Machine Listening for Music and Sound Analysis

Lecture 4 - Environmental Sound Analysis

Dr.-Ing. Jakob Abeßer
Fraunhofer IDMT

Jakob.abesser@idmt.fraunhofer.de

https://machinelistening.github.io

Fraunhofer IDMT

Overview & Learning Objectives

- Introduction
- Sound Event Detection
- Acoustic Scene Classification
- Acoustic Anomaly Detection
- Application Scenarios

Introduction
Motivation

- Sound carries information about our environment
- Challenging attempt to mimic the human's abilities
    - Environment perception
    - Context-awareness & localization of sound sources
    - Acoustic scene understanding
- Complementary sensory path to vision → multimodality
- Related to other content analysis domains (speech, music)

Introduction
Environmental Sounds (Recap)

- Sound sources
  - Nature, climate, humans, machines, etc.
- Sound characteristics
  - Structured or unstructured, stationary or non-stationary, repetitive or without any predictable nature
- Sound duration
  - From very short (gun shot, door knock, shouts) to very long and almost stationary (running machines , wind, rain)

https://freesound.org/people/{InspectorJ/sounds/416529, prometheus888/sounds/458461, MrAuralization/sounds/317361}
https://ccsearch-dev.creativecommons.org/photos/{39451123-ee45-4ec3-ad8d-b42d856bca06, c69d3b07-76bd-43e2-a44e-8742edc8447a, ab3062ab-fe0f-420d-b93d-7451db166b4e}

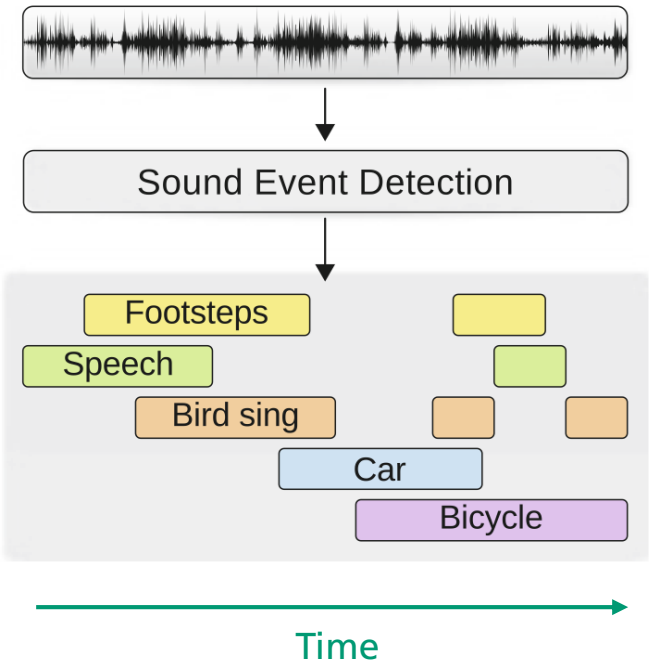
Fraunhofer IDMT

Introduction
Tasks / Categories

- Sound event detection (SED)
- Acoustic scene classification (ASC)
- Acoustic anomaly detection (AAD)

Sound Event Detection
Introduction

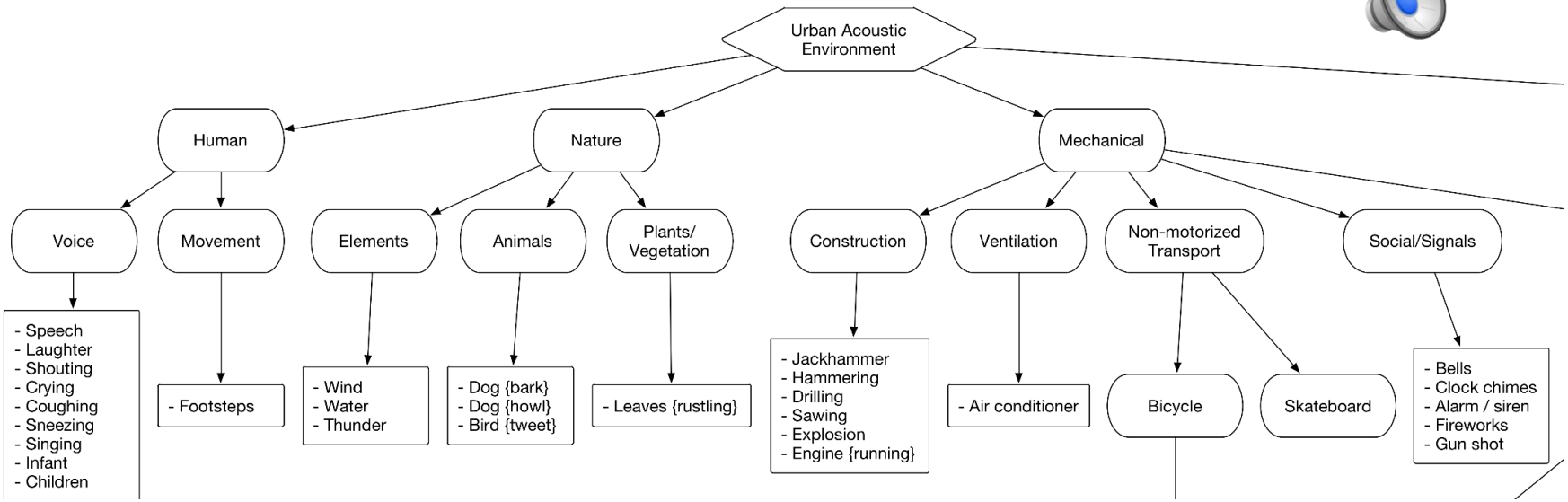
- Sound event detection → 2 simultaneous tasks
  - Segmentation (detection of temporal boundaries)
  - Classification (type of sound)
- Sound polyphony
  - Number of simultaneous sounds
  - Depends on the acoustic scene composition & sound sources



Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 157, Fig. 6.3

Sound Event Detection
Introduction

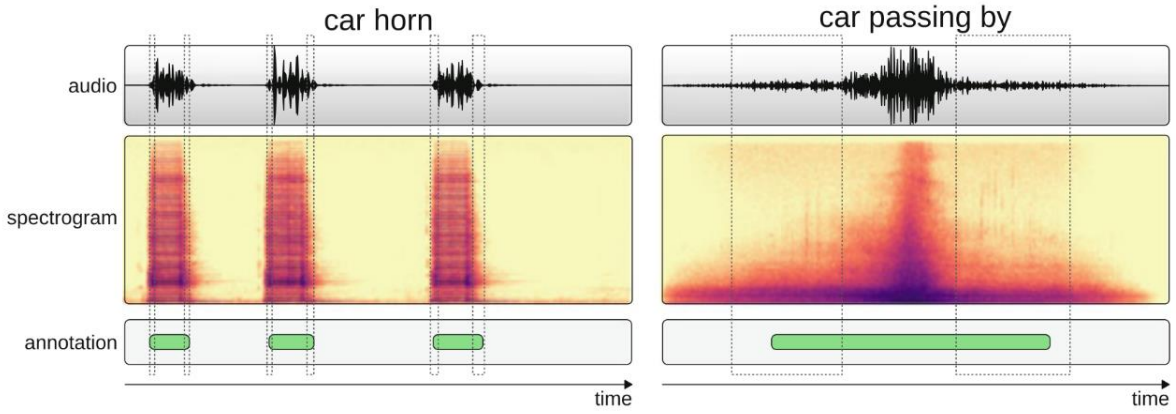
- Sound source categories
  - Humans, animals, vehicles, tools, machines, climate, …
- Hierarchies of sounds (e.g. urban sounds)
  - Based on origin & characteristics



https://urbansounddataset.weebly.com/taxonomy.html#
https://freesound.org/people/G_M_D_THREE/sounds/424404/

Sound Event Detection
Challenges

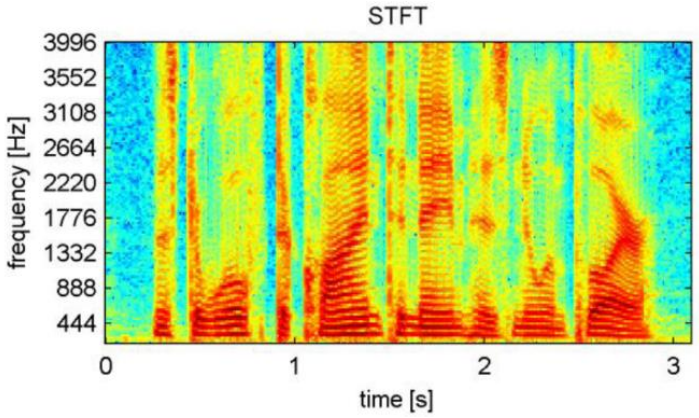
- Large range of different timbre characteristics
  - Short transients, noise-like signals, harmonic / inharmonic signals
- Different sound durations
  - Short (gun shot, door knock) → long / stationary (machines, wind)
- Ill-defined temporal boundaries
  - Complicates annotation & detection



Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 157, Fig. 6.3

Sound Event Detection
Challenges

- Sound appear in the foreground & background
  - depending on relative sound source position
- Non-local / sparse energy distribution
  - Fundamental frequency & overtones

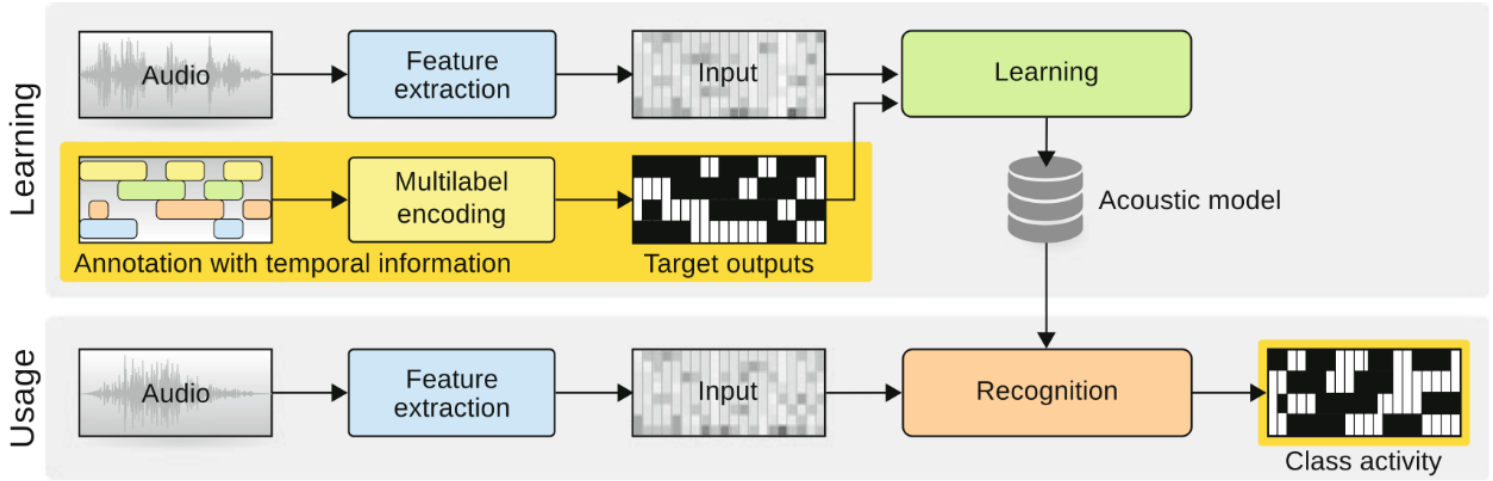


- Sounds are „transparent“
  - Phase-dependent overlap, possible cancelations

https://towardsdatascience.com/whats-wrong-with-spectrograms-and-cnns-for-audio-processing-311377d7ccd

