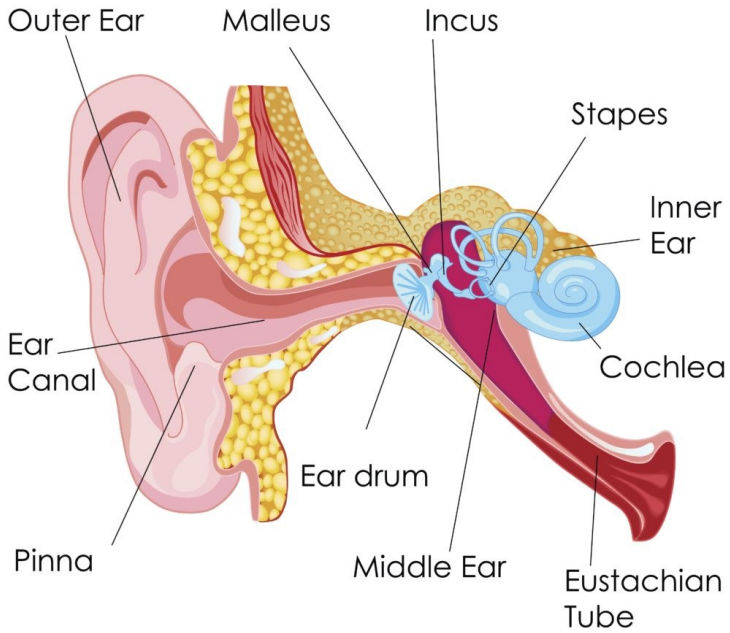
Reconstructing Ear Canal Channels for Fine-Grained Detection of Tympanic Membrane Changes

[*ACM IMWUT'25*]

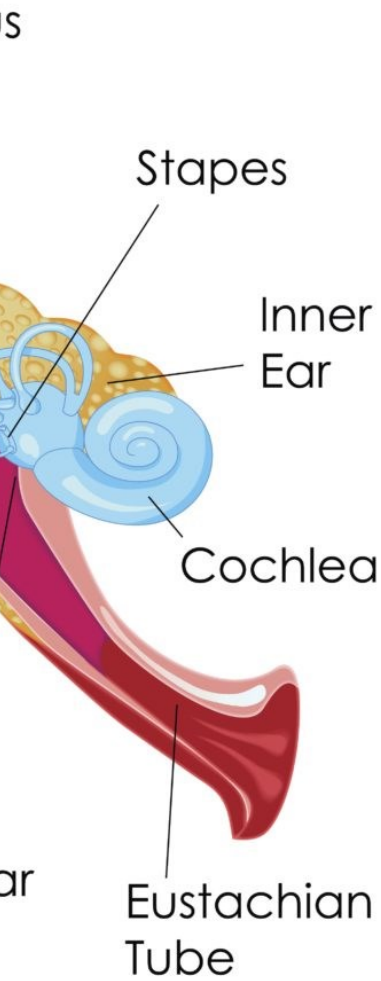
YONGZHI HUANG, JIAYI ZHAO, and KAISHUN WU,
Data Science and Analytics Thrust (DSA),
Information Hub (INF), The Hong Kong University of Science and Technology (Guangzhou), China

