Outline of Tutorial : (9am – 10:30am)

I. 3D Sound and Headphones

II. Natural Sound Rendering for Headphone
   - Virtualization
   - Sound scene decomposition
   - Individualization
   - Equalization
   - Head tracking
   - Integration
   - 3D Audio Headphones
   - Demo
III. An Overview and Applications of ANC Headsets

IV. Natural Augmented Reality Audio in Headsets
   • Signal processing and practical challenges
   • ARA headset for mobile and wearable devices
   • Multichannel headphone sound reproduction
   • Natural augmented reality (NAR) headset using adaptive techniques

V. Assisted Listening in Hearing aids
   • Hearing loss and hearing aids
   • Noise reduction and speech enhancement
   • Integration of ANC and noise reduction in hearing aids

VI. Summary and Demo
Module I

3D Sound and Headphones
(Assisted) Binaural Listening

W. S. Gan, J. He (NTU)

Assisted Listening for Headphones and Hearing Aids

16th Dec 2015
Selected Kickstarter projects on assisted listening

Hooke: Wireless 3D Audio Headphones

Wireless headphones with built in binaural microphones that let you listen to music, take calls and capture 3D Audio on any device at any time.

Follow along

Created by Hooke

1,300 backers pledged $193,196 to help bring this project to life

OpenEars

Start the sound of your life whenever you want!

First Bluetooth In-Ear Headphones to record your video with 3D audio to share & stream. GoPro connection, great sound. HearThrough mode

227 backers

€34,336 pledged of €50,000 goal

19 days to go

MUSIC, sound, technology, where the waves meet.

Brionic ME

Bionic Membrane Ears - 3 backed

bioear.com

Find me on bioear.com

Wireless Smart Headphones

Sleep better with the world’s first sleep sensing EEG headphones. Perfect peace and comfort with audio that responds to your sleep.

2,802 backers

$622,698 pledged of $100,000 goal

35 days to go

London, UK

The Dash – Wireless Smart In Ear Headphones


Pre-order now

Created by BRAGI LLC.

15,598 backers pledged $1,300,000 to help bring this project to life.

Rondo Motion: Bring your headphones to life

Add motion sensing technology to your headphones and get an unparalleled immersive sound experience

created by Dynomic Inc.

199 backers pledged $45,831 to help bring this project to life.

HELIX

The World's First Wearable with Wireless Bluetooth Headphones

HELIX by Ashley Chloe is the first wearable with headphones on your wrist. HELIX ensures that fashion and tech are always at hand.

PRE-ORDER now

Created by Ashley Chloe Inc.

1,858 backers pledged $294,265 to help bring this project to life

Aivy Q: Smart Headphones Caches Personalized Music For You

High-end headphones, a streaming app and personalized streaming music all in one magic device Aivy Q.

Pre-order now

Created by the aivy team

811 backers pledged $168,573 to help bring this project to life.
A realistic and engaging experience
Speakers vs Headphones

Speaker systems

Headphones

Surround Sound system

Speakers vs Headphones
Strong headphone market

2013 Top 5 Revenue Categories

Growth: ~10%

Source from

Source from
Market Share

AV Headphone Market by Type: Worldwide

Worldwide AV Headphone Aftermarket, 2011: 228 million units

- Over-Ear 16%
- On-Ear 20%
- In-Ear 64%

© 2012 Futuresource Consulting Ltd
## Prices of Over-Ear Headphones

<table>
<thead>
<tr>
<th>Headphone Type</th>
<th>Target Consumers</th>
<th>Headphones Brands</th>
<th>Features and comments</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stereo</strong></td>
<td>General Pro-s user</td>
<td>AKG, BeyeDynamic, Denon, German Maestro, Grado, Koss, Ultrasone</td>
<td>High fidelity stereo, mostly requires power amplification for optimal sound</td>
<td>Hundreds to Thousands</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-End Consumer</strong></td>
<td></td>
<td>Bose, Fanny Wang, Jays, Monster (beats), Sony, Shure</td>
<td>Stereo, simple to drive, can feature good fidelity stereo</td>
<td>Hundreds (&lt; $500)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>General Consumer</strong></td>
<td></td>
<td>Beats, Creative, Eskuche, Hed Kandi, Goldring, iFrogz, Jays, Marshall, Ministry of Sound, Philips, Pioneer, Shure, SkullCandy, SonicGear, Sumajin, TDK</td>
<td>Low cost, stereo, meant for general usage, styling of headphone is general more critical than sound quality</td>
<td>Hundreds (&lt; $200)</td>
</tr>
<tr>
<td><strong>Stereo with Virtual Surround</strong></td>
<td>General Consumer/Gamers</td>
<td>Acoustic Research, Creative, Corsair, Philips, Logitech, Pioneer, Razer, SteelSeries, Sony, Turtle Beach, Tritton, Ultrasone, Yamaha, Zulman</td>
<td>Virtual surround achieved with HRTF processing, Dolby headphone, S-Logic (ultrasone), CMSS-3D (creative), Dolby Ex (turtle beach)</td>
<td>Hundreds (&lt; $200)</td>
</tr>
<tr>
<td><strong>Discrete Surround</strong></td>
<td>Consumer for gaming, general entertainment</td>
<td>Creative, Logitech, Mentor, Razer, Psyko, Razor, Turtle Beach, Tritton, Zalman</td>
<td></td>
<td>Hundreds (&lt; $400)</td>
</tr>
</tbody>
</table>
## Hearings Aids

### DC (deep-canal)
- Mild to moderate hearing losses
- Fits deep inside the ear canal, making it **invisible**
- Less occlusion
- Not suitable for people with narrow ear canals
- Size prevents the use of directional microphones

### CIC (completely-in-canal)
- Mild to moderate hearing losses
- Very small case
- Fits inside the ear canal, making it practically invisible

### ITC (in-the-canal)
- Mild to moderately-severe hearing losses
- Small, one piece case
- Fits inside the ear canal
- Directional microphones are possible with this model

### RIE (receiver in ear)
- Mild to moderately-severe hearing losses
- Ear canal open for a natural sound quality
- Smallest external hearing aid, as the receiver is located in the end of the tube inside the ear
- Very small case that sits behind the ear, making it practically invisible

### Open (open ear)
- Mild to moderately-severe hearing losses
- Ear canal is open for a natural sound quality
- Very small case that sits behind the ear, making it practically invisible
- Many colour options

### BTE (behind-the-ear)
- Mild to severe losses
- Fully featured hearing aids
- Larger case can be easier for wearers with dexterity considerations
- Case contains all features and sits behind the ear
- Many colour options

### Power (high powered)
- Profound hearing losses
- More powerful solutions that provide the greatest levels of amplification
- Larger case worn behind the ear

Hearing Aids market

“The hearing aids market is expected to reach USD 8,374 million by 2020 from USD 6,183 million in 2015, at a CAGR of 6.3%.” - Marketsandmarkets reports

“Hearing Aid Sales Increase by 8.8% in First Half of 2015.” - Hearing Industries Association (HIA)

Source from http://www.bdti.com/InsideDSP/2014/10/16/ONSemi
Pursuing natural 3D sound in Headphones Industry


Source from https://www.youtube.com/watch?v=RW-JDy2uSyY

Source from http://www.razerzone.com/referral/invite/surround/?ref=Adrian%20Wong&ref_email=awsh%40techarp.com


Preferred Headphone Target Response

Source from http://seanolive.blogspot.sg/2014/01/the-perception-and-measurement-of.html
Experimenting 3D audio

• Designing an experimentation platform using headphones
  – With user interaction (touch and spatial audio)
  – Sonification / Auralization
  – User preference / hearing profile
  – Personalized HRTF / anthropometry
  – Computation and resources demand in portable devices

• Creating new mobile and wearable apps
  – Assistive Applications
  – Enhanced Telepresence
  – Remote Monitoring

• Goal: design a headphone system which is perceptually indistinguishable from real listening.
Existing Apps: Binaural Recordings

3D Sounds Illusions

- Package name: sounds3d.soundboard

Retail: free (with ads)

Real-time 3D Audio Processing: ✗

HRTF Processing: ✗

Features:

- A collection of binaural recording files
- Playback binaural recordings.
Existing 3D audio Apps in the Market

Parrot headphones apps

Headquakes apps

Audio-3D Player Headphones HD 7.1

They are mainly stereo extender and virtual downmix apps
3DA³

App Statistics

<table>
<thead>
<tr>
<th>CURRENT / TOTAL</th>
<th>AVG. RATING / INSTALLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,048 / 15,676</td>
<td>★ 4.11 / 132</td>
</tr>
</tbody>
</table>

Main Features

1. Real-time DSP on Android platform
2. Fully customized user interface and design.
3. Audio Widening (Externalization)
4. 3D Audio using HRTF filtering
5. Near Field 3D (audio depth) with recording
6. Elevation function
7. Auto-rotate function
8. Up to four channels simultaneous processing
9. Virtual Bass System integration
10. Ffmpeg Audio Decoding

