**Motivation**

- Augmented Reality (AR) audio needs real sounds unaltered and fusion of real and virtual sounds
- Hear-through (HT) used for real sounds and binaural rendering for virtual sounds
- Different playback devices have certain advantages and disadvantages for AR audio:
  - Open back headphones
    - Advantages: Passive hear-through
    - Disadvantages: Comb filtering at high frequencies
  - Closed in-ear headphones
    - Advantages: Single Pinna cues preserved naturally
    - Disadvantages: Low fitting
  - Closed back headphones
    - Advantages: Open ear canal listening
    - Disadvantages: Need to embed pinna cues in EQ
  - Open ear emitter
    - Advantages: No need for HT EQ
    - Disadvantages: Poor isolation and bass leakage effects

**Proposed system**

AR audio headset prototype

- AR audio headset prototype shown in previous study [1]
- External microphones denoted by \( r_{EXT,U}(t) \) and processed ear signals by \( u_{R}(t) \)
- Aim of the system: DoA estimation and directional HT filtering in time-frequency domain

**Data estimation using neural network**

- NN based model tested for frontal source directions (-90° to 90°)
- Model trained using 10 s white noise filtered with HRTFs measured at external microphone
- Simple network topology:

  ![Network Diagram]

- Input layer: 128 nodes with 102 dimension input vector
- Hidden layer: Single hidden layer with 128-13 nodes
- Output layer: 25 samples; maximum 100 epochs

**Parameter-hear-through processing**

- EQ filters precomputed to cover 360° at resolution of 15° called idealHT
- Zone based EQ filters for three zones (GroupedHT): frontal (-60° to 60°), lateral (60°-120°, -60° to 120°), and rear (120° to 180°, -120° to -180°)
- AvgHT: Average across all directional EQ filters
- Filtering: STFT of captured signal \( r_{EXT,U}(k,n) \) filtered by sub-band directional filter \( h_{GROUPED}(k,\theta(k,n)) \) chosen for each direction

**Results and analysis**

- Signal synthesis
  - 2 uncorrelated pink noise signals of 2s each filtered by 3 bandpass filters: 0.1-1 kHz (low), 1.5-6 kHz (middle), and 5-16 kHz (high)
  - Obtained signals filtered with impulse response for two direction pairs: (0°, 30°) and (-15°, 75°)
  - All combinations taken (total 12, 6 each for overlapping and non-overlapping frequency bands)
  - Real sound: broadband music and narrowband speech signal (4s each) convolved with 2 directions chosen randomly from set of 13 azimuthal positions (total 156 soundtracks)

- Hear-through equalization results
  - \( SD_{combined} = \frac{\sum_{k=1}^{N} \sum_{n=1}^{M} \left( r_{EXT,U}(k) - r_{REF}(k) \right)^2}{\sum_{k=1}^{N} \sum_{n=1}^{M} r_{EXT,U}(k)^2} \)

- \( SD_{L/M} = \frac{1}{r_{REF}(k)} \sum_{k=1}^{N} \sum_{n=1}^{M} \left[ \frac{\left( r_{EXT,U}(k) - r_{REF}(k) \right)^2}{\left( r_{REF}(k) \right)^2} \right] \)

**Conclusion and future work**

- Directive EQ filters (IdealHT and groupedHT) show close match to reference for all cases, including real and overlapping sounds
- NN based DoA approach using IACC and ILD features shows good localization performance

**Conclusions**

- Directive EQ filters (IdealHT and groupedHT) show close match to reference for all cases, including real and overlapping sounds
- NN based DoA approach using IACC and ILD features shows good localization performance

**Future work**

- Real time system with NN based DoA estimation and parametric HT filtering
- Sound classification and HT for diffuse sounds

**References**


**Acknowledgement**

This work is supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research grant MOE2017-T2-2-060.