

EarLog: Opportunistic Ear Physiology Sensing toward Personalized Hearing Risk Awareness

Yongzhi Huang 
huangyongzhi@email.szu.edu.cn

Jiayi Zhao 
jzhao238@connect.hkust-gz.edu.cn

Kaishun Wu 
wuks@hkust-gz.edu.cn

Data Science and Analytics Thrust (DSA), Information Hub (INF)
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Abstract

Many headphone users risk irreversible hearing damage during high-load listening, yet current protection still relies on sound-pressure limits. Clinical tests detect damage after it becomes measurable, leaving the risk-accumulating process unobserved. The acoustic reflex (AR), an involuntary response to high acoustic load, marks when the ear enters a protective state. AR traces are weak, and existing measurements require dedicated probes and controlled conditions that are difficult to maintain during everyday audio. We present *EarLog*, which turns playback, the dominant interference, into an opportunistic probe. *EarLog* introduces content-conditioned residual sensing to suppress variations caused by nonstationary content, individual ear acoustics, and device response. It identifies weak AR through loudness-dependent changes and localized spectral irregularities in the residual. A resonance-constrained sparse finite impulse response (FIR) model captures the short-window response. Dual-microphone observations, mild-distortion recovery, and cross-session reference updates keep detection usable under low signal-to-noise ratio (SNR), fit changes, and environmental drift. *EarLog* runs in real time on an embedded controller using a low-cost dual-microphone add-on. We evaluate it with 81 participants, 16 headphone models, and more than 400 hours of recordings. *EarLog* achieves 99.0% sensitivity and 98.7% specificity on a clinically referenced benchmark. *EarLog* records listener-specific AR events during everyday listening, providing data for future longitudinal studies of hearing risk.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Applied computing** → **Consumer health**; • **Hardware** → *Sensor devices and platforms*.

Keywords

Earable computing, acoustic reflex, in-ear acoustic sensing, opportunistic sensing, physiological-load logging, personalized hearing risk awareness, commodity headphones

ACM Reference Format:

Yongzhi Huang, Jiayi Zhao, and Kaishun Wu. 2026. *EarLog: Opportunistic Ear Physiology Sensing toward Personalized Hearing Risk Awareness*. In *The 32nd Annual International Conference on Mobile Computing and Networking (MobiCom '26)*, October 26–30, 2026, Austin, TX, USA. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3795866.3844456>

1 Introduction

More than one billion young people are at risk of irreversible hearing damage from unsafe headphone listening [51]. Current protection covers two endpoints: device-side control of sound pressure level (SPL) and listening exposure [14, 34, 47, 51], and clinical hearing tests that detect damage after it occurs [21, 26, 36, 37, 49]. Between these endpoints, how risk accumulates during high-load listening remains a blind spot in hearing protection.

This blind spot matters because hearing risk varies across listeners and from moment to moment. Physiological tolerance differs widely, and brief high-energy peaks in modern audio can push the auditory system into overload even when the volume setting appears safe [9, 18, 40]. The ear signals these moments through its own protective response. When the load reaches a listener-specific threshold, the brainstem contracts the stapedius muscle in the middle ear. This contraction stiffens the sound-conduction path and reduces the acoustic energy transmitted to the cochlea [46, 48]. This involuntary protective response, known as the acoustic reflex



This work is licensed under a Creative Commons Attribution 4.0 International License.

MobiCom '26, Austin, TX, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2505-0/26/10

<https://doi.org/10.1145/3795866.3844456>

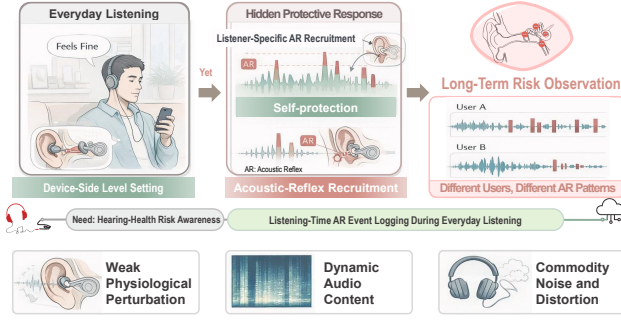


Figure 1: EarLog overview. EarLog turns everyday playback into an in-ear probe and records listener-specific AR events that remain invisible to device-side volume rules.

(AR), also leaves a minute perturbation in ear-canal reflections that a headphone can capture. These activations can occur even at ordinary listening levels. Recording these AR activations during everyday listening would allow hearing protection to track how each ear responds over time.

However, this stable measurement setup is unavailable during everyday listening on commodity headphones. Clinical systems use a dedicated probe microphone, a fixed probe tone, a separate elicitor, and a stationary subject. Together, these controls provide a stable acoustic environment for measuring small changes in admittance [19, 32, 45]. Everyday listening provides far less stability and can both trigger AR and mask its minute trace. Changes driven by its dynamic content can be much larger than the reflex trace and shift as the content evolves. Ear-canal multipath, headphone frequency response, and fit changes introduce additional variations at the same scale as AR. Reliable clinical AR labels are also scarce in this scenario, making it difficult to train an end-to-end model directly from raw signals.

EarLog turns playback, the dominant interference, into the probe. It uses the known input to build a no-AR baseline for the expected in-ear response. EarLog then adapts the baseline using playback energy in each frequency band, allowing the estimate to follow changes in content. Because AR typically lasts 100–200 ms, EarLog estimates the response within a short window to avoid temporal smearing of this transient response [2, 23, 38]. EarLog uses ear-canal resonances to constrain the plausible delays of reflected sound. A sparse finite impulse response (FIR) model then captures the in-ear response within the short window. With this baseline, EarLog applies content-conditioned residual sensing. The measured response is compared with the baseline to suppress variations caused by nonstationary content, individual ear acoustics, and device response. Loudness-dependent changes and localized spectral irregularities in the residual

serve as evidence of the weak AR trace. Fit changes, motion, leakage, and measurement saturation can still distort the residual. EarLog therefore uses dual-microphone observations to reject unreliable windows, recovers from mild distortion, and updates the baseline across sessions. These designs let EarLog record AR events during everyday listening without interrupting playback, forming a longitudinal record of each ear’s protective response (Fig. 1).

We implement EarLog as an approximately \$5 dual-microphone add-on. The module follows the inward- and outward-facing microphone layout of a simplified active-noise-cancellation (ANC) front end, and the full pipeline runs in real time on an embedded controller. We evaluate EarLog with 81 participants, 16 headphone models, and more than 400 hours of recordings. On a clinically referenced benchmark, EarLog achieves 99.0% sensitivity and 98.7% specificity.

This paper makes three contributions.

A new sensing paradigm for everyday hearing protection. EarLog moves observation from device output and episodic hearing screening to AR events during everyday listening. These event logs provide a process-level record that could support individualized, long-term hearing-protection research.

Opportunistic sensing under dynamic playback. EarLog turns the user’s playback into the probe and introduces content-conditioned residual sensing to separate weak AR traces from observations shaped jointly by nonstationary content, individual ear acoustics, and device response. This formulation makes AR during dynamic playback a measurable earable sensing target even though ordinary playback does not provide per-window AR labels.

A complete sensing system for commodity headphones. EarLog combines short-window modeling constrained by ear-canal resonance with dual-microphone observation, recovery from mild distortion, and cross-session reference updates. These components keep weak AR detection usable under low SNR, dynamic content, fit changes, and environmental drift. The system uses low-cost commercial components and is validated across users, headphone models, everyday behaviors, and cross-day conditions.

2 Related Work

At a broader level, commodity mobile sensing has used vibration and optical responses to infer latent physical states, including liquid identity and gradual quality changes [27–30]. In mobile health, pulmonary-function systems have extended spirometry from clinical instruments to phones and earphones, using breathing, coughing, and speech to

reduce testing burden [8, 22, 43, 52–55]. Earphones and hearables have also become important physiological sensing interfaces. Existing systems use in-ear microphones, active-noise-cancellation (ANC) paths, and repurposed headphone transducers to sense heart rate, heart sounds, respiration, and other weak physiological signals [3, 6, 7, 15–17, 44]. Related commodity-device research has expanded to voluntary middle-ear activity, ear-condition screening, cochlear responses, otoacoustic emissions, and tympanometry [10–13, 33, 41]. Software-defined and interference-aware platforms further support acoustic sensing across devices and in the presence of competing sound [4, 5]. These studies establish earphones as practical interfaces for periodic, repeatable, or task-elicited physiological sensing. Most closely related, EarCSI reconstructs user-specific ear-canal channels from passive broadband audio for fine-grained sensing of tympanic-membrane changes [31]. *EarLog* targets weak, brief, and unpredictable non-periodic events that arise opportunistically during arbitrary dynamic playback, where periodic structure, preset tasks, clear raw observables, and scalable label collection offer limited help.

Clinical AR measurement, by contrast, detects the reflex with a dedicated probe microphone, a fixed probe tone, a separate elicitor, and stable measurement conditions [19, 32, 45]. These conditions disappear during arbitrary playback on commodity headphones. The next section shows why clinical AR criteria and direct classification of raw in-ear signals do not readily transfer to this setting.