Sound Event Detection
Pipeline

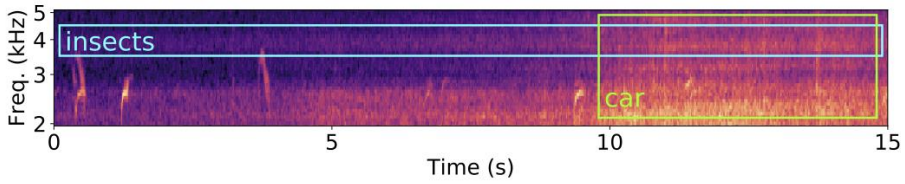
- Supervised learning pipeline
  - Feature extraction & pre-processing
  - Label encoding
  - Acoustic modeling



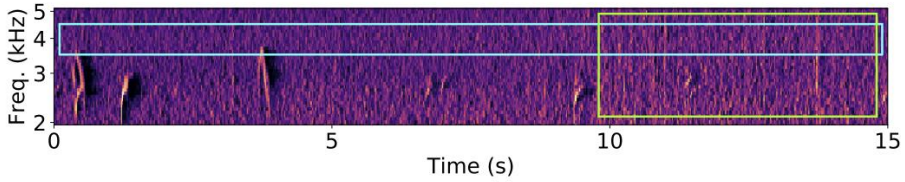
Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 31, Fig. 2.11

Sound Event Detection
Pipeline

- Feature extraction
  - 1D features (audio samples) → “end-to-end learning”
  - 2D features (mel-spectrogram, STFT)
- Feature pre-processing
  - Log-magnitude scaling
  - Per-channel energy (PCEN)
    - Dynamic range compression
    - Adaptive gain control
    - Suppresses stationary (background) noise



(a) Logarithmic transformation.

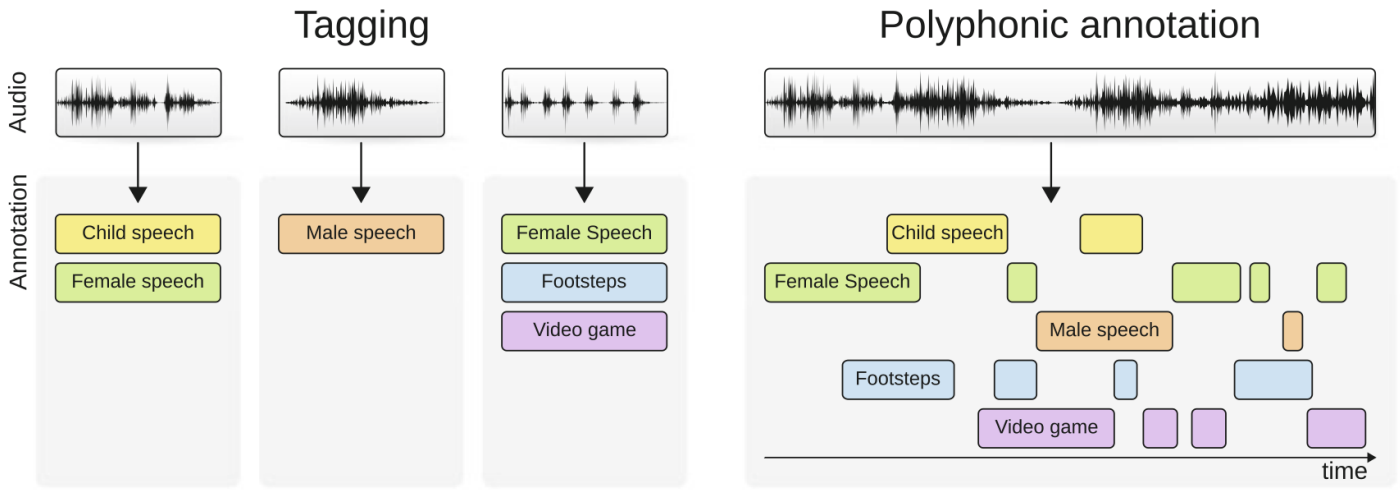


(b) Per-channel energy normalization (PCEN).

Lostanlen et al., "Per-Channel Energy Normalization: Why and How," IEEE Signal Processing Letters, vol. 26, no. 1, pp. 39-43, Jan. 2019, Fig. 1

Sound Event Detection
Pipeline

- Annotation
  - Quality of “ground truth”? (limited agreement / reliability)
  - Different granularities
    - Tagging / Global level (“weak” labels) → cheap
    - Event-level (“strong" labels) → expensive

