Machine Listening for Music and Sound Analysis

Lecture 6 - Environmental Sound Analysis 2

Dr.-Ing. Jakob Abeßer

Fraunhofer IDMT

Jakob.abesser@idmt.fraunhofer.de

https://machinelistening.github.io

Fraunhofer IDMT

Overview

- Acoustic Scene Classification
- Acoustic Anomaly Detection
- Real-World Deployment
    - Process Steps
    - Challenges
- Use-Cases
    - Urban Noise Monitoring
    - Traffic Monitoring
    - Industrial Sound Analysis
    - Context-sensitive Hearables
    - Bioacoustic Monitoring

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

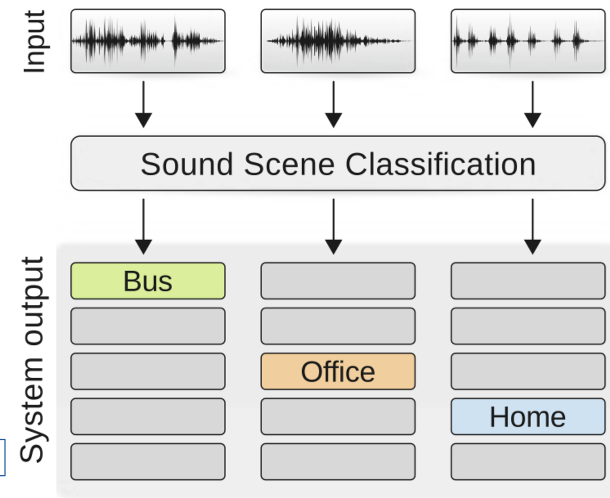


Fig. 1

AUD-1

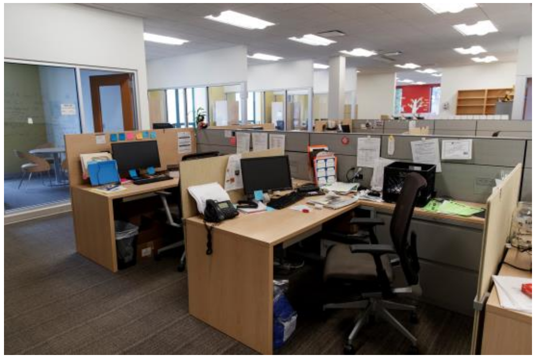
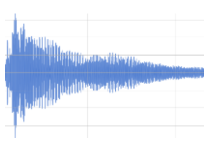


Fig. 2

Acoustic Scene Classification
Pipeline

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



$$\underbrace{[\;0\;\;1\;\;0\;\;0\;]}_{\text{Sound classes}}$$

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 [Abeßer, 2020]
  - 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
    - Is emitted sound from target object 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…)

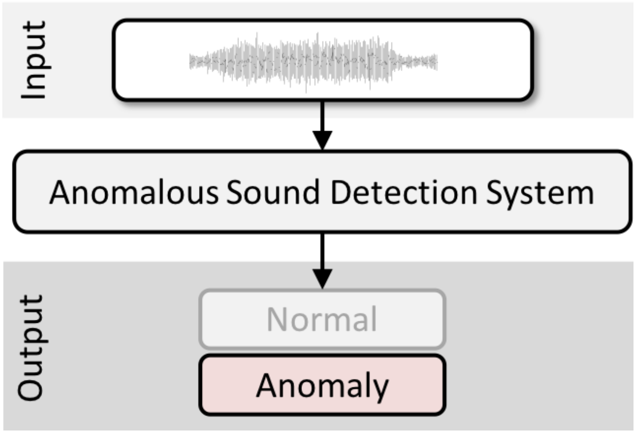


Fig. 3

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 prediction error
    - E.g.: Autoregressive models, Hidden-Markov-Models (HMM)