3 Problem Formulation

3.1 Why Direct AR Sensing Fails

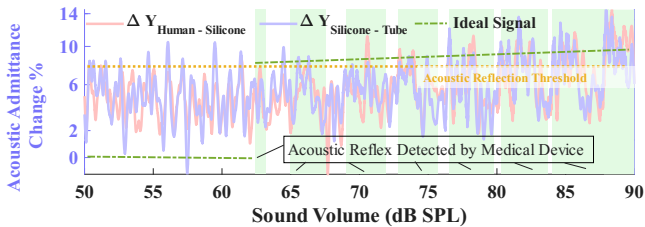


Figure 2: The clinical AR criterion does not transfer to commodity headphones. The contrast between two passive fixtures crosses the criterion although neither can generate AR.

Clinical acoustic-reflex tests infer AR from changes in absolute acoustic admittance [35, 39]. This criterion does not transfer to commodity headphones. In Fig. 2, the silicone artificial ear and IEC 60318-4 ear simulator cannot generate AR, yet their response contrast varies with sound pressure level and crosses the Titan clinical criterion. The change must therefore come from device response and acoustic propagation. Clinical systems avoid this ambiguity with

high-precision instrumentation, a dedicated probe, and fixed excitation. Commodity headphones lack this stable reference, so changes in absolute admittance or reflection cannot provide direct evidence of AR. Appendix A.2 describes the control setup and contrast calculation.

A fixed threshold calibrated for commodity headphones also fails because the in-ear response changes with playback even when AR is absent. As Fig. 3 shows, the Do and Re notes produce different reflection curves in the same ear canal under nominally identical conditions. The inward microphone receives overlapping ear-canal propagation paths whose contributions depend on the current spectrum and playback level. A fixed threshold or one-time reference therefore becomes stale as the content changes. *EarLog* needs a content-conditioned baseline that follows the current playback. Appendix A.3 shows how playback content and ear-canal multipath jointly shape the observation.

3.2 *EarLog*’s Sensing Core

Dynamic playback shifts the baseline, but it does not erase the local structure that AR leaves in the response. Figure 4 shows that, in the absence of AR, response changes across adjacent loudness levels remain relatively smooth. When AR engages, some bands exhibit localized gain enhancement, slight shifts in spectral peaks, and irregular gain changes across adjacent levels. Removing the expected response for the current content reveals these local deviations that are otherwise masked by the much stronger playback-driven response. This observation motivates content-conditioned residual sensing. *EarLog* predicts the no-AR response to the current playback, compares this baseline with the measured response, and identifies weak AR from loudness-dependent changes and localized spectral irregularities in the residual. Sections 4.1 and 4.2 describe the baseline model and residual features, respectively.

Figure 5 summarizes *EarLog*’s window-level pipeline. The system first checks whether the playback and dual-microphone observations support reliable AR detection. Windows that fail this check are logged as *invalid* or *overload-only* and excluded from AR detection. For each valid window, *EarLog* predicts the no-AR baseline response for the current playback (Sec. 4.1). When the observation contains mild clipping or recoverable noise, this baseline also guides signal repair (Sec. 4.3). *EarLog* then compares the measured response with the baseline to form the residual (Sec. 4.1.3) and examines the residual for loudness-dependent features and localized spectral irregularities (Sec. 4.2). A lightweight verifier outputs *valid AR* or *valid no-AR*, and the system records the result. Across sessions, *EarLog* checks baseline consistency and updates the reference when fit or environmental conditions shift (Sec. 4.4). The resulting log records

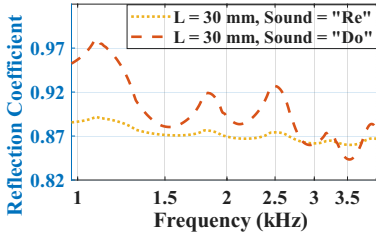
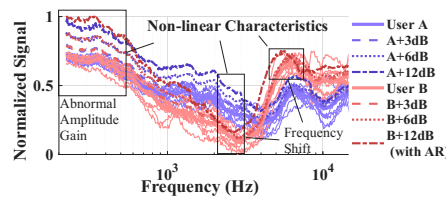
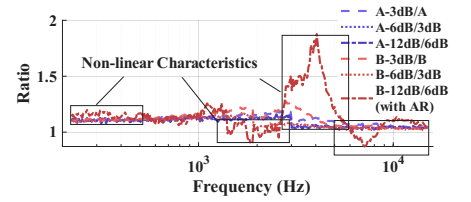


Figure 3: Reflection coefficient curves for two notes (Do and Re) in the same ear canal.



(a) Responses at different sound levels.



(b) Signal enhancement ratio.

Figure 4: Frequency shifts and nonlinear effects associated with acoustic reflex.

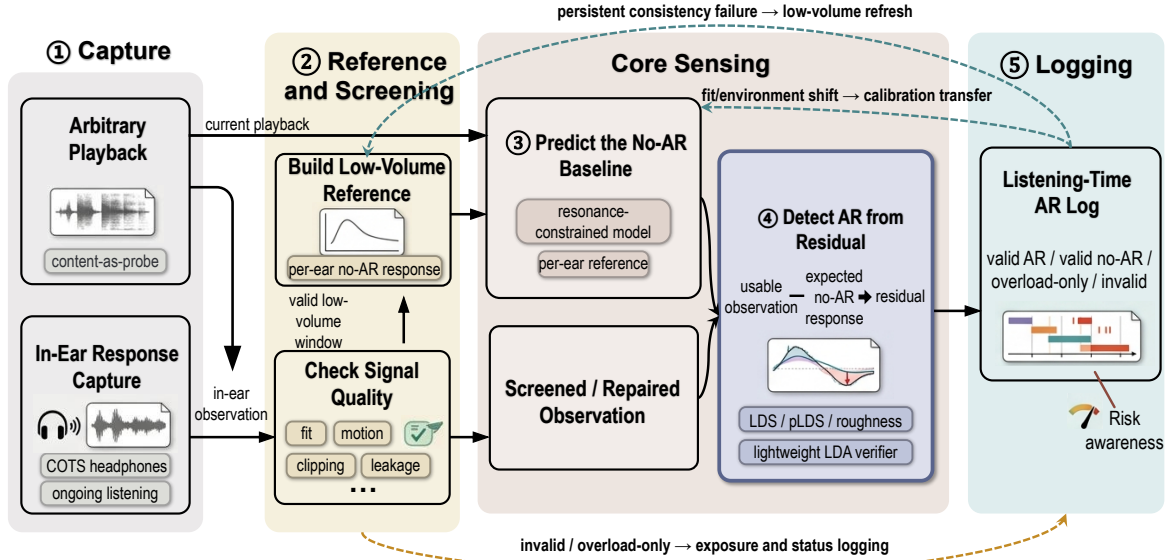


Figure 5: *EarLog* workflow. *EarLog* screens each window, predicts the no-AR baseline for the current playback, detects AR from the residual, and records the resulting window state.

AR events during playback together with each window's sensing status.

4 *EarLog* Design

4.1 Predicting the No-AR Baseline

Section 3 shows that the absolute in-ear response alone does not provide reliable evidence of AR during dynamic playback. *EarLog* therefore asks what response the same ear would produce under the current playback if AR were absent. This expected response defines the no-AR baseline. Because the current window may already contain AR, *EarLog* learns the no-AR response from low-volume segments and predicts the baseline for the current playback condition.

4.1.1 Building the Low-Volume Reference. Low-volume playback provides the reference needed by *EarLog*. In this regime, non-silent audio still excites the in-ear acoustic response,

while the response follows the smooth no-AR pattern illustrated in Fig. 4. *EarLog* therefore builds its per-ear no-AR reference from non-silent low-volume segments without requiring a clinical WB-ART measurement. Silent blocks provide an independent negative control, but without excitation, they cannot be used to identify the acoustic response.

The digital playback samples are known, but the in-ear observation is not a copy of the source. The headphone transducer, insertion depth, seal, and ear-canal multipath transform the sound before the inward-facing microphone records it. *EarLog* pairs the known playback with the corresponding inward-microphone response from the same ear and device configuration. Over a short, low-volume segment, this input–output relation can be approximated as locally linear.

A single nominal volume is insufficient for grouping these calibration segments. Two audio windows with the same

overall level can distribute their energy across different frequency bands and produce different ear-canal responses. *EarLog* therefore computes a playback-energy vector over B analysis bands: $\mathbf{V} = [E_1, \dots, E_B]$, where E_b is the playback energy in band b , and B is the number of analysis bands. Windows with similar $\mathbf{V}$ enter the same band-energy bucket. Each bucket represents a local no-AR operating condition under similar content and gain, instead of a single reference intended to cover all playback.