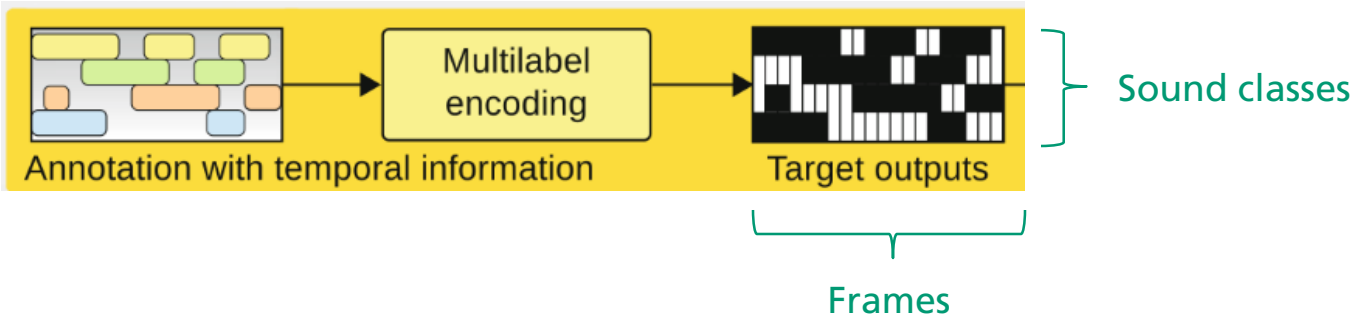


Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 154, Fig. 6.2

Fraunhofer IDMT

Sound Event Detection
Pipeline

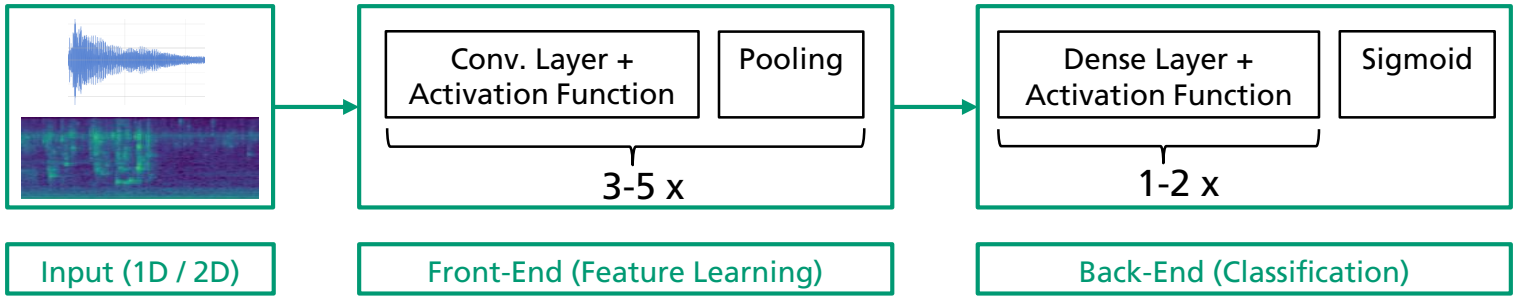
- Label encoding
  - Binarized sound activity (0/1)
    - Multilabel classification
    - 1 (independent) binary detector per class
  - Temporal resolution (duration of each annotated time frame)



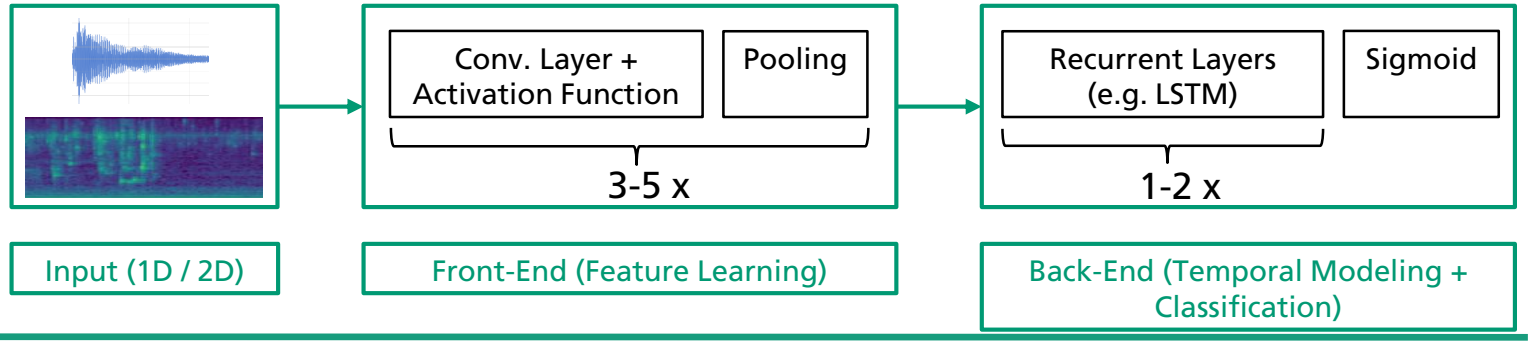
Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 31, Fig. 2.11 (excerpt)

Sound Event Detection
Pipeline

- Typical neural network architectures
  - CNN



  - CRNN



Fraunhofer IDMT

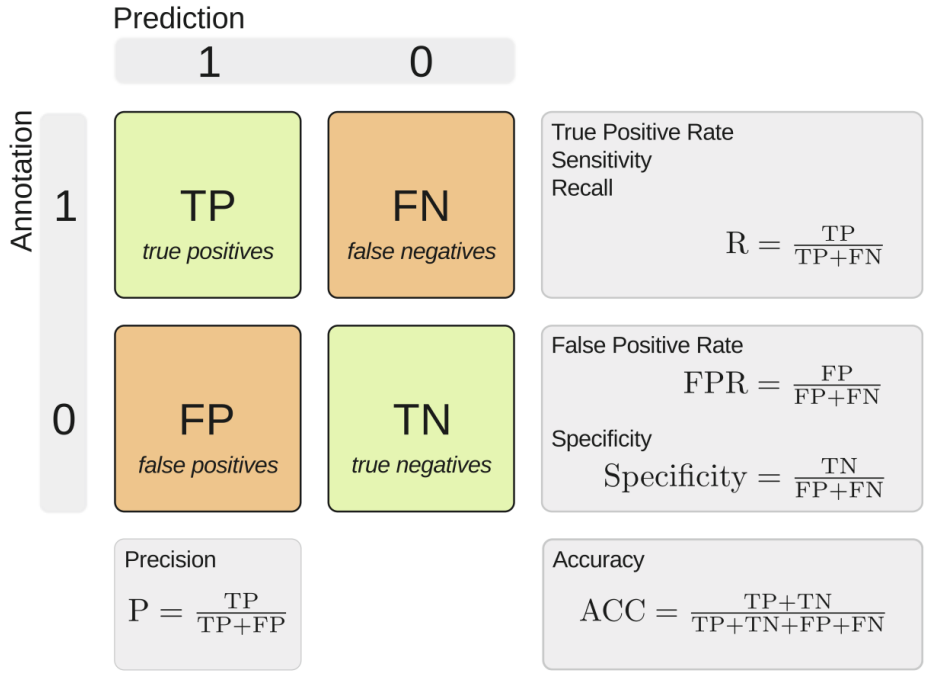
Sound Event Detection
Pipeline

- Data Augmentation
  - Cope with limited amount of training data
  - Approach 1: Apply signal transformations
    - Adding noise
    - Time-stretching/Pitch-shifting
    - Mix-up data augmentation [Zhang]
      - Random mix of two items (audio & targets)
      - Example: 0.7 x car + 0.3 x speaker
  - Approach 2: Data synthesis
    - Generative Neural Network Models (e.g. Generative Adversarial Networks (GAN), SampleRNNs)