Download here:
http://tinyurl.com/kboq2g7

Technical Specifications

<table>
<thead>
<tr>
<th>Supported Platform</th>
<th>Android</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supported OS Versions</td>
<td>4.0 and up</td>
</tr>
<tr>
<td>Minimum CPU requirement (Chipset)</td>
<td>Snapdragon 400, Nvidia Tegra 3 T30L, Mediatek MT6589T, Exynos 4210</td>
</tr>
<tr>
<td>Minimum RAM requirement</td>
<td>512 MB</td>
</tr>
<tr>
<td>App Size</td>
<td>6.84 MB</td>
</tr>
<tr>
<td>Audio Decoder</td>
<td>FFMPEG library</td>
</tr>
<tr>
<td>Supported formats</td>
<td>MP3, WAV</td>
</tr>
<tr>
<td>Supported Sampling rate</td>
<td>44100Hz</td>
</tr>
<tr>
<td>Frame Size</td>
<td>Stereo: 1152 (Mp3), 1024 (WAV)</td>
</tr>
<tr>
<td>HRTF Database</td>
<td>CIPIC HRTF Database</td>
</tr>
<tr>
<td>HRTF taps</td>
<td>200</td>
</tr>
<tr>
<td>Data type</td>
<td>double-precision 64-bit IEEE 754 floating point</td>
</tr>
<tr>
<td>HRTF sets (48 points/set)</td>
<td>Azimuth: 6 sets, Elevation: 2 sets, Near Field: 10 sets</td>
</tr>
<tr>
<td>Memory usage (HRTF)</td>
<td>6.75 KB</td>
</tr>
<tr>
<td>Azimuth HRTF Resolution</td>
<td>7.5°</td>
</tr>
<tr>
<td>Elevation HRTF Resolution</td>
<td>7.5°</td>
</tr>
<tr>
<td>Near Field depth</td>
<td>25cm, 50cm, 75 cm, 100cm, 125cm</td>
</tr>
<tr>
<td>Devcies tested</td>
<td>Asus Transformer Pad TF300T, Xiaomi Redmi, Samsung Galaxy S2, S3, S4, S5, Asus Padfone Infinity</td>
</tr>
</tbody>
</table>

W. S. Gan, J. He (NTU) Assisted Listening for Headphones and Hearing Aids 16th Dec 2015
Demo video on 3DA
Quick Overview

- Techniques to capture 3D sound (Spatial Audio Recording)?
- **Types of processing required**
- Rendering over **headphones** and loudspeakers.
Spatial Audio Recording

- Binaural recording
  - Dummy head
  - Human subjects

- Stereo/5.1

- B-format

- Microphone arrays
Spatial Audio Processing

- Spatial audio coding
  - Directional audio coding (DirAC)
  - Spatial audio scene coding (SASC)
  - Binaural cue coding (BCC)
Spatial Audio Processing

- Spatial audio coding
  - Directional audio coding (DirAC)
  - Spatial audio scene coding (SASC)
  - Binaural cue coding (BCC)

- Audio mixing (Down-mix, Up-mix)
- Sound Scene Decomposition
Spatial Audio Processing

- Spatial audio coding
  - Directional audio coding (DirAC)
  - Spatial audio scene coding (SASC)
  - Binaural cue coding (BCC)

- Audio mixing (Down-mix, Up-mix)
- Sound Scene Decomposition
- Binaural synthesis
- Artificial reverberation
- Equalization, Decorrelation, Crosstalk cancellation etc.
Spatial Audio Standards

- **MPEG Surround**

- **MPEG SAOC**

- **MPEG-H 3D Audio**
Spatial Audio Reproduction

- Stereo/ Multichannel Loudspeakers (5.1, 7.1, 22.2)

- Stereo Headphone
  - Binaural Headphone

- Wave Field Synthesis (WFS)

- Ambisonics
Emerging Spatial Audio Reproduction

- Dolby Atmos
- DTS:X
- AURO 3D Audio

Layer 3 Top
Layer 2 Height
Layer 1 Lower

Layer 3 Top
Layer 2 Height
Layer 1 Lower

AURO 11.1 Height and Top Layer
5.1 Surround Layer
Binaural Technology

- Encompasses a set of tools for recording or synthesizing and rendering binaural signals at the listener’s ear

- Deals with the natural cues of auditory localization which results from the reflection and diffraction of the acoustical waves with the human torso, shoulders, and the external ears

* Picture from logicomusic.com.br
From Real to Virtual Reality

Real

Capture

Virtual

Reproduction

Transmission
How do we perceive sound location?

- Compare sound received at two ears
  - **Interaural Level Differences (ILD)**
    - Effective for high frequencies above 1.5 kHz
    - Head size > wavelength
    - Smallest detectable ILD = 0.5 dB

- **Interaural Time Differences (ITD)**
  - Effective for low frequencies below 1.5 kHz
  - Rayleigh’s duplex theory of ILD and ITD
  - Smallest detectable ITD = 13 μs

Inadequacy of Interaural Difference Cues

- Similar ITD and ILD due to:
  - Cone of confusion
  Sources A & B; Sources C & D have identical ITD and ILD

- Media Plane (extreme case of cone of confusion)

- We need another sound localization cue!
Sound interacts with torso, head, external ears and arrives at the two ear canals.
Main Binaural Cues due to Head Related Transfer Function (HRTF)

- Sound wave scatters of torso (~45 cm), head (~20 cm) and ears (~4 cm)
- Also scatter off surrounding (>2 m)
- Model scattering effect independently
- Ears act as directional acoustic probes.
Head Related Transfer Functions (HRTF)

Many high-frequency details due to pinna scattering

HRTF of KEMAR dummy head for an angle of 30 degree azimuth*


Sound of car is slight louder and faster on left ear as compared to right ear

W. S. Gan, J. He (NTU) Assisted Listening for Headphones and Hearing Aids 16th Dec 2015
• Head related impulse response (HRIR)
• Encodes the acoustic propagation between the sound source and the listener’s ears
• Characteristic to the position of the sound source with respect to the listener
• Highly dependent on the morphology of listener.
Binaural Recording can be done either at the eardrum (or at blocked ear canal) of a listener or a dummy head.

- Played back using a stereo set of loudspeakers or a headphone
- Encapsulates all the directional information generated by the interaction of the sound with the listener’s morphology

Binaural Recording on human subjects and dummy heads
Individual Sound Filtering (Earprint)

Variation in Pinna morphology

Pinna of human subjects taken from the CIPIC database

- Human pinna is found to be as idiosyncratic as the fingerprint
- HRTFs are highly individual and differs substantially from one subject to the other
- For perfect 3D audio playback, individualized recordings/HRTFs and individualized headphone equalization are required
Highly Individualized Ear’s Response

![Diagram of Frequency (kHz) vs. Elevation (deg) for Transfer (dB)]

Picture extracted from Paul M. Hofman, “Relearning sound localization with new ears,” nature neuroscience • volume 1 no 5 • september 1998
CIPIC Anthropometry Measurements

Figure 2: Head and torso measurements

<table>
<thead>
<tr>
<th>Var</th>
<th>Measurement</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>head width</td>
<td>14.49</td>
<td>0.95</td>
<td>13</td>
</tr>
<tr>
<td>$x_2$</td>
<td>head height</td>
<td>21.46</td>
<td>1.24</td>
<td>12</td>
</tr>
<tr>
<td>$x_3$</td>
<td>head depth</td>
<td>19.96</td>
<td>1.29</td>
<td>13</td>
</tr>
<tr>
<td>$x_4$</td>
<td>pinna offset down</td>
<td>3.03</td>
<td>0.66</td>
<td>43</td>
</tr>
<tr>
<td>$x_5$</td>
<td>pinna offset back</td>
<td>0.46</td>
<td>0.59</td>
<td>254</td>
</tr>
<tr>
<td>$x_6$</td>
<td>neck width</td>
<td>11.68</td>
<td>1.11</td>
<td>19</td>
</tr>
<tr>
<td>$x_7$</td>
<td>neck height</td>
<td>6.26</td>
<td>1.69</td>
<td>54</td>
</tr>
<tr>
<td>$x_8$</td>
<td>neck depth</td>
<td>10.52</td>
<td>1.22</td>
<td>23</td>
</tr>
</tbody>
</table>

Figure 3: Pinna measurements
3D audio reproduction using headphones is degraded due to front-back confusion and in-the-head localization. By overcoming these issues, we are able to faithfully recreate 3D audio using headphones (as in natural listening).
Binaural audio reproduction over headphones

Binaural audio is highly idiosyncratic

- Front-Back Confusions
- Non-Individualized Headphone Equalization
- Non-Individualized HRTFs
- Imperfect 3D Sound
- In-head localization

Use of non-individual HRTFs degrades the veracity of the perception of 3D sound

**Solution:**
Individualizing the non-individual HRTFs can improve the 3D audio playback using headphones
Other Challenges and New Applications

- Natural Sound Rendering
- Active Noise Reduction
- Extension to Augmented Reality Applications
- Assisted Listening
- Binaural Signal Processing...

Real

Augmented Reality (AR)

Virtual

Capture

Sony Project Morpheus
Oculus rift
Benefits of Spatial (binaural) Listening

- **Binaural Localization**
  - Improve situation awareness and provide augmented information to visual

- **Cocktail party effect**
  - Improve intelligibility for remote sensing and better discriminate or selective attention

- **Active Noise Control**
  - Better conversation and ability to hear others in telecommunication.

- **Natural Listening**
  - Improve immersive listening and enhance “being there” experience

Adapted from Begault
Key References on Fundamentals of 3D Sound

Module II

Natural Sound Rendering for Headphones

To achieve natural sound rendering in headphones

**Natural sound rendering** essentially refers to rendering of the spatial sound using headphones to create an immersive listening experience and the sensation of “being there” at the venue of the acoustic event.

- **Differences** between natural listening and headphone listening;
- **Challenges** for rendering sound in headphone to mimic natural listening;
- How can **signal processing** techniques help?
- How to **integrate** these techniques?
- And **subjective evaluation**
## Challenges and solutions

<table>
<thead>
<tr>
<th>Listening</th>
<th>Headphone listening</th>
<th>Signal processing</th>
<th>Natural listening</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td><img src="image-source.png" alt="Source" /></td>
<td>Virtualization</td>
<td><img src="image-natural-source.png" alt="Natural Source" /></td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td><img src="image-medium.png" alt="Medium" /></td>
<td>Sound scene decomposition</td>
<td><img src="image-natural-medium.png" alt="Natural Medium" /></td>
</tr>
<tr>
<td><strong>Receiver</strong></td>
<td><img src="image-receiver.png" alt="Receiver" /></td>
<td>Equalization</td>
<td><img src="image-natural-receiver.png" alt="Natural Receiver" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Individualization</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Head tracking</td>
<td></td>
</tr>
</tbody>
</table>

Signal processing techniques

1. **Virtualization:**
   - to match the desired playback for the digital media content;

2. **Sound scene decomposition:**
   - to optimally facilitate the separate rendering of sound sources and sound environment;

3. **Individualization:**
   - to compensate for the lost or altered individual filtering of sound in headphone listening;

4. **Equalization:**
   - to preserve the original timbral quality of the source and alleviate the adverse effect of the inherent headphone response;

5. **Head tracking:**
   - to adapt to the dynamic head movements of the listener.
Virtualization

(a) Virtualization of multichannel loudspeaker signals

(b) Virtualization of source and environment signals

Virtualization with head tracking

- Head movement information is tracked by a sensor (e.g., accelerometer, gyroscope, magnetometer, camera);
- Adapt to the changes of sound scene with respect to head movements;
- Reduce front-back confusions, azimuth localization errors;
- Concern of head tracking latency (80ms).
Virtualization: further considerations

- **Add reverberation**
  - Externalization of the sound sources, and enhance depth perception;
  - Rendering of the sound environment;
  - How to select correct amount of reverberation.