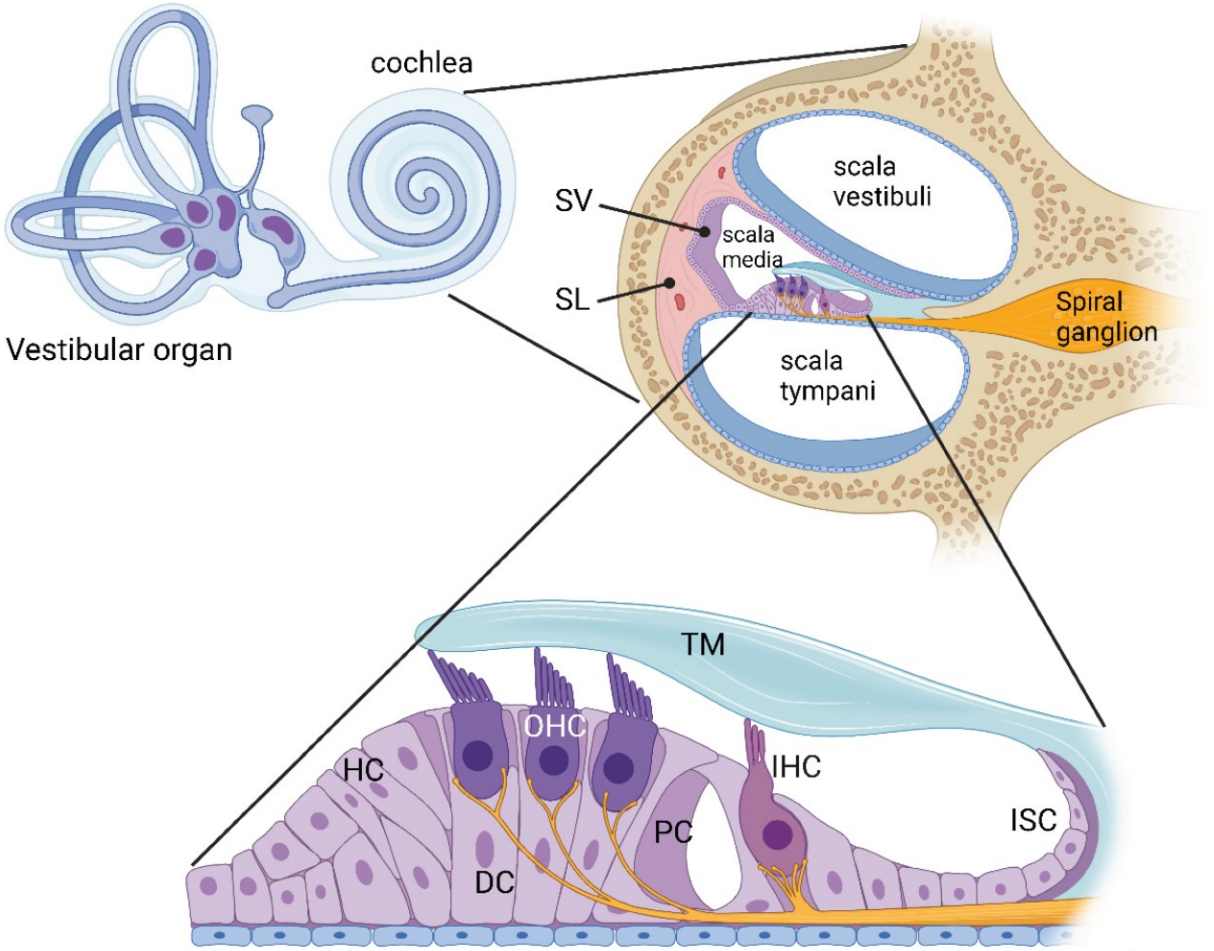
Background

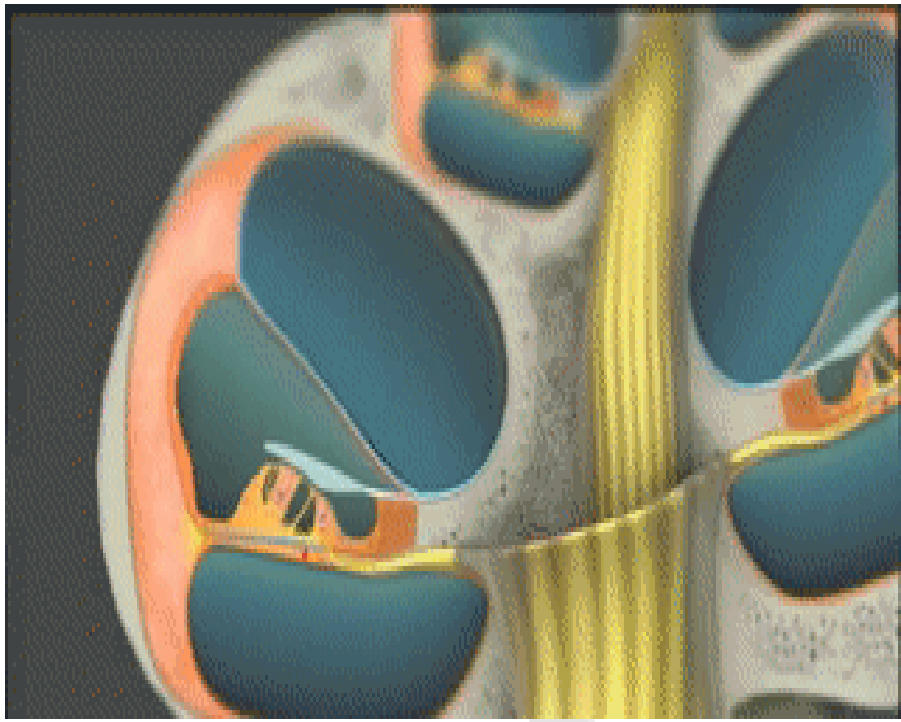
Vision: Ear-Canal Digital Twin

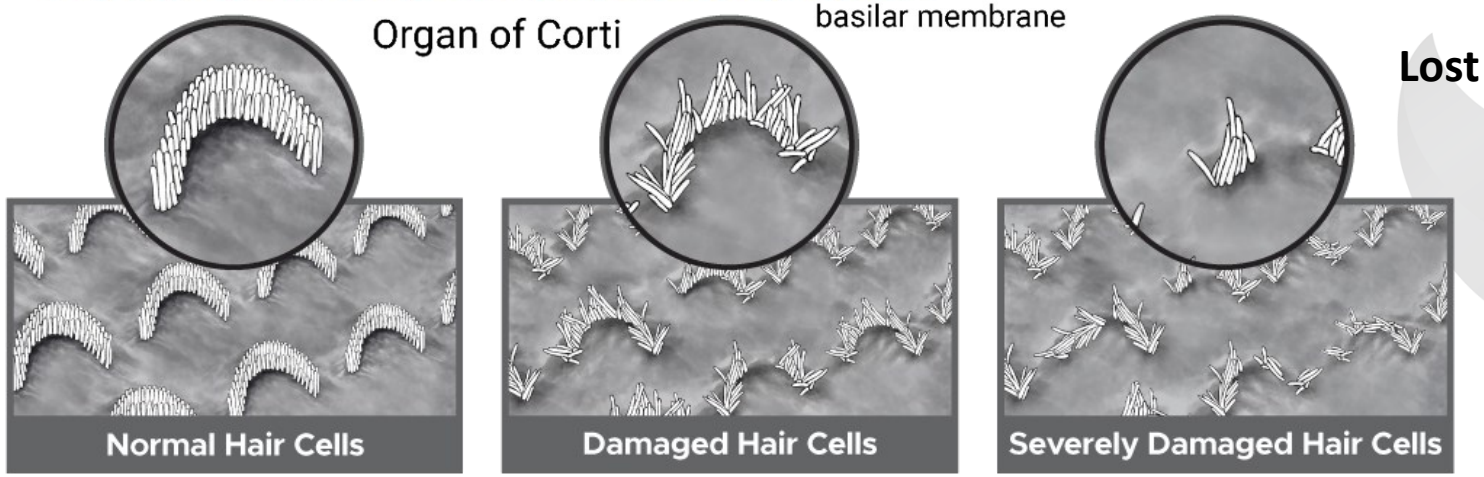


A personal digital twin of the ear canal





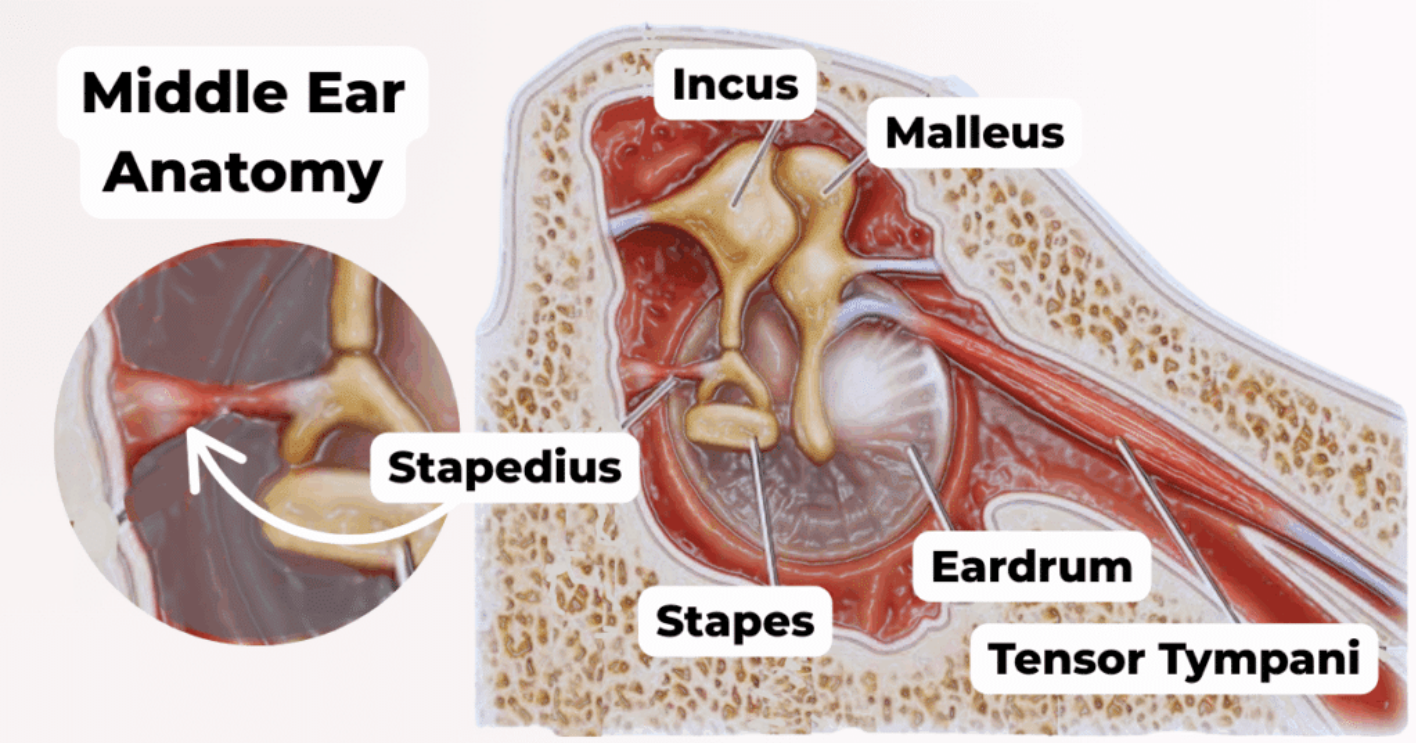




Lost hair cells don't regenerate

Permanent loss in hearing.

Background - Acoustic Reflex



Background

- **Subtle tympanic membrane (TM) deformation** is an **early signal** of long-term hearing risk.

- **Problem**: Raw loudness is confounded by device, fit, and environment.

- **Solution:** Reconstruct a **per-ear canal channel (CSI)**—a digital twin—to strip confounds and expose tiny changes.

Steps

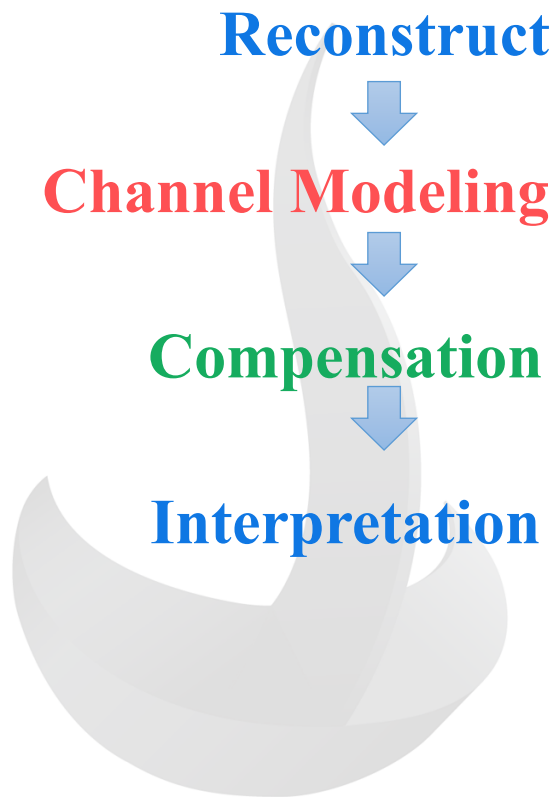
Calculate the CSI of the ear canal!

01 1. Reconstruct the canal **Length** **Shape**

02 2. Audio-driven inference

03 3. Robust to the real world

04 4. Health interpretation



Challenge 1——How to obtain fine ear canal structure?

- Traditional **ToF ranging method** has insufficient accuracy, system latency and thread synchronization issues.

Length **Shape**

Low-precision microphones

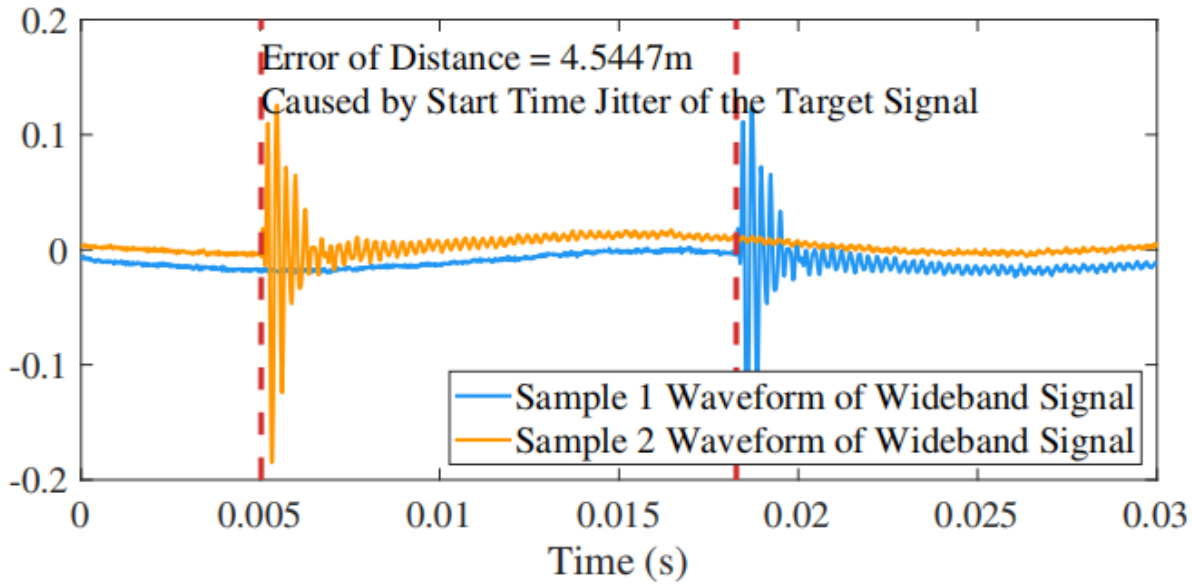
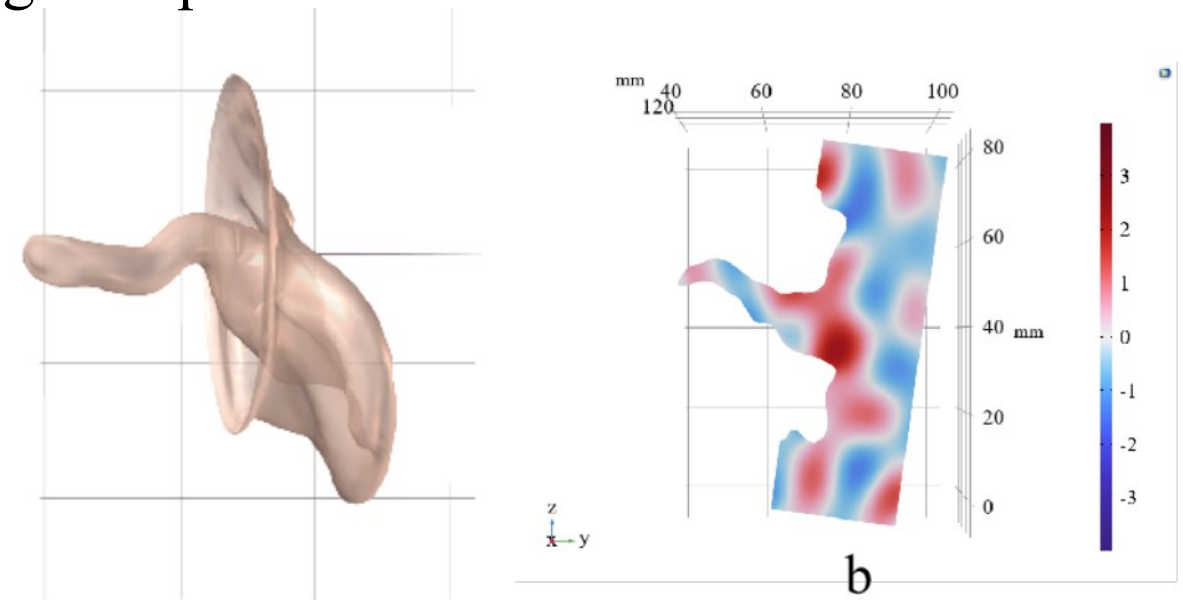