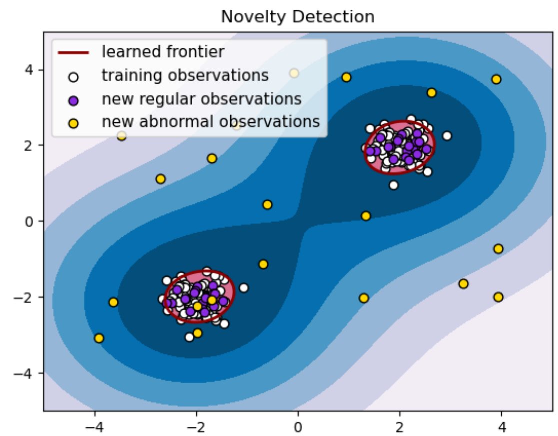


Fig. 4

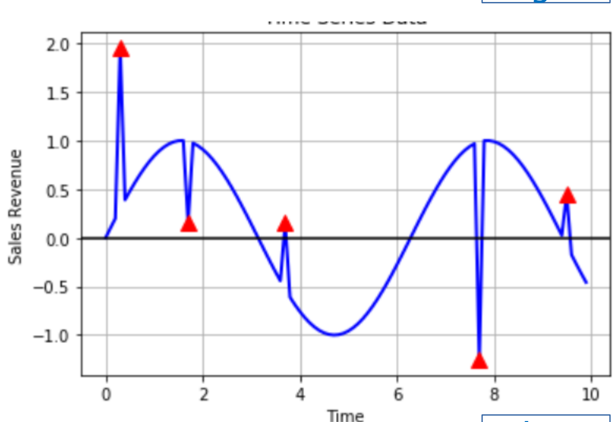
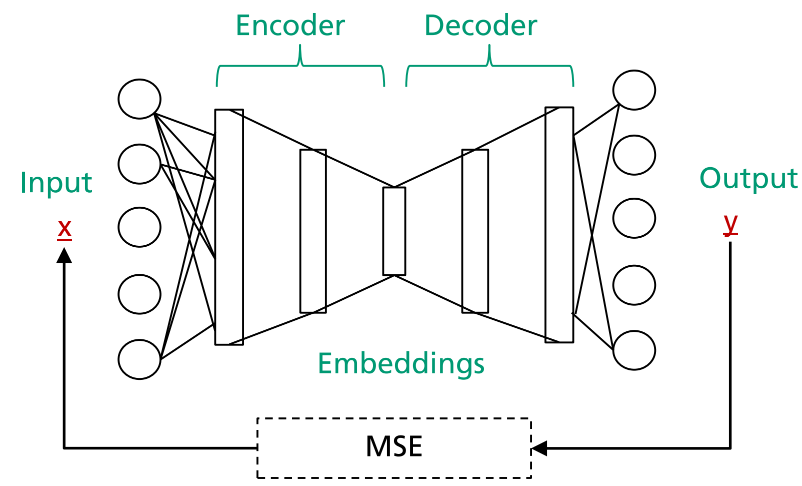


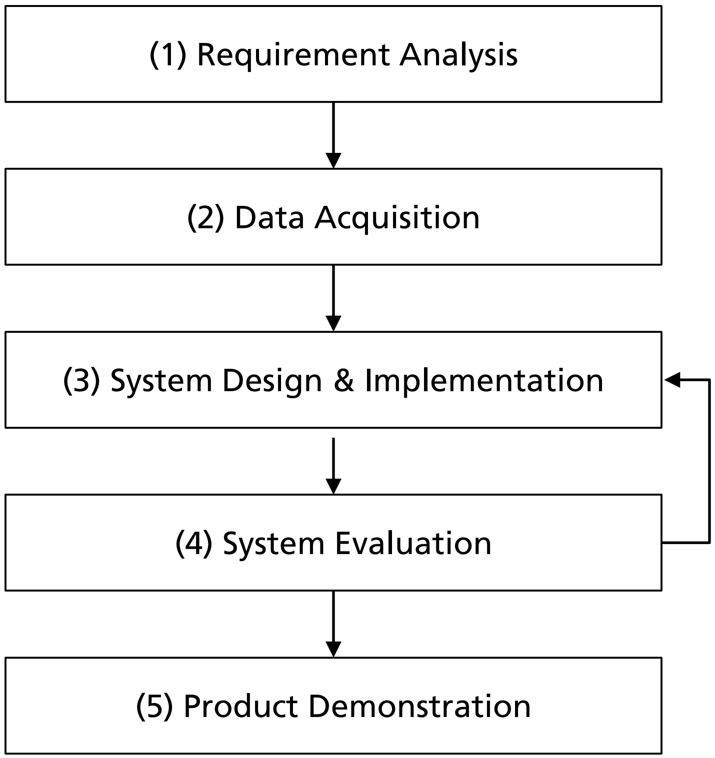
Fig. 5

Acoustic Anomaly Detection
Approaches

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



Real-World Deployment
Project Phases



Real-World Deployment
(1) Requirement Analysis

- Target application
    - Main characteristics
- Relevant ESA research problem
    - Sound classes
- Performance requirements
    - Analysis window size
    - Metrics (accuracy, recall, precision, f-score, etc.)
- User Experience
    - Error type categorization / prioritization

Real-World Deployment
(1) Requirement Analysis (Example)

- Target application (context-aware cell phones)
  - Main characteristics (ringtone type & loudness adapts to user's environment)
- Relevant ESA research problem (acoustic scene classification)
  - Sound classes (at home, opera, traffic …)
- Performance requirements
  - Analysis window size (5s)
  - Metrics (accuracy, recall, precision, f-score, etc.) (F > 0.85)
- User Experience
  - Error type categorization / prioritization
    - (confusion opera ⟷ traffic worse than traffic ⟷ at home)

Real-World Deployment
(1) Requirement Analysis

- Performance constraints
  - Computer platform (Raspberry 4, Jetson Nano, etc.)
  - Memory, CPU / GPU performance
  - Inference time vs. real-time
- Model constraints
  - Architecture
  - # Parameters
  - # Layers
  - Model size
  - Floating-point resolution

Fig. 6

Rhaspberry 4

Fig. 7

Jetson Nano

Real-World Deployment
(2) Data Acquisition

- Preliminary considerations
  - Acoustic conditions at deployment scenario / target use-case
    - Room size / characteristics, echoes / feedback, background noises
  - Target sound variability
  - Sensor placement
  - Recording procedure
    - Microphone type / setup
    - (Background noise removal)
  - Security / Privacy
  - Data transmission / storage

Fig. 8

Real-World Deployment
(2) Data Acquisition

- Audio Recording
- Annotation
    - Time / labor expensive
    - Contextual metadata (time, location, …)
    - Granularity (segment vs. file-level)
    - Subjectivity (annotator agreement)
    - Use existing tools (e.g., Sonic Visualiser)
- Data split
    - Train / Validation / Test