We use V to denote the bucket associated with one such band-energy pattern. Within that bucket, *EarLog* models the local no-AR input–output relation with a finite impulse response (FIR):

$$y^{(V)}(n) = \sum_{k=0}^{M-1} h_k^{(V)} x(n-k) + \eta(n).$$

Here, $x(n)$ is the n -th digital playback sample, $y^{(V)}(n)$ is the inward-microphone observation, $h_k^{(V)}$ is the k -th FIR coefficient for bucket V , M is the FIR length, and $\eta(n)$ is additive noise. *EarLog* estimates the bucket-specific coefficients with the least mean squares (LMS) algorithm. This step identifies the local no-AR response. AR detection occurs later. Bucketing reduces content variation, but a short calibration segment still cannot reliably estimate a generic high-order FIR with many free coefficients. *EarLog* next constrains the tap locations with ear-canal physics.

4.1.2 Resonance-Constrained Sparse FIR Modeling. The response model must be estimated within a short window. A longer window provides more samples for FIR estimation, but it also spreads a short AR transition across unrelated content and causes temporal smearing [2, 23, 38]. A generic high-order FIR therefore has too many free coefficients for the available window. Ear-canal reflections provide the missing constraint because their main arrival times occupy a limited set of propagation delays.

EarLog approximates the ear canal as a quarter-wave resonator. Its lowest credible resonance f_0 provides a coarse effective propagation length $L \simeq c/(4f_0)$, where c is the speed of sound. The corresponding one-way delay is $t_0 = L/c$. *EarLog* restricts nonzero FIR taps to narrow neighborhoods around the direct arrival and successive round-trip delays $k_i \simeq 2it_0f_s$, where f_s is the sampling rate and i is the reflection order. The effective length serves as a structural prior on tap locations. Constraining the FIR support preserves the main propagation components while reducing the number of coefficients that must be estimated in each short window.

Each band-energy bucket maintains its own LMS state. Let $e^{(V)}(n)$ be the difference between the inward-microphone sample and the bucket-specific FIR prediction at sample n .

For a tap index k inside the resonance-constrained support $\mathcal{K}_V$, the online update is

$$h_k^{(V)}(n+1) = h_k^{(V)}(n) + \mu e^{(V)}(n)x(n-k), \quad k \in \mathcal{K}_V, \quad (1)$$

where μ is the LMS step size. Buckets update independently, and coefficients outside $\mathcal{K}_V$ remain zero. The support constraint therefore limits adaptation to propagation delays supported by the ear-canal model rather than allowing an unconstrained filter to absorb the short-window observation.

After sparse FIR identification, a gammatone filterbank maps the response into the same B analysis bands used by the playback-energy vector. The band-wise baseline parameters and residual features in Sec. 4.2 use this same frequency representation. The shared band structure connects low-volume calibration, baseline prediction, and residual-based AR detection without repeatedly converting between incompatible representations.

4.1.3 Predicting the Baseline for Current Playback. The low-volume buckets cover only part of the content and energy states encountered during everyday playback. Once playback leaves those states, the measured response may already contain AR and can no longer serve as its own no-AR reference. *EarLog* therefore predicts the baseline for the current playback from the responses learned during low-volume calibration.

For each analysis band b , *EarLog* represents the bucket response with a baseline parameter $\theta_b(V)$. Here, $\theta_b(V)$ is the log-magnitude of the identified no-AR transfer response in band b for playback state V . *EarLog* assumes that this parameter varies smoothly across nearby playback conditions and fits a low-order model:

$$\theta_b(V) = \sum_{m=0}^P a_{b,m} \phi_m(V).$$

Here, $\phi_m(V)$ is the m -th basis function, $a_{b,m}$ is its fitted coefficient, and P is the model order. This low-order relation avoids a separate calibration for every content–level combination and limits the variation introduced by the finite low-volume dataset.

At runtime, *EarLog* computes the current playback state V_t and reconstructs the corresponding no-AR baseline $H_{\text{baseline}}(f, V_t)$. Comparing it with the observed in-ear response $H_{\text{obs}}(f, t)$ yields the log-magnitude residual:

$$\Delta \log H(f, t) = \log |H_{\text{obs}}(f, t)| - \log |H_{\text{baseline}}(f, V_t)|.$$

This subtraction removes most of the variation explained by the current playback and the modeled in-ear response. The residual may still contain baseline mismatch, fit drift, leakage, noise, or distortion. The next subsection therefore tests whether the residual shows the loudness dependence and localized spectral structure associated with AR.

Removing the physics-constrained no-AR baseline reduces accuracy from 98.71% to 90.35%, a loss of 8.36 percentage points (Table 1). The 8.36-point loss confirms the role of baseline prediction in residual-based AR detection. Appendix C gives bucketed identification details and runtime construction, while Appendix E gives the resonance and sparse-tap derivation.

4.2 Detecting AR from the Residual

Subtracting the no-AR baseline removes most of the response explained by the current playback, but a nonzero residual alone is insufficient evidence of AR. Baseline mismatch, fit drift, cross-band content coupling, and recovery artifacts can all produce a nonzero residual. The observations in Sec. 3.2 suggest two tests that are more specific to AR. When AR occurs, the residual is expected to become loudness-dependent in a subset of bands and to exhibit a localized spectral irregularity rather than a smooth shift across frequency. *EarLog* uses three residual cues for these tests. The Loudness Differential Signature (LDS) measures how the residual changes with loudness, the partial Loudness Differential Signature (pLDS) accounts for cross-band contributions before estimating the local slope, and local spectral roughness describes whether the remaining change is localized in frequency.

4.2.1 Loudness Differential Signature. The first cue tests whether the residual varies with playback loudness. Let $r_b(t)$ denote the residual log response in analysis band b , derived from $\Delta \log H(f, t)$. Let $L_{s,b}(t)$ denote the source-side log loudness in the same band. The Loudness Differential Signature (LDS) is the local sensitivity

$$\text{LDS}_b = \frac{\partial r_b(t)}{\partial L_{s,b}(t)}.$$

The no-AR baseline already accounts for the ordinary level-dependent response expected from the current playback. Without AR, the residual therefore remains approximately flat with respect to source-side loudness, and LDS_b stays close to zero. When AR engages, stapedius contraction changes middle-ear transmission. Near ear-canal resonances, this transmission change appears as a new local slope in a subset of bands. LDS measures the loudness-dependent response that remains after baseline subtraction, without relying on the absolute response magnitude used by clinical thresholds.

A slope measured in one band, however, may include variation from neighboring bands. Dynamic audio excites many frequencies at once, and harmonics, ear-canal multipath, and filterbank leakage couple observations across bands. *EarLog* therefore accounts for these cross-band contributions before interpreting the local slope.

4.2.2 Removing Cross-Band Coupling with pLDS. To separate the band-local response from cross-band content variation, let B denote the number of analysis bands. For frames $t = t_1, \dots, t_W$ in a short estimation window, *EarLog* jointly models the residuals as

$$\mathbf{r}(t) = \mathbf{A}\mathbf{L}_s(t) + \mathbf{c} + \boldsymbol{\epsilon}(t).$$

Here, $\mathbf{r}(t) \in \mathbb{R}^B$ is the residual log-response vector, $\mathbf{L}_s(t) \in \mathbb{R}^B$ is the source-side loudness vector, $\mathbf{A} \in \mathbb{R}^{B \times B}$ is the cross-band coefficient matrix, $\mathbf{c}$ is an intercept, and $\boldsymbol{\epsilon}(t)$ is the remaining modeling error. Dynamic playback often produces strongly correlated loudness trajectories across bands, so *EarLog* estimates the coefficients with

$$(\hat{\mathbf{A}}, \hat{\mathbf{c}}) = \arg \min_{\mathbf{A}, \mathbf{c}} \sum_{t=t_1}^{t_W} \|\mathbf{r}(t) - \mathbf{A}\mathbf{L}_s(t) - \mathbf{c}\|_2^2 + \lambda_A \|\mathbf{A}\|_F^2, \quad (2)$$

where $\lambda_A \geq 0$ controls the ridge penalty.

Each entry $\hat{A}_{ij}$ describes how source-side loudness in band j contributes to the residual in band i . The off-diagonal coefficients account for cross-band contributions, while the diagonal coefficient estimates the band-local loudness-to-residual relation after conditioning on the other excited bands, giving $\text{pLDS}_b = \hat{A}_{bb}$. pLDS therefore reduces the cross-band contamination that makes a single-band LDS sensitive to rich broadband content unrelated to the band-local AR change. Ridge regularization stabilizes the joint estimate when source-band loudness trajectories are correlated. Because the current playback does not reliably excite every band, *EarLog* computes pLDS only where playback energy is sufficient. This support rule prevents leakage, noise, or numerical instability in weakly excited bands from being interpreted as a local loudness response.

pLDS estimates the band-local loudness dependence, but a local slope can still arise from smooth gain drift, mild fit variation, or baseline extrapolation error. These effects differ from AR mainly in how they vary across frequency.