Fig. 8. Error in traditional acoustic ranging methods

Reasons for inaccurate TOF

- **Sound propagation path** inside the ear canal is **complex**. (multipath reflection, diffraction, and scattering phenomena)

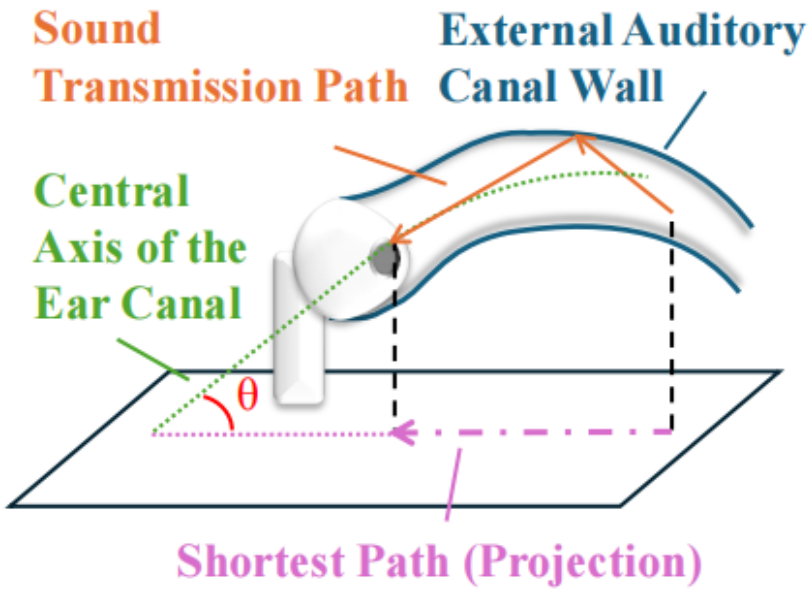
- **Nonlinear relationship** between the peak spacing (Δf) and the actual propagation path.

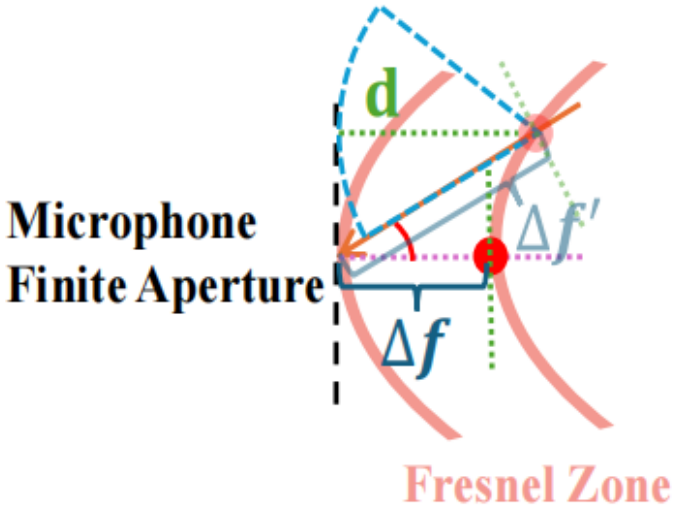


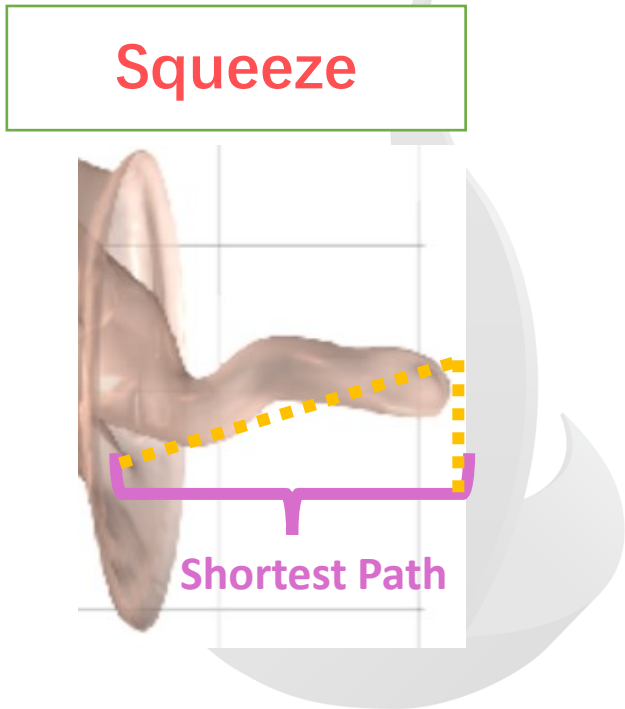
Ear canal simulation

Method 1: We design the Spectral angle and shortest path estimation method

- Propose a **shortest path estimation model** based on spectral peak spacing (Δf) and angle.







Method 1

- Establish a geometric projection model using the **finite aperture effect** of microphone diaphragm:

$$\Delta f_{\text{ideal}}(\theta) = \frac{8 f_1 f_2 d}{c(2n+1)} \sec\theta \equiv k \sec\theta,$$

- Introduce systematic error modeling, calibrate multipath reflection error and directional error, and establish a **robust spectrum space mapping model**:

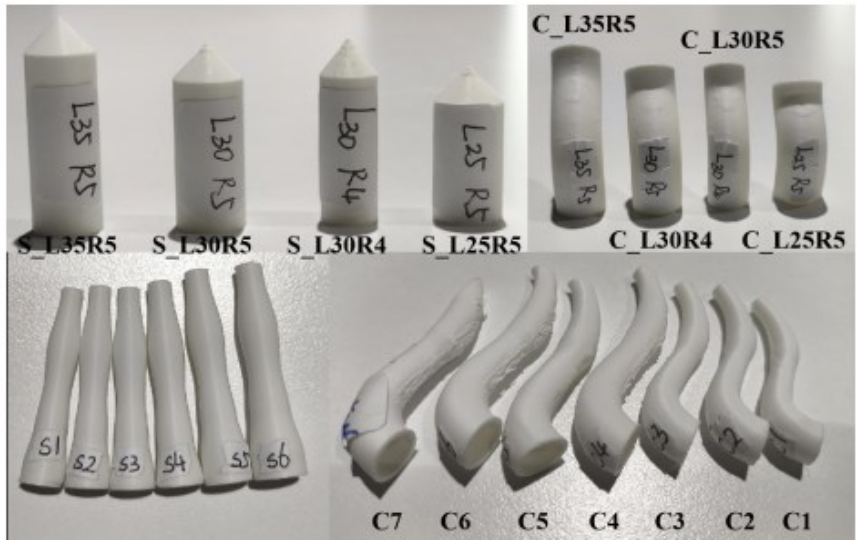
$$\Delta f_{\text{obs}}(\theta) = k \sec\theta \left[1 + \varepsilon_{\text{dir}}\, g_{\text{dir}}(\theta) + \varepsilon_{\text{mp}}\, g_{\text{mp}}(\theta) \right] + \eta_{\text{sys}},$$

Method 1

Innovation Point

- Utilizing frequency domain signals to achieve spatial angle and path estimation **without the need for precise time synchronization**.

- Effectively **avoiding the bottleneck** of traditional time-domain ranging errors.



Silicone simulation of the ear canal

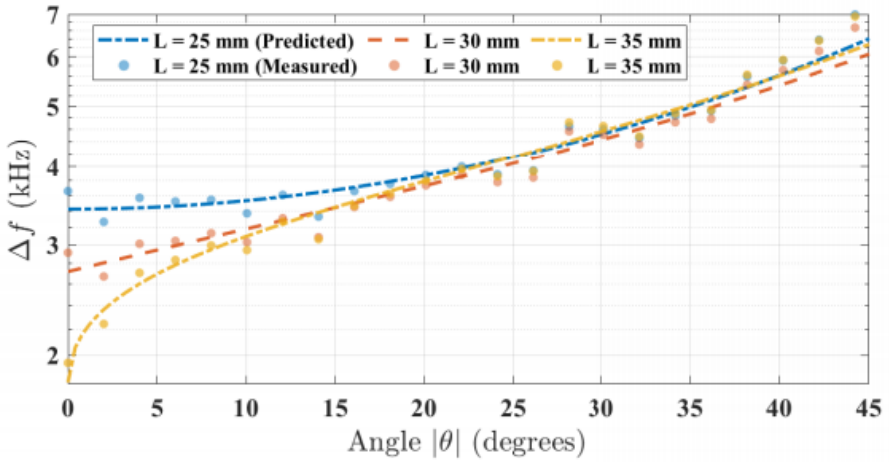


Fig. 11. Relationship between propagation angle θ and spectral peak spacing Δf. Theoretical predictions for L = 25, 30, and 35 mm are compared with measured values obtained using curved silicone ear canal models.

Challenge 2——Impact of changes in ear canal cross-sectional area

- Merely considering the path length is insufficient to accurately model the acoustic characteristics of the ear canal.

- The actual cross-sectional area of the ear canal significantly affects acoustic transmission, and traditional fixed area models have large errors.

Shape

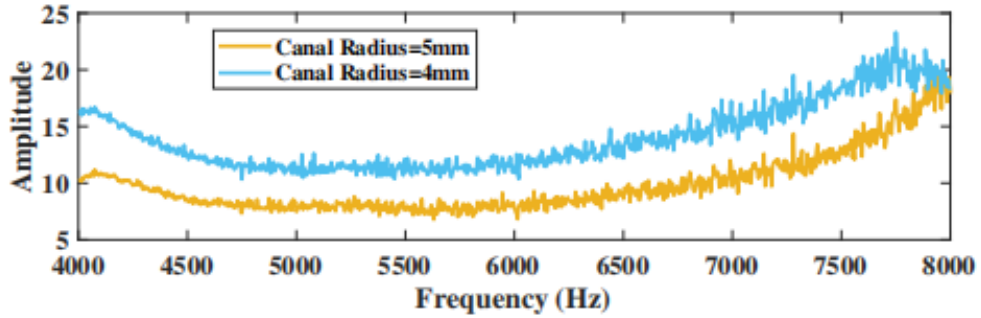


Fig. 12. Frequency response of two ear-canal phantoms with identical lengths but different radii (4 mm vs. 5 mm).

Spectral variations caused by different ear canal radii

Method 2: We design the Frequency domain reconstruction method based on the cross-sectional area function

How to infer the cross-sectional area of the ear canal merely from sound signals?

■ Propose a **segment by segment inverse estimation model** for ear canal cross-sectional area based on reflection coefficient:

Helmholtz equation

$$\frac{\mathrm{d}^2 P(x)}{\mathrm{d}x^2} + k^2 P(x) = 0, \quad \Longrightarrow \quad R_n = \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n},$$

By applying the principles of fluid mechanics!