Zhang et al.: mixup: Beyond Empirical Risk Minimization, ICLR 2018

Sound Event Detection
Evaluation

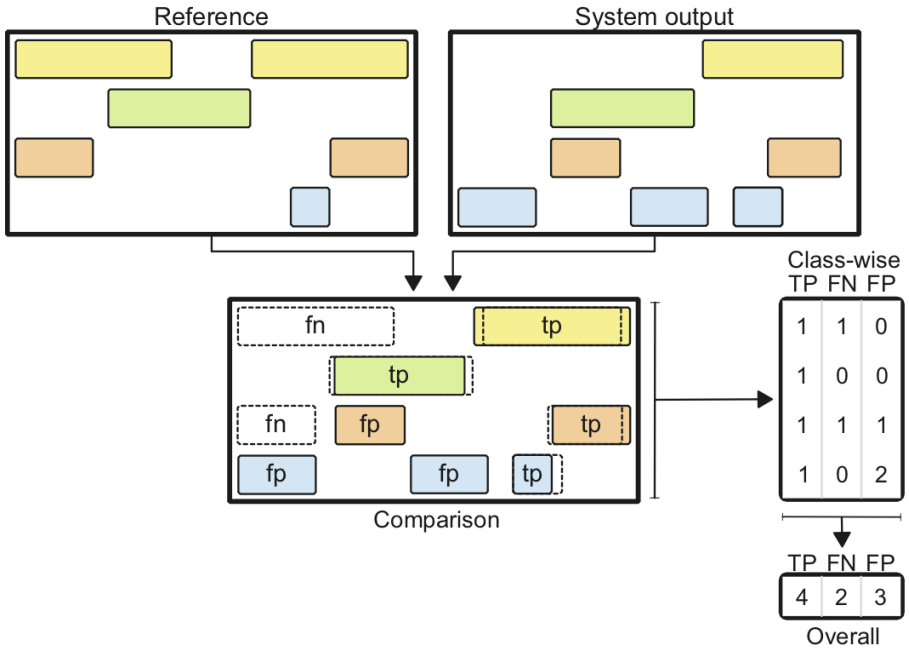
- Recap: Binary classification evaluation
  - True/false positives (TP/FP)
  - True/false negatives (TN/FN)
  - Metrics
    - Precision
    - Recall
    - Accuracy
    - F-score



Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 170, Fig. 6.7

Sound Event Detection
Pipeline

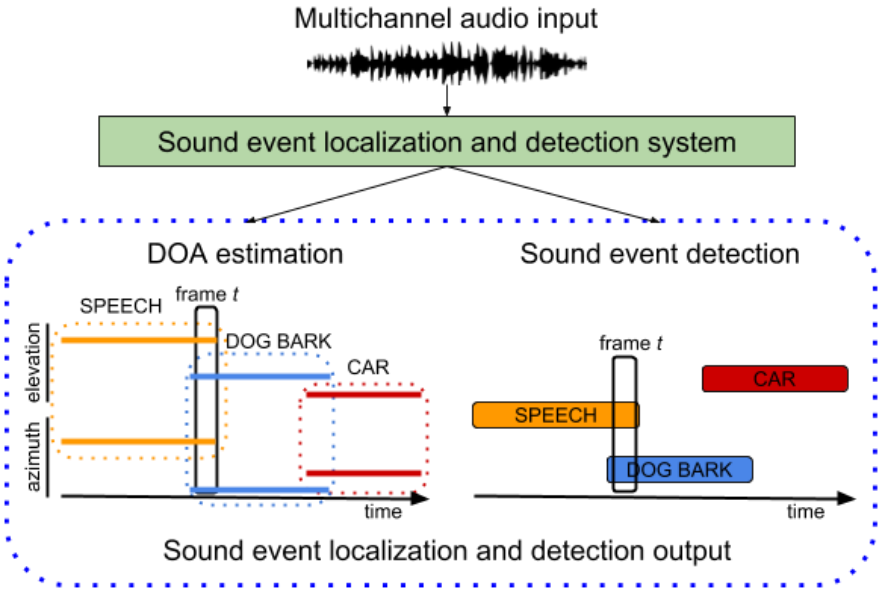
- Evaluate SED → binary classification results on a frame-level
- Compare reference with predictions
- Count TP/FN/FP→ aggregate over time → compute metrics

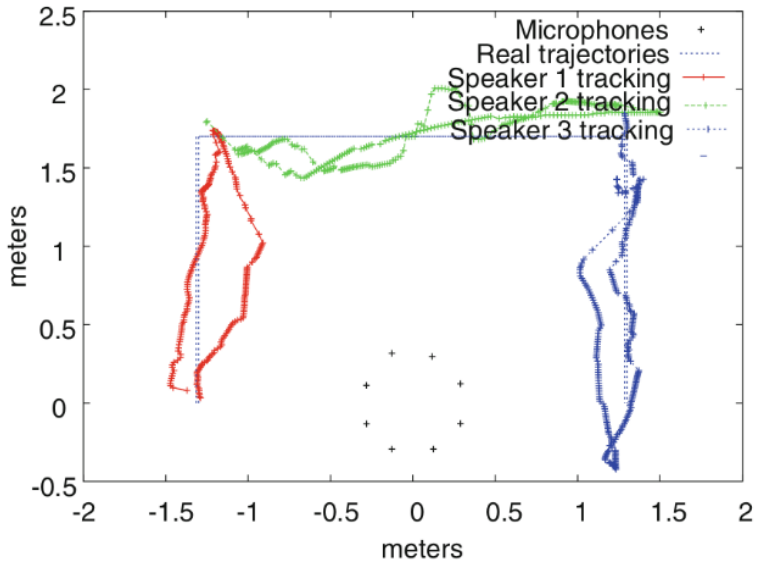


Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 169, Fig. 6.6

Sound Event Detection
Related tasks

- Sound event localization & tracking
  - Multichannel audio recordings (e.g., first-order ambisonic microphones)
  - Estimate direction-of-arrival (DOA) & track source movement



