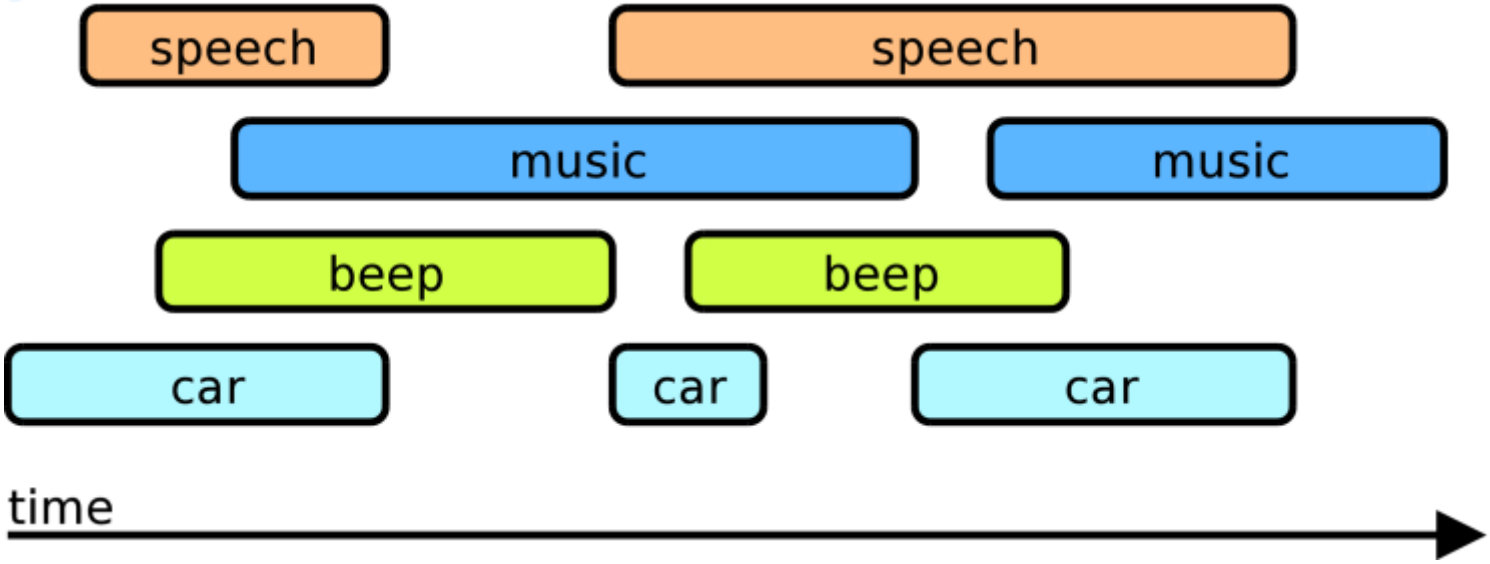
Sound Event Detection in Multisource Environments Using Source Separation

Toni Heittola[1], Annamaria Mesaros[1], Tuomas Virtanen[1], Antti Eronen[2]
[1]Tampere University of Technology, Department of Signal Processing
[2]Nokia Research Center Tampere
1st September, 2011

Sound event detection

- Aims at detecting acoustic events in an audio signal
- Predefined event classes = supervised classification
- Estimate the start and end time of each event



Environmental audio data

- Audio from everyday environments: street, office, grocery store, in a car, etc.
- Application areas of environmental sound event detection: context-aware devices, automatic annotation of videos

Outline of the presentation

- Sound event detection and environmental audio data
- Monophonic detection system
- Sound source separation based polyphonic detection system
- Evaluation & demonstration