Source from http://www.torgny.biz/Recording%20sound_2.htm
**Sound scene decomposition: overview**

**Aim:** to obtain useful information about the original sound scene from given mixtures, and facilitate natural sound rendering.

- **Blind source separation**
  - “Sum of sources”

- **Primary ambient extraction**
  - “Sum of primary and ambient components”

## Sound scene decomposition: comparison

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Blind Source Separation</th>
<th>Primary Ambient Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective</strong></td>
<td>To obtain useful information about the original sound scene from given mixtures, and facilitate natural sound rendering</td>
<td></td>
</tr>
<tr>
<td><strong>Basic model</strong></td>
<td>1. Multiple sources sum together</td>
<td>1. Dominant sources + Environmental signal</td>
</tr>
<tr>
<td></td>
<td>2. Sources are independent</td>
<td>2. Primary components are highly correlated;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Ambient components are uncorrelated</td>
</tr>
<tr>
<td><strong>Common characteristics</strong></td>
<td>1. Usually no prior information, only mixture signals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Perform extraction/separation based on various signal models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Require objective as well as subjective evaluation</td>
<td></td>
</tr>
<tr>
<td><strong>Typical applications</strong></td>
<td>Speech, music</td>
<td>Movie, gaming</td>
</tr>
<tr>
<td><strong>Limitations</strong></td>
<td>1. Small number of sources</td>
<td>1. Small number of sources</td>
</tr>
<tr>
<td></td>
<td>2. Sparseness/disjoint</td>
<td>2. Sparseness/disjoint</td>
</tr>
<tr>
<td></td>
<td>3. No/simple environment</td>
<td>3. Low ambient power</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Primary ambient uncorrelated</td>
</tr>
</tbody>
</table>

Sound scene decomposition: BSS

Objective:

to extract the K sources from M mixtures

Mixtures = function (gain, source, time difference, model error )

\[ x_m(n) = \sum_{k=1}^{K} g_{mk} s_k(n - \tau_{mk}) + e_m(n), \quad \forall m \in \{1, 2, \ldots, M\} \]
Sound scene decomposition: BSS

Objective:
to extract the K sources from M mixtures

<table>
<thead>
<tr>
<th>Case</th>
<th>Typical techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>M = K</td>
<td>ICA</td>
</tr>
<tr>
<td>M &gt; K</td>
<td>ICA with PCA, Least-squares</td>
</tr>
<tr>
<td>M = 2</td>
<td>Time-frequency masking</td>
</tr>
<tr>
<td>M = 1</td>
<td>NMF, CASA</td>
</tr>
</tbody>
</table>

ICA: Independent component analysis
PCA: Principal component analysis
NMF: Non-negative matrix factorization
CASA: Computational auditory scene analysis
One example using NMF

Sound scene decomposition: PAE

Objective:

to extract the primary and ambient components from M (M = 2, stereo) mixtures

Mixtures = primary component + ambient component

\[ x_m(n) = p_m(n) + a_m(n) \]
### Sound scene decomposition: PAE

**Objective:**

to extract the primary and ambient components from $M$ ($M = 2$, stereo) mixtures

---

<table>
<thead>
<tr>
<th>Case</th>
<th>Typical techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic model</td>
<td>Channel-wise: Time frequency masking</td>
</tr>
<tr>
<td></td>
<td>Combine channels: Linear estimation, Ambient spectrum estimation</td>
</tr>
<tr>
<td>Complex model</td>
<td>Time/phase shifting, Classification, Subband, Pairing up two channels, etc.</td>
</tr>
</tbody>
</table>
Definitions with Stereo Signal Model

**Signal** = **Primary** + **Ambient**

\[ x_0 = p_0 + a_0 \]
\[ x_1 = p_1 + a_1 \]

**Assumptions**

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary components highly correlated</td>
<td>( p_1 = k p_0 )</td>
</tr>
<tr>
<td>Ambient components uncorrelated</td>
<td>( a_0 \perp a_1 )</td>
</tr>
<tr>
<td>Primary ambient components uncorrelated</td>
<td>( p_i \perp a_j )</td>
</tr>
<tr>
<td>Ambient power balanced</td>
<td>( P_{a_0} = P_{a_i} )</td>
</tr>
</tbody>
</table>

PAE: time frequency masking

Mask can be constructed using

- Inter-channel coherence [Avendano and Jot, 2004]
- Pairwise correlation [Thompson et al., 2012]
- Equal level of ambience [Merimaa et al., 2007]
- Diffuseness [Pulkki, 2007]
**PAE: linear estimation**

**Objectives and relationships of four linear estimation based PAE approaches.**

- **Blue** solid lines represent the relationships in the **primary** component;
- **Green** dotted lines represent the relationships in the **ambient** component.
- **MLLS**: minimum leakage LS
- **MDLS**: minimum distortion LS

PAE: an example from least-squares
PAE: ambient spectrum estimation

**Signal model**

\[ X_0 = P_0 + A_0, \quad X_1 = P_1 + A_1 \]
\[ A_c = |A| e^{j\theta_c}, \quad \forall c \in \{0,1\}, \]

**Ambient Phase Estimation (APE)**

\[ A_c = (X_1 - kX_0) / (e^{j\theta_1} - ke^{j\theta_0}) \cap e^{j\theta_c}, \]
\[ P_c = X_c - A_c, \quad \forall c \in \{0,1\}. \]

**Find** \( \theta_0, \theta_1 \)

**Ambient Magnitude Estimation (AME)**

**Find** \( r = |A| \)

**Sparsity constraint**

\[ \text{APES} : \hat{\theta}_1^* = \arg \min_{\hat{\theta}_1} \| \hat{P}_1 \|_1, \quad \text{or} \quad \text{AMES} : \hat{r}^* = \arg \min \| \hat{P}_1 \|_1 \]

---

Approximate efficient solution APEX: \( \hat{\theta}_1^* \) = 
\[
\begin{cases} 
\angle X_1, & \forall k > 1 \\
\angle (X_1 - X_0), & \forall k = 1 
\end{cases}
\]

Performance of Ambient Spectrum Estimation approaches:
- Lower estimation error (ESR reduction: 3-6 dB average);
- Robust to variation on ambient magnitude difference (up to 10 dB);
- Validated in subjective listening tests.
For mixture signals with partially correlated primary components

- More accurate estimation of model parameter;
- Lower extraction error;
- Closer estimation of the spatial attributes;
- Increase of computational load.

PAE: multiple sources

Multi-shifting PAE with ICC based output weighting

Subband PAE with frequency bin partitioning

J. He, and W. S. Gan, “Multi-shift principal component analysis based primary component extraction for spatial audio reproduction,” in Proc. ICASSP, Brisbane, Australia, Apr. 2015, pp. 350-354.

PAE: from stereo to multiple channels

1. Using down-mix

2. Using pairing

3. Direct
Related: Parametric spatial sound processing

**Signal extraction:**
1. Single-channel wiener filters;
2. Multiple-channel filters: LCMV;

Individualization: the need

Variation of HRTFs (Idiosyncratic)

Overview of HRTF individualization techniques

To obtain individualized HRTF/perception

- Acoustical measurements
- Anthropometry
- Training/tuning
- Frontal projection
Individualization: acoustical measurements

Air Force Research Laboratory, US
Nagaoka University of Technology, Japan
ISVR, University of Southampton, UK
South China University of Technology, China
Tohoku University, Japan
## Summary of popular HRTF databases

<table>
<thead>
<tr>
<th>Databases</th>
<th>(Subjects, Directions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRCAM France <a href="http://recherche.ircam.fr/equipes/salles/listen">http://recherche.ircam.fr/equipes/salles/listen</a></td>
<td>(51, 187)</td>
</tr>
<tr>
<td>CIPICT, UC Davis <a href="http://interface.cipic.ucdavis.edu/sound/hrf.html">http://interface.cipic.ucdavis.edu/sound/hrf.html</a></td>
<td>(45,1250)</td>
</tr>
<tr>
<td>University of Maryland <a href="http://www.isr.umd.edu/Labs/NSL/Software.htm">http://www.isr.umd.edu/Labs/NSL/Software.htm</a></td>
<td>(7,2093)</td>
</tr>
<tr>
<td>Tohoku University, Japan <a href="http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf">http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf</a></td>
<td>(3,454)</td>
</tr>
<tr>
<td>Nagoya University, Japan <a href="http://www.sp.m.is.nagoya-u.ac.jp/HRTF/database.html">http://www.sp.m.is.nagoya-u.ac.jp/HRTF/database.html</a></td>
<td>(96,72)</td>
</tr>
<tr>
<td>TU Berlin (3m,2m,1m,0.5m) <a href="http://audio.qu.tu-berlin.de/?p=641">http://audio.qu.tu-berlin.de/?p=641</a></td>
<td>(KEMAR,360)</td>
</tr>
<tr>
<td>MIT Lab <a href="http://sound.media.mit.edu/resources/KEMAR.html">http://sound.media.mit.edu/resources/KEMAR.html</a></td>
<td>(KEMAR,710)</td>
</tr>
<tr>
<td>Oldenburg University (0.8m,3m) <a href="http://medi.uni-oldenburg.de/hrir/html/documentation.html">http://medi.uni-oldenburg.de/hrir/html/documentation.html</a></td>
<td>(HATS,365)</td>
</tr>
<tr>
<td>SDAC, KAIST (0.2,0.6,1m) <a href="http://sdac.kaist.ac.kr/research/index.php?mode=area&amp;act=DownHRTFDatabase">http://sdac.kaist.ac.kr/research/index.php?mode=area&amp;act=DownHRTFDatabase</a></td>
<td>(HATS, 100)</td>
</tr>
<tr>
<td>Nagoaka University (1.5 m) <a href="http://www.nagaoka-ct.ac.jp/ee/lab_syano/index_e.html">http://www.nagaoka-ct.ac.jp/ee/lab_syano/index_e.html</a></td>
<td>(SAMRAI dummy head + 3 subjects, 72 azim, 8 elev)</td>
</tr>
<tr>
<td>DSP Lab @ NTU (0.35,0.45,0.50,0.60,0.75,0.8,1,1.4m) <a href="http://eeeweba.ntu.edu.sg/DSPLab/dsplabhrtf/">http://eeeweba.ntu.edu.sg/DSPLab/dsplabhrtf/</a></td>
<td>(HATS + 3 subjects, 72)</td>
</tr>
</tbody>
</table>
Individualization: anthropometry

