Detection and classification of acoustic scenes and events: an *IEEE AASP* challenge

Dimitrios Giannoulis[1], Emmanouil Benetos[3], Dan Stowell[1], Mathias Rossignol[2], Mathieu Lagrange[2], and Mark Plumbley[1]

[1]Centre for Digital Music
School of Electronic Engineering & Computer Science
Queen Mary University of London
{dimitrios,dan.stowell,mark.plumbley}@eecs.qmul.ac.uk

[2]Analysis-Synthesis team
IRCAM, CNRS-STMS
{mathieu.lagrange,matthias.rossignol}@ircam.fr

[3]Music Informatics group
City University London
emmanouil.benetos.1@city.ac.uk

Abstract: What is this challenge about?

A public evaluation on the performance of systems dealing with two important subtasks in Computational Auditory Scene Analysis: the classification of acoustic scenes and the detection of acoustic events. Main challenge aims:

- Help the research community move a step forward in better defining the specific tasks.
- Offer a comprehensive evaluation framework and methodology for the tasks.
- Provide incentive for researchers to pursue research in this field by making available a well-structured and fully documented datset.
- Help to shed light on controversies that currently exist in the tasks.

Challenge results

Task SC: Acoustic Scene Classification

A train/test classification task: 10 classes, 10 x 30-sec audio recordings per class. Five-fold crossvalidation.

Participants	Code	Method	Lang
Chum et al.	CHR	Various features at 2 frame sizes, classified either: (a) per-frame SVM + majority voting; (b) HMM	Matlab
Elizalde	ELF	**See poster:** "An I-Vector Based Approach ..."	Matlab
Geiger et al.	GSR	**See poster:** "Large-Scale Audio Feature Extraction and SVM ..."	Weka/ HTK
Krijnders and ten Holt	KH	"A Tone-Fit Feature Representation ..."	Python
Li et al.	LTT	Wavelets, MFCCs and others, classified in 5-second windows by treebagger, majority voting	Matlab
Nam et al.	NHL	Feature learning by sparse RBM, then event detection and max-pooling, classified by SVM	Matlab
Nogueira et al.	NR1	MFCCs + MFCC temporal modulations + event density estimation + binaural modelling features, feature selection, classified by SVM	Matlab
Olivetti	OE	Normalised compression distance (Vorbis), Euclidean embedding, classified by Random Forest	Python
Patil and Elhilali	PE	**See poster:** "Multiresolution Auditory Representations ..."	Matlab
Rakotomamonjy and Gasso	RG	**See poster:** "Histogram of Gradients of Time-Frequency Representations ..."	Matlab
Roma et al.	RNH	**See poster:** "Recurrence Quantification Analysis Features ..."	Matlab
Baseline		MFCCs, classified with a bag-of-frames approach	Python

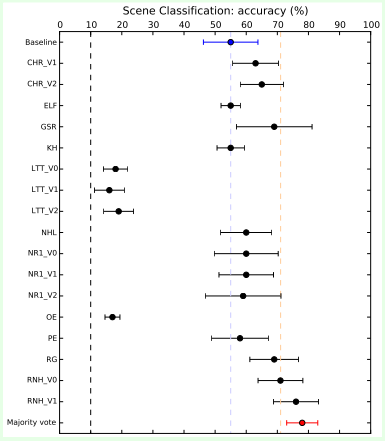

Figure: Classification accuracies (mean and 95% CI). Dashed lines indicate: chance performance (black); baseline (light blue); mean human accuracy (orange). "Majority vote" is a meta-classifier using the majority decision over all submissions.

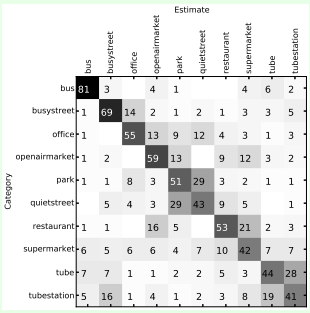

Table: Confusion matrix, aggregated over all submitted algorithms.

Task OL: Acoustic Event Detection (Live Monophonic Audio)

Task: identify the start, end and class of events in an office scene. Monophonic sequences, real recordings. 16 classes.