Monophonic event detection system

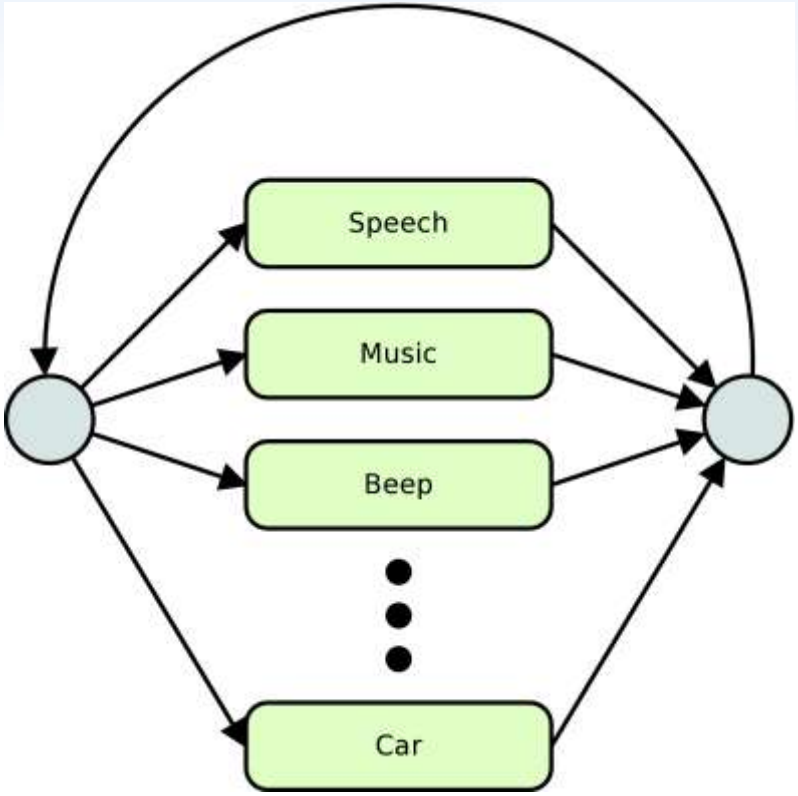
A. Mesaros, T. Heittola, A. Eronen, T. Virtanen. Acoustic event detection in real life recordings. In proc. EUSIPCO 2010.

- HMM classifier
- 61 event classes: (e.g. speech, music, beep, car, car door, bird, dog barking, footsteps, keyboard, coughing…)
- Each class modeled with a 3-state HMM (16 Gaussians per state, MFCC features).
- Train model for each event class separately using audio segments that are annotated to include the event

Monophonic event detection system

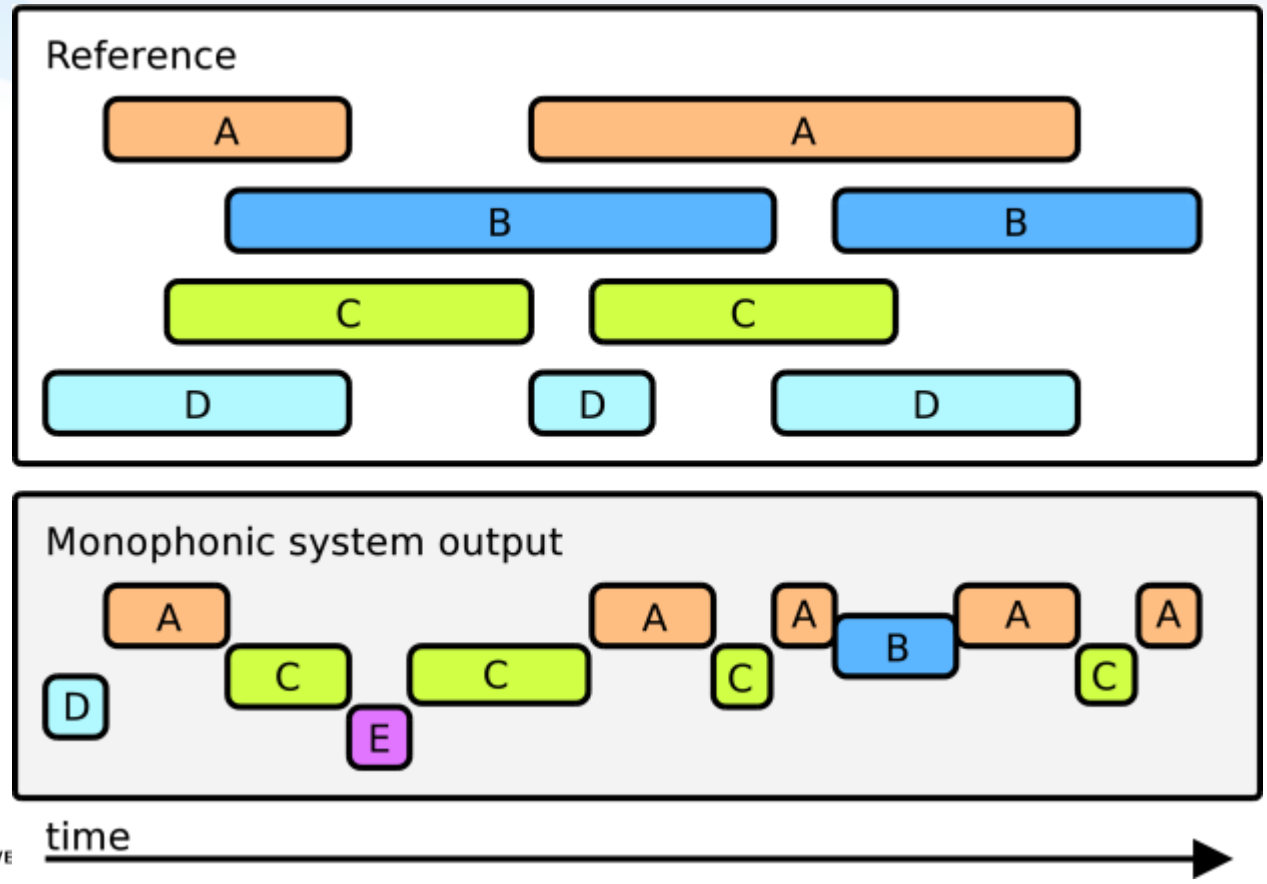
A. Mesaros, T. Heittola, A. Eronen, T. Virtanen. Acoustic event detection in real life recordings. In proc. EUSIPCO 2010.