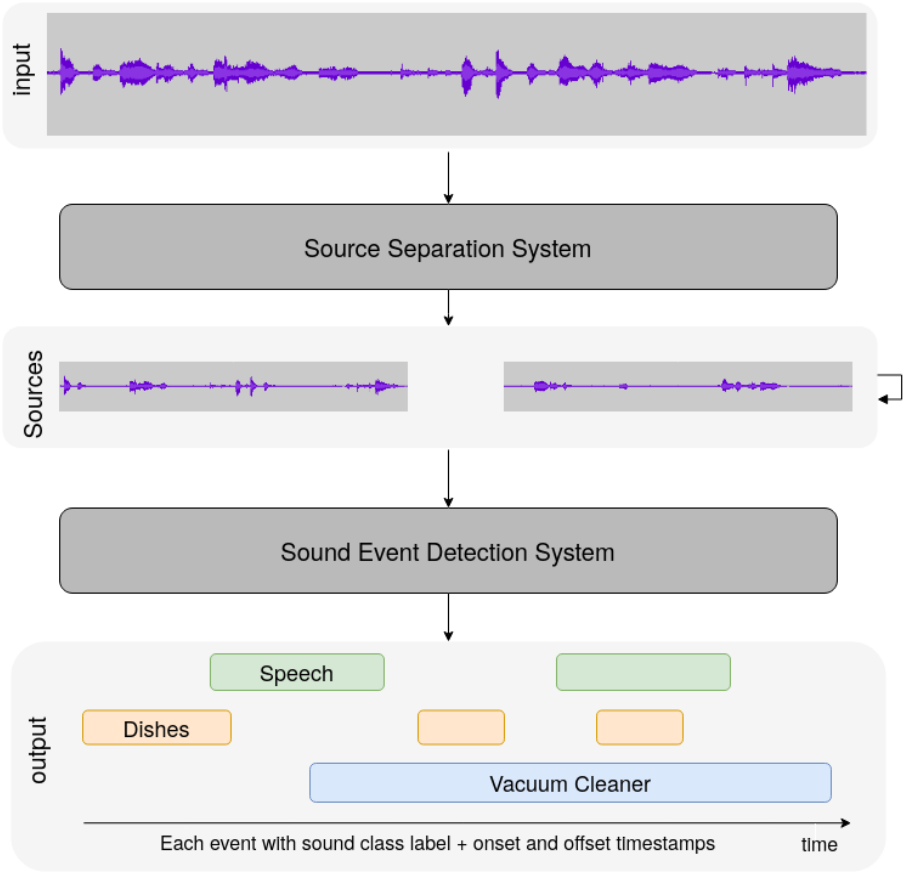


http://dcase.community/challenge2019/task-sound-event-localization-and-detection
Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 267, Fig. 9.7

Fraunhofer IDMT

Sound Event Detection
Related tasks

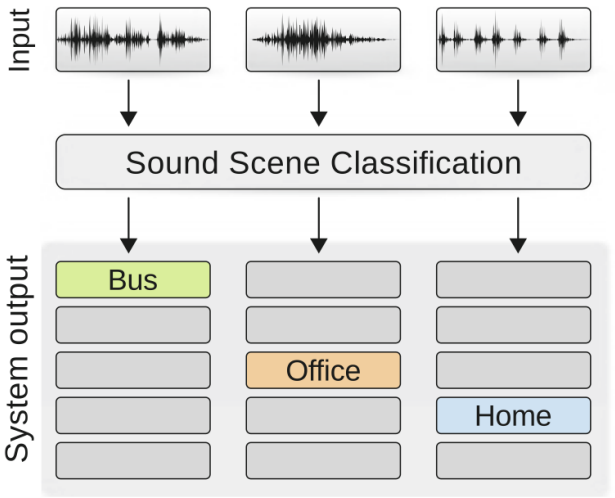
- Source separation
  - Prior to sound event detection
- Chicken-egg problem
  - Alternative: sound-informed source-separation

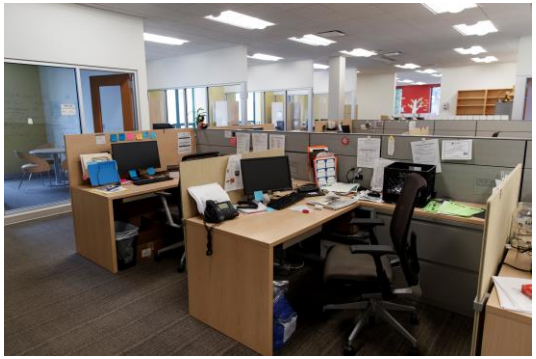


http://dcase.community/challenge2020/task-sound-event-detection-and-separation-in-domestic-environments

Acoustic Scene Classification
Task

- Acoustic scene classification (ASC)
  - Multi-class (1 of N) classification scenario
  - Summative label (tagging)
- Acoustic scene
  - Typical set of sounds
  - Example: office
    - Keyboard clicks
    - Human conversations
    - Printer
    - Air conditioner

