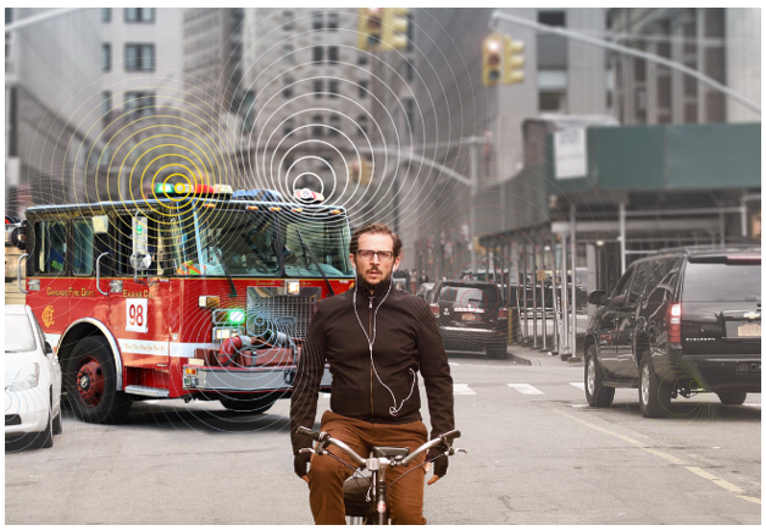
Machine Learning Based Context-Sensitive Acoustic Transparent Headphone

Christoph Sladeczek, Hanna Lukashevich, Jakob Abeßer, Patrick Kramer, Mario Seideneck, Wolfgang Lorenz, Josua Hagedorn
Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau

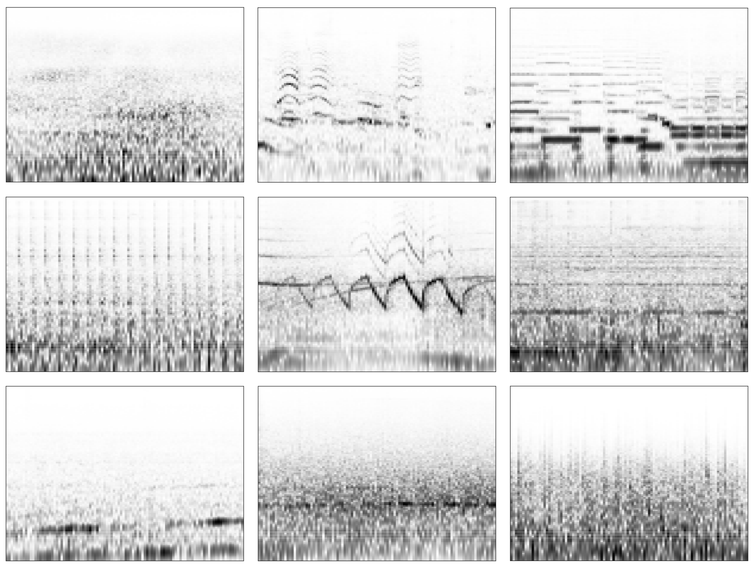
Motivation



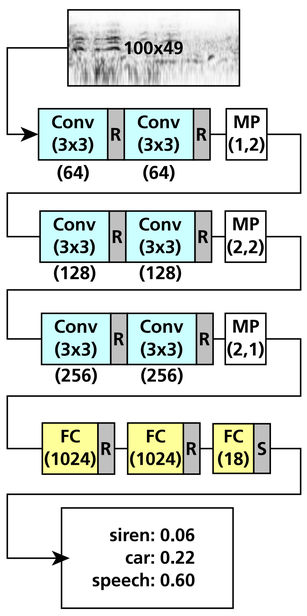
- Headphones are worn in (nearly) every life situation
- Safety-relevant noises can be overheard having appropriate loudness levels
- **Goal: Adding smart safety function to headphones by using adaptive acoustic transparency based on machine learning**