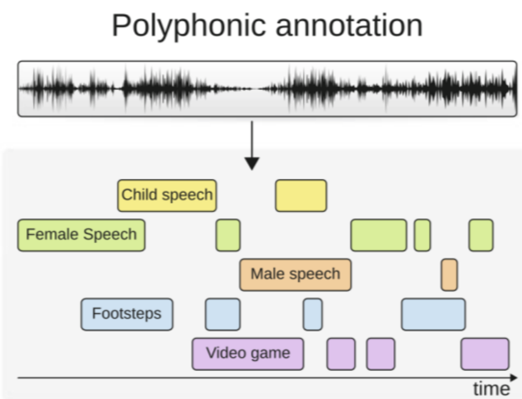


Fig. 9

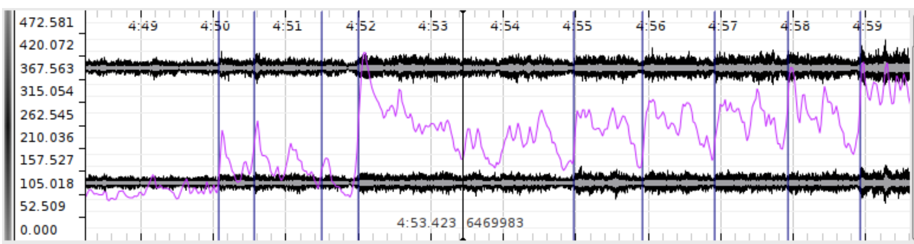


Fig. 10

Fraunhofer IDMT

Real-World Deployment
(3) System Design & Implementation

- Goal → Proof-of-Concept (PoC)
  - Solves defined problem
  - Demonstrate capability / feasibility under laboratory environment (datasets)
- Quickly implement baseline system (reference point)
- Iterative improvement of system components
  - Audio processing (pre-processing, feature extraction)
  - Machine learning (learning / recognition / detection)

Real-World Deployment
(4) System Evaluation

- Goal → Realistic performance estimate
  - Ideally test condition & target application are similar
  - Compare to baseline system / state-of-the-art methods
- Incremental changes & evaluation
  - Identify most important factors that influence the system's performance
- Evaluation
  - Offline (pre-recorded audio) vs. online (real-time recordings)
  - Objective (test dataset, defined metrics) vs. subjective (user tests)

Real-World Deployment
(5) Product Demonstration

- Goal → Develop PoC further into a Prototype
    - Key features according to requirement analysis
    - Tested in realistic use-cases (technology validation)
    - Tested with real users (user experience / perception of good performing system)
- Iterative development until ready for deployment
    - Problem examples: too high latency, too low noise-robustness
- Finally
    - System integration (user interface etc.)
    - Deployed to the market (small scale pilot -> full scale)

Real-World Deployment
Challenges

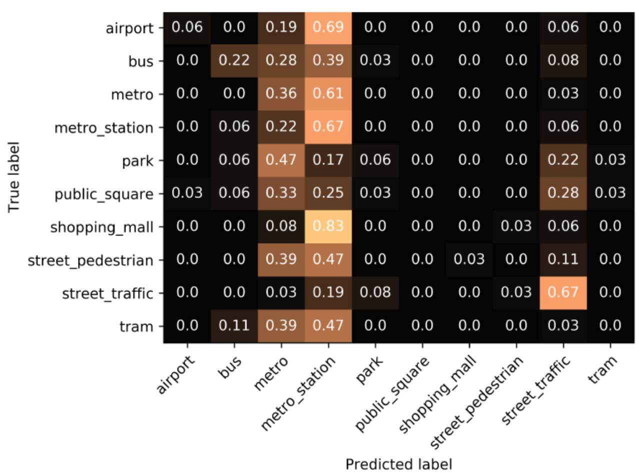
- Data Mismatch / Domain Shift
- Model Complexity
- Privacy / Security

Real-World Deployment
(1) Data Mismatch / Domain Shift

- Differences in data distribution due to
  - Room acoustics (reverb, reflections)
  - Microphone characteristics (frequency response, directionality)
- Domain adaptation
  - Adapt model / feature mapping from source to target domain
    - Unsupervised: adversarial training [Gharib, 2018]
    - Supervised: transfer learning
- Data augmentation
  - Increase model robustness by increasing data variability
- Data normalization [Johnson, 2020]
  - Align source and target data distributions

Real-World Deployment
(1) Data Mismatch / Domain Shift

- Domain adaptation (DA)
  - Unsupervised DA via adversarial training [Gharib, 2018]