P(x)： **acoustic pressure**

k=2πf/ceff： **wavenumber**

How to obtain it?

$$A_{n+1} = A_n \cdot \frac{1 - R_n}{1 + R_n} \quad \Longleftarrow \quad R_n = \frac{A_n - A_{n+1}}{A_n + A_{n+1}}$$

Slice derivation!

Method 2——Frequency domain reconstruction based on cross-sectional area function

- The Webster equation is introduced to establish a frequency domain model of sound wave propagation in the ear canal that is related to the area. ：

$$H(\omega) = \frac{U_{\text{out}}(x,\omega)}{P_{\text{in}}(x,\omega)} = -\frac{A(x)\,\dfrac{\partial}{\partial x}\left(-\dfrac{P(x)}{j\,\omega\,\rho_0}\right)}{P_{\text{in}}(x,\omega)}.$$

- A method for inferring the changes in the cross-sectional area of the ear canal from frequency domain signals is proposed.

$$\frac{\partial^2 p(x,t)}{\partial x^2} - \frac{1}{A(x)}\frac{\partial A(x)}{\partial x}\frac{\partial p(x,t)}{\partial x} = \frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2},$$

$$\frac{\mathrm{d}^2 P(x)}{\mathrm{d}x^2} - \frac{1}{A(x)}\frac{\mathrm{d}A(x)}{\mathrm{d}x}\frac{\mathrm{d}P(x)}{\mathrm{d}x} + \frac{\omega^2}{c^2}P(x) = 0.$$

- It effectively overcomes the errors caused by the fixed cross-sectional area model and provides accurate ear canal structure information.

$$H(\omega) = \frac{U_{\text{out}}(x,\omega)}{P_{\text{in}}(x,\omega)} = -\frac{A(x)\,\dfrac{\partial}{\partial x}\left(-\dfrac{P(x)}{j\,\omega\,\rho_0}\right)}{P_{\text{in}}(x,\omega)}.$$

Challenge 3: Efficiency of Segment-by-Segment Cross-Sectional Reconstruction

- While the recursive reconstruction based on reflection coefficients described above is theoretically feasible, it is computationally intensive at high-resolution frequency response.

Large amounts of Computation

- How to improve the computational efficiency of the algorithm while ensuring accuracy?

Use the multiplication operator

Linear approximation

Method 3: Geometric Inference in Fast Reflection Mode

- We design A fast approximation algorithm based on Taylor expansion is proposed:

$$\frac{1 - R_n}{1 + R_n} \approx 1 - 2R_n$$

- A second-order Taylor correction strategy is used. When the reflection coefficient is large ($R\ n > 0.05$), the following method is used:

$$A_{n+1} \approx A_n \left(1 - 2R_n + 2R_n^2\right)$$

Use the multiplication operator

Linear approximation

Challenge 4: Stability and Global Consistency Issues

- While local Taylor expansion can accelerate computation, the recursive process is still inherently local.

- In cases with long ear canals or complex anatomical structures, local errors can still propagate layer by layer, resulting in insufficient overall model accuracy.

Method 4: Define a global objective

- A robust ear canal cross-sectional area reconstruction framework based on global optimization is proposed by defining the objective function:

$$\Phi\big(A(x)\big) = \sum_i \big[R_{\text{measured}}(f_i) - R\big(f_i, A(x)\big)\big]^2$$

- Adding length constraints and smoothing regularization:

- Length constraints: Ensure that the estimated ear canal length is within a physically reasonable range;
- Smoothing regularization: Ensure that the variation of the ear canal cross-sectional area is continuous and physically reasonable.

- Optimize using a gradient descent algorithm to iteratively solve the cross-sectional area function $A\ (x)$.

Results

Compare with Original CSI

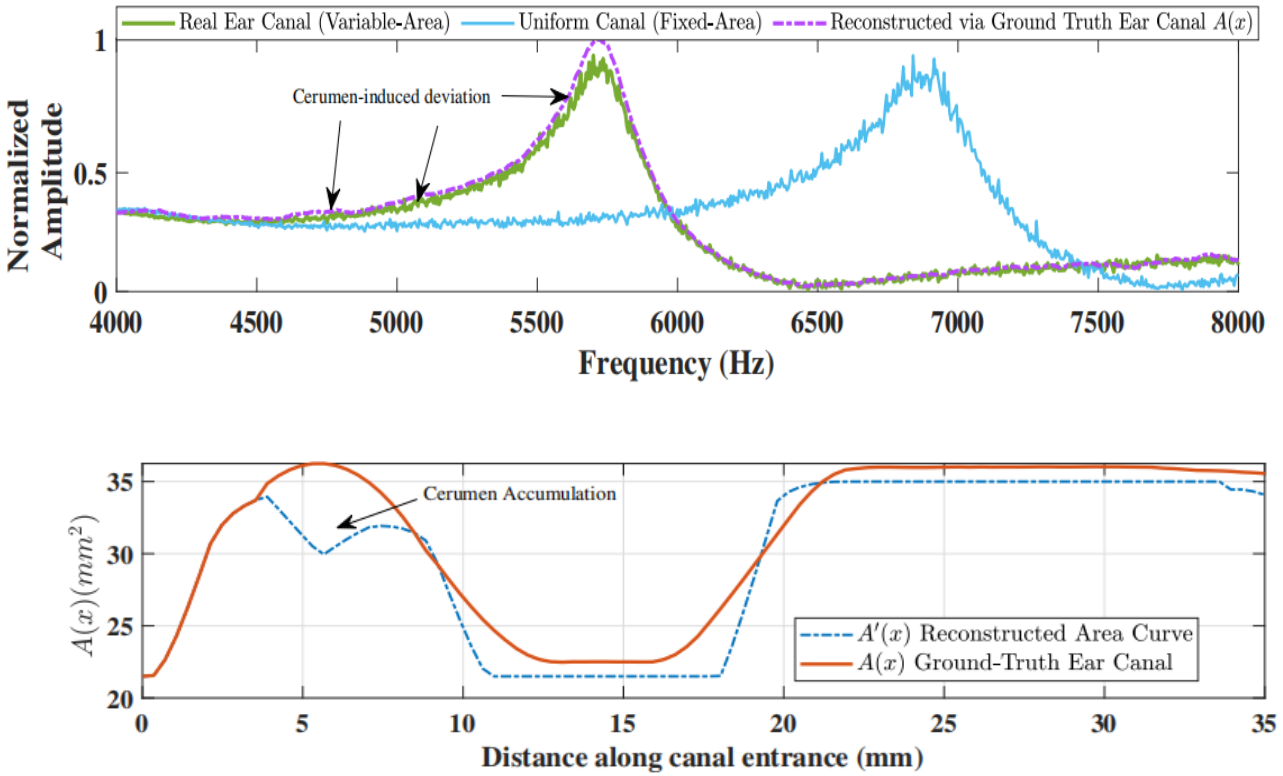


Fig. 14. Comparison Between the Ground-Truth and Reconstructed Cross-Sectional Area of the Ear Canal, while deviations reflect minor in-vivo morphological variation

Innovation Point

- Greatly improve the efficiency of cross-sectional area reconstruction algorithm and reduce computational complexity.
- By using adaptive **high-order approximation**, the accumulation of errors is effectively suppressed.

Result in different ears

- The solid curve of the true cross-sectional area and the estimated dashed curve are nearly identical, with an extremely low overall error.
- The maximum absolute error is < 6 mm²; the average error across age groups is 2.0–5.2 mm², with no significant change with age.

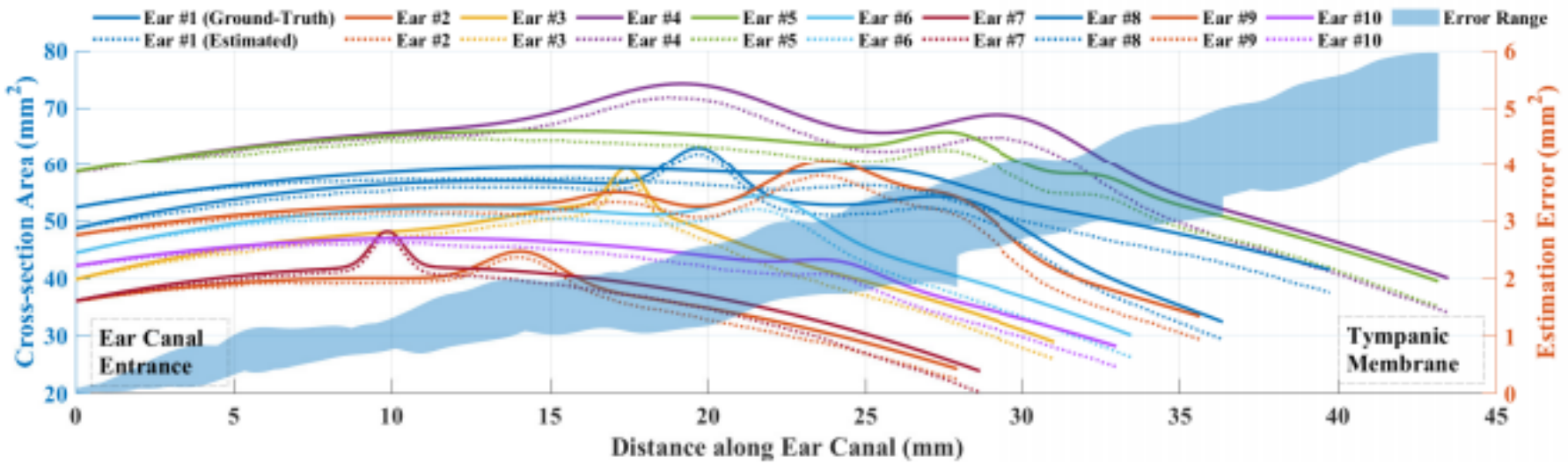


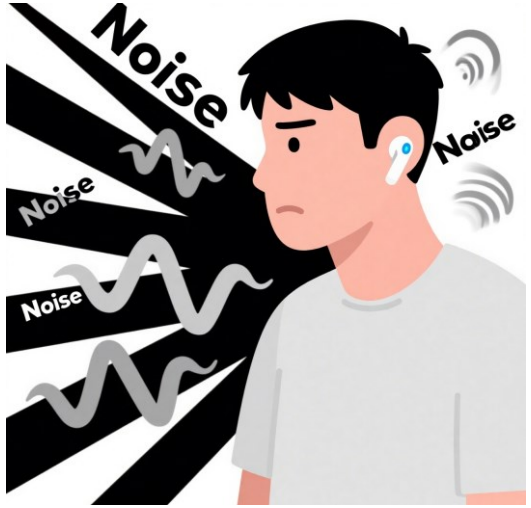
Fig. 15. Reconstruction of Cross-Sectional Area $A(x)$ for different Ear Canal Models. Dashed lines denote estimated profiles; shaded regions indicate distance-dependent reconstruction error range.