4.2.3 Local Spectral Roughness. AR most strongly perturbs a subset of bands near ear-canal resonances. The residual can therefore contain localized gain changes, small spectral-peak shifts, or abrupt differences between adjacent bands, whereas device gain drift and mild fit changes more often vary smoothly across frequency. For each window, *EarLog* robustly normalizes the residual as

$$\tilde{r}_b(t) = \frac{r_b(t) - \text{median}_{b'} r_{b'}(t)}{\text{MAD}_{b'}(\mathbf{r}(t)) + \epsilon}, \quad \epsilon > 0. \quad (3)$$

The median removes a global offset. The median absolute deviation (MAD), computed across the B bands around that median, supplies a robust within-window scale; ϵ keeps the normalization defined when the residual is nearly flat. This centering and scaling prevents a broadband offset or uniform

gain change from appearing as localized structure. *EarLog* then computes

$$\begin{aligned} \text{TV}(t) &= \sum_{b=1}^{B-1} |\tilde{r}_{b+1}(t) - \tilde{r}_b(t)|, \\ \text{Curv}(t) &= \sum_{b=2}^{B-1} |\tilde{r}_{b+1}(t) - 2\tilde{r}_b(t) + \tilde{r}_{b-1}(t)|, \\ J_{\max}(t) &= \max_{1 \leq b < B} |\tilde{r}_{b+1}(t) - \tilde{r}_b(t)|. \end{aligned} \quad (4)$$

The three descriptors capture complementary aspects of the normalized residual. Total variation accumulates adjacent-band changes, curvature responds to local changes in spectral slope, and $J_{\max}$ isolates the strongest adjacent-band discontinuity. Smooth gain or fit drift generally keeps all three small, whereas a localized AR perturbation can increase them. LDS and pLDS test whether the residual changes with loudness. The roughness descriptors test whether that change is concentrated in a subset of bands rather than appearing as a smooth global shift.

4.2.4 Lightweight AR Verifier. Each valid window yields

$$\mathbf{z}(t) = [\text{LDS}(t)^\top, \text{pLDS}(t)^\top, \text{TV}(t), \text{Curv}(t), J_{\max}(t)]^\top. \quad (5)$$

The implementation normalizes $\mathbf{z}(t)$ using z-score statistics computed from the training participants. A shrinkage-regularized linear discriminant analysis (LDA) model learns $\mathbf{w}$ and b from the clinical labels and produces

$$\begin{aligned} s(t) &= \mathbf{w}^\top \mathbf{z}(t) + b, \\ \bar{s}(t) &= \lambda_s \bar{s}(t-1) + (1 - \lambda_s) s(t), \quad \lambda_s \in [0, 1). \end{aligned} \quad (6)$$

Covariance shrinkage stabilizes the discriminant under the available labels. Score smoothing suppresses frame-level jitter; it does not replace the residual evidence in $\mathbf{z}(t)$.

The z-score statistics, $\mathbf{w}$, and b are learned from the training participants and remain fixed during online inference. Runtime processing therefore consists of constructing $\mathbf{z}(t)$, evaluating the linear score, and updating the smoothed score for eligible windows. This separation prevents the verifier from adapting its decision boundary to a transient residual or to the current user's event history.

A two-threshold hysteresis rule converts $\bar{s}(t)$ into the window state. *EarLog* enters the AR state only after the score stays above the higher threshold for K_+ consecutive valid windows, and exits only after it stays below the lower threshold for K_- consecutive valid windows. Invalid and overload-only windows receive no AR decision and do not advance the AR state or event count, although they remain in the listening-exposure and signal-quality log. Consecutive valid windows assigned to the AR state are merged into one event.

The ablation results separate the roles of pLDS and roughness. Replacing the multiband pLDS representation with single-band LDS reduces accuracy from 98.71% to 86.55%, a loss of 12.16 percentage points. Removing the roughness descriptors and roughness-based gating reduces accuracy to 92.20%, a loss of 6.51 percentage points (Table 1). The 12.16-point loss reflects the value of cross-band decoupling under dynamic broadband playback. The 6.51-point loss reflects the additional separation between localized AR structure and smooth drift.

Appendix B derives the residual relation and records pLDS window construction, regularization details, training-time feature normalization, and the hysteresis parameters. The residual cues above assume that the playback and microphone observations remain physically meaningful. Large motion, leakage, saturation, and environmental noise can invalidate that assumption or distort the residual. The next subsection explains how *EarLog* rejects unusable windows and repairs mild artifacts before verification.

4.3 Screening and Repairing In-Ear Observations

Sections 4.1 and 4.2 assume that the inward-facing microphone still captures a meaningful response to the known playback under the current headphone fit and ear-canal condition. Everyday disturbances violate this assumption in two different ways. Severe leakage, large motion, abrupt fit changes, and hard saturation can break the playback-response relation. Mild clipping and localized background noise corrupt only part of the response. As shown in Fig. 5, *EarLog* routes these cases through screening and repair before computing residual features. Unusable windows are excluded from the AR-detection path and are logged as *invalid* or *overload-only*. Recoverable windows are repaired using the current no-AR baseline as a constraint before residual detection. The repair is deliberately conservative because the verifier relies on loudness-dependent changes and local spectral roughness. Preprocessing must preserve these patterns without introducing artifacts that resemble them.

4.3.1 Checking Window Validity. *EarLog* begins by checking whether the current window can be compared reliably with its no-AR baseline. Four complementary indicators screen for different failure modes. The clipped-sample ratio $r_{\text{clip}}(t)$ detects saturation in the microphone or analog front end. The leakage ratio $\lambda_{\text{leak}}(t)$ compares outer- and inner-microphone energy to reveal seal loss or abnormal leakage of external sound. The source-coherence score $c_{\text{src}}(t)$ checks whether the inward observation still follows the known playback, exposing large motion or interference unrelated to the playback source. The baseline-consistency score $m_{\text{fit}}(t)$ checks whether the current response is consistent with the recent

no-AR baseline under the present fit. These indicators determine whether the captured playback–response relation remains interpretable.

A window remains eligible for further processing only while all four indicators stay within their calibrated ranges. To prevent rapid switching near a boundary, the gate changes state only after either a violation or a recovery persists across consecutive windows. Windows dominated by hard front-end saturation are logged as *overload-only*, because the playback–response relation has been destroyed and cannot constrain a reconstruction. Other windows that cannot be interpreted reliably are logged as *invalid*. Windows in either state are excluded from AR detection and do not contribute to the AR event count, although they remain in the listening-exposure and signal-quality records described in Sec. 6. A transient baseline inconsistency causes the current window to be rejected. Persistent inconsistency is handled in Sec. 4.4, which adapts or refreshes the no-AR reference.

The gate removes observations whose input–output relation has been lost. Only locally corrupted windows that retain an interpretable playback–response relation are repaired. For mild clipping, *EarLog* uses the current no-AR baseline to reconstruct the dominant linear response and applies residual compensation only in bands whose reconstruction error remains elevated. For background interference, it suppresses only components with low correlation to the known playback. Both operations are constrained to preserve the local structure used by LDS, pLDS, and roughness, and candidate AR windows never update the repair parameters. Removing software artifact/interference mitigation reduces accuracy from 98.71% to 75.90%, a loss of 22.81 percentage points (Table 1). These mechanisms address short-lived corruption within individual windows; Sec. 4.4 instead handles path drift that persists across windows.

4.4 Maintaining the Baseline over Time

The preceding sections handle content variation and transient observation failures. A slower source of error remains: drift in the no-AR baseline. Refitting changes insertion depth and seal, while humidity and prolonged wear gradually alter the acoustic response between the headphone and the ear. A stale baseline can then leave a persistent residual offset even when every window passes the validity checks. *EarLog* accounts for this drift in the no-AR response model and keeps the AR verifier unchanged.

4.4.1 Calibration Transfer. Changes in fit, seal, and humidity do not affect all frequency bands independently. Across paired low-volume no-AR observations, their effects recur as a small set of band-wise patterns. *EarLog* extracts the leading patterns and forms a compact drift basis Ψ . For playback

state V_t and current fit or environmental condition u_t , the adapted baseline parameters are

$$\theta(V_t, u_t) = \theta(V_t) + \Psi \alpha_t. \quad (7)$$

Here, $\theta(V_t)$ is the baseline predicted for the current playback, the columns of Ψ describe recurring band-wise drift patterns, and α_t gives the current shift in that basis.

Let $\mathbf{q}_t \in \mathbb{R}^B$ be the band-wise average log-transfer observation from the current valid low-volume support windows, expressed in the same representation as θ . Let $\mathcal{B}$ be the stored low-volume baseline buckets, and let $\mathbf{P}_{\mathcal{A}}$ select the stable anchor bands. *EarLog* first matches the current support to a stored reference:

$$\hat{V}_t = \arg \min_{V \in \mathcal{B}} \|\mathbf{P}_{\mathcal{A}} [\mathbf{q}_t - \theta(V)]\|_2^2. \quad (8)$$

After matching, *EarLog* estimates the drift coefficients from the remaining band-wise mismatch:

$$\hat{\alpha}_t = \arg \min_{\alpha} \|\mathbf{W} [\mathbf{q}_t - \theta(\hat{V}_t) - \Psi \alpha]\|_2^2 + \lambda_{\alpha} \|\alpha\|_2^2, \quad (9)$$

where $\mathbf{W} = \mathbf{I} - \mathbf{P}_{\mathcal{A}}$ selects the non-anchor bands that carry most of the drift, and λ_{α} controls the ridge penalty. The matched bucket and $\hat{\alpha}_t$ provide the stored reference and current drift term in Eq. 7; the resulting baseline replaces the stale reference used by Secs. 4.1.3 and 4.2.

