Selective Hearing: A Machine Listening Perspective

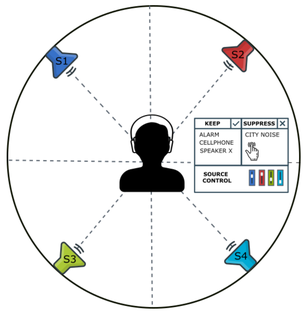
Estefanía Cano
Fraunhofer IDMT
cano@idmt.fraunhofer.de

Hanna Lukashevich
Fraunhofer IDMT
lkh@idmt.fraunhofer.de

DEFINITION

Selective Hearing

The possibility to selectively enhance, attenuate, suppress or modify sound sources in the auditory scene by means of a hearing device such as headphones, earbuds, etc.



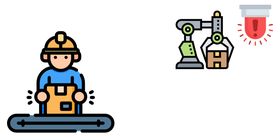
USE-CASE 1: BICYCLE COMMUTE

The biker is commuting home while playing music through his headphones

- Noise control blocks the background city noise
- The system detects and localizes cars approaching
- The system warns the biker of possible dangers

ANC DETECTION LOCALIZATION

USE-CASE 2: INDUSTRIAL SAFETY



The worker is performing a task while wearing noise protection earmuffs.

- Noise control blocks background noise from the machinery.
- A machine behind the worker is malfunctioning
- The alarm goes off to alert the workers of the danger
- The system detects the alarm sound and alerts the worker

ANC DETECTION

USE-CASE 3: CONFERENCE ENHANCEMENT

There is a speaker giving a talk at a conference. There is a person with minor hearing loss in the audience. A loud noise source can be heard outside of the conference room.

- The voice of the speaker is enhanced to improve intelligibility for the audience
- The noise source is localized
- The noise source is suppressed using separation technologies.

ENHANCEMENT SEPARATION LOCALIZATION

A LOOK AT THE STATE-OF-THE-ART

Localization

- Direction of arrival (DOA) either in 2D (azimuth) or 3D (including elevation) [39], [40]
- Number of source in the scene required
- Joint estimation of the number of sources and locations [41]
- Efforts to reduce computational requirements [42-45]

Detection & Classification

AUDIO EVENTS
- Relevant work comes from the field of detection and classification of acoustic scenes and events (DCASE) [18]
- Monophonic [20] and polyphonic [21] sound event detection
- Robustness to weak and noisy labels [23], [25]

SPEECH
- Voice activity detection [28-30]
- Speaker recognition [31]
- Focus on increased robustness to different conditions [32-34]

MUSIC
- Singing voice activity detection in music [38], musical instrument activity detection [37]
- Musical instrument classification [35]

Separation

- Speaker-independent separation [46] and lead instrument separation in the music domain [52] are the most common applications
- Use of spatial location information to help separation performance [4], [47]
- Efforts to reduce computational demands of algorithms [48-51]

Active Noise Control (ANC)

- Very strict computational requirements (ms)
- Work on specific types of noises: automobile cabins [56], industrial scenarios [57]
- ANC on a specific spatial region [58]

Enhancement

- The most common application is enhancement of speech corrupted by noise.
- Single channel speech enhancement [59-61]
- Music remixing applications also use enhancement to increase loudness of a musical instrument in the mix [63]
- Enhancement is often used in conjunction with sound separation methods [6] [64]

ecanoc/SelectiveHearing

CURRENT TRENDS & CHALLENGES

Current Trends:

1. Joint optimization of more than one machine listening tasks
   - Sound event localization and detection [27]
   - Sound event separation and classification [65]

2. Efforts to increase accuracy and robustness
   - Robustness to noise [25], [26]
   - Embeddings [32] and domain adaptation [67]

3. Computational Efficiency
   - Low-latency source separation [48]
   - End-to-end systems for localization [45], separation [50], speech enhancement [61]

Challenges:

1. Methods capable of joint optimization of more than one machine listening task
2. Real-time processing is required for realistic selective hearing applications

Fraunhofer IDMT