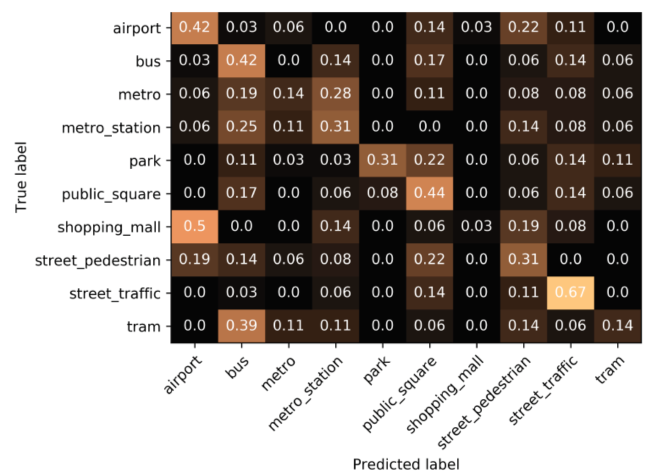


Fig. 11

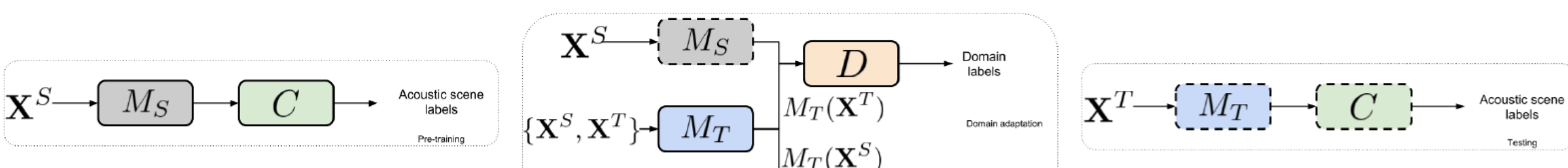


Fig. 12

Real-World Deployment
(1) Data Mismatch / Domain Shift

- Data normalization
  - Align source and target data distribution (zero mean & standard deviations) [Johnson, 2020]
    - Reduce domain shift

Metal ball surface classification (colors = classes, shadings = recordings)

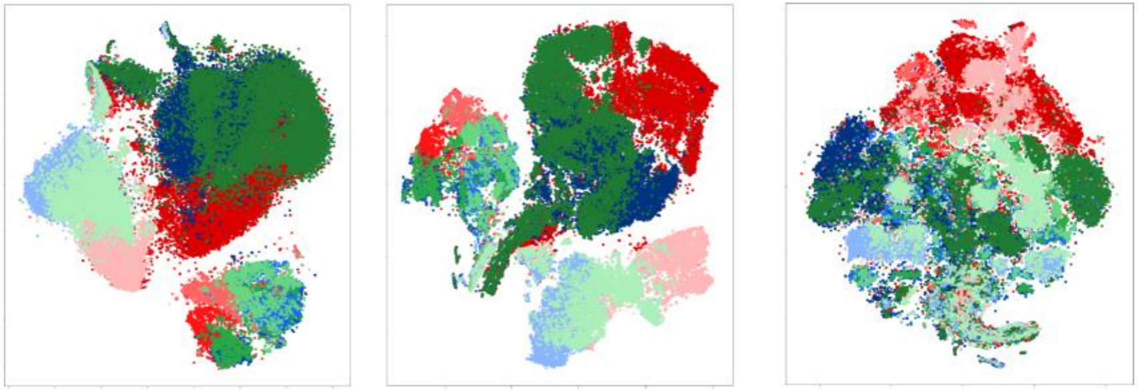


Fig. 13

(a) No Norm (b) Global Norm (c) Adaptive Norm

Real-World Deployment
(2) Model Complexity

- Goals
  - Reduce model size – fewer parameters, less memory required
  - Reduce latency (inference time) / lower energy consumption

- Approaches ([Wang, 2021])
  - Pruning
    - Identify & remove redundant connections / neurons

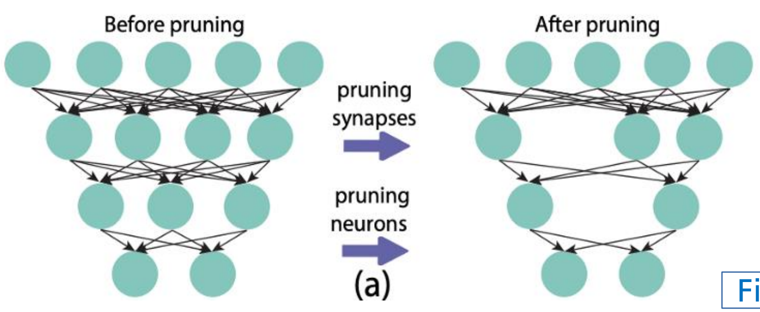


Fig. 14

Real-World Deployment
(2) Model Complexity

- Approaches
  - Quantization
    - Reduce numeric precision while minimize information loss
      - Ex.: 32-bit floating point -> 8-bit fixed point (256 values)
    - Reduce memory footprint of network weights
  - Low-rank tensor decompositions
    - Replace (many) redundant filters by a linear combination of fewer filters
  - Knowledge Distillation
    - Transfer knowledge from complex (teacher) to simpler (student) model

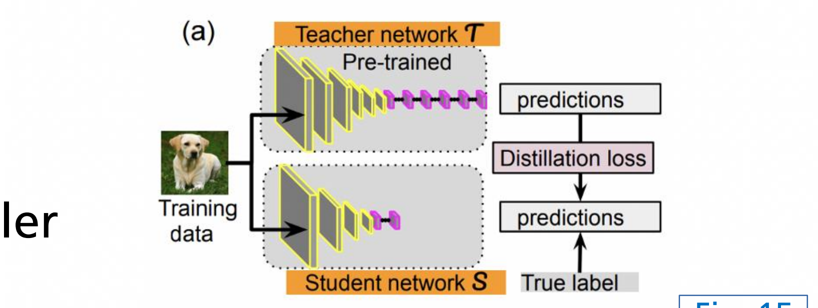


Fig. 15

Real-World Deployment
(3) Privacy / Data Protection and Data Security

- Depending on the specific application, challenges include e.g.
  - Avoiding processing and storage of speech content and speaker characteristics (person-related information)
  - Ensuring authenticity of recordings, and recording time / location
  - Ensuring confidentiality of recordings, annotations and models during storage, transmission and (sometimes) training
  - Avoiding replay attacks

Real-World Deployment
(3) Privacy / Data Protection and Data Security

- Countermeasures
    - Data anonymization (speech filtering / scrambling, etc.)
    - Data authentication, encryption and key management (based on security standards and cryptography)
    - Secure Federated Learning (incl. FHE and Differential Privacy)
    - Replay detection

