Detection and classification of acoustic scenes and events - an IEEE AASP challenge

[1]Centre for Digital Music
School of Electronic Engineering & Computer Science
Queen Mary University of London

{dimitrios,emmanouilb,dans,markp}@eecs.qmul.ac.uk

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

[2]Analysis-Synthesis team
IRCAM, CNRS-STMS

{mathieu.lagrange,matthias.rossignol}@ircam.fr

Challenge URL: http://www.elec.qmul.ac.uk/digitalmusic/sceneseventschallenge/

Abstract: What is this challenge about?

It is 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 shedding light on controversies that currently exist in the tasks.

Task 1 - Acoustic Scene Classification

A train/test task aiming to characterise or “label” the acoustic environment (soundscape) in which the audio was recorded.

- Two datasets: public and private set
- Dataset size: ten 30-sec audio recordings per class, ten classes
- Recording specs: Soundman binaural microphones (OKM II), PCM, 44100 Hz, 16 bit
- Evaluation metrics: classification accuracy, confusion matrix

busy street	park
quiet street	bus
supermarket	tube/metro
restaurant	tube station
office	open-air market

Table: Class names for scene classification task.

Task 2 - Acoustic Event Detection

The second challenge will address the problem of identifying individual sound events that are prominent in an acoustic scene.

- Environment: office
- Two experiments: monophonic sequences (real recordings) and polyphonic sequences (artificially contacenated events)
- Three subsets per experiment: training, development, testing
- Recording specs: Soundfield SPS422B, PCM, 44100 Hz, 24 bit
- Evaluation tracks: 4-channel and stereo
- Evaluation procedure: frame-based, event-based, classwise event-based
- Evaluation metrics: acoustic event error rate, precision, recall, F-measure

door knock	keyboard click
door slam	mouse click
speech	object dropped on table
laughter	switch
clearing throat	keys
coughing	phone ringing
drawer	short alert
printer	page turning

Table: Class names for event detection task.

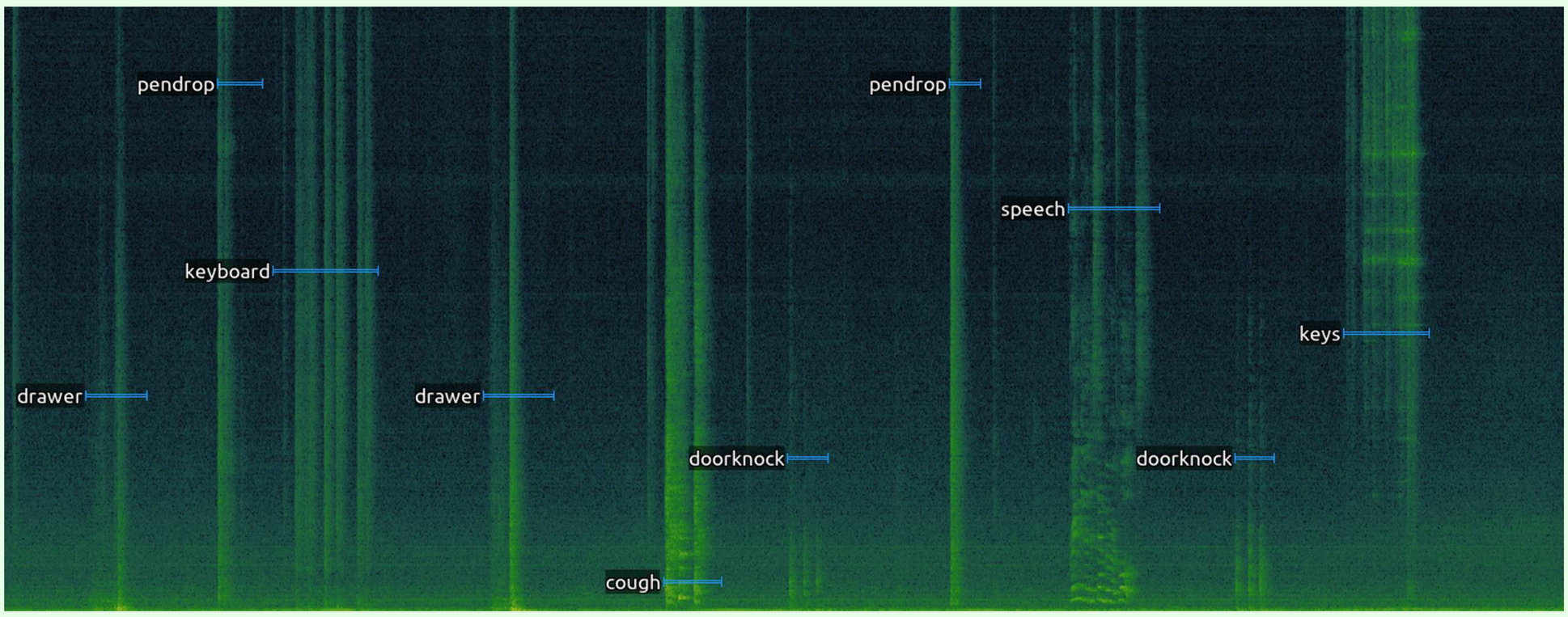

Figure: Spectrogram of a scripted sequence with overlayed annotation.