Obtain HRTFs numerically by:
1. Solving of acoustic equation
2. Numerical methods: Finite element method (FEM), Boundary element method (BEM)
Individualization: anthropometry
Individualization: training/tuning

Razer

University of Minho

DTS

Vivo (DTS)
Individualization: training/tuning

Source from https://www.youtube.com/watch?v=pOtN-KMWTeM
Individualization: frontal projection

- No additional measurements and listening experiments required
- Reduce front-back confusion by > 50%
- Zero user effort, plug and play (automatic during playback)

## Individualization: summary

<table>
<thead>
<tr>
<th>Obtain Individual Features</th>
<th>Techniques</th>
<th>Pros and Cons</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acoustical measurements</strong></td>
<td>Individual measurements, IRCAM France, CIPIC, Tohuko Univ., etc.</td>
<td>Ideal, accurate</td>
<td>Reference for individualization techniques</td>
</tr>
<tr>
<td><strong>Anthropometry</strong></td>
<td>3D mesh, pictures; Numerical Solutions: PCA, FEM, BEM, ANN</td>
<td>Need a large database; Requires high resolution imaging; Expensive</td>
<td>Uses the correlation between individual HRTF and anthropometric data</td>
</tr>
<tr>
<td><strong>Training/tuning</strong></td>
<td>PCA weight tuning, Tune magnitude Spectrum, Selection from HRTF database</td>
<td>Directly relates to perception; requires regular training;</td>
<td>Obtains the best HRTFs perceptually</td>
</tr>
<tr>
<td><strong>Frontal projection</strong></td>
<td>Frontal Projection Headphones</td>
<td>No additional measurement, Type-2 EQ</td>
<td>Automatic customization, reduced front-back confusions</td>
</tr>
<tr>
<td><strong>Non-individualized HRTF</strong></td>
<td>Generalized HRTF</td>
<td>Easy to implement, Poor localization</td>
<td>Not an individualization technique</td>
</tr>
</tbody>
</table>
Equalization

Headphone is not acoustically transparent:

- Headphone colors the input sound spectrum;
- Affects the free-field characteristics of the sound pressure at the ear

Breakdown of headphone transfer function (HPTF)
Decoupled equalization for binaural/stereo

Aim: Emulate the reproduction in a reference field

- **Free-field:**
  - Target: free-field front loudspeaker response

- **Diffuse-field and other reference curves:**
  - Target: response of diffuse-field, or a reference room
  - Lesser inter-individual variability

Source from http://seanolive.blogspot.sg/2014/01/the-perception-and-measurement-of.html
Non-decoupled equalization for binaural

**Aim: Spectrum at eardrum is the individual HRTF features**

- For conventional headphone
- For front projection headphone
Conventional equalization (Type 1 EQ)

Headphone is not acoustically transparent, therefore the effect of the headphone must be removed.

Equalization process: Removing the headphone transfer function

\[ Y(\omega) = S(\omega) \cdot HRTF(\omega) \cdot \frac{1}{HPTF(\omega)} \cdot HPTF(\omega) \]

Where,
- \( Y(\omega) \) = Equalized Binaural Signal
- \( S(\omega) \) = Source Signal Spectrum
- \( HRTF(\omega) \) = Head Related Transfer Function (Left/Right)
- \( HPTF(\omega) \) = Headphone Transfer Function (Left/Right)

And, \( \frac{1}{HPTF(\omega)} \) = Equalization Filter

Dependent on individual pinna feature
Type 2 EQ (for frontal emitters)

- Reflections/diffractions created by the interactions with the pinna due to the frontal projection are important and should be retained.
- Does not include headphone-ear coupling.
- Equalizing to the free field response of the headphone with the ear-cup.

Measurement of Free Field Response

Type 2 EQ (for frontal emitters)

- Reflections/diffractions created by the interactions with the pinna due to the frontal projection are important and should be retained.
- Does not include headphone-ear coupling.
- Equalizing to the free field response of the headphone with the ear-cup.

\[
Y(\omega) = S(\omega) \cdot HRTF(\omega) \cdot \frac{1}{FFR(\omega)} \cdot HPTF(\omega)
\]

\[
HPTF(\omega) = FFR(\omega) \cdot PC(\omega)
\]

Where,
- \(Y(\omega)\) = Equalized Binaural Signal
- \(S(\omega)\) = Source Signal Spectrum
- \(HRTF(\omega)\) = Head Related Transfer Function (Left/Right)
- \(HPTF(\omega)\) = Headphone Transfer Function (Left/Right)
- \(FFR(\omega)\) = Free Field Response of the Frontal Emitter
- \(PC(\omega)\) = Personalized Pinna Cues generated by frontal projection
Type-2 EQ works best for the frontal emitter playback

ANOVA results show
- Type of emitter has a significant effect on rate of reversal
- Type-2 EQ has a significant effect in reducing F/B reversal
Integration

Sound sources → Sound Mixing → Sound environment

Sound sources → Head movement → Head tracking

Individual parameters → Individualization

Decomposition using BSS/PAE

Source rendering

Environment rendering

Virtualization

Equalization

Headphone playback

Rendering of natural sound

3D Audio Headphone: an example

Key features

- Patented structure with strategic-positioned emitters;
- Individualization via frontal projection; no measurements or training required;
- Recreate an immersive perception of sound objects with surrounding ambience;
- Compatible with all existing sound formats.

Subjective evaluation

- Conventional stereo system: stereo headphone
- Natural sound rendering system: 3D headphone
- **Stimuli**: binaural, movie and gaming tracks;
- **4 measures**: Sense of direction, externalization, ambience, and timbral quality;
- 18 subjects, score of 0-100, and overall preference.

![Subjective evaluation graph](image)

\[ p \text{-value} \ll 0.01 \]
Advent of high speed, low power, and low cost embedded processors fueling a strong growth in portable and wearable applications.

Many opportunities for new innovations in spatial audio rendering for assistive technologies; being-there communications; immersive AR/VR gaming; and interactive entertainments.

Seamless integration of real and virtual sound objects to achieve natural listening.

“... future headphones are becoming more content-aware, location-aware, listener-aware, and hence become more intelligent and assistive.”

IEEE Signal Processing Magazine, March ‘15
Key References on Natural Sound Rendering

Key References on Natural Sound Rendering

Key References on Natural Sound Rendering


Demo of 3D audio headphones
Module III
ANC Headphones
Active Noise Control

- Active noise control (ANC) uses additional secondary sources to produce anti-noise that cancel the undesired noise.

- Principle:
  1. Mathematics: $x + (-x) = 0$
  2. Physics (Superposition): Anti-noise of equal amplitude and opposite phase is combined with the primary noise, resulting in the cancellation of both noises

![Figure 1. Physical Concept of Active Noise Cancellation](image)

Leug, 1933
Sound Field Interaction of 2 Point Sources

Destructive

Constructive

Normalised sound pressure

-1  -0.8  -0.6  -0.4  -0.2  0  0.2  0.4  0.6  0.8  1
金庸小说“倚天屠龙记”

张无忌一惊，不及趋避，足尖使劲，拔身急起，斜飞而上，只听得飕飕两声轻响，跟着“啊”的一下长声呼叫。他在半空中转过头来，只见何太冲和班淑娴的两柄长剑并排插在鲜于通胸口。原来何氏夫妇纵横半生，却当众败在一个后辈手底，无论如何咽不下这口气去，两人拾起长剑，眼见张无忌正俯身在点鲜于通的穴道，对望一眼，心意相通，点了点头，突然使出一招“无声无色”，同时疾向他背后刺去。这招“无声无色”是昆仑派剑学中的绝招，必须两人同使，两人功力相若，内劲相同，当剑招之出，劲力恰恰相反，于是两柄长剑上所生的震荡之力、破空之声，一齐相互抵消。这路剑招本是用于夜战，黑暗中令对方难以听声辨器，事先绝无半分朕兆，白刃已然加身，但若白日用之背后偷袭，也令人无法防备。
ANC is 80+ old!

Hype Cycle

1933 Conover

1960 Wagner

1990 Wider

ANC is 80+ old!