Each transfer is estimated directly against the stored calibration reference. Previous transfer coefficients do not enter the next objective, so transfer error does not accumulate recursively across sessions. Several factors may act together, such as insertion depth, humidity, and seal evolution; their combined effect remains correctable while its band-wise displacement lies within the learned drift basis. Appendix D describes paired-difference construction, the drift and anchor sets, support-window formation, frequency-domain transport, and consistency-triggered refresh.

4.4.2 Consistency Check and Low-Volume Refresh. *EarLog* retains a transferred baseline only while it remains consistent with current low-volume observations. On valid low-volume support windows, *EarLog* monitors residual energy and the low-volume LDS, pLDS, and roughness statistics. A single mismatch causes the validity gate in Sec. 4.3 to reject the current window. A mismatch that persists across consecutive windows indicates that the current shift is no longer explained by the learned drift patterns.

When this occurs, *EarLog* pauses AR detection and refreshes the band-energy buckets and no-AR baseline using naturally occurring low-volume playback. Listening exposure and signal-quality status continue to be recorded during the refresh. AR detection resumes after the new baseline passes the same consistency check.

The seven-day evaluation in Sec. 7.4.3 quantifies how often this fallback is required and how much low-volume playback

it consumes. The results show that calibration transfer absorbs most day-to-day drift and that refresh is needed only occasionally.

5 Implementation

Figure 6 shows the *EarLog* hardware and its placement. The prototype is a 12 mm × 9 mm dual-microphone add-on with a bill-of-materials cost of approximately \$5. When mounted in an in-ear assembly, the add-on sits between the headphone driver and the ear-canal entrance. Two analog microelectromechanical systems (MEMS) microphones (ICS-40720) face opposite directions. The inner microphone records the response shaped by playback, headphone fit, and the ear canal, supporting no-AR baseline prediction and residual detection. The outer microphone provides the environmental and leakage reference used by the validity checks in Sec. 4.3. The two channels are conditioned separately and digitized as independent 16-bit streams at 48 kHz. To reduce slow temperature-dependent gain drift in the microphone and analog front end, *EarLog* places an NTC thermistor in the op-amp feedback path. Calibration transfer handles the remaining frequency-dependent changes caused by fit and environment. Appendix G.1 details the circuit.

A fixed-point implementation on the Arduino Portenta H7 runs the complete pipeline at 1.05× real time. Relative to audio playback alone, the *EarLog* processing and sensing path adds 16 mW on average. This corresponds to approximately 0.016 Wh for one hour of listening, or about 8% of a representative 0.20 Wh earbud battery. Appendix G separates this incremental overhead from whole-board development-platform power and reports the fallback-refresh measurement.

The prototype provides the four capabilities required by *EarLog* in a uniform form across headphone models: an inward reflection microphone, an outward environmental reference, synchronized access to playback and microphone streams, and stable local signal processing. Many active-noise-cancellation (ANC) headphones already contain comparable inward and outward microphones together with on-board DSP [16]. A product implementation can reuse these components when synchronized raw access is available. For headphones without an accessible reference path, the add-on supplies the same observations and enables controlled evaluation across the 16 headphone models reported in this paper.

6 Experimental Setup

Participant-disjoint evaluation. All main results use leave-one-subject-out (LOSO) evaluation. Every window from one participant remains in the same split, so the test participant contributes no window to supervised training. The short

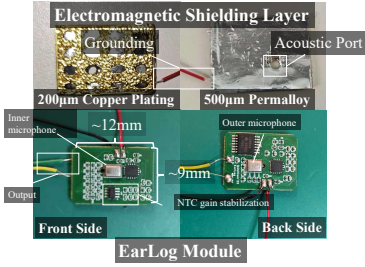
low-volume segment used to build that participant’s no-AR baseline is an inference-time, per-ear calibration step. All LDA parameters, feature-normalization statistics, and other supervised parameters are learned from the remaining participants. We report sensitivity, specificity, precision, recall, F1 score, and area under the precision-recall curve (AUPRC). Confidence intervals, when reported, use participant-level bootstrapping rather than window-level resampling (Appendix H.4).

Clinical reference and synchronization. During synchronized trials, the clinical reference neither enters nor stimulates the *EarLog* ear. Each participant first undergoes a separate standard examination with an Interacoustics Titan tympanometer. This examination measures the evaluation ear’s wideband acoustic-reflex threshold (WB-ART), the lowest wideband stimulus level at which Titan detects reflex activation [19, 32]. In our evaluation, the measured WB-ART is used to schedule clinically verified sub-threshold control blocks and to stratify the cohort in Sec. 7.1. Its numerical value is not provided to *EarLog*’s baseline model or AR verifier.

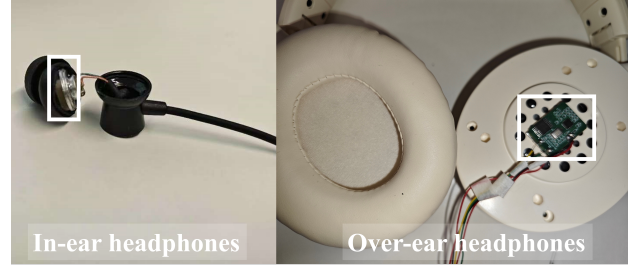
In subsequent trials, the headphone stimulates the *EarLog*-instrumented target ear while Titan monitors time-varying admittance in the contralateral ear. The acoustic-reflex pathway is bilateral, so sufficient stimulation of the target ear can produce a synchronized reflex response in the opposite ear. The Titan trace therefore supplies an independent time-resolved reference without contributing to the headphone-side acoustic observation. The Titan and *EarLog* signals are recorded as separate streams and aligned to the stimulus timeline by cross-correlation. The maximum measured misalignment across participants is 10 ms. The passive Titan-versus-commodity comparison in Sec. 3.1 is a separate benchmark experiment and does not generate participant labels.

Independent reference labels. Reference labels are generated before any *EarLog* inference. An independent script reads only the Titan trace and stimulus timestamps, then marks intervals as AR-active or AR-inactive. A window is reference-positive when at least $0.5T_{\text{win}}$ overlaps a Titan-active interval. A reference-negative window lies within a Titan-inactive interval, including silent and below-threshold blocks. The overlap rule associates analysis windows with the clinical trace; physiological onset latency is not counted as part of the labeled event duration.

Both silent and non-silent below-threshold blocks provide reference-negative windows. Baseline identification uses only the non-silent blocks because input-output estimation requires a known excitation. No baseline value, residual feature, or *EarLog* prediction enters the reference-label script. The unilateral music-validation trials use the same frozen classifier, stream alignment, and synchronized contralateral



(a) Dual-microphone add-on module.



(b) In-ear and over-ear deployment geometry.

Figure 6: EarLog hardware and deployment. The $12\text{ mm} \times 9\text{ mm}$ add-on orients the inner microphone toward the ear canal and the outer microphone toward the environment; the deployment view shows how it sits inside representative in-ear and over-ear headphones.

Titan reference. The intervals shown in Fig. 12 follow this protocol.

Window states, events, and coverage. Evaluation follows the four runtime states defined in Sec. 4.2.4. Only valid windows enter the clinical confusion matrix. Consecutive valid-AR windows within one uninterrupted AR state are merged into one event, and event rates are normalized by valid listening time.

The valid-window ratio is the fraction of all recorded windows that pass the validity gate and remain eligible for AR detection. It measures sensing coverage, not classification accuracy. Invalid and overload-only windows do not contribute to TP, TN, FP, or FN counts and are never merged into AR events. They remain in the listening-exposure and signal-quality logs. Evaluation therefore reports accepted-window errors together with valid-window coverage when disturbances cause abstention.

Corpus and metric denominators. The complete corpus contains one evaluation ear from each of 81 participants, spans 16 headphone models, and exceeds 400 hours. It includes controlled measurements, robustness tests, and ordinary-listening deployments with wideband noise, piano tones, instrument timbres, and music excerpts at 50–90 dB SPL.

Clinical accuracy uses the synchronized and valid subset of this corpus. This benchmark contains 145,800 non-overlapping windows with $T_{\text{win}} = 250\text{ ms}$, corresponding to 10.1 hours of Titan-synchronized valid time. Only these valid referenced windows enter sensitivity, specificity, precision, recall, F1, and AUPRC. The unilateral music validation is analyzed separately but uses the same frozen classifier and contralateral Titan reference. Ordinary-listening records without synchronized Titan observation support standalone logging analyses and do not enter the clinical accuracy denominator. The study was approved by the institutional review board. Participant statistics, stimulus schedules, device configurations, and robustness protocols appear in Appendix H.