- To model the whole signal, any event is allowed to follow any event



Output of the monophonic system

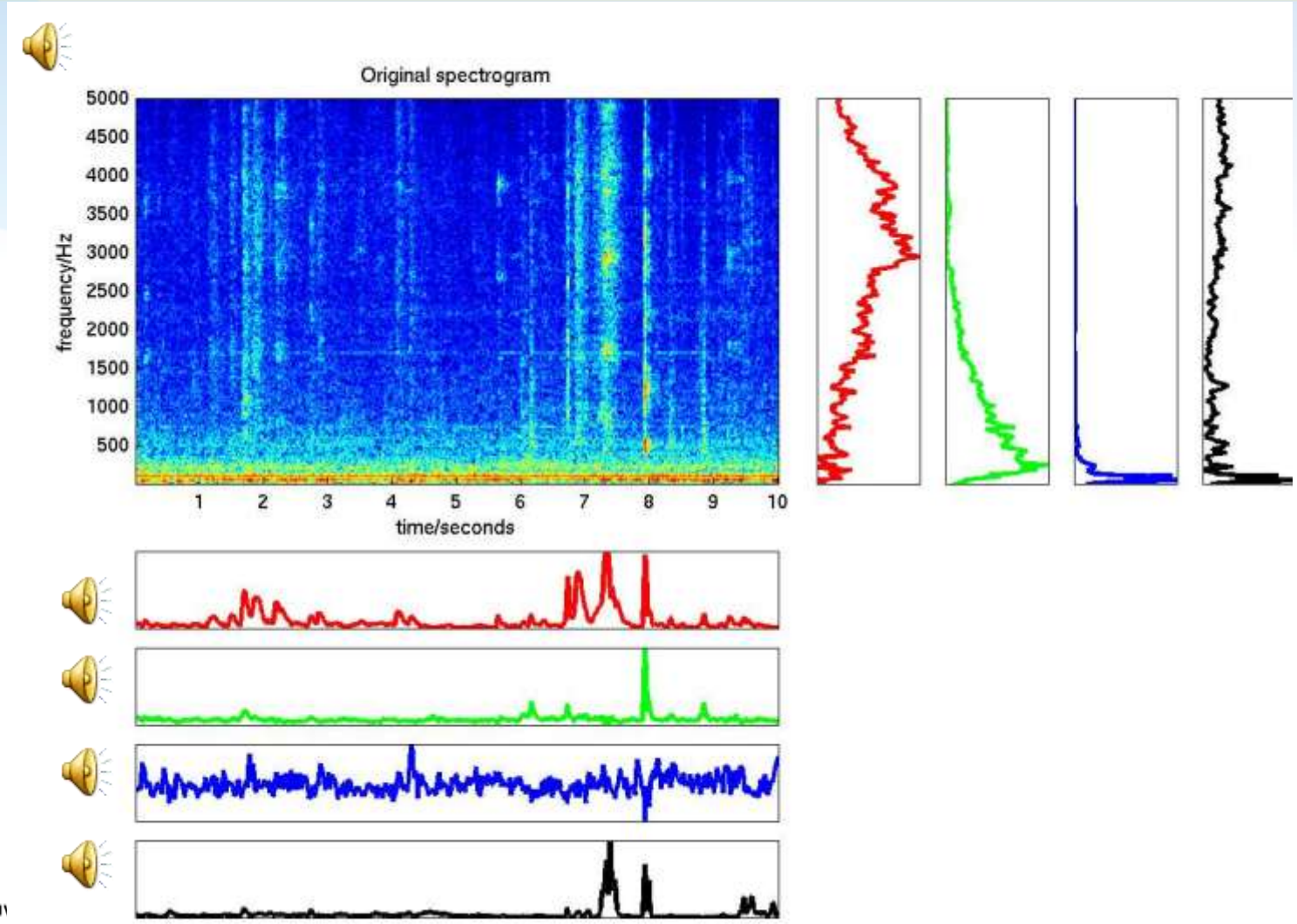
- The output is a sequence of non-overlapping events