Application Scenarios
(1) Urban Noise Monitoring

StadtLärm

Fig. 16

- 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)

Fig. 17

Fig. 18

Application Scenarios

(1) 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)

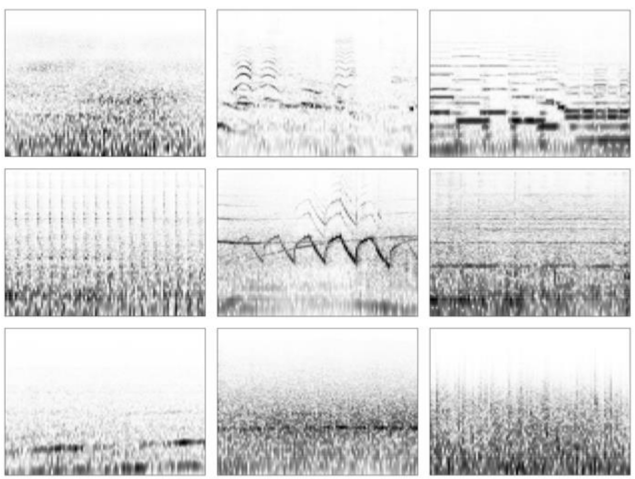


Fig. 19

- CNN architecture

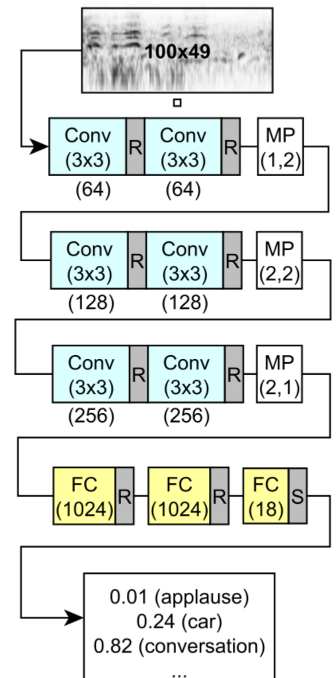


Fig. 20

Application Scenarios
(2) Traffic Monitoring

- Tasks
  - Vehicle detection
  - Direction of movement estimation
  - Speed estimation
  - Vehicle type classification
    - Car, truck, bus, motorcycle etc.
- Challenges
  - Microphone type
  - Local acoustic conditions
  - Vehicle speed
  - Street surface quality & weather conditions

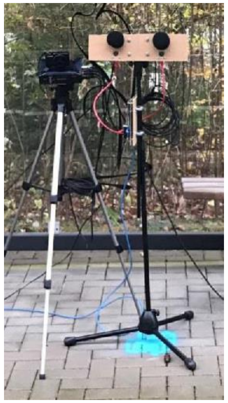


Fig. 21

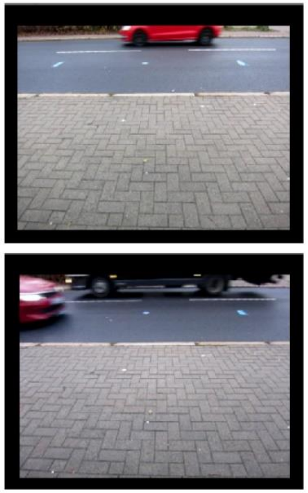
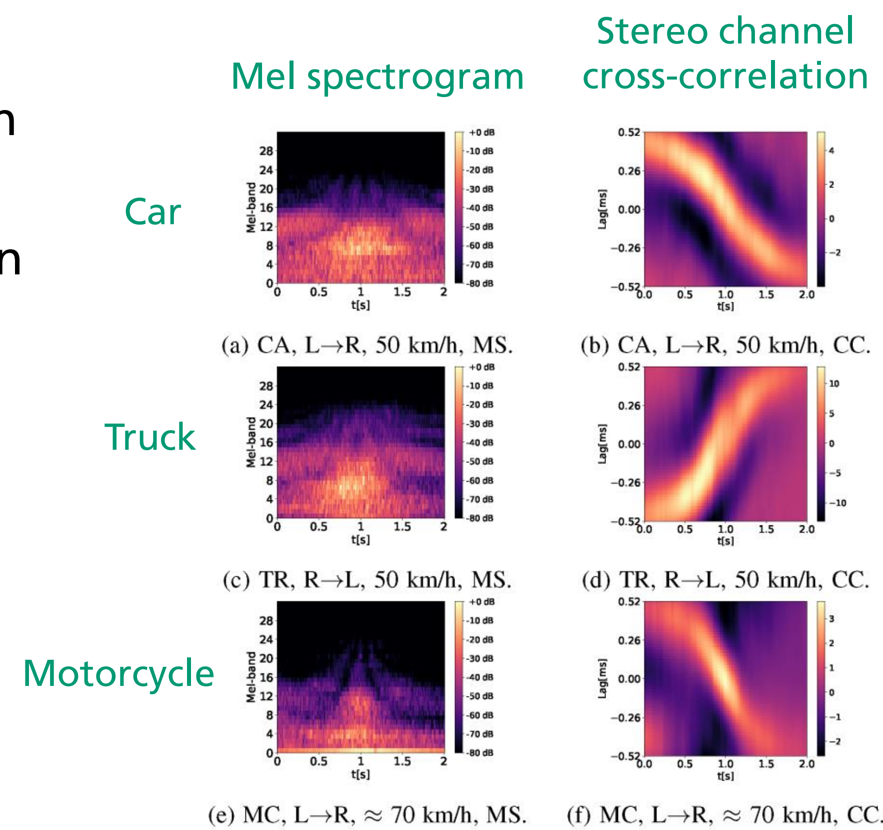