Participants	Code	Method	Lang
Chauhan et al.	CPS	Feature extraction - Segmentation - Likelihood ratio test classification	Matlab
Diment et al.	DHV	MFCCs (features) - HMMs (detection)	Matlab
Gemmeke et al.	GVV	**See poster:** "An Exemplar-Based NMF Approach ..."	Matlab
Niessen et al.	NVM	**See poster:** "Hierarchical Modeling Using Automated Sub-Clustering ..."	Matlab
Nogueira et al.	NR2	**See poster:** "Automatic event classification using front end single channel noise reduction, ..."	Matlab
Schröder et al.	SCS	**See poster:** "On the Use of Spectro-Temporal Features ..."	Matlab
Vuegen et al.	VVK	**See poster:** "An MFCC-GMM Approach ..."	Matlab
Baseline		NMF with pre-extracted bases (detection)	Matlab

	Evaluation Method									
	Event-Based				Class-Wise Event-Based				Frame-Based	
System	F (%)	F_{offset} (%)	$AEER$	$AEER_{offset}$	F (%)	F_{offset} (%)	$AEER$	$AEER_{offset}$	F (%)	$AEER$
CPS	2.23	1.65	2.285	2.301	0.65	0.49	1.872	1.891	3.82	2.116
DHV	26.67	22.43	2.519	2.676	30.72	25.29	2.182	2.370	26.0	3.128
GVV	15.52	13.46	1.779	1.831	13.21	12.03	1.556	**1.606**	31.94	1.084
NVM_1	32.57	24.95	1.864	2.095	29.37	21.80	1.639	1.899	40.85	1.115
NVM_2	34.16	26.28	1.852	2.095	33.05	24.88	1.602	1.877	42.76	1.102
NVM_3	34.51	27.01	1.827	2.052	33.52	24.65	1.575	1.846	45.50	1.212
NVM_4	30.47	24.68	1.906	2.083	28.17	21.62	1.650	1.849	42.86	1.360
NR2	19.21	15.26	3.076	3.244	21.54	17.64	2.857	3.010	34.66	1.885
SCS_1	39.47	36.74	1.669	1.749	36.33	34.20	1.579	1.677	53.02	1.167
SCS_2	**45.17**	**41.06**	**1.601**	**1.727**	**41.51**	**38.32**	**1.511**	1.646	**61.52**	1.016
VVK	30.77	25.40	2.054	2.224	24.55	20.36	1.762	1.949	43.42	**1.001**
Baseline	7.38	1.58	5.900	6.318	9.00	1.86	5.960	6.462	10.72	2.590

Table: Evaluation measures: the F-measure and the *acoustic event error rate* (AEER). Each is calculated three ways: per event, per event (normalised for class prevalence), and per audio frame.

Task OS: Acoustic Event Detection (Synthetic Polyphonic Audio)

Task: identify the start, end and class of events in an office scene. Synthetic mixtures of recorded events, with controlled polyphony. 16 classes.

Participants	Code	Method	Lang
Diment et al.	DHV	MFCCs (features) - HMMs (detection)	Matlab
Gemmeke et al.	GVV	**See poster:** "An Exemplar-Based NMF Approach ..."	Matlab
Vuegen et al.	VVK	**See poster:** "An MFCC-GMM Approach ..."	Matlab
Baseline		NMF with pre-extracted bases (detection)	Matlab

	Event-Based				Class-Wise Event-Based				Frame-Based	
System	F (%)	F_{offset} (%)	$AEER$	$AEER_{offset}$	F (%)	F_{offset} (%)	$AEER$	$AEER_{offset}$	F (%)	$AEER$
DHV	**8.45**	6.18	4.741	4.860	**9.73**	**7.58**	4.028	4.147	**13.08**	8.426
GVV	7.69	**7.33**	1.913	1.920	6.69	6.51	1.584	1.591	10.30	**1.553**
VVK	5.80	5.28	**1.885**	**1.895**	5.10	4.77	**1.436**	**1.445**	5.77	2.106
Baseline	4.98	0.24	6.507	6.895	6.69	0.18	5.389	5.782	6.88	3.047

Table: Evaluation measures (for details see OL task above).

Full details: http://c4dm.eecs.qmul.ac.uk/sceneseventschallenge/