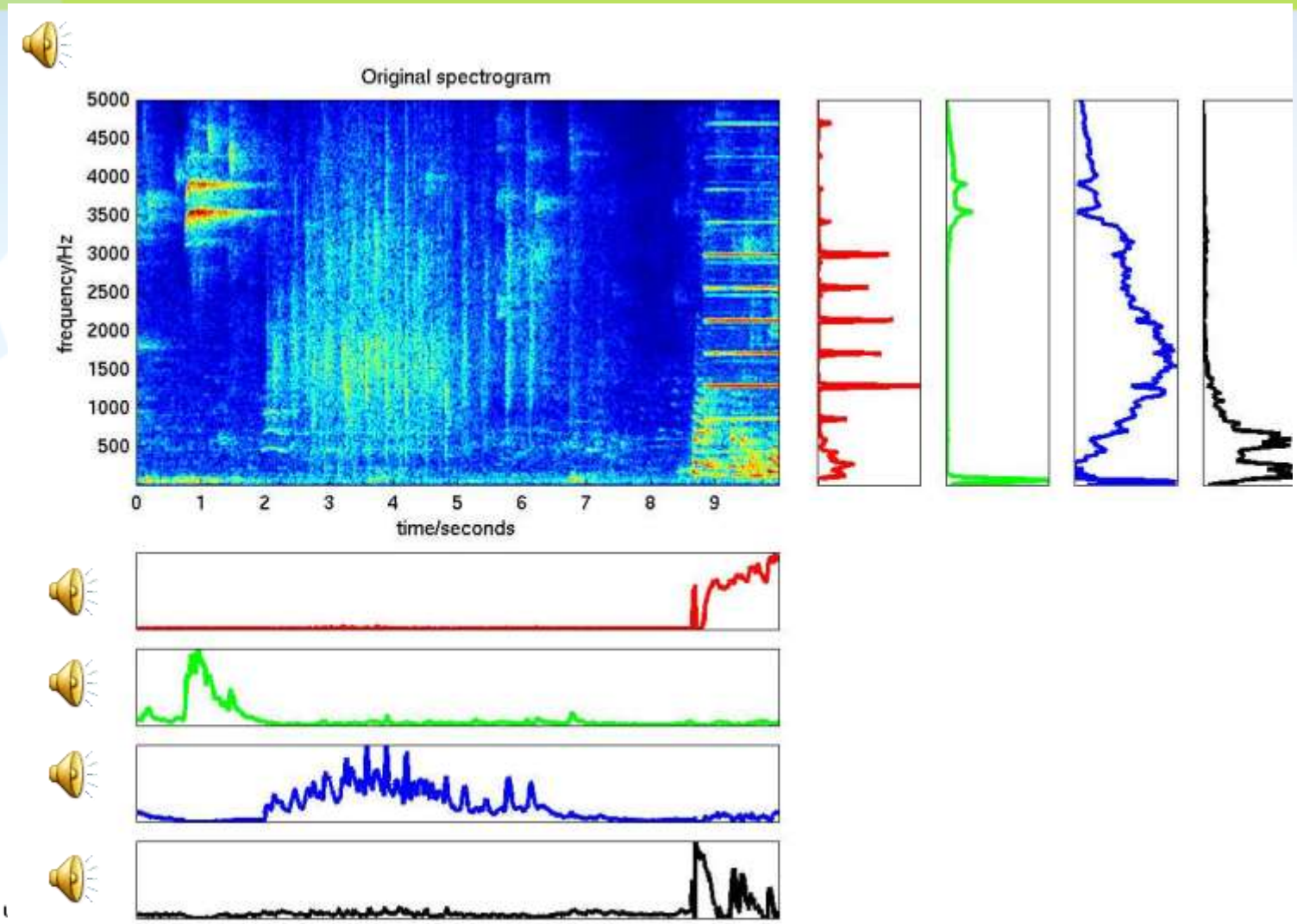
Non-negative spectrogram factorization based signal separation