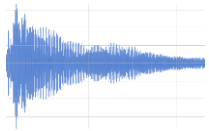


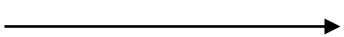


Virtanen et al., Computational Analysis of Sound Scenes and Events, p. 267, Fig. 9.7
https://freesound.org/people/16HPanskaTyllova_Terezie/sounds/497363
https://images.theconversation.com/files/349387/original/file-20200724-15-ldrybi.jpg

Acoustic Scene Classification
Pipeline

- Label encoding
  - One-hot-encoded (global) target
- Example
  - 4 scene classes (bus, office, home, forest)
  - Encoding of an office recording





[0 1 0 0]

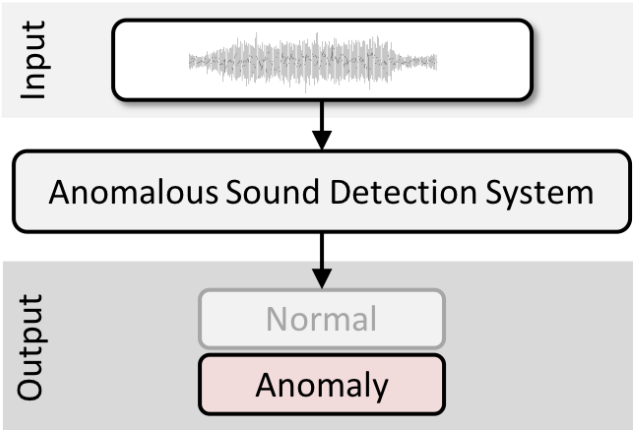
Sound classes

Acoustic Scene Classification
Pipeline

- Network architectures
  - Similar to SED (CNN & CRNN)
- Differences
  - Temporal result aggregation within network
  - Dense layer / pooling
  - Final layer: softmax activation function (multiclass classification)
- Current Research Topics
  - Attention → learn to focus on spectrogram regions
  - Open-set classification → detect unknown classes
  - Transfer learning → fine-tune pre-trained models with less data

Acoustic Anomaly Detection
Task

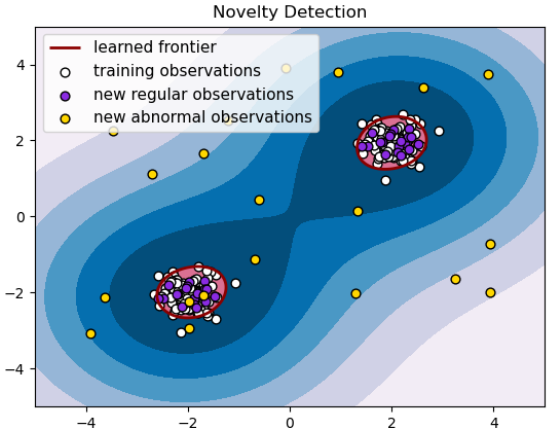
- Goal
  - Detect deviations from “normal” state
  - Identify whether emitted sound from target object is normal or anomalous
- Challenges
  - Often only training examples for normal state available
  - Acoustic anomalies are often subtle compared to louder background noise
- Application Scenarios
  - Detecting machine failures
  - Intrusion detection (glass break…)

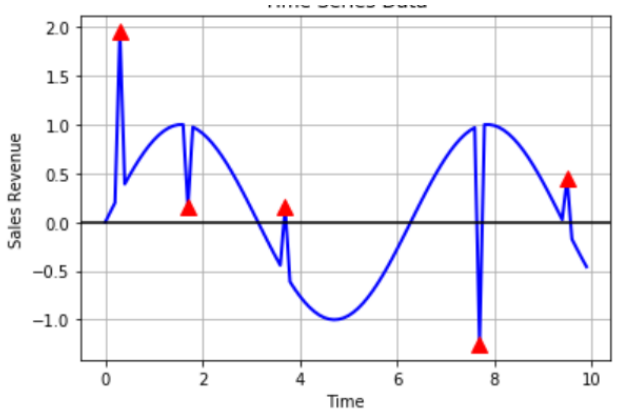


http://dcase.community/challenge2020/task-unsupervised-detection-of-anomalous-sounds

Acoustic Anomaly Detection
Approaches

- Traditional methods
  - Distribution outlier detection
    - Modelling normal state distribution
    - Detect distribution outliers
    - E.g.: One-class GMM / SVM
  - Time-series analysis
    - AD via local deviation from prediction
    - E.g.: Autoregressive models, Hidden-Markov-Models (HMM)

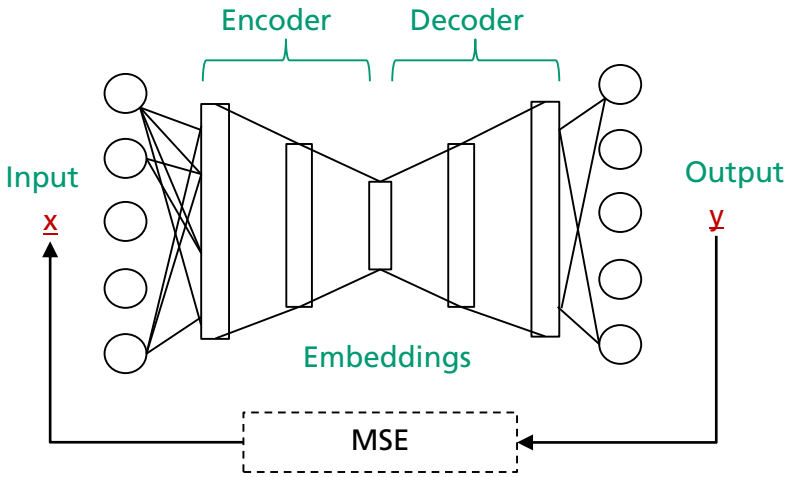




https://scikit-learn.org/stable/_images/sphx_glr_plot_oneclass_0011.png
https://miro.medium.com/max/722/1*TvZ9jI9vGX-fWwc3AHwNDw.png