Fig. 22

Fraunhofer IDMT

Application Scenarios
(2) Traffic Monitoring

- Audio Features
  - Vehicle detection & direction of movement & speed
    - Channel cross-correlation
  - Vehicle type classification
    - Mel spectrogram
- Neural network architectures (#parameters)
  - CNNs (1,1 – 3,2 mio.)
  - MobileNetMini (15,000)
- Example (truck, car, motorcycle)
  - 2s clips (IDMT-Traffic dataset)



AUD-2

Fig. 23

Fraunhofer IDMT

Application Scenarios
(3) Industrial Sound Analysis

- 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

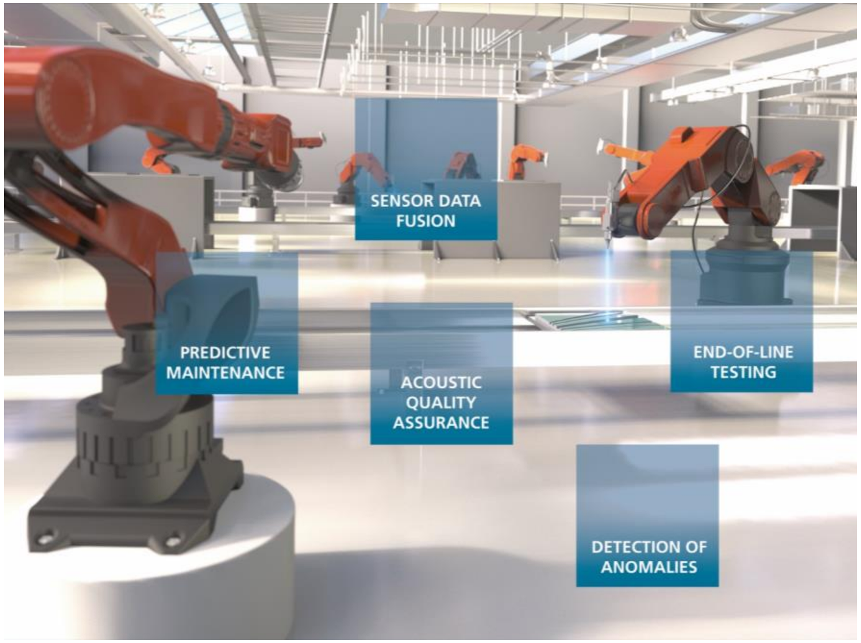
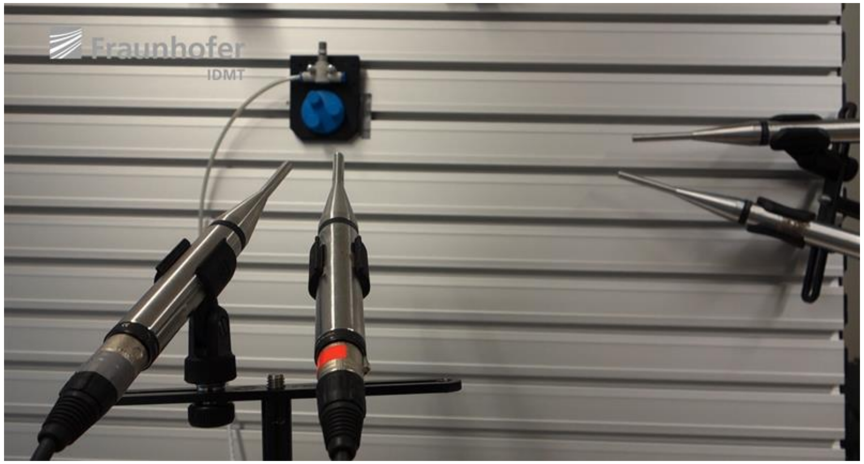


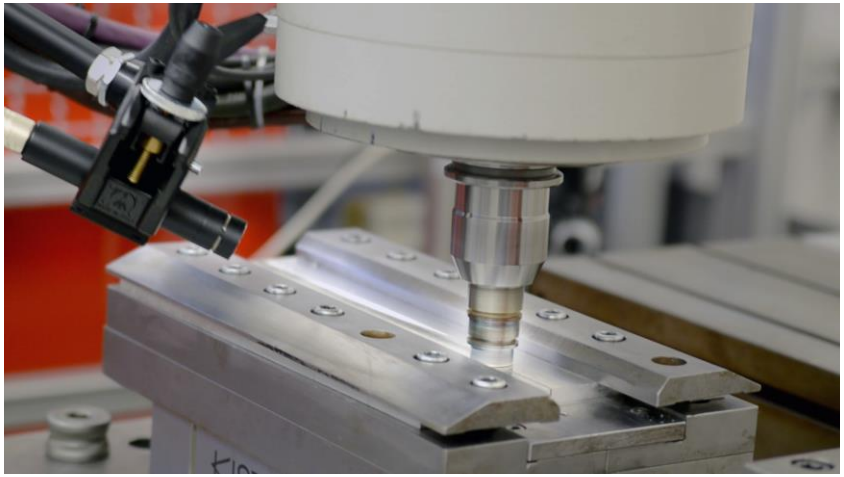
Fig. 24

Application Scenarios
(3) Industrial Sound Analysis

- Example use-cases @ Industrial Media Applications (Fraunhofer IDMT)