Start of ANC; lots of activities & patents

Dormant for 30 years

Developing ANC theory / New Apps

*: communication with Prof. Elliott

W. S. Gan, J. He (NTU) Natural sound rendering for Headphones

16th Dec 2015
Applications

Active Noise Control Headset

Image of Active Noise Control Headset

Diagram of noise-canceling headphones with labels:
- Electronics
- Speaker
- Microphone

Diagram text:
- Sound waves created by headphone speaker
- Noise created by external source

Caption:
©2007 HowStuffWorks
©2007 CNET Networks, Inc.
Commercial ANC Headphones

- **Bose** Quiet Comfort 15 Acoustic Noise Canceling Headphones (US$365)
- **Sony** MDR-NC100D Digital Noise Canceling Earbuds (US$131)
- **Creative** HN-900 Noise Canceling Headphones (US$159)
- **Sennheiser** MM 550 Bluetooth Wireless Headphones (US$475)
- **Digital Silence** Ambient Noise Canceling Earbud (US$62)
- **Blackbox** i10 Active Noise Rejection Earbud-powered by iPod battery (US$127)
ANC Headphones

- With many promising applications, we are witnessing the Golden Age of ANC
  - Low cost and accurate sensors and actuators
  - High speed, low cost and low power consumption embedded processors

- For successful consumer applications:
  - ANC needs to be combined or integrate with other functions
  - Allow sharing hardware resources (e.g. amplifier, loudspeakers, microphones etc.)
  - Reuse software code or tap on existing audio functions in digital implementation.
  - Many opportunities to innovate

- Outline some of the work carried out by the research community to integrate ANC with other functions
1st Paper to look into Integration of ANC with Audio System (from Kuo et al.)

- Appears in the 1993 IEEE Transaction on Consumer Electronics
- Won the 1st place in the IEEE Consumer Electronics Society Chester Sall Awards in 1993.

Using multiple adaptive filters to integrate:
(i) Active noise control
(ii) Acoustic echo cancellation
(iii) Adaptive noise cancellation
(iv) Adaptive musical interference suppression

Figure 1. Integrated Hands-Free Cellular Phone, Active Noise Control and Audio System

This paper lays the foundation of many integrated ANC + Audio systems/papers

(picture from)

Basic Definitions in ANC: A Simple Duct Application

FXLMS Algorithm: \[ \mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{x}'(n)e(n) \]
**Offline Secondary Path Modeling**

Experimental setup for off-line modeling of secondary-path $S(z)$.

**Note:**
- Offline secondary path modeling performed before the actual operation of ANC.
- The estimated model is used in actual operation.
- $S_{est}(z)$ can be fixed or adaptively update.
- Within the limit of slow adaptation, the algorithm will converge with nearly of 90° phase error between $S_{est}(z)$ and $S(z)$.

Picture courtesy of Sen M Kuo
ANC System Layout with On-line Modeling

Picture courtesy of Colin Hansen
## Configurations of Adaptive ANC System

<table>
<thead>
<tr>
<th>Feedforward ANC</th>
<th>Feedback ANC</th>
</tr>
</thead>
</table>

### Feedforward ANC
- Broadband and narrowband FFANC
- 2 sensors and 1 actuator (1-dim case)

\[
W^o(z) = \frac{P(z)}{S(z)}
\]

### Feedback ANC
- Also known as internal model control (IMC) feedback ANC.
- 1 sensor and 1 actuator (1-dim case)
- Estimate the primary noise based on error signal and adaptive filter output.
- Adaptive filter loop & Synthesis loop
- Generally good in cancelling out predictable tonal noise.
Hybrid FF and FB ANC

- Dual role in canceling primary noise from reference sensor & residual noise pickup by error sensor.
- Good performance for both narrow and broadband noise and offer flexibility in ANC design.
- May come with extra computational cost.

More info in Prof. Kuo’s classic textbook on Active Noise Control
Performance Hierarchy of ANC

Figure 2.5 Performance hierarchy of an active noise control system.

Picture extracted from Hansen’s book
A. Wearable Devices
   - Headsets
   - Motorcycle Helmet
   - Hearing Aids

B. Other Applications:
   - Automotive
   - Snoring cancellation
   - Incubator
   - MRI

Integrated ANC with other functions

Use carefully fitted, and if possible, noise-cancelling earphones/headphones. If suited to the individual user, earphones and headphones allow music to be heard clearly at lower levels of volume. Noise-cancelling earphones and headphones cut down background noise, so that you can hear sounds at lower volumes than otherwise needed.

International Ear Car Day
on 3rd March '15

Make Listening Safe. Once you lose your hearing, it won't come back!

World Health Organization
Applying ANC to Headphones/Earphones

- Music playback or received speech signal
- Communication
- Perceptual Compensation
- Noise Reduction
- Augmented Reality Headsets
Earphones: Adaptive vs Static ANC

In general, the adaptive filter always achieves about 20 dB attenuation in the frequency range between 100 Hz and 1000 Hz; but static ANC results in only 10 dB reduction.

### Commercial ANC Headphones with digital adaptation

- Some of the commercial ANC headphones are still based on analog controller
- Recent years have seen more digital adaptive ANC in commercial products and patents:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Model Number/Name</th>
<th>Weblink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sony</td>
<td>MDR-10R Noise Cancelling Headphones</td>
<td><a href="http://www.sony.com.sg/product/mdr-10rnc">http://www.sony.com.sg/product/mdr-10rnc</a></td>
</tr>
<tr>
<td>Samsung</td>
<td>Samsung Level Over</td>
<td><a href="http://www.amazon.com/Samsung-Cancelling-Wireless-Headphones-Smartphones/dp/B00KGGK738">http://www.amazon.com/Samsung-Cancelling-Wireless-Headphones-Smartphones/dp/B00KGGK738</a></td>
</tr>
<tr>
<td>Beats</td>
<td>Beats Noise Cancelling Headphones</td>
<td><a href="http://sg.beatsbydre.com/headphones/">http://sg.beatsbydre.com/headphones/</a></td>
</tr>
</tbody>
</table>
ANC Headset with audio playback

**Fig. 1 Integrated audio and ANC headsets**

\[
E'(z) = D(z) - A(z)S(z) - Y(z)S(z) + \hat{S}(z)A(z).
\]

Assuming \( \hat{S}(z) = S(z) \)

\[
E'(z) = D(z) - Y(z)S(z).
\]

Possible error microphone locations corresponding to ear surface

Magnitude responses of secondary path $S(z)$ at microphone locations #3 and #8

Frequency responses of secondary-path $S(z)$ at microphone locations #1 to #8

Optimum location #8 for the error microphone
Results

Fig. 4. Noise spectrum for the error signal with (dotted line) and without (solid line) using IFBANC under repeating siren disturbance

Fig. 5. Noise spectral for the error signals with (dotted line) and without (solid line) using IFBANC under an engine disturbance.
Advantage of the integrated approach

- Good estimation of the true residual noise $e'(n)$ without interfering with the audio signal $a(n)$;
- **Large step size** can be used in adapting the cancellation filter $W(z)$;
- The adaptive feedback ANC technique provides a more accurate noise cancellation since the microphone is placed inside the ear-cup of the headset;
- The system uses **single microphone per ear cup**, thus produces a compact, lower power consumption and a cheaper solution;
- The audio signal can be neatly used to drive both **on-line and offline modelling** of the secondary path transfer function.
Include communication features in integrated ANC + Audio playback.

Our paper also compare the noise cancellation performance with commercial-off-the-shelf ANC headphones.

Our adaptive ANC integrated headphones outperformed the commercial one by 15-20 dB at low frequency below 183 Hz.

More objective and subjective evaluation can be conducted to test out other parameters of the ANC headphones.

Figure 8. Net noise cancellation comparison for the proposed AFANC and a commercial noise canceling headset for engine noise at (a) 2,200 rpm and (b) 3,700 rpm

ANC Headset with audio playback and speech communication

To remove near-end noise before transmission. Use an adaptive noise canceller. **Note:** a purely electrical cancellation; not acoustic.

A correlated noise input used to train Noise canceller

Acoustic path

Figure 13. The top and bottom plots show the original signal corrupted by engine noise and enhanced near-end speech respectively.
Extension to a Hybrid ANC system

- Combined both feedforward and feedback ANC techniques to further improved on its broadband noise cancellation performance.

- Also perform better than the feedback ANC for narrow frequency separation engine noise. (~6 to 10 dB better)

---

Yong-Kim Chong, Liang Wang, See-Chiat Ting and Woon-Seng Gan, "INTEGRATED HEADSETS USING THE ADAPTIVE HYBRID ACTIVE NOISE CONTROL SYSTEM", ICICS 2005
Broadband Feedback ANC

Fig. 2. Model for a classical feedback ANC system.

Fig. 3. Model for an adaptive feedback ANC system involving the FxLMS approach.

Combined Analog-Digital

In general, attenuation of noise in lower frequency areas as well as periodic noise components in all frequency areas

Patent application on a Feedforward ANC system for use in a portable audio device using an adaptive digital filter and a reference/error microphones.

Novelty is adding a **non-adaptive pre-shaping filter (minimum phase)** and has > 2dB more gain over low frequency.

Integrating speech/audio signal into the FF ANC system.
To control the ambient acoustic noise outside the device that may be heard by a user of the device.

Depending on the signal strength of the sound emitted from the earpiece speaker and the ambient acoustic noise to **activate/deactivate ANC**.

- **Objective**: Restore the original audio loudness and timbre
- Improve perceived quality of sound reproduction of an audio
- Include a perceptual noise compensator to generate noise-compensated signal based on **audio signal** and **residual noise characteristic**.
Conclusions on ANC Headphones

- Fundamental theories, algorithms, and experiments of ANC have been well established over the last few decades. **Focus now on how we can deploy and integrate ANC in existing and other new applications.**
- Witnessing an increased activities from the industry applying ANC techniques into their products (especially in **ANC headsets**, **hearing aids** and **automotive** applications)
- Issues with maintainability and costs become lesser concern with **better, reliable, small form factor, and low cost** sensors ➔ suited for wearable consumer devices.
Key References on ANC Headphones

Module IV
Natural Augmented Reality Headsets
Natural Listening for VR and AR

Natural listening via headset

Virtual Reality (VR)

Augmented Reality (AR)
Natural Listening in Virtual Reality

- **Real**
  - Capture
  - Transmission
  - Reproduction

- **Virtual**
  - Capture
  - Transmission
  - Reproduction

Sony Project Morpheus
Oculus rift
Natural Listening in Augmented reality

Real

Virtual

Augmented Reality (AR)

Capture

Microsoft Hololens
Augmented Reality

Augmented reality is changing the way we live in real world

Wearable VR/AR devices:

- Google-glass
- Oculus rift
- Microsoft Hololens

VR/AR applications:

- Navigation
- VR and AR world
- Gaming
Related Works

1. **Augmented reality audio (ARA) headset** using in-ear headphones and external binaural microphones to assist the listener with pseudo-acoustic scenes[1].