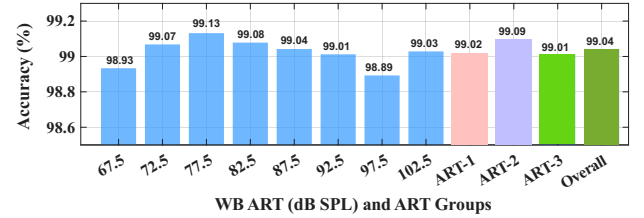


Figure 7: EarLog maintains high recognition accuracy across listener-specific WB-ART levels.

7 Evaluation

7.1 Accuracy Across Listener-Specific Reflex Thresholds

We first verify that *EarLog* operates across listeners with different physiological protection thresholds. All results use leave-one-subject-out evaluation, so no window from the test participant contributes to supervised training. We group the 81 evaluation ears by wideband acoustic-reflex threshold (WB-ART): ART-1 spans 65–75 dB SPL ($n = 16$), ART-2 spans 75–85 dB SPL ($n = 29$), and ART-3 is above 85 dB SPL ($n = 36$). Appendix H.1 gives the cohort distribution.

Figure 7 shows consistently high recognition accuracy across the measured WB-ART range. The clinically referenced benchmark contains 145,800 valid windows, including 7,883 true positives, 136,036 true negatives, 1,801 false positives, and 80 false negatives. These counts yield **99.0% sensitivity, 98.7% specificity, 81.4% precision, 89.3% F1, and 98.71% accuracy**. The 95% confidence intervals for sensitivity and specificity are ± 0.8 percentage points.

The same participant-disjoint model therefore covers ears with substantially different reflex thresholds. This result shows that *EarLog* can log listener-specific AR events across a broad range of reflex thresholds rather than operating only within one narrow threshold range. Appendix H.5 provides the interval-wise and cross-content subgroup results.

7.2 Label Efficiency of the Residual Representation

The clinical benchmark establishes detection accuracy, but it does not show whether a generic classifier could obtain the same result directly from raw spectra. We therefore train a support vector machine (SVM), a decision tree, and a lightweight deep neural network (DNN) under the same participant-disjoint supervision used by *EarLog*.

Ordinary playback provides abundant audio, but clinically confirmed AR-positive windows cannot be expanded in the same way. Repeatedly eliciting positive reflexes is constrained by participant safety and ethics. The raw spectral appearance of the same physiological event also changes with playback content, ear acoustics, headphone response, and fit. Under limited supervision, a raw-spectrum model can therefore rely on content or fit correlations. *EarLog* first removes the expected playback and in-ear response, so each clinical label describes a more focused residual difference.

Figure 8a shows that the median availability of clinically confirmed positive windows is 62 per evaluation ear. Figure 8b varies the AR-positive training-label budget and marks 67 labels as the current operating point.

At the marked *EarLog* operating point in Fig. 8c, sensitivity, specificity, precision, F1, and false-positive rate (FPR) are 98.9%, 98.7%, 81.4%, 89.3%, and 1.3%, respectively. Its pooled AUPRC is 0.807, compared with 0.147 for SVM, 0.085 for the decision tree, and 0.071 for the DNN. Figure 8d reports the cumulative distribution of per-participant AUPRC. Both views show that *EarLog* improves the precision–recall trade-off and shifts the participant distribution toward higher AUPRC. The advantage comes from the content-conditioned residual representation, which makes each clinically confirmed label more informative.

7.3 Ablation Study

We next verify that the measured accuracy follows from the coupled sensing design. The first comparison replaces the sensing formulation or removes the outward reference microphone. The second removes complete processing blocks from the residual pipeline.

The energy-only baseline uses short-time in-ear energy and a fixed threshold. It does not predict the no-AR response or use LDS, pLDS, and roughness. Figure 9 shows that this baseline reaches 47% accuracy. The single-microphone variant retains the inward channel, no-AR baseline, and residual verifier but removes the outward environmental reference. It reaches 72% accuracy. Full *EarLog* reaches 98.71%. The comparison shows that the residual formulation handles dynamic playback, while the outward microphone enables leakage, environment, and window-validity checks.

Table 1 isolates the internal design blocks. Removing artifact and interference handling reduces accuracy by 22.81 percentage points, showing the value of validity screening, clipping recovery, and selective denoising on commodity hardware. Replacing multiband pLDS with single-band LDS costs 12.16 percentage points, confirming the role of cross-band decoupling under broadband playback. Removing the physics-constrained no-AR baseline costs 8.36 percentage points, while removing roughness descriptors costs 6.51 percentage points. Together, the four ablations isolate content variation, cross-band coupling, smooth drift, and real-world interference as distinct error sources.

7.4 Generalization Across Everyday Conditions

The preceding experiments establish the core detector and its design dependencies. We further examine whether the same frozen system continues to operate as the headphone, playback content, environment, fit, behavior, and recording day vary. The tests progress from external device and content variation, to active changes in the fitted acoustic path, and finally to seven-day reference drift.

7.4.1 Devices, Ambient Noise, and Unseen Content. To verify hardware and content generalization without retraining, we keep the wideband-trained classifier fixed. We test 16 headphone models and four ambient-noise levels. Appendix H.5 extends the subgroup analysis to piano and instrument playback.

Across over-ear, in-ear, and semi-in-ear devices, *EarLog* reaches 97.0%, 100.0%, and 94.5% accuracy, respectively, as shown in Fig. 11. Table 2 shows that accuracy remains 95.15% at 70 dB ambient noise and changes by only 4.22 percentage points across the tested 45–70 dB range. The outward reference, validity gate, and selective noise mitigation preserve a usable playback–response observation across these hardware and environmental conditions.

Appendix H.5 combines wideband, piano, and instrument playback in a noisy environment. Without additional fine-tuning, accuracy reaches 95.9% for ART-1 ears, ranges from 88.3% to 92.5% across the ART-2 transition range, and reaches 87.7% for ART-3 ears, preserving the cross-threshold pattern observed in the clinically referenced benchmark.

7.4.2 Refitting and Everyday Behaviors. The preceding tests preserve the fitted headphone–ear interface. We next deliberately change that interface to verify whether fit and behavior are incorrectly written into the AR log.

Playback remains 10 dB below each ear’s WB-ART, and synchronized Titan-inactive intervals define an AR-negative condition. Every positive output in this experiment is therefore a false alarm.

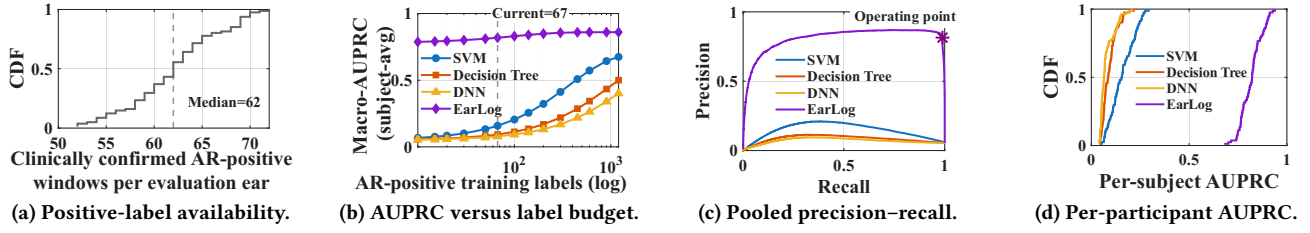


Figure 8: The residual representation improves label efficiency and participant-disjoint precision-recall performance over raw-spectrum classifiers.

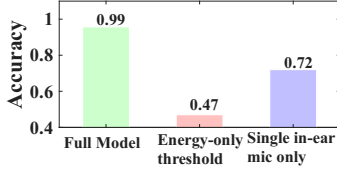


Figure 9: Full *EarLog*, energy-only, and single-microphone variants.

Variant	Acc. (%)	Change (pp)
Full <i>EarLog</i>	98.71	0.00
Without physics-constrained no-AR baseline	90.35	-8.36
Single-band LDS variant	86.55	-12.16
Without roughness descriptors and gating	92.20	-6.51
Without artifact and interference handling	75.90	-22.81

Table 1: Each processing block addresses a distinct source of error in the residual pipeline.

Environment	Acc. (%)
Quiet dormitory (45 dB)	99.37
Classroom (55 dB)	97.78
Office with talking (65 dB)	96.21
Outdoor road (70 dB)	95.15

Table 2: *EarLog* retains high accuracy as ambient noise increases.