Compressed Air Leakage Detection Fig. 25



Friction Stir Welding Fig. 26

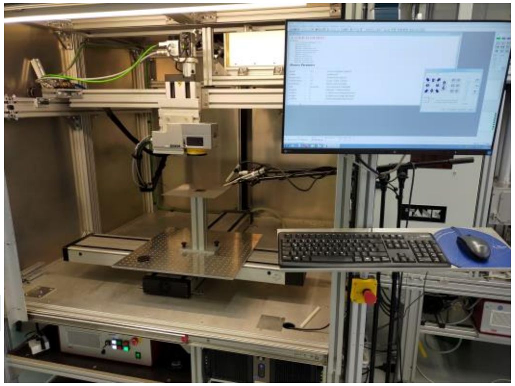


Fig. 27 Laser Ablation Machine

Application Scenarios
(4) Context-Sensitive Hearables

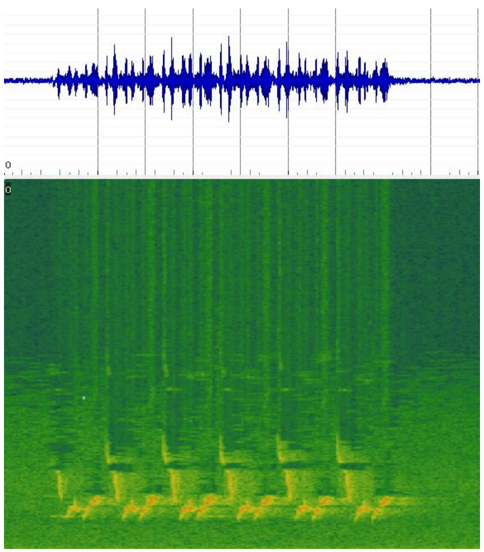
- Wireless earbuds, hearing aids
- Functionality
    - Context-awareness
        - Detect listeners location / activity (ASC)
            - E.g.: At home, traffic, subway, restaurant, sport
        - Detect relevant sound events (SED):
            - E.g.: Siren, honking, scream
    - Background noise reduction
    - Dynamic volume adjustments
    - (Immersive listening experience)

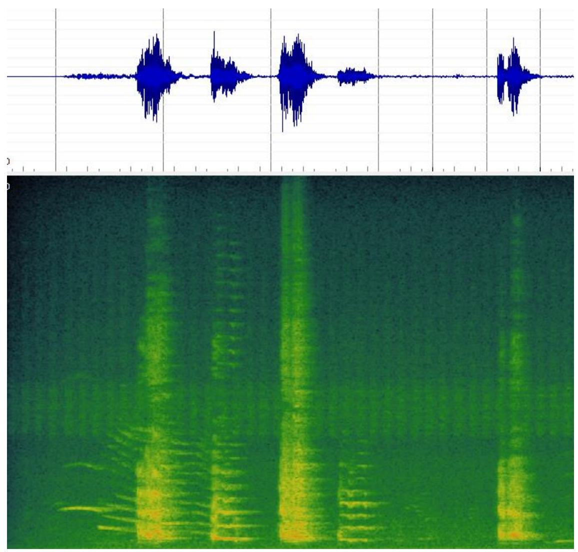
Application Scenarios
(5) Bioacoustic Monitoring

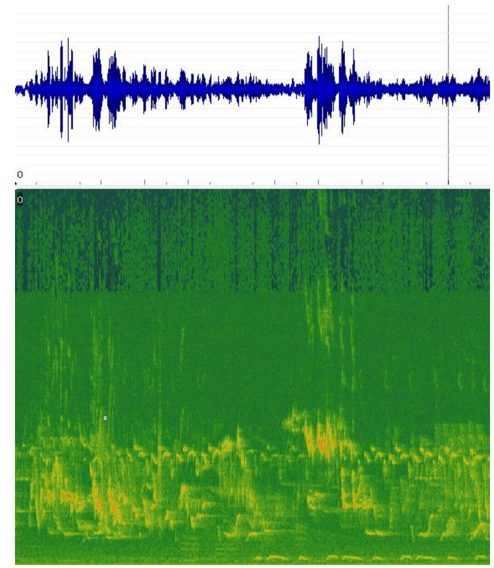
- Autonomous acoustic sensors
  - Non-intrusive
  - Allow for long-term recordings (days / weeks …)
- Monitored species: birds, primates, bees, marine mammals, etc.
- Monitor
  - Population sizes / migration patterns
- Challenges for SED
  - High variability even within sounds classes
  - Large amounts of unlabelled data (annotation requires expert knowledge)
    - Few-shot learning (DCASE 2021, task 5)

Application Scenarios
(5) Bioacoustic Monitoring

- Bird sound detection → detection / classification / counting







AUD-3

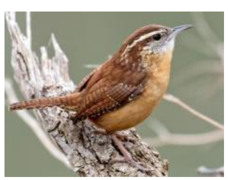


Fig. 28

Carolina Wren

AUD-4

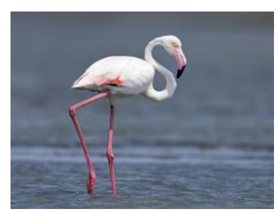


Fig. 29

Flamingo

AUD-5

Dawn chorus
(bird ensemble)

Summary

- Acoustic Scene Classification
- Acoustic Anomaly Detection
- Real-World Deployment
    - Process Steps
    - Challenges
- Use-Cases
    - Urban Noise Monitoring
    - Traffic Monitoring
    - Industrial Sound Analysis
    - Context-sensitive Hearables
    - Bioacoustic Monitoring

References

Abeßer, J. et al. (2018). A Distributed Sensor Network for Monitoring Noise Level and Noise Sources in Urban Environments. *Proceedings of the 6th IEEE International Conference on Future Internet of Things and Cloud (FiCloud)*, 318–324. Barcelona, Spain.