   **Problem addressed:** Blockage of natural sounds coming from outside. Using binaural microphones to capture, process and playback so as to make ARA headset acoustically transparent.

2. Surround sound reproduction over headphones with **binaural microphones positioned inside** ear cup near ear opening [2].

   **Problem addressed:** Large localization errors due to non-individualized equalization of headphones. Using ANC technique to calibrate the system for every individual to achieve sound reproduction same as a multichannel setup.


ARA Headset - Overview

- Use of closed in-earphones to capture the external sound, mix with the virtual sounds and playback
- Basic idea is to relay the external sounds unaltered with minimum latency (<1ms)

Left and middle: ARA headset (Philips SHN2500)
Right: Prototype ARA mixer [2]

ARA Headset (challenges)

- Blockage of external sounds by ARA headset

- Closed in-ear headphones with **good fittings** is necessary for good performance of ARA headset

- Loose fittings can dramatically change the attenuation

---

• Closed in-ear phones modify the ear canal resonance (pressure chamber principle)

• Difficult to predict the headphone response if loosely fitted

ARA Headset (Design of ARA mixer)

- ARA headset equalization

ARA Headset

- Generic ARA equalization based on 4 individual measurements

![Graph showing frequency response](image)

Masking in ARA Headset

3D headphone sound reproduction using ANC

- Emulating a 5 channel sound reproduction setup through headset with two microphones

Offline calibration process using adaptive process FxLMS

Acoustic-Hear-Through AR (open-air headphones)

Real

Listener

Virtual

Capture

Augmented Reality (AR)

Signal Processing Techniques

Open-air / Open-ear headsets
NAR headset potential applications:

**Augmented Audio tour**

*Welcome to virtual audio tour of world war memorial*

*Hey, buddy!!!!*
NAR headset potential applications:

**Augmented Assistive listening for Visually Impaired**

Assistive Listening is also needed for the normal person!
NAR headset potential applications:

Augmented Multi-party Conferencing
Using HRTF in Gaming to Create Audio Depth

Enemy 1 & 2 sound closer compared to enemy 3 & 4

Image from: CRYSIS 2 game
Proposed Natural Augmented Reality (NAR) Headset

Proposed a headset structure with two pairs (int/ext) binaural microphones attached to the earcups.

- NAR headset with **2 pairs** of binaural microphones
- Adaptive equalization methods to compensate for sonic difference between real and headphone playback virtual signals
- Natural mixing of real and virtual sources

**NAR headset prototype**

- **Internal microphone** used as *error microphone* to adapt the virtual sound at ear canal to natural sound
- **External microphone** used as *reference microphone* to capture the sound event from the surroundings (real sounds)
NAR Headset Overview Block Diagram

- Head movement
- Virtual source position
- Virtual source
- Environment Parameters (from $m_{ext}$)

Head Tracking

Individualization (HRTF database)

Adaptive Headphone Equalizer

Environment adaptation

Adaptive estimation

Adaptive Equalization for virtual sources

Natural augmented reality

Open earcup

$m_{int}$ to $m_{ext}$

Real sources
Headphones Isolation Characteristics

• NAR headsets should not block the **direct sounds** coming from physical sound sources

• We analyze the headphone isolation characteristics of different type of headphones

The performance of the measured HRTFs can be evaluated by a spectral distortion (SD) score in dB given by

\[
SD = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left( 20 \log \frac{|H(f_k)|}{|\hat{H}(f_k)|} \right)^2}
\]

where

\( |H(f_k)| \) is the magnitude response of reference HRTF i.e., measured without headphones from loudspeakers

\( |\hat{H}(f_k)| \) is the magnitude response of HRTFs measured with the headphones from loudspeakers

*Smaller Sd score indicates a closer magnitude response to the reference*
Headphones Isolation Characteristics

- Spectral distortion scores for 4 different types of headphones

Contralateral (left) and ipsilateral (right) ears

0 20 40 60 80 100 120 140 160 180

[0.1-1.5 kHz]

[1.5-7.0 kHz]

[7.0-16.0 kHz]
NAR headset: **(A) Normal Mode**

**Normal mode** – Only external sound source present (No additional processing in NAR headset)

We choose **open headphones** so as to allow direct sound without much attenuation

![Diagram](image)

\[
\begin{align*}
  r(n) & \quad h_{ext}(n) \quad h_{he}(n) \\
  & \quad r_{ext}(n) \quad r_{int}(n) \\
  r(n) & \quad h_{int}(n) \quad r_{int}(n)
\end{align*}
\]

\[
h_{int}(n) = h_{ext}(n) \ast h_{he}(n)
\]

\[
H_{he}(z) = \frac{H_{int}(z)}{H_{ext}(z)}
\]

Account for the ear cup plus pinnae

Natural sounds pass through the open ear cup of the AR headset and reach the ear opening.
NAR headset: **(B) Binaural Synthesis Mode**

**Binaural Synthesis using headphones playback**

- Virtual monaural signal convolves with HRTF of the virtual object.
- Individual HPTF effect must be removed using compensation filter.
- Compensation can be done by:
  - Direct inversion of HPTF, which may not be available [1]
  - Using an adaptive algorithm like FxLMS, which is more effective [2]

Augmented reality mode – virtual sound reproduction in the presence of external signals

Aim: To reproduce virtual sources as if they sound similar to physical sources, without being affected by external sounds

**Virtual source**
\[ x(n) \]

**Real source**
\[ r(n) \]

**Personalized HRTF database**

**Open earcup**

**HRTF**

**Desired signal path**
\[ x(n) \rightarrow h_{\text{int}}(n) \rightarrow y_{\text{int}}(n) \rightarrow d(n) \]

**Actual signal path**
\[ x(n) \rightarrow w(n) \rightarrow h_{\text{hp}}(n) \rightarrow l(n) \rightarrow y_{\text{ext}}(n) \rightarrow y_{\text{int}}(n) \]

\[ x_{\text{int}}(n) = w(n) \ast h_{\text{hp}}(n) \ast x(n) \]

\[ W(z) = \frac{H_{\text{int}}(z)}{H_{\text{hp}}(z)} \]

**Desired signal path**

**Actual signal path**

**Open earcup**

**Personalized HRTF database**

**leakage**

**signal**

**x(n)**

**Real source**

**h_{\text{ext}}(n)**

**m_{\text{ext}}**

**y_{\text{ext}}(n)**

**y_{\text{int}}(n)**

**m_{\text{int}}**

**l(n): Leakage from headphone to external microphone, m_{\text{ext}}**

**W. S. Gan, J. He (NTU) Natural sound rendering for Headphones 16th Dec 2015 IV.26 / 47**
NAR headset: \textit{int} and \textit{ext} Transfer Functions

- $H_{\text{int}}(z)$ is an approximate HRTF with additional headphone effects
- $H_{\text{ext}}(z)$ contains all individual related characteristics and environment minus the pinnae specific notch and headphone shell reflections
NAR headset: Mixing virtual augmented signal with real ext source

Augmented reality mode – virtual sound reproduction in the presence of external sounds: Hybrid Adaptive Equalizer (Assuming negligible leakage signal power, $l(n) = 0$)

**Hybrid adaptive equalizer:** simple combination of conventional and modified FxNLMS

Augmented reality mode – virtual sound reproduction in the presence of external sounds: Hybrid Adaptive Equalizer (Assuming negligible leakage signal power, \( l(n) = 0 \))

**NAR headset:** Mixing virtual augmented signal with real ext source

\[
\begin{align*}
\sum w_1(n) &\quad \text{Adaptive filter corresponding to } \text{conventional FXNLMS} \\
\quad &\quad \text{• Slower convergence rate due to presence of secondary path transfer function} \\
\sum w_2(n) &\quad \text{Adaptive filter corresponding to } \text{Modified FxNLMS} \\
\quad &\quad \text{• Faster convergence rate by introducing spatial filter, } h^v_{ext}(n) \text{ in the secondary path but slightly higher steady state error (shorter filter taps)}
\end{align*}
\]
Augmented reality mode – virtual sound reproduction in the presence of external sounds: Hybrid Adaptive Equalizer (Assuming negligible leakage signal power, \( l(n) = 0 \))

Hybrid adaptive filters:

\[
\begin{align*}
    w(n) &= w_1(n) + h^{v}_{ext}(n) * w_2(n) \\
    W(z) &= W_1(z) + H^{v}_{ext}(z) W_2(z) \\
    \text{Spatial information retained in } h^{v}_{ext}(n) \text{ results in faster convergences and smaller MSE using hybrid adaptive filters.}
\end{align*}
\]
**NAR headset:** Mixing virtual augmented signal with real ext source

**Augmented reality mode** – virtual sound reproduction in the presence of external sounds

**Weight update equations:**

\[
\begin{align*}
\mathbf{w}_1(n+1) &= \mathbf{w}_1(n) + \mu \frac{\mathbf{x}'(n)}{\|\mathbf{x}'(n)\|^2} e'(n) \\
\mathbf{w}_2(n+1) &= \mathbf{w}_2(n) + \mu \frac{\mathbf{x}_{\text{ext}}'(n)}{\|\mathbf{x}_{\text{ext}}'(n)\|^2} e'(n) \\
\mathbf{w}_r(n+1) &= \mathbf{w}_r(n) - \mu \mathbf{r}_{\text{ext}}(n) e'(n)
\end{align*}
\]