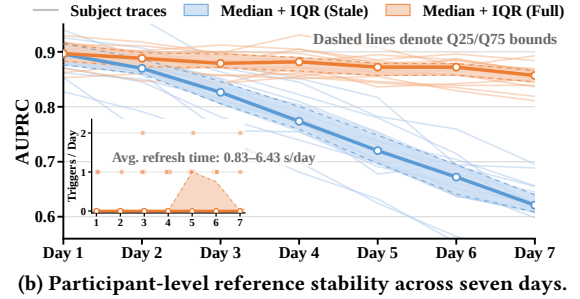
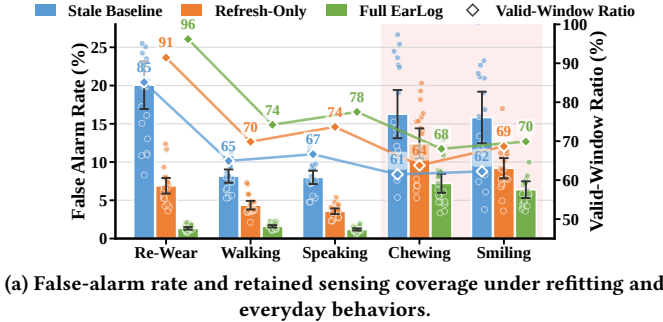


Figure 10: Quality-aware robustness under refitting, everyday behaviors, and cross-day drift.

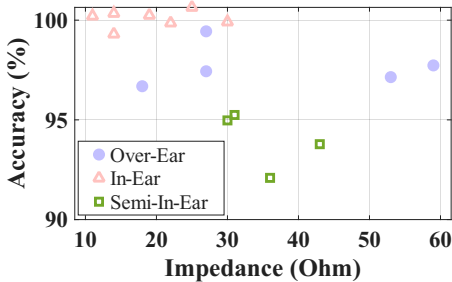


Figure 11: *EarLog* maintains high recognition accuracy across headphone form factors.

We compare a stale baseline, low-volume refresh alone, and the full calibration-transfer-plus-refresh pipeline. Figure 10a reports false-alarm rate among valid windows together with the valid-window ratio. Reporting both quantities shows how effectively *EarLog* controls erroneous events while retaining sensing coverage.

After re-wearing, a stale baseline produces 20.05% false alarms. Low-volume refresh reduces this value to 6.89%, while the full pipeline reaches 1.33% and retains 96.20% of the recorded windows. Calibration transfer therefore aligns

the new fit while preserving most of the listening period for sensing.

Walking and speaking retain 74.21% and 77.52% of recorded windows, with false-alarm rates of 1.60% and 1.19%. During chewing and smiling, *EarLog* retains 68.05% and 69.90% of windows and limits false alarms to 7.22% and 6.39%. The quality-aware output preserves roughly seven in ten observations even when facial motion changes both the seal and canal geometry. Appendix H.6 gives the detailed protocol.

7.4.3 Cross-Day Reference Maintenance. Window-level screening handles short disturbances. We next ask whether the same personalized baseline remains useful across days. Multiple participants repeat the experiment for seven days with the same evaluation ear and headphone model. We compare continued reuse of the Day-1 baseline with the full reference-maintenance pipeline.

With the initial baseline reused unchanged, participant-level median AUPRC changes from 0.896 on Day 1 to 0.621 on Day 7. Full *EarLog* starts at 0.897 and reaches 0.857 on Day 7. Reference maintenance therefore provides a 0.236 AUPRC gain over the stale baseline by the end of the study.

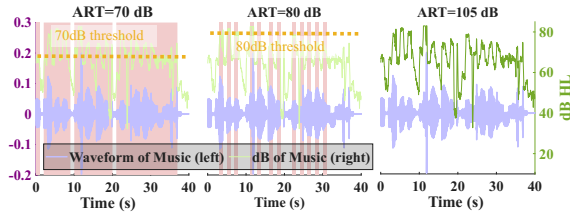


Figure 12: Titan-confirmed AR-event logs reveal listener-specific recruitment under the same device-recommended playback setting.

Calibration transfer accounts for most of the observed cross-day stability. The median refresh count remains zero on every day, and the mean ranges from 0.06 to 0.39 triggers per day. Across days, accumulated low-volume refresh time ranges from 0.83 to 6.43 s per day. Each triggered refresh lasts approximately 13.8–16.5 s. The stored reference therefore remains useful through transfer, and refresh is needed only occasionally. Appendix H.7 gives the session protocol and participant-level aggregation.

7.5 What the Event Log Reveals

The previous experiments establish accuracy and robustness. We finally use *EarLog* in its intended logging mode to examine whether ordinary playback produces listener-specific recruitment patterns. We keep the classifier frozen, play music unilaterally through the target headphone, and monitor the contralateral ear with Titan. This design provides an independent clinical reference while leaving the headphone-side acoustic observation unchanged.

Figure 12 shows one representative evaluation ear from each WB-ART range. Every red interval is an *EarLog* detection that overlaps a synchronized contralateral Titan-active interval. Confirmed events cluster around transient high-energy and spectrally intense passages. The same device-recommended playback setting produces different recruitment densities across listeners. Lower-threshold ears show denser confirmed intervals, while higher-threshold ears show sparser recruitment under the same playback regime.

EarLog aggregates these momentary events against valid listening time and co-reports sensing coverage. The resulting log describes how each ear recruits protection across content and sessions, revealing listener-specific patterns that remain invisible to a common device setting. Section 6 defines the reference-event protocol, and Appendix H gives the participant, stimulus, and deployment details.

8 Conclusion

This paper presents *EarLog*, which enables commodity headphones to record acoustic-reflex events without interrupting everyday listening. The challenge is that AR is weak and

short-lived, whereas nonstationary content, individual ear acoustics, device response, fit, and environmental disturbances introduce stronger variations in the in-ear observation. *EarLog* turns playback into an opportunistic probe and uses content-conditioned residual sensing to suppress these dominant variations. Resonance-constrained short-window modeling, dual-microphone observations, and cross-session reference updates allow the system to run in real time on commodity hardware. We evaluate *EarLog* with 81 participants, 16 headphone models, and more than 400 hours of recordings. *EarLog* achieves 99.0% sensitivity and 98.7% specificity on a clinically referenced benchmark and remains usable across content, headphone models, everyday behaviors, refitting, and cross-day conditions. The resulting logs also show that the same audio played at the same nominally safe volume can produce markedly different AR recruitment patterns across listeners. These listener-specific recruitment logs can support long-term risk observation and future personalized safe-listening strategies.

Acknowledgments

This work was supported by the Guangdong Provincial Key Laboratory of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No. 2023B1212 010007), the National Natural Science Foundation of China (Grant No. 62472366), the Department of Education of Guangdong Province (Nos. 2023KCXTD042, 2024GCZX003, 2024 KCXTD008, and 2025KCXTD056), the 111 Center (No. D25008), and the Shenzhen Science and Technology Foundation (No. ZDSYS20190902092853047).

References

- [1] Abbie Baricevich, Danielle Bassett, Sophia Chan, Shayna Lavi, and Jonathan Siegel. 2025. Frequency and level dependence of the middle ear acoustic reflex and its decay measured in wideband absorbance with contralateral narrowband noise elicitors. *Hearing Research* 459 (April 2025), 109225. doi:10.1016/j.heares.2025.109225
- [2] Erik Borg. 1982. Time Course of the Human Acoustic Stapedius Reflex: A Comparison of Eight Different Measures in Normal-Hearing Subjects. *Scandinavian Audiology* 11, 4 (1982), 237–242. doi:10.3109/01050398209087473
- [3] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. hEART: Motion-resilient Heart Rate Monitoring with In-ear Microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (Atlanta, GA, USA). IEEE, Piscataway, NJ, USA, 200–209. doi:10.1109/PERCOM56429.2023.10099317
- [4] Chao Cai, Henglin Pu, Menglan Hu, Rong Zheng, and Jun Luo. 2023. Acoustic Software Defined Platform: A Versatile Sensing and General Benchmarking Platform. *IEEE Transactions on Mobile Computing* 22, 2 (Feb. 2023), 647–660. doi:10.1109/TMC.2021.3093259
- [5] Chao Cai, Henglin Pu, Liyuan Ye, Hongbo Jiang, and Jun Luo. 2023. Active Acoustic Sensing for “Hearing” Temperature Under Acoustic Interference. *IEEE Transactions on Mobile Computing* 22, 2 (Feb. 2023), 661–673. doi:10.1109/TMC.2021.3096792