- One-channel input signal is separated into multiple *tracks*
- NMF-based separation: magnitude spectrogram matrix represented as a of product of two non-negative matrices
- Represents the signal as a sum of components having fixed spectrum and time-varying gain
- Unsupervised separation: no prior knowledge about the sounds

Example of separated signals: kitchen

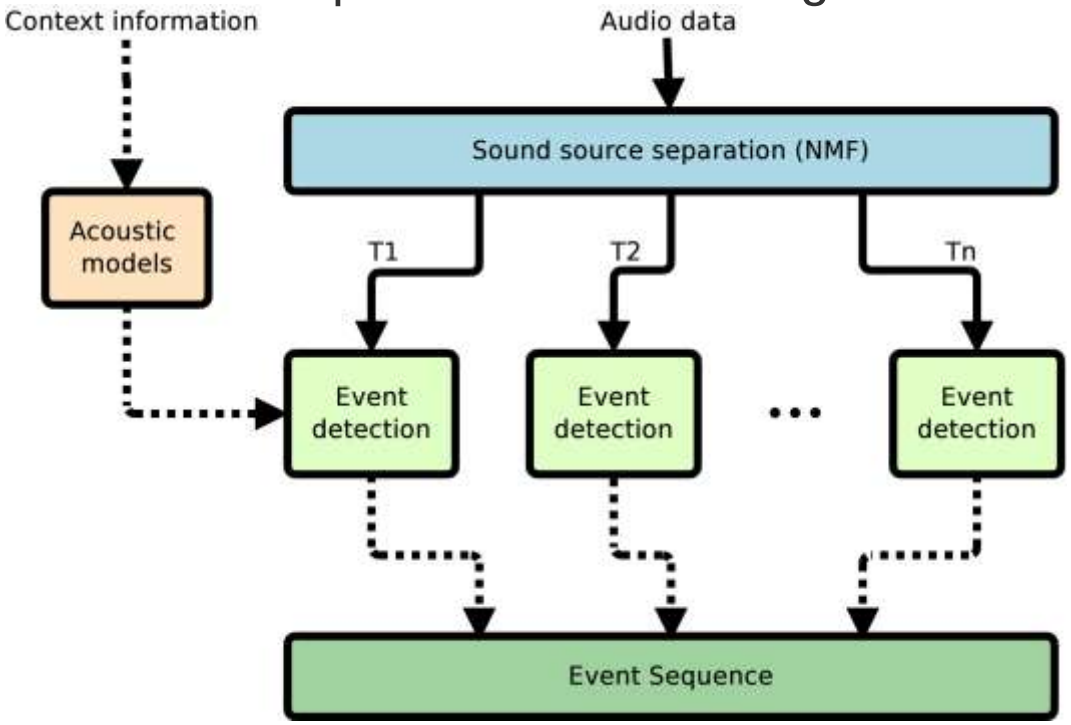


Example of separated signals: basketball game



Polyphonic event detection system

- Separation used as a preprocessing step
- Monophonic recognizer applied on all the separated tracks separately -> events obtained from the tracks are combined
- Training: All the tracks are pooled to the training data of an events