**Optimal solution for hybrid adaptive filter:**

**Optimal solution for adaptive estimation filter:**

\[
W^o(z) = \alpha W^o_1(z) + \beta H^o_{\text{ext}}(z)W^o_2(z),
\]

where,

\[
\begin{align*}
W^o_1(z) &= \frac{R_{\text{int}}(z)}{H_{\text{hp}}(z)} \\
W^o_2(z) &= \frac{H_{\text{int}}(z)}{H_{\text{ext}}(z)}
\end{align*}
\]

\[
W^o_r(z) = \frac{R_{\text{int}}(z)}{H_{\text{hp}}(z)} = H_{\text{he}}(z)
\]

\[
\forall \alpha + \beta = 1 \quad 0 \leq \alpha, \beta \leq 1
\]

The optimal solution for adaptive estimation filter is the **headphone effect transfer function** between the two microphone positions

**Optimal solution for hybrid adaptive filter can be expressed as linear combination of optimal solutions for two FxNLMS adaptive filters**
NAR headset: Adaptive estimation of external source signal

Augmented superimposed signal:

\[ y_{int}(n) = x_{int}(n) + r_{int}(n), \]

where,

\[ x_{int}(n) = h_{hp}(n) \ast u(n) \]

Error signal:

\[ e'(n) = \{d(n) + \hat{r}_{int}(n)\} - y_{int}(n) \]
\[ = \{d(n) - x_{int}(n)\} + \{- (\hat{r}_{int}(n) - \hat{r}_{int}(n))\} \]
\[ = e_v(n) + e_r(n) \]
Results: Augmented Reality mode

Hybrid FxNLMS performance comparison with or without external sound source present but no adaptive estimation process i.e., $w_r(n) = 0$

Clearly, presence of external sound sources affect the adaptive process and results in larger steady state error
Results: Augmented Reality mode

Hybrid FxNLMS performance with adaptive estimation i.e., adaptive $w_r(n)$ Vs off-line estimation i.e, fixed $w_r(n)$ as average $h_{he}(n)$ filter to estimate the real signals

Hybrid FxNLMS with adaptive estimation works equally well even in the presence of real sounds reproducing the virtual sources as close as possible to real sources.
Results: Augmented Reality mode

Spectral distortion score comparison for Hybrid Adaptive Equalizer (100-16000 Hz)

Clearly, Hybrid FxNLMS is optimal for all the azimuths as compared to conventional and modified FxNLMS.
Listening Test

Three main objectives:

- **Naturalness**: Does virtual playback sound source feels natural to you?
- **Sound similarity**: Does virtual sounds similar to real speaker sound?
- **Source position similarity**: Does virtual sound source coming from position in 3D space as real source?
Experiment setup

- 7 loudspeakers: 5 in horizontal plane and 2 in median plane

Listening Test Set up (.FileReader 'Elevated speaker; FileReader 'Azimuth speaker)
Listening test overview

Conducted in two phases:

1. Listeners’ BRIRs (automated) measurements
   - Head Tracker to ensure “look ahead” and still head position

2. Listening test based on individual’s BRIRs measurements
   - Fixed listening position
   - No head movements allowed (future: headtracker)
Three different listening sets are carried out as follows:

- **SET 1**: Perceptual similarity test of a male speech signal (5 sec)
- **SET 2**: Perceptual similarity test for playback of two male speech signals from two different directions (3.5 sec each)
- **SET 3**: Perceptual similarity test for superposition of a speech signal with ambient sound (6 sec)
SET 1:
- Total seven speaker positions
- Two hidden anchors (coming from the same physical speakers)

SET 2:
- Total 6 speaker pairs
- Two hidden anchors

SET 3:
- Total 4 pairs
- Speech from F and ambience from surroundings (2ch and 4ch)
- Two methods: Adaptive eq. with and without adaptive estimation
- One hidden anchor
Listening Test Results: Source Confusion

Source confusion measure in percentage as responses for subjects who marked either virtual sound as real or both sounds as real.

- Clearly, more than 75% source confusion for the case where subjects marked both sounds as real implying virtual source perceived natural by most of the listener.
- Source confusion further increases when more source present.
Listening Test Results: Sound Similarity

Sound similarity: Subjective score of 0 to 10

Mean subjective score with 99% confidence interval for sound similarity

<table>
<thead>
<tr>
<th></th>
<th>SET 1</th>
<th>SET 2</th>
<th>SET 3</th>
</tr>
</thead>
</table>

Speaker Position
F A B C D E1 E2

Speaker Pairs
rA-vB vA-rB rC-vD vC-rD rE1-vE2 vE1-rE2

Case II (A-B)
Case III (A-B-C-D)

r: real speaker
v: virtual speaker

Barely different
Same

10
8
6
4
2
0
Listening Test Results: Source position Similarity

Source position similarity: Subjective score of 0 to 10

Mean subjective score with 99% confidence interval for Source position similarity

<table>
<thead>
<tr>
<th></th>
<th>SET 1</th>
<th>SET 2</th>
<th>SET 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>8.26 (7.33 - 9.19)</td>
<td>9.09 (8.32 - 9.87)</td>
<td>9.28 (8.77 - 9.78)</td>
</tr>
<tr>
<td>Very close</td>
<td>7.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very different</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

r: real speaker          v: virtual speaker
### NAR Headset Extension

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Extensions based on current limitations of the NAR headset</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>NAR headset should adapt to the change in external environment ($H_{int}$ and $H_{ext}$)</td>
</tr>
<tr>
<td>E2</td>
<td>Individualized HRIR acquisition using NAR headset</td>
</tr>
<tr>
<td>E3</td>
<td>Adaptive equalization for any type of virtual signals (speech, music, etc.)</td>
</tr>
<tr>
<td>E4</td>
<td>Detection and Fast Estimation of Headphone Transfer Function in NAR Headset</td>
</tr>
<tr>
<td>E5</td>
<td>Adaptive Equalization for Non-stationary Virtual signals with Adaptive Estimation of External signals</td>
</tr>
<tr>
<td>E6</td>
<td>ANC mode for the NAR headset in presence of unwanted/unpleasant ambient sound</td>
</tr>
<tr>
<td>E7</td>
<td>NAR headset with head tracking to include dynamic motion cues</td>
</tr>
</tbody>
</table>
Conclusions on NAR Headsets

- **Adaptive filtering techniques** presented to enable natural listening in augmented reality environments using **NAR headset** with two pairs of binaural microphones.

- **Hybrid adaptive equalizer** based on FxNLMS for virtual sound reproduction to equalize the NAR headset for an individual.

- **Faster convergence and smaller steady state residual error.**

- Proposed hybrid adaptive equalizer able to attenuate the error signal by more than **25 dB**, and implying virtual signals are perceptually reproduced very close to the real signals.

- Listening test result shows **very high source confusion %** for virtual sources.

- **Very high sound similarity and source position similarity** were observed between real and augmented sounds.
Key References on NAR Headsets

Key References on NAR Headsets


Module V
Assisted Listening in Hearing Aids
Outline for Assisted Listening in Hearing Aids

1. Hearing loss and hearing aids

2. Noise reduction and speech enhancement
   • Acquisition
   • Localization
   • Enhancement
   • Presentation

3. Integration of ANC in hearing aids
Global Prevalence of Hearing Loss

• “Over 5% of the world’s population – 360 million people – has disabling hearing loss.”
• “Approximately 1/3 of persons over 65 years are affected by disabling hearing loss.”

- WHO, Mar 2015

Figure from http://www.who.int/pbd/deafness/WHO_GE_HL.pdf?ua=1
Hearing aids: for everyone to hear well
Hearing loss

https://www.youtube.com/watch?v=2rhRo73F324
What it sounds like with hearing loss

https://www.youtube.com/watch?v=Bcz7AeBMLSc
Inside a hearing aids

Digital Signal Processing (DSP) system in hybrid packaging

Includes:

- DSP Chip
- Software (speech processing algorithms)
- And peripherals such as radio IC, EEPROM, Power Management IC for rechargeable batteries,…

Source from:
http://www.bdti.com/InsideDSP/2014/10/16/ONSem
How hearing aids work

1. Sound goes in the Microphone.
2. Sound gets amplified.
3. Sound comes out the Speaker into your Ear.

Introduction to Hearing Aid Features by Steve Barber, HLAA-Wake Chapter
Style

**DC** (deep-canal)
- Mild to moderate hearing losses
- Fits deep inside the ear canal, making it **invisible**
- Less occlusion
- Not suitable for people with narrow ear canals
- Size prevents the use of directional microphones

**CIC** (completely-in-canal)
- Mild to moderate hearing losses
- Very small case
- Fits inside the ear canal, making it practically invisible
- Directional microphones are possible with this model

**ITC** (in-the-canal)
- Mild to moderately-severe hearing losses
- Small, one piece case
- Fits inside the ear canal
- Directional microphones are possible with this model

**RIE** (receiver in ear)
- Mild to moderately-severe hearing losses
- Ear canal open for a natural sound quality
- Smallest external hearing aid, as the receiver is located in the end of the tube inside the ear
- Very small case that sits behind the ear, making it practically invisible

**Open** (open ear)
- Mild to moderately-severe hearing losses
- Ear canal is open for a natural sound quality
- Very small case that sits behind the ear, making it practically invisible
- Many colour options

**BTE** (behind-the-ear)
- Mild to severe losses
- Fully featured hearing aids
- Larger case can be easier for wearers with dexterity considerations
- Case contains all features and sits behind the ear
- Many colour options

**Power** (high powered)
- Profound hearing losses
- More powerful solutions that provide the greatest levels of amplification
- Larger case worn behind the ear
- Many colour options
Technology

- **Analog**: Settings and Sound are both processed via analog technology.
- **Digital Programmable**: Settings are processed digitally, Sound is processed via analog technology.
- **Full Digital**: Both Settings and Sound are processed digitally.
Possible features

- Volume Control
- Telecoil
- Multiple Microphone Directionality
- Compression
- Clipping
- Direct Audio Input
- FM
- Bluetooth
- Programmability
- Speech Enhancement/Noise Reduction
- Frequency Shifting
- Earmold/Vent
- Remote Control
## Hearing aid features

