy	crossentropy	nll_var	nr_obs	accuracy	crossentropy	nll_var
	Mean	Median	STD	Mean	Median	STD								
EP1	Jeff Ens,Philippe Pasquier	-	-	-	-	-	500	1.000	0.313	0.000	500	0.998	0.318	0.001
TD1	Timothy de Reuse	0.596	0.612	0.292	0.594	0.598	0.238	-	-	-	-	-	-	-
YB2	Adrien Ycart,Emmanouil Benetos	-	-	-	-	-	-	-	-	-	499	0.703	0.668	0.003
YB5	Adrien Ycart,Emmanouil Benetos	-	-	-	-	-	-	-	-	-	499	0.731	0.646	0.010
	Task Captains	0.583	0.608	0.195	0.506	0.508	0.176	-	-	-	-	-	-	-
Task Captains: Iris Ren, Berit Janssen, Tom Collins														

## Patterns for Prediction – Part. 1

SubID ▾	Participants	Explicit Task					
		symMono (Modulo12Pitch)			symPoly (Pitch)		
		Mean	Median	STD	Mean	Median	STD
TD1	Timothy de Reuse	0.596	0.612	0.292	0.594	0.598	0.238
	Task Captains	0.583	0.608	0.195	0.506	0.508	0.176
Task Captains: Iris Ren, Berit Janssen, Tom Collins							

## Patterns for Prediction – Part. 2

SubID ▾ Participants	data	Implicit Task			
		Oberservations	Accuracy	Mean Probability	Variance
EP1	Jeff Ens,Philippe Pasquier	mono	500	1	1
EP1	Jeff Ens,Philippe Pasquier	poly	500	0.998	0.992
YB2	Adrien Ycart,Emmanouil Benetos	poly	499	0.703	0.527
YB5	Adrien Ycart,Emmanouil Benetos	poly	499	0.731	0.554
Task Captains: Iris Ren, Berit Janssen, Tom Collins					

### Audio Onset Detection

SubID ▾	Participants	Avg. F-measure
		(2017) 0.8737
BS1	Sebastian Böck, Jan Schlüter	
CC1	Chris Cannam	
CS1	Chris Cannam, Dan Stowell	
ML1	Mingtai lin	
ML2	Mingtai lin	
SB4	Sebastian Böck	
SB5	Sebastian Böck	
SB6	Sebastian Böck	
SB7	Sebastian Böck	
Task Captain: Sebastian Böck		

### 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	<b>0.907</b>	<b>0.860</b>	<b>0.369</b>	<b>0.538</b>	<b>0.635</b>	<b>0.970</b>	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	<b>0.663</b>
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)
<b>MIREX06 Dataset</b>			
		(2016) 0.8981	
CD3	Chris Cannam, Matthew Davies	0.7386	
GK1	Aggelos Gkiokas, Vassilis Katsouros	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	<b>0.8977</b>	
ZL4	Zeyu Li	0.6064	
<b>GiantSteps</b>			
CD3	Chris Cannam, Matthew Davies	0.3962	
GK1	Aggelos Gkiokas, Vassilis Katsouros	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			