Abeßer, J., et al. (2019). Urban Noise Monitoring in the Stadtlärm Project - A Field Report. *Proceedings of the Detection and Classification of Acoustic Scenes and Events (DCASE)*. New York, NY, USA.

Abeßer, J. (2020). A Review of Deep Learning Based Methods for Acoustic Scene Classification. *Applied Sciences*, *10*(6).

Abeßer, J., Gourishetti, S., Kátai, A., Clauß, T., Sharma, P., & Liebetrau, J. (2021). IDMT-Traffic: An Open Benchmark Dataset for Acoustic Traffic Monitoring Research. *Proceedings of the European Signal Processing Conference (EUSIPCO)*. Dublin, Ireland.

Gharib, S., Drossos, K., Emre, C., Serdyuk, D., & Virtanen, T. (2018). Unsupervised Adversarial Domain Adaptation for Acoustic Scene Classification. *Proceedings of the Detection and Classification of Acoustic Scenes and Events (DCASE)*. Surrey, UK.

Howard, A. G., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv Preprint ArXiv:1704.04861*.

Johnson, D. S., & Grollmisch, S. (2021). Techniques improving the robustness of deep learning models for industrial sound analysis. *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 81–85. Amsterdam, The Netherlands.

References

Stowell, D., Petrusková, T., Šálek, M., & Linhart, P. (2018). Automatic acoustic identification of individual animals: Improving generalisation across species and recording conditions. *ArXiv Preprint ArXiv:1810.09273*.

Virtanen, T., Plumbley, M. D., & Ellis, D. (Eds.). (2018). *Computational Analysis of Sound Scenes and Events*. Cham, Switzerland: Springer International Publishing.

Wang, L., & Yoon, K. J. (2021). Knowledge Distillation and Student-Teacher Learning for Visual Intelligence: A Review and New Outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(8), 1–40.

Images

Fig. 1: [Virtanen, 2018], p. 267, fig. 9.7

Fig. 2: https://images.theconversation.com/files/349387/original/file-20200724-15-ldrybi.jpg

Fig. 3: http://dcase.community/challenge2020/task-unsupervised-detection-of-anomalous-sounds (Figure 1)

Fig. 4: https://scikit-learn.org/stable/_images/sphx_glr_plot_oneclass_0011.png

Fig. 5: https://miro.medium.com/max/722/1*TvZ9jI9vGX-fWwc3AHwNDw.png

Fig. 6: https://en.wikipedia.org/wiki/Raspberry_Pi#/media/File:Raspberry_Pi_4_Model_B_-_Top.jpg

Fig. 7: https://developer.nvidia.com/sites/default/files/akamai/embedded/images/jetsonNano/JetsonNano-DevKit_Front-Top_Right_trimmed.jpg

Fig. 8: https://www.idmt.fraunhofer.de/content/dam/idmt/documents/IL/IMA/AI4Edge_DE.pdf (cover image)

Fig. 9: [Virtanen, 2018], p. 154, fig. 6.2, right

Fig. 10: https://www.sonicvisualiser.org/doc/reference/1.7.2/en/images/pane-layers.png

Fig. 11: [Gharib, 2018], p. 3., fig. 2 (a) & (b)

Fig. 12: [Gharib, 2018], p. 2., fig. 1

Images

Fig. 13: IMG-13: Johnson & Grollmisch: Techniques improving the robustness of deep learning models for industrial sound analysis, EUSIPCO 2021, Fig. 1, p.82

Fig. 14: https://miro.medium.com/max/955/1*C3rR1-qzZfgYE_QA7WvLOQ.png

Fig. 15: [Wang, 2021], p. 2, fig. 1 (a)

Fig. 16: https://stadtlaerm.de/pics/talaerm.svg

Fig. 17: [Abeßer, 2019], p. 2, fig. 2

Fig. 18: [Abeßer, 2018], p. 3, fig. 2

Fig. 19: [Abeßer, 2019], p.3, fig. 3

Fig. 20: [Abeßer, 2018], p.5, fig. 4

Fig. 21 & 22: [Abeßer, 2021], p.3, fig. 1, (b, c, d) source images

Fig. 23: [Abeßer, 2021], p.3, fig. 2

Fig. 24-27: Fraunhofer IDMT

Fig. 28: https://www.allaboutbirds.org/guide/assets/photo/304470861-1280px.jpg

Fig. 29: https://cdn.download.ams.birds.cornell.edu/api/v1/asset/54167691/1800

Sounds

AUD-1: https://freesound.org/people/16HPanskaTyllova_Terezie/sounds/497363

AUD-2: Three clips from IDMT-Traffic dataset [Abeßer, 2021]

AUD-3: https://freesound.org/people/IFartInUrGeneralDirection/sounds/96195/

AUD-4: https://freesound.org/people/InspectorJ/sounds/400860/

AUD-5: https://freesound.org/people/Simon%20Spiers/sounds/516876/

Thank you!

- Any questions?

Dr.-Ing. Jakob Abeßer
Fraunhofer IDMT

Jakob.abesser@idmt.fraunhofer.de

https://machinelistening.github.io

Fraunhofer IDMT