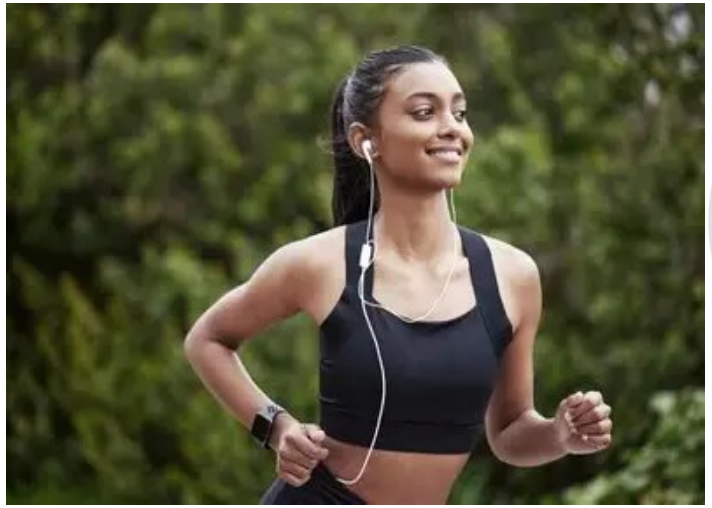
Challenges and Methods

Challenge 5——Stability issue of reconstructing cross-sectional area segment by segment

- In daily use, there are inevitably factors such as background noise, motion interference, unstable headphone wearing position, and nonlinear distortion.

These will seriously **affect the reliability** of frequency domain features, leading to **unstable results** in ear canal reconstruction.





Method 5: Multi-Level Interference Compensation Strategy (Increase Robustness)

Design an interference compensation module to suppress interference from the signal and distribution layers:

- **5.1** Time Window and Link Dispersion Constraints:
  - Due to hardware and system dispersion (transducer group delay, codec buffer, etc.).

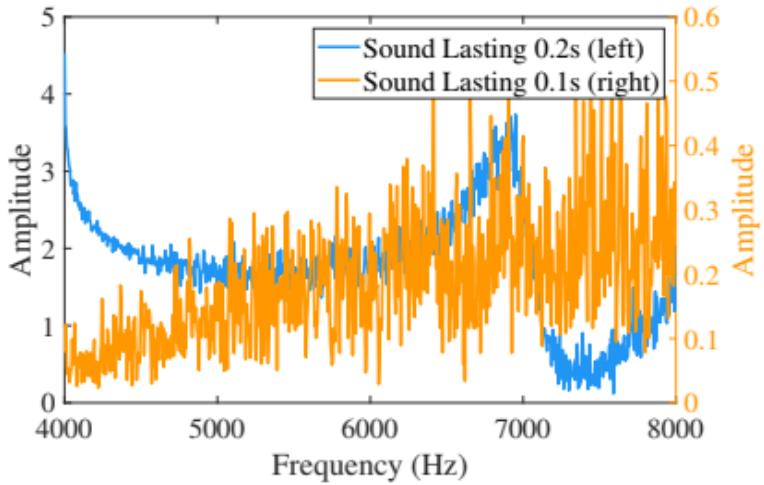


Fig. 18. Temporal Constraint for Reliable Frequency Response Capture in Ear-Canal Sensing

Method 5: Multi-Stage Interference Compensation Strategy

- **5.2** Background Noise and Headphone Leakage Compensation:

  - Frequency-domain adaptive denoising technology is used to extract stable frequency components from the signal;

  - Short-time spectral subtraction is used to suppress the influence of background noise.

Method 5

5.3 Headphone Wearing Instability and Motion Artifacts:

- We design **an instability detection algorithm** based on the frequency response phase change rate is proposed.

Normalized Phase Slope:

$$Y_I(\omega) = \frac{\Im\{H(\omega)\}}{\omega}$$

- Detects headphone offset and performs real-time dynamic calibration.

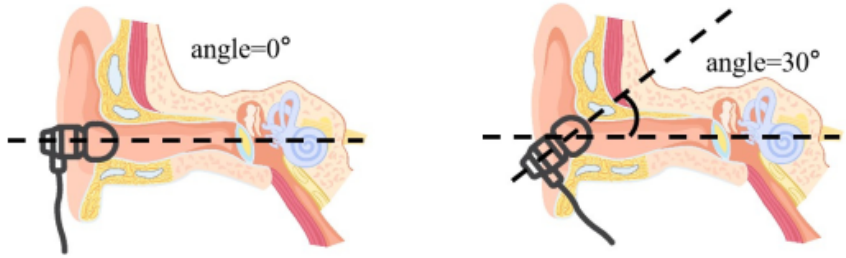


Fig. 21. Illustration of headphone misalignment: Proper vs rotated orientation.

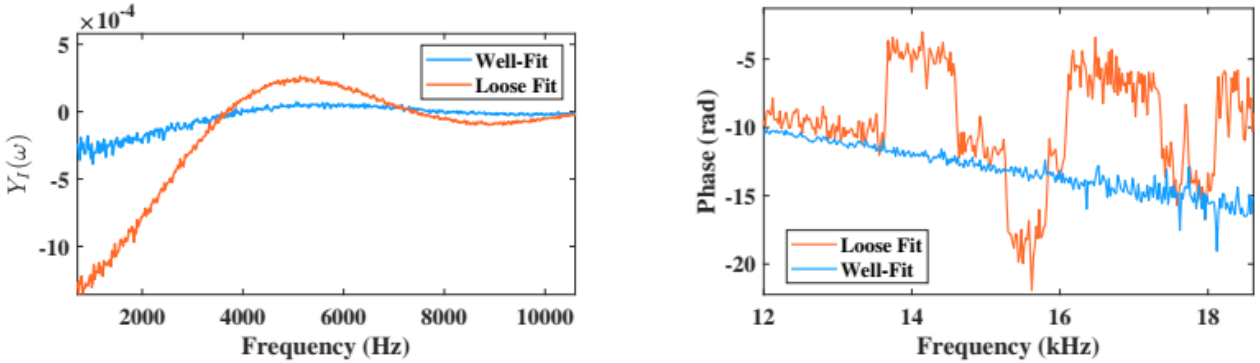


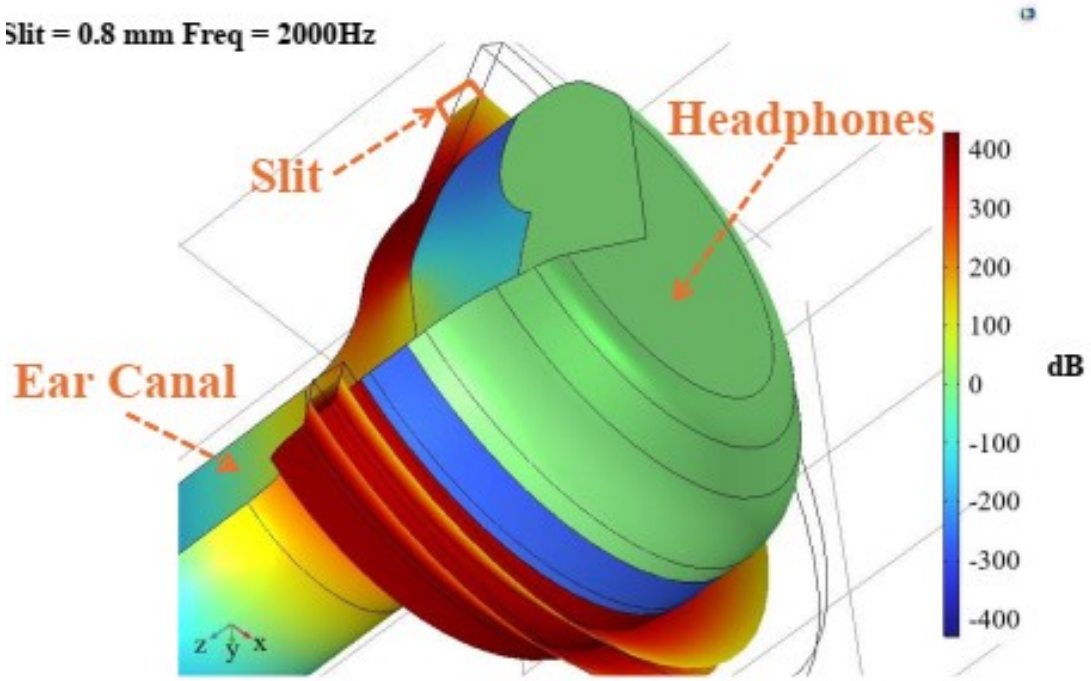
Fig. 22. Phase behavior under motion-induced misalignment. Left: $Y_I(\omega)$ remains flat in well-fit condition, but fluctuates when loose-fit. Right: The phase response deviates from linearity under loose fitting.

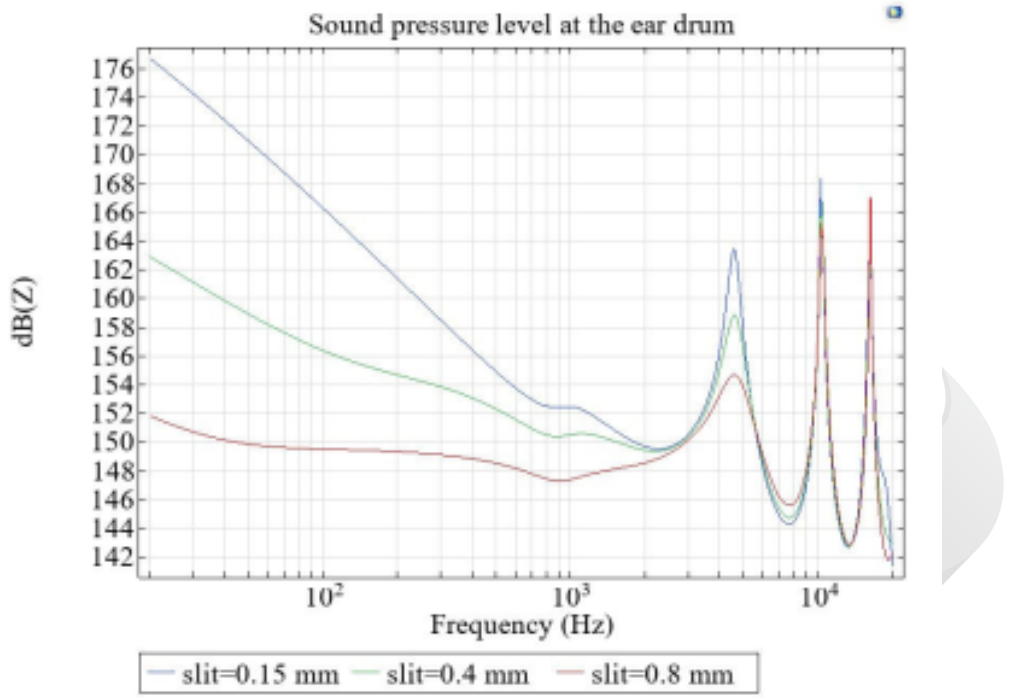
Because of looseness/rotation, angular multipath causes the signal to fluctuate with frequency and the phase curve loses linearity.