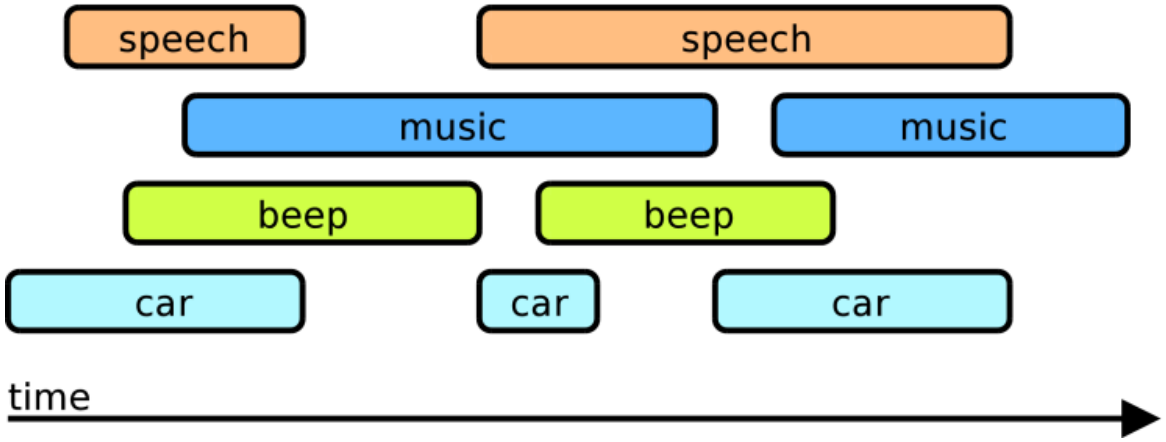


Acoustic database

- Material for the database was gathered from ten contexts
  - basketball game, beach, inside a bus, inside a car, hallway, office, restaurant, grocery store, street and stadium with track and field sports
- Each context is represented by 8 to 14 recordings, to a total of 103 recordings included in the database.
- In total ~19 hours of audio
- In total ~10.000 annotated events

Annotations

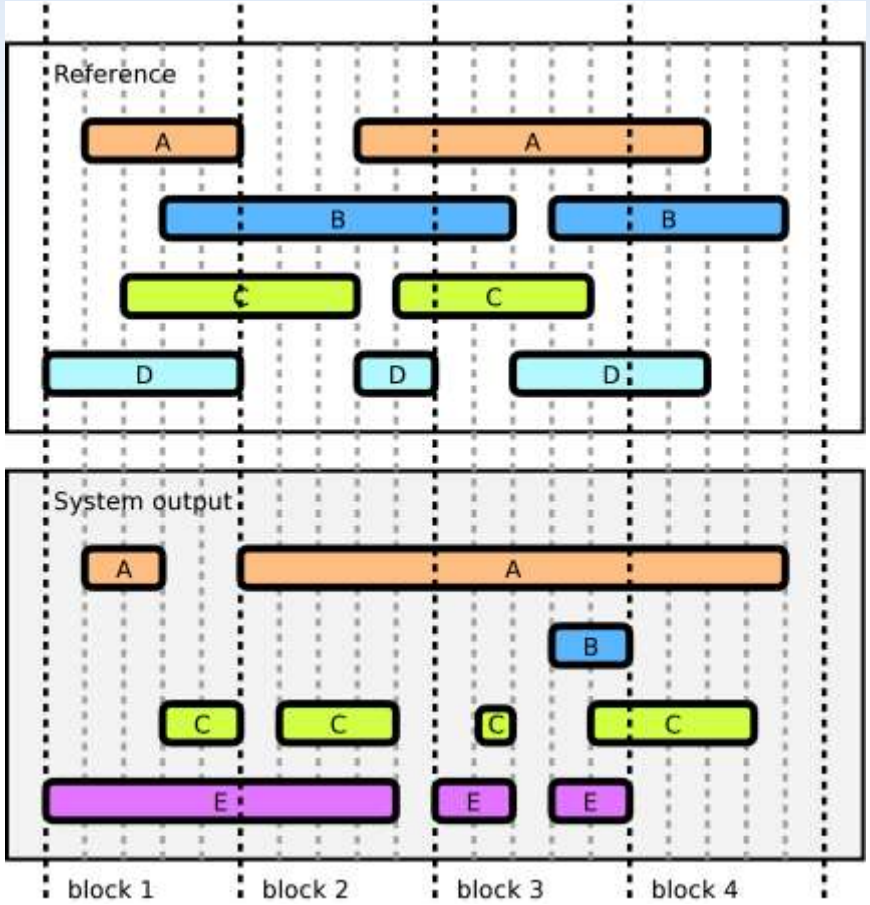
- Recordings were manually annotated indicating the start and end time of all clearly audible sound events
- Annotated sound events present in the recordings were grouped into 61 event classes
- Event classes include e.g. speech, laughter, applause, car door, road, dishes, door, chair, music, and footsteps



Demonstration

Evaluation metrics

- Detected events are regarded only at the block level, within 30 seconds
- Precision and recall is calculated inside the blocks, and combined into F-score
- Data divided into 70% training / 30% testing sets, 5 folds



Event detection performance (average F-score %)

	Monophonic	Polyphonic
Overall	28.2	52.6
Context		
Basketball	30.3	68.2
Beach	23.0	38.7
Bus	24.4	57.6
Car	18.8	46.7
Hallway	37.0	51.1
Office	30.1	49.7
Restaurant	25.4	54.2
Shop	27.7	56.2
Street	26.4	50.1
Track&Field	41.7	57.4

Conclusions

- NMF-based sound source separation can be used to do polyphonic event detection
- It improves significantly the performance of a monophonic event detection system
- It is possible to detect prominent sound events even in diverse real-world environments to some degree