Acoustic Event Detection

- Characteristic patterns of multiple acoustic events are learned and recognized in short-term spectrograms



- Hybrid deep convolutional neural network (CNN) architecture with dense classification layers allow for a simultaneous detection of multiple event types

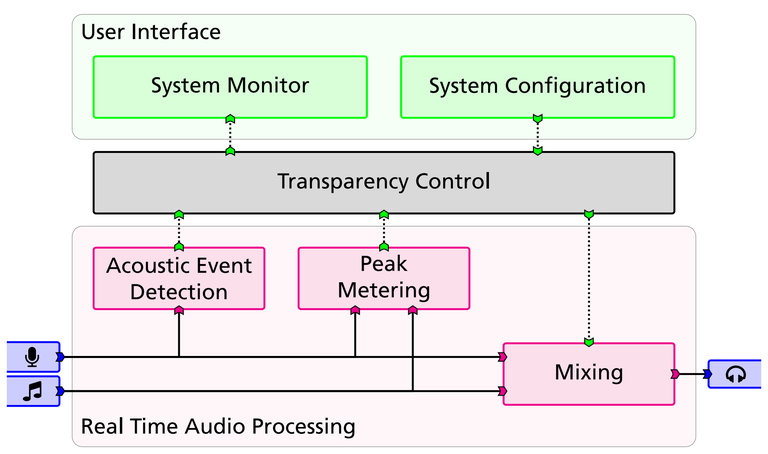


- Hierarchical feature learning via pairs of convolution layers with intermediate downsampling using max pooling

Training Data

- Audio recordings and acoustic event annotations compiled from publicly avaible datasets
  - Urban Sound Dataset (https://urbansounddataset.weebly.com/)
  - Tampere University of Technology (TUT) Sound Events dataset (http://www.cs.tut.fi/ heittolt/datasets)
- Applied data augmentation methods for enlarging the training dataset:
  - Time stretching
  - Pitch shifting
  - Dynamic range compression
  - Mixing with environmental background noises

Prototype Implementation



- Implementation of acoustic event detection module as portable C++ library
- Recognized acoustic events: siren, car, speech
- Signal processing:
  - Input: recorded environment signal, user audio signal (e.g. music)
  - Confidences of acoustic event detection are transmitted to Transparency Control block
  - Transparency Control decides further actions (ignore/set transparency)
  - Weighted mix of and
  - Output: mixed audio signal
- Communication with web based user interface

- Decision about signal class based action to take, configurable in user interface
- Monitoring of several system properties:
  - Input levels
  - Event detection confidence
  - Transparency control gains

Prototype Demonstration Data

- Goal: Test of system under realistic conditions
- Sound field recordings for capturing environmental sounds
- Recording of traffic noises using a microphone array

- Microphone arrangement compliant with object-based reproduction

Demonstration Environment

- Object-based sound (re)-production system developed by Fraunhofer IDMT
  - Combines the ability for interactive spatial sound design and acoustic room enhancement in one device
  - 64 audio objects
  - Support of (nearly) every loudspeaker setup; no need to place loudspeakers at fixed positions

- Acoustic laboratory that was build according to recommendation BS.1116
- Three-dimensional, multi-channel loudspeaker system consisting of 120 loudspeakers
- Spatial audio reproduction system rendered recorded traffic noises using audio objects
- Sounds to be detected can be rendered at arbitrary positions using audio objects or introduced as real objects
- Headphone equipped with omnidirectional microphone, used as input signal for acoustic event detection ()
- Informal tests showed, the concept is working
- Further evaluation has still to be carried out

Acknowledgement

Parts of this work were carried out in the project **KISH - AI-based selective hearing for headphones**, which was funded by the Federal Ministry of Economics and Technology (BMWi) within the framework of the AI Innovation Competition.
More information can be found on the webpage: **https://www.idmt.fraunhofer.de/kish**.

Fraunhofer IDMT