Method 5

5.4 Signal leakage:

■ We found there is Signal leakage: Since earphones can fall off or become loose during exercise, low-frequency anomaly detection is designed to eliminate interference.





Method 5

- **5.5** Nonlinear Distortion:

  - Continuous amplitude threshold crossings can lead to nonlinear segments of microphone/preamplifier saturation.

  - **Traditional Methods, such as direct polynomial fitting, can produce Runge oscillations.**

Solution:

Lightweight bandpass decomposition

+

Signal Reconstruction

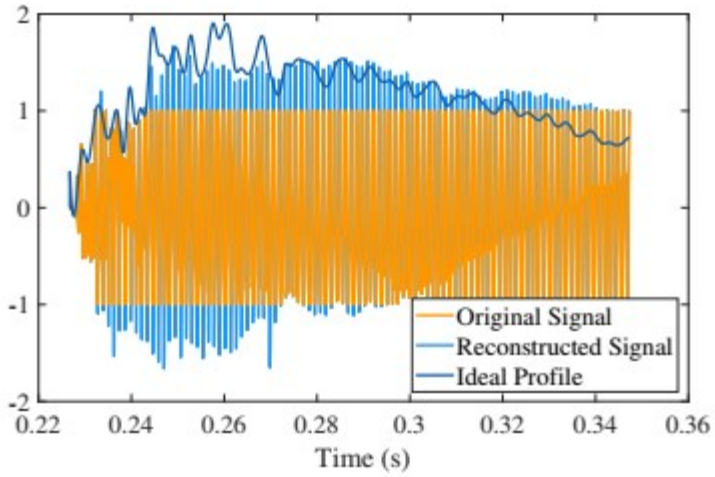


Fig. 24. Comparison Between Reconstructed Nonlinear Signal and Ground Truth Signals Profiles

Method 5

■ 5.6 Nonlinear distortion and environmental drift:

■ We design FID (Feature Invariant Distribution) transfer method to dynamically calibrate frequency domain features;

■Effectively alleviate spectral changes caused by environmental drift such as temperature and humidity changes.

Reconstructing Ear Canal Channels for Fine-Grained Detection of Tympanic Membrane Changes

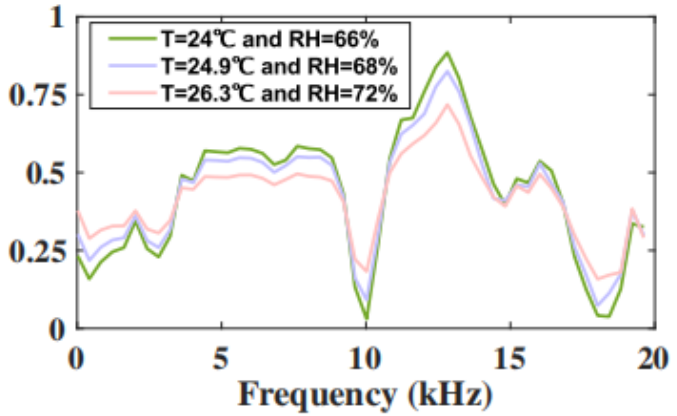


Fig. 25. Impact of Environmental Variability

System Implementation

- **Hardware**:
  - Smartphone (Android 13+Snapdragon 8 Gen 2+8GB RAM)
  - Customized in ear headphones (integrated microphone)
  - USB-C → 3.5mm interface (USB Audio Class 2.0)
- **Software**:
  - Linear Discriminant Analysis (LDA)
  - Optimize the projection vector ω:

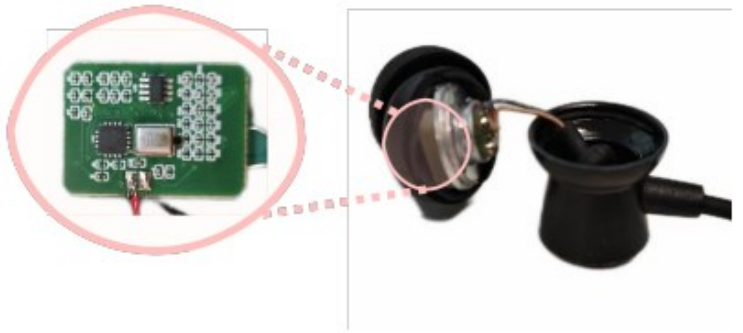
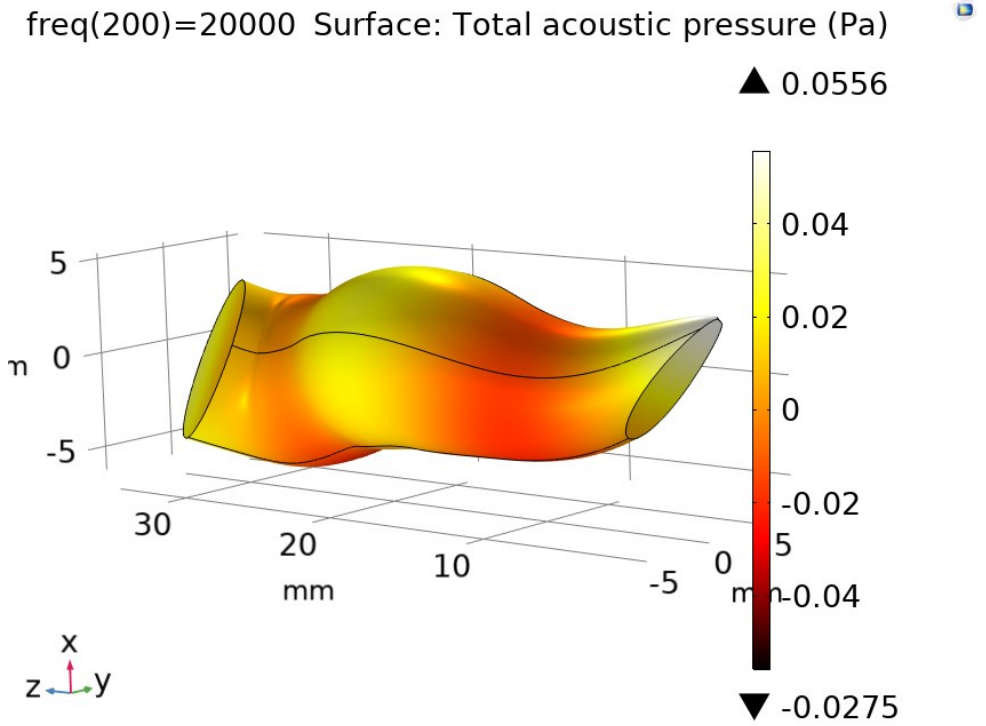


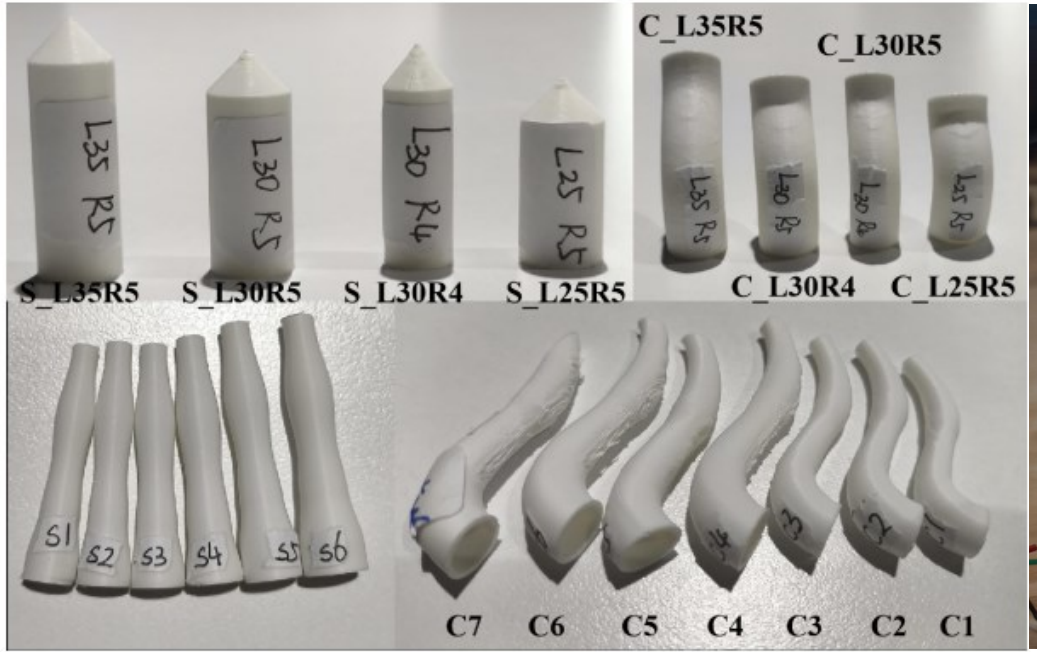
Fig. 27. The In-Ear Microphone Deployment

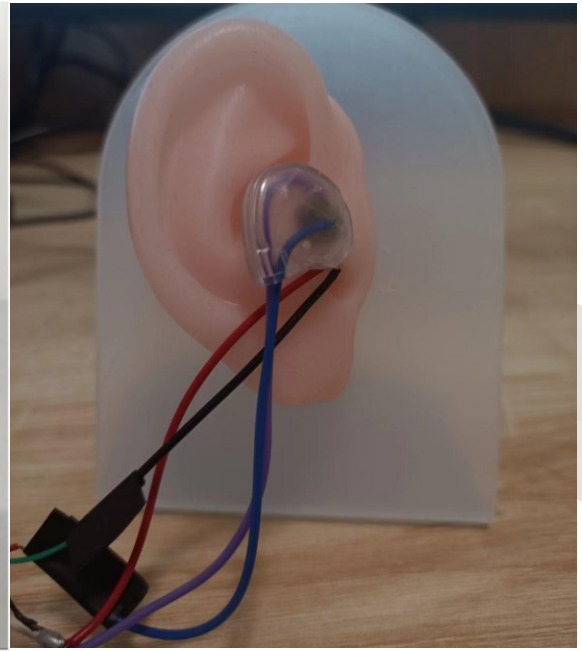
Experimental Tools

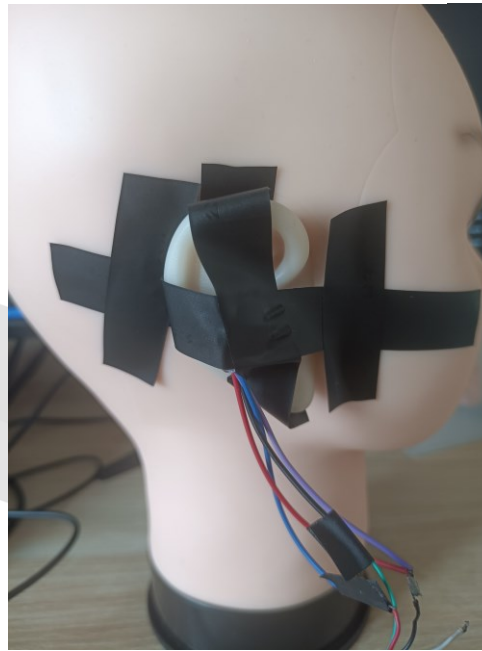
Testing Environment:

- COMSOL
- Sound Pressure Sensor
- Simulated Silicone Ears
- Head Model with Headphones

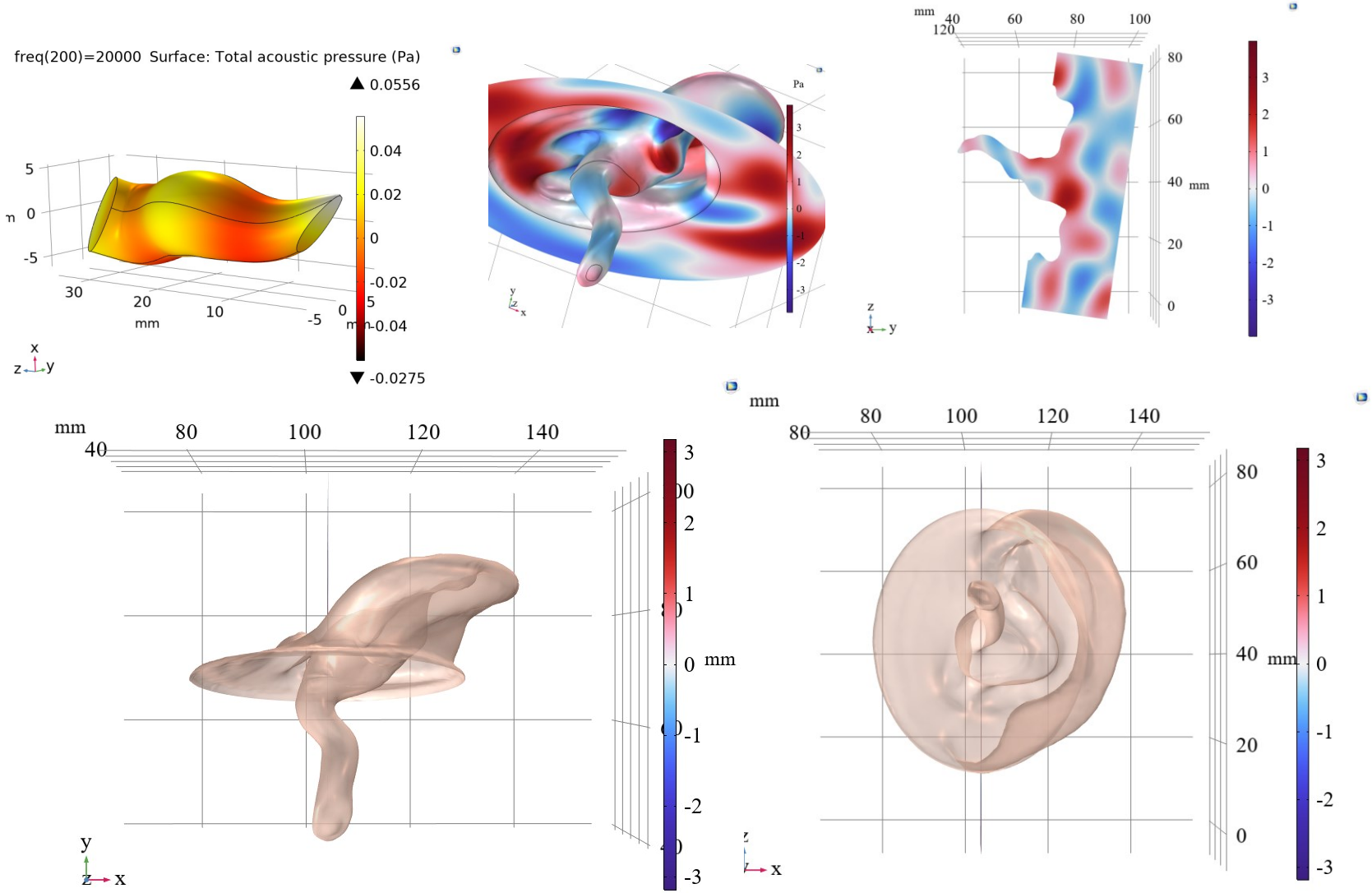








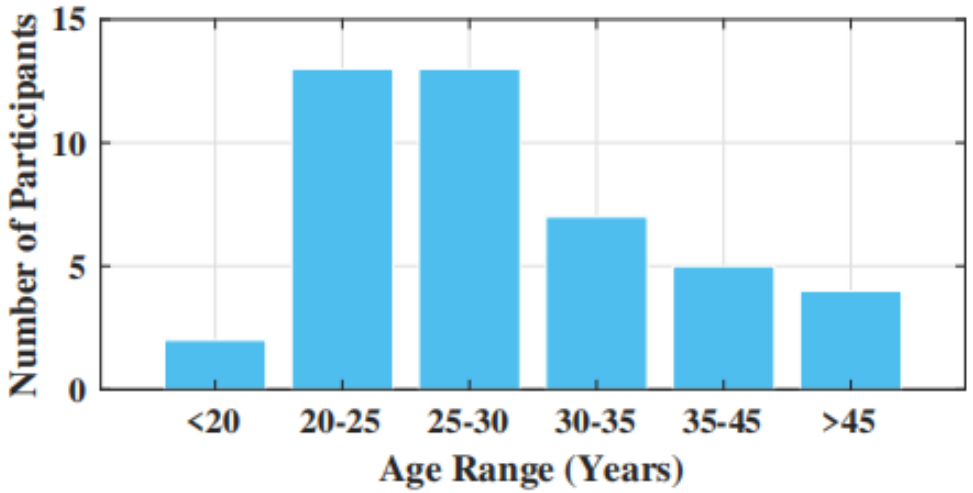
Experimental



Data Collection

- **Participants:**
  - 44 volunteers (24 ♂ / 20 ♀)
  - Age distribution

- **Equipment comparison:**
  - Verification device: MEMS microphone (ICS-40720, SNR 70dB)

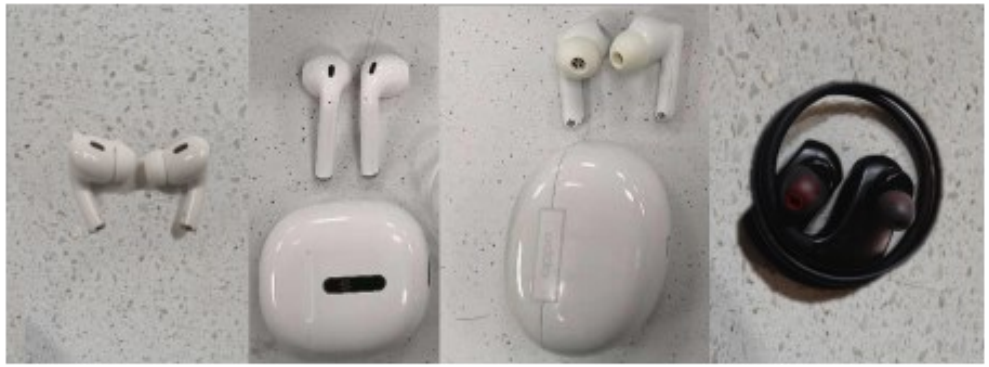


Real-time Processing

DATA: 1-second audio clip

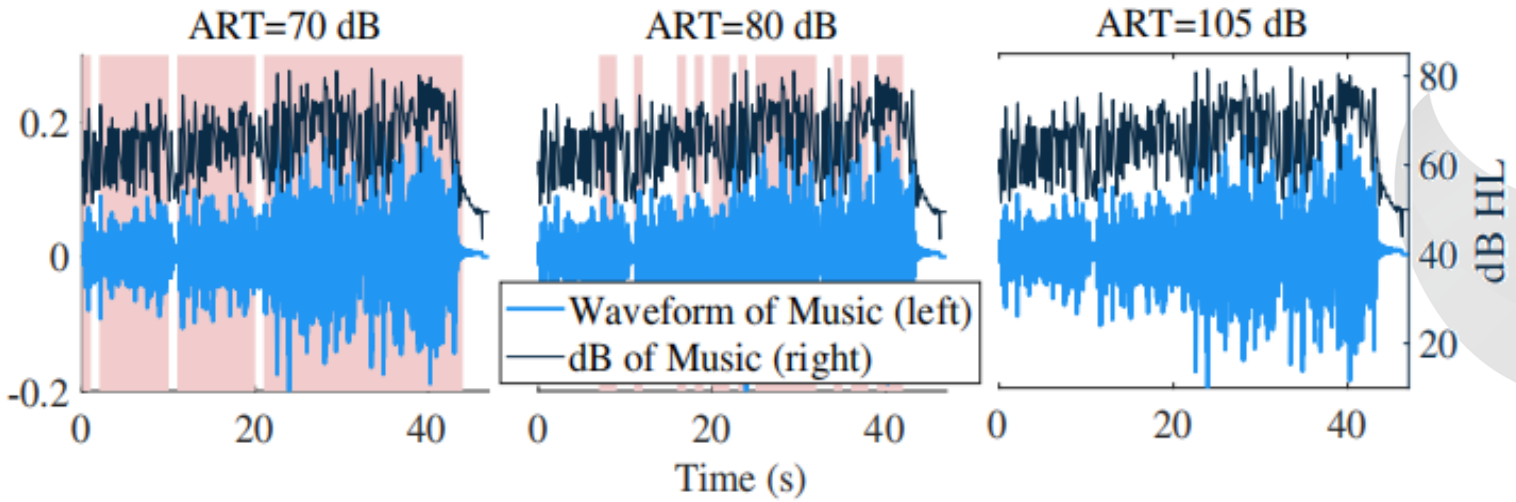
Performance: 168.38ms/ inference

Evaluation of safe listening levels



Medical Device

Individual Variations in Cochlear Response to Music at Normal Volume.