Acoustic Anomaly Detection
Approaches

- Novel methods
  - Autoencoder (encoder → decoder) models
    - Idea: normal sounds can be better reconstructed than unknown anomalous sounds
    - Dense, convolutional, variational AE
    - Interpolation DNN
      - Interpolate spectrogram frame from surrounding frames



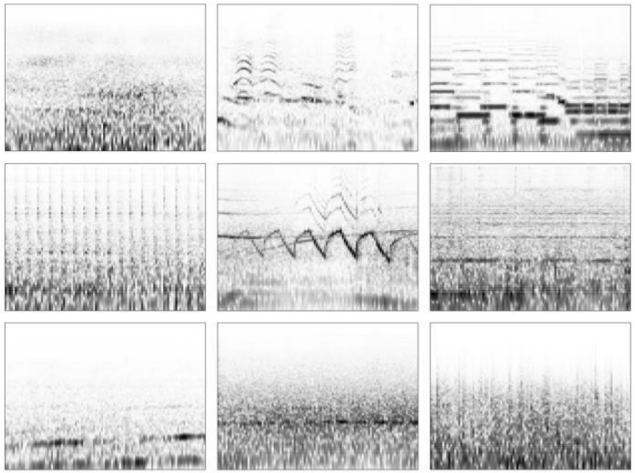
Application Scenarios
Urban Noise Monitoring

- Joint R&D project (2016 – 2018)
  - Fraunhofer IDMT, IMMS, SSJ GmbH, BE
- Goal
  - Develop distributed sensor network for
    - Sound level measurement
    - Sound classification
- Approach
  - Mobile sensor units
    - Raspberry Pi 3, quad-core ARM, 1GB RAM
    - Battery + MEMS microphones
    - Sensor locations (light poles)

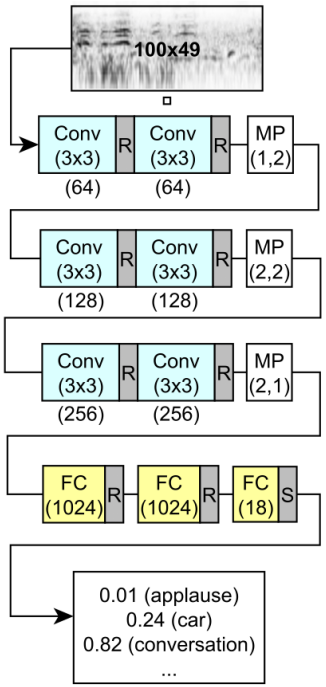
Application Scenarios
Urban Noise Monitoring

- Measurements
  - Different loudness values (8/s)
  - Sound event detection (1/s)
    - 9 sound event classes (car, conversation, music, roadworks, siren, train, tram, truck, wind)

Spectrogram examples (2 s long)



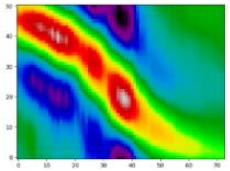
- CNN architecture



Fraunhofer IDMT

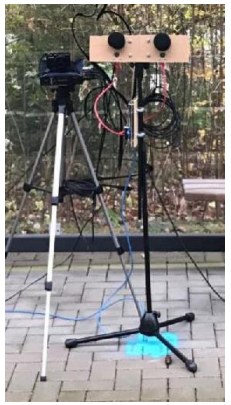
Application Scenarios
Traffic Monitoring

- Traffic monitoring
  - Vehicle detection & direction of movement
    - Stereo channel cross-correlation



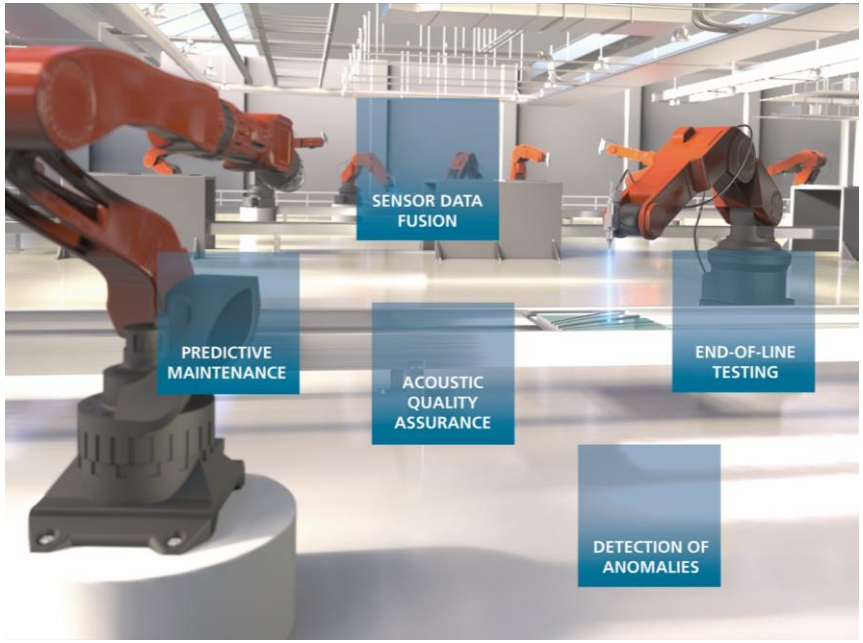
Example: movement left → right

  - Vehicle type classification (car, truck, bus, motorcycle)
- Challenges
  - Microphone type
  - Vehicle speed
  - Street surface quality & weather conditions



Application Scenarios
Production Chain Monitoring

- Challenges
  - Real-time analysis & classification of industrial sounds
  - Energy-efficient AI algorithms
  - Sound variations due to different machine states
  - Acoustic anomalies subtle compared to background noises



Fraunhofer IDMT

Summary

- Introduction
- Sound Event Detection
  - Challenges
  - Pipeline
  - Evaluation
- Acoustic Scene Classification
- Acoustic Anomaly Detection
- Application Scenarios
  - Urban Noise Monitoring
  - Traffic Monitoring
  - Product Chain Monitoring