<table>
<thead>
<tr>
<th>Feature</th>
<th>What does it do?</th>
<th>How does it help?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channels</td>
<td>Separates the sound signal into discrete sections for processing.</td>
<td>Allows the hearing aid prescription to be customized across all pitches. The more channels in the hearing aid, the greater the ability to customize the frequency response.</td>
</tr>
<tr>
<td>Directional microphone systems</td>
<td>Gives preference to sounds coming from the front of the wearer and reduces sounds coming from other directions.</td>
<td>Improves speech understanding in background noise. Satisfaction is higher for hearing aids with directional microphone systems than for hearing aids without them.</td>
</tr>
<tr>
<td>Digital noise reduction</td>
<td>Determines if the signal contains unwanted noise and reduces the level of noise if present.</td>
<td>Makes the background noise less annoying and increases listening comfort. Digital noise reduction has been shown to be effective and preferred by hearing aid wearers.</td>
</tr>
<tr>
<td>Impulse noise reduction</td>
<td>Smooths quick impulse noises such as car keys rattling, typing on a keyboard and dishes rattling.</td>
<td>Improves listening comfort.</td>
</tr>
<tr>
<td>Feedback management</td>
<td>Reduces or eliminates whistling that can sometimes occur.</td>
<td>Improves listening comfort. Basic feedback management systems may reduce the overall amplification in order to remove the whistling. Advanced feedback management systems reduce or eliminate whistling without affecting overall amplification of the hearing aid.</td>
</tr>
<tr>
<td>Telecoil</td>
<td>Picks up signal from a compatible telephone or other electromagnetically looped system.</td>
<td>Improves signal to noise ratio and eliminates feedback because the signal bypasses the microphone and directly enters the processor. Commonly available in public places, like theatres and places of worship.</td>
</tr>
<tr>
<td>FM compatibility</td>
<td>Enables hearing aids to wirelessly connect with FM systems, sometimes via a special attachment called a boot.</td>
<td>Improves signal to noise ratio because the signal bypasses the microphone and directly enters the processor. Commonly used with children in educational settings.</td>
</tr>
<tr>
<td>Bluetooth compatibility</td>
<td>Enables hearing aids to wirelessly connect to mobile phones, MP3 players and other Bluetooth devices.</td>
<td>Improves signal to noise ratio and eliminates feedback or interference because the signal bypasses the microphone and directly enters the processor.</td>
</tr>
<tr>
<td>Wind noise reduction</td>
<td>Reduces the whooshing noise of wind blowing across the hearing aid microphone(s).</td>
<td>Improves listening comfort for people who spend time outdoors—such as golfers, boaters and walkers.</td>
</tr>
<tr>
<td>Data logging</td>
<td>Stores data about listening environments and user preferences.</td>
<td>Data can be viewed by hearing healthcare professional to improve fitting at follow-up.</td>
</tr>
<tr>
<td>Learning features</td>
<td>Logs settings that are set by the wearer for certain environments and then begins to make these changes automatically.</td>
<td>Gradually, the wearer will find that they need to adjust the volume or program less frequently, as the hearing aids become able to make these changes based on the sound environment.</td>
</tr>
<tr>
<td>Binaural processing</td>
<td>The two hearing aids communicate with each other.</td>
<td>This can be used to keep the hearing aids operating synchronously or to stream auditory signals from one aid to the other.</td>
</tr>
</tbody>
</table>

Noise reduction + binaural processing

- Single microphone
  - Adaptive analog filters
  - Spectral subtraction
  - Spectral enhancement

- Multi-microphone
  - Directional microphone elements
  - Two-microphone adaptive noise cancellation
  - Array with time-invariant weights
  - Two-microphone adaptive arrays
A general framework

![Diagram of signal acquisition, enhancement, and presentation processes]

Picture from

A general framework: acquisition

\[ x = h_0s_0 + \sum_{p=1}^{P-1} h_ps_p + n = h_0s_0 + v, \]
A general framework: localization

- Steered-response power (SRP);
- MUSIC;
- TDOA;
- GCC-PHAT; SRP-PHAT;
- BSS-based
A general framework: enhancement

Data-independent beamformer
- Delay-and-sum
- Superdirective
- Differential

Need to know
- Target DOA
- Complete microphone topology

More suitable for monaural devices
A general framework: enhancement

Data-dependent beamformer: statistically optimum
- Minimum Variance Distortionless Response (MVDR)
- Multichannel Wiener Filtering (MWF)
- Blind Source Separation (BSS)

Need to estimate the interference and noise statistics
A general framework: enhancement

MVDR beamformer

$$\min_w w^H \Phi_{xx} w, \quad \text{s.t. } w^H h_0 = 1$$

$$w_{MVDR} = \frac{\Phi_{vv}^{-1} h_0}{h_0^H \Phi_{vv}^{-1} h_0}$$
A general framework: enhancement

MVDR beamformer with relative transfer function

\[
\min_w \mathbf{w}^H \Phi_{xx} \mathbf{w}, \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{h}_0 = h_{0,r} \quad \Rightarrow \quad \hat{\mathbf{w}}_{\text{MVDR}} = \frac{\Phi_{vv}^{-1} \tilde{h}_0}{\tilde{h}_0^H \Phi_{vv}^{-1} \tilde{h}_0} \]

\[
\tilde{h}_0 = \frac{h_0}{h_{0,r}}
\]
A general framework: enhancement

Multichannel Wiener Filter (MWF)

$$\min_w E \left\{ \left| s_0 - w^H x \right|^2 \right\} \rightarrow w_{MWF} = \Phi^{-1}_{xx} h_0 \phi_{s_0 s_0}$$

PSD of $s_0$
A general framework: enhancement

MWF with relative transfer function

\[
\min_w E \left\{ \left\| h_{0,r}s_0 - w^H x \right\|^2 \right\} \quad \tilde{w}_{\text{MWF}} = \left( \phi_{s_0s_0} h_0 h_0^H + \Phi_{vv}^{-1} \right) \phi_{s_0s_0} h_0 h_0^* 
\]
A general framework: enhancement

MWF with speech-distortion-weighted

$$\min_w E \left\{ \left| h_{0,r} s_0 - w^H h_0 s_0 \right|^2 \right\} + \mu E \left\{ \left| w^H v \right|^2 \right\} \quad \Rightarrow \quad \tilde{w}_{\text{MWF}} = \left( \phi_{s_0 s_0} h_0 h_0^H + \mu \Phi_{vv}^{-1} \right) \phi_{s_0 s_0} h_0 \tilde{h}_{0,r}$$
A general framework: presentation

Binaural extraction:
Also important to preserve the binaural cues of the residue noise to allow binaural unmasking
ANC in hearing aids with occlusion reduction (ITC hearing aids)

- Increased SPL at low frequencies when ear canal is blocked (~15 dB from 100-300 Hz)
- Occlusion reduction using FBANC.
- Need a fast adaptive algorithm to achieve good occlusion reduction.

Fig. 2 Configuration of the proposed system.

Picture from

Integrating ANC in Hearing Aids

- No earmold to prevent ambient sound from leakage into the ear canal → Lower SNR
- Attenuate leakage noise signal using ANC.
- Need an ear canal microphone

Picture from

Summary on hearing aids

- Various signal processing approaches have been applied in noise reduction for hearing aids.
- Incorporating multiple microphones and binaural processing have been shown to be beneficial and will be the key research topics in hearing aids.
- Personalization embedded with intuitive interface (in phones): hearing loss profile, or listening preference.


Module VI
Summary
Aims of Assisted Listening

- Correct spatial cues
- Good audio quality
- Only wanted signals
Summary on Assisted Listening

- Real Sound
- Spatial filtering
- Adaptive filtering
- Head tracking
- Individualization
- Decomposition using BSS/PAE
- Source rendering
- Environment rendering
- Virtualization
- Rendering of natural sound

Sound sources → Sound Mixing → Sound mixture

Head movement → Individual parameters → Head tracking → Individualization

Sound environment → Sound mixture → Decomposition using BSS/PAE → Equalization

Equalization → Spatial filtering → Adaptive filtering

Headphone Hearing aids Headset
### The Future of Assisted Listening

<table>
<thead>
<tr>
<th>Action/Technique</th>
<th>You hear</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Virtual sound</td>
<td>Real sound</td>
</tr>
<tr>
<td>Put on headphone</td>
<td>Normal</td>
<td>Modified by headphone</td>
</tr>
<tr>
<td>Masking</td>
<td>Enhanced</td>
<td>Modified by headphone</td>
</tr>
<tr>
<td>ANC</td>
<td>Normal</td>
<td>Reduced</td>
</tr>
<tr>
<td>Hear-through</td>
<td>Normal</td>
<td>Recovered</td>
</tr>
<tr>
<td>Acoustic SP</td>
<td>Normal</td>
<td>Personally modified</td>
</tr>
<tr>
<td>Acoustic &amp; Audio SP</td>
<td>Natural</td>
<td>Blocked</td>
</tr>
<tr>
<td>Acoustic &amp; Audio SP</td>
<td>Natural</td>
<td>Recovered</td>
</tr>
<tr>
<td>Acoustic &amp; Audio SP</td>
<td>Natural</td>
<td>Selective</td>
</tr>
</tbody>
</table>
Acknowledgements

And all researchers working in assisted listening, 3D audio, active noise control, virtual reality, augmented reality!

Mr. Rishabh Ranjan (Researcher @NTU)
Dr. Kaushik Sunder (Visiting research fellow @NASA)
Dr. Joseph Tan Ee-Leng (CTO @Beijing Sesame World)

Mr. Physo Ko Ko (Researcher @NTU)
Mr. Nguyen Duy Hai (Researcher @NTU)
Contact us

Dr. Woon-Seng Gan: ewsgan@ntu.edu.sg

Mr. Jianjun He: jhe007@e.ntu.edu.sg

DSP Lab, School of EEE, Nanyang Technological University, Singapore