Evaluation

1. Ablation study

Table 3. Accuracy and Inference Latency for Different System Configurations

Configuration	Accuracy (%)	Latency (ms)	Accuracy Gain
Raw Input (No Processing)	34.01	37.9	–
+ Channel Modeling Only	80.45	93.5	↑ 46.44%
+ Interference Mitigation Only	59.20	81.6	↑ 25.19%
+ Full System (All Modules)	**95.91**	**168.4**	↑ 61.90%

Table 4. Accuracy Drop When Removing Individual Interference Mitigation Submodules

Removed Submodule	Accuracy Drop (%)	Remaining Accuracy
FID-Based Environmental Adaptation	↓ 9.95	85.96
Nonlinear Distortion Correction	↓ 8.71	87.20
Background Noise Suppression	↓ 4.38	91.53

Evaluation

2. Accuracy - latency trade-off

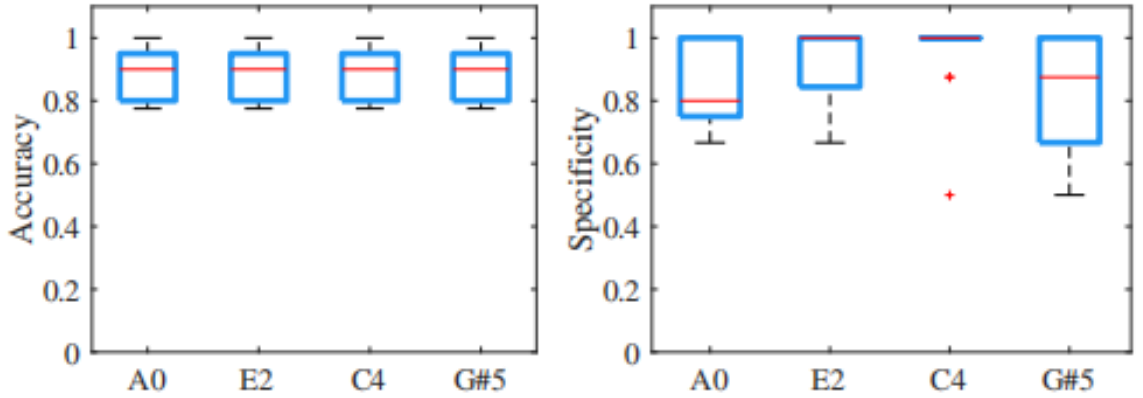
Latency faults per module (Snapdragon 8 Gen 2)

Module	Max (ms)	Average (ms)
Channel Reconstruction	250	70
Noise Suppression	125	50
Leakage Detection	25	10
Motion Artifact Rejection	25	10
Nonlinear Distortion Compensation	42	15
Environmental Adaptation (FID)	17	5
Classification	17	8
Total System (Average)	–	**168**
Worst-Case Runtime	**500**	–

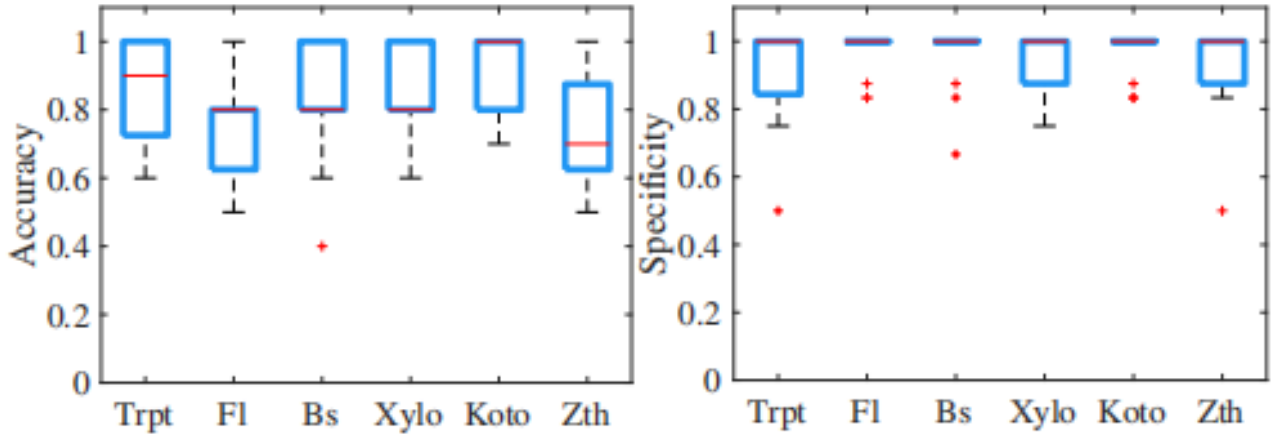
Evaluation

3. Model generalization for music input

Detection performance of five piano timbres (A0, E2, C4, G#5, B6).



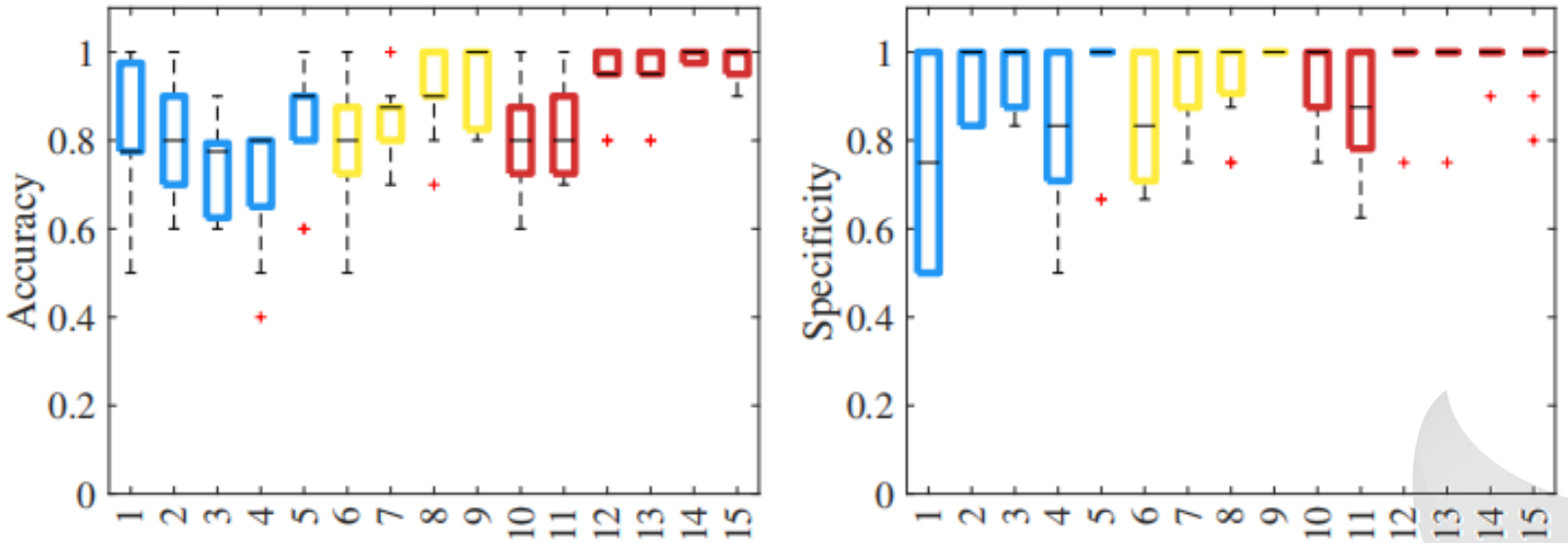
Detection performance of monotone E3 played using different instruments.



Evaluation

4. Assessment under different health conditions

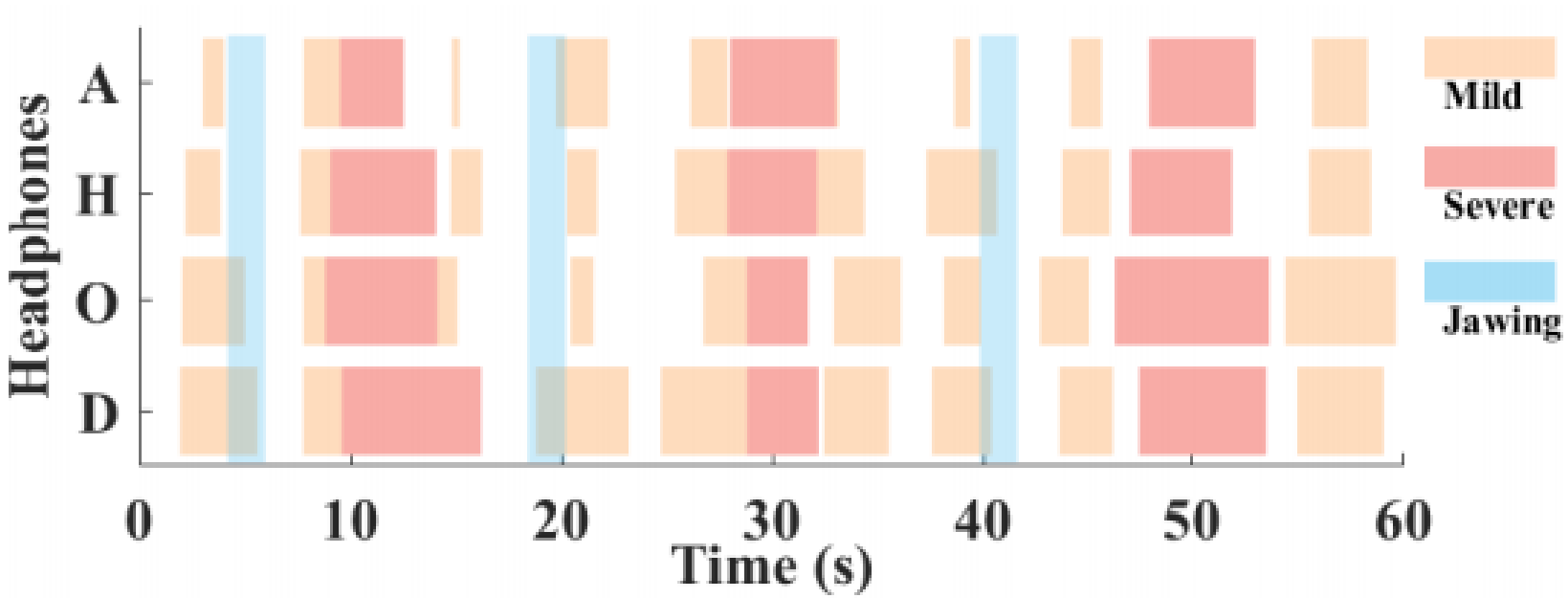
Model accuracy and specificity for 15 representative ears spanning three healthy groups: ART-1 (normal reflex), ART-2 (mildly elevated), and ART-3 (high threshold).Each ear was tested with broadband, piano, and musical instrument stimuli (50-90dB).



The median accuracy remained above 77.5% for all ear canals, and the specificity always exceeded 0.8.

Evaluation

The tympanic membrane contractions of the four headset models were detected for temporal alignment under the same playback conditions. Reflex events are marked in peach and coral, and jaw opening interventions are marked in blue.



香港科技大學(廣州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Thank you!

yhuang849@connect.hkust-gz.edu.cn