- [6] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Jun Luo. 2025. Enabling Passive User Authentication via Heart Sounds on In-Ear Microphones. *IEEE Transactions on Dependable and Secure Computing* 22, 2 (2025), 1195–1209. doi:10.1109/TDSC.2024.3429574
- [7] Yetong Cao, Chao Cai, Anbo Yu, Fan Li, and Jun Luo. 2023. Ear-AcE: Empowering Versatile Acoustic Sensing via Earable Active Noise Cancellation Platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (June 2023), 47:1–47:23. doi:10.1145/3596242
- [8] Yetong Cao, Dong Ma, Wentao Xie, Qian Zhang, and Jun Luo. 2025. ES-PIRO: Natural Pulmonary Function Monitoring via Earphone-Acquired Speech. In *Proceedings of the 31st Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 329–344. doi:10.1145/3680207.3723477
- [9] Andrew Causon, Kevin J. Munro, Christopher J. Plack, and Garreth Prendergast. 2020. The Role of the Clinically Obtained Acoustic Reflex as a Research Tool for Subclinical Hearing Pathologies. *Trends in Hearing* 24 (2020), 2331216520972860. doi:10.1177/2331216520972860
- [10] Justin Chan, Nada Ali, Ali Najafi, Anna Meehan, Lisa R Mancl, Emily Gallagher, Randall Bly, and Shyamath Gollakota. 2022. An off-the-shelf otoacoustic-emission probe for hearing screening via a smart-phone. *Nature biomedical engineering* 6, 11 (2022), 1203–1213.
- [11] Justin Chan, Antonio Glenn, Malek Itani, Lisa R. Mancl, Emily Gallagher, Randall Bly, Shwetak Patel, and Shyamath Gollakota. 2023. Wireless earbuds for low-cost hearing screening. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. Association for Computing Machinery, New York, NY, USA, 84–95. doi:10.1145/3581791.3596856
- [12] Justin Chan and Shyamath Gollakota. 2022. Inner-ear cochlea testing with earphones. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. Association for Computing Machinery, New York, NY, USA, 609–610. doi:10.1145/3498361.3538657
- [13] Justin Chan, Ali Najafi, Mallory Baker, Julie Kinsman, Lisa R Mancl, Susan Norton, Randall Bly, and Shyamath Gollakota. 2022. Performing tympanometry using smartphones. *Communications medicine* 2, 1 (2022), 57.
- [14] Fei Chen, Hui Xue, Meng Wang, Zhiling Cai, and Shipeng Zhu. 2023. Hearing care: safe listening method and system for personal listening devices. *International Journal of Environmental Research and Public Health* 20, 3 (2023), 2161.
- [15] Tao Chen, Yongjie Yang, Xiaoran Fan, Xiuzhen Guo, Jie Xiong, and Longfei Shangguan. 2024. Exploring the Feasibility of Remote Cardiac Auscultation Using Earphones. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 357–372. doi:10.1145/3636534.3649366
- [16] Xiaoran Fan, David Pearl, Richard Howard, Longfei Shangguan, and Trausti Thormundsson. 2023. APG: Audioplethysmography for Cardiac Monitoring in Hearables. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, Article 67, 15 pages. doi:10.1145/3570361.3613281
- [17] Xiaoran Fan, Longfei Shangguan, Siddharth Rupavatharam, Yanyong Zhang, Jie Xiong, Yunfei Ma, and Richard Howard. 2021. HeadFi: bringing intelligence to all headphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 147–159. doi:10.1145/3447993.3448624
- [18] M. Patrick Feeney. 2023. Wideband Acoustic Reflex Measurement. *Seminars in Hearing* 44, 1 (2023), 84–92. doi:10.1055/s-0042-1758752
- [19] M. P. Feeney et al. 2017. Normative Wideband Reflectance, Equivalent Admittance at the Tympanic Membrane, and Acoustic Stapedius Reflex Threshold in Adults. *Ear and Hearing* 38 (2017), e84–e98. doi:10.1097/AUD.0000000000000399
- [20] M Patrick Feeney, Douglas H Keefe, Lisa L Hunter, Denis F Fitzpatrick, Angela C Garinis, Daniel B Putterman, and Garnett P McMillan. 2017. Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic stapedius reflex threshold in adults. *Ear and hearing* 38, 3 (2017), e142–e160.
- [21] Katharine A. Fernandez, Penelope W. C. Jeffers, Kumud Lall, M. Charles Liberman, and Sharon G. Kujawa. 2015. Aging after Noise Exposure: Acceleration of Cochlear Synaptopathy in "Recovered" Ears. *Journal of Neuroscience* 35, 19 (2015), 7509–7520. doi:10.1523/JNEUROSCI.5138-14.2015
- [22] Mayank Goel, Elliot Saba, Maia Stiber, Eric Whitmire, Josh Fromm, Eric C. Larson, Gaetano Borriello, and Shwetak N. Patel. 2016. Spiro-Call: Measuring Lung Function over a Phone Call. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5675–5685. doi:10.1145/2858036.2858401
- [23] Michael P. Gorga and Patricia G. Stelmachowicz. 1983. Temporal Characteristics of the Acoustic Reflex. *Audiology* 22, 2 (1983), 120–127. doi:10.3109/00206098309072774
- [24] GRAS. 2020. GRAS 43AC-S1 CCP Ear Simulator Kit According to IEC 60318-4. <https://www.grasacoustics.com/products/ear-simulator-kit/product/235-43ac-s1>.
- [25] GRAS. 2020. GRAS 45BB-15 KEMAR with Hi-Res Ear Simulators for Test of Ear. <https://www.grasacoustics.com/products/head-torso-simulators-kemar/kemar-for-ear-headphone-test-2-ch/product/811-45bb-15>.
- [26] Kayla Hill, Hu Yuan, Xianren Wang, and Su-Hua Sha. 2016. Noise-Induced Loss of Hair Cells and Cochlear Synaptopathy Are Mediated by the Activation of AMPK. *The Journal of Neuroscience* 36, 28 (2016), 7497–7510. doi:10.1523/JNEUROSCI.0782-16.2016
- [27] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. A portable and convenient system for unknown liquid identification with smartphone vibration. *IEEE Transactions on Mobile Computing* 22, 4 (2021), 1894–1911.
- [28] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. Vi-liquid: unknown liquid identification with your smartphone vibration. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (New Orleans, Louisiana) (MobiCom '21). Association for Computing Machinery, New York, NY, USA, 174–187. doi:10.1145/3447993.3448621
- [29] Yongzhi Huang, Kaixin Chen, Lu Wang, Yinying Dong, Qianyi Huang, and Kaishun Wu. 2021. Lili: liquor quality monitoring based on light signals. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (New Orleans, Louisiana) (MobiCom '21). Association for Computing Machinery, New York, NY, USA, 256–268. doi:10.1145/3447993.3483246
- [30] Yongzhi Huang, Kaixin Chen, Jiayi Zhao, Lu Wang, and Kaishun Wu. 2023. Beverage deterioration monitoring based on surface tension dynamics and absorption spectrum analysis. *IEEE Transactions on Mobile Computing* 23, 5 (2023), 3722–3740.
- [31] Yongzhi Huang, Jiayi Zhao, and Kaishun Wu. 2025. Reconstructing Ear Canal Channels for Fine-Grained Detection of Tympanic Membrane Changes. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 89 (Sept. 2025), 48 pages. doi:10.1145/3749516
- [32] Interacoustics. 2020. Titan - A Modular and Portable Tympanometer. <https://www.interacoustics.com/tympanometers/titan>.

- [33] Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, and Zhanpeng Jin. 2022. EarHealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. Association for Computing Machinery, New York, NY, USA, 397–408. doi:10.1145/3498361.3538935
- [34] Megan Knoetze, Faheema Mahomed-Asmail, Vinaya Manchaiah, and De Wet Swanepoel. 2021. Sound-level monitoring earphones with smartphone feedback as an intervention to promote healthy listening behaviors in young adults. *Ear and Hearing* 42, 5 (2021), 1173–1182.
- [35] S. Kujawa and M. Liberman. 2009. Adding Insult to Injury: Cochlear Nerve Degeneration after “Temporary” Noise-Induced Hearing Loss. *The Journal of Neuroscience* 29 (2009), 14077 – 14085. doi:10.1523/JNEUROSCI.2845-09.2009
- [36] Sharon G. Kujawa and M. Charles Liberman. 2009. Adding Insult to Injury: Cochlear Nerve Degeneration after “Temporary” Noise-Induced Hearing Loss. *Journal of Neuroscience* 29, 45 (2009), 14077–14085. doi:10.1523/JNEUROSCI.2845-09.2009
- [37] M. Charles Liberman, Michael J. Epstein, Sandra S. Cleveland, Haobing Wang, and Stéphane F. Maison. 2016. Toward a Differential Diagnosis of Hidden Hearing Loss in Humans. *PLOS ONE* 11, 9 (2016), e0162726. doi:10.1371/journal.pone.0162726
- [38] W. W. Qiu and F. Stucker. 1998. Characterization of Acoustic Reflex Latency in Normal-Hearing Subjects. *Scandinavian Audiology* 27, 1 (1998), 43–49. doi:10.1080/010503998419696
- [39] W. W. Qiu and F. Stucker. 1998. Characterization of acoustic reflex latency in normal-hearing subjects. *Scandinavian audiology* 27 1 (1998), 43–9. doi:10.1080/010503998419696
- [40] Alison Reynolds and Eric C Bielefeld. 2023. Music as a unique source of noise-induced hearing loss. *Hearing Research* 430 (2023), 108706.
- [41] Tobias Röddiger, Christopher Clarke, Daniel Wolfram, Matthias Budde, and Michael Beigl. 2021. EarRumble: Discreet Hands- and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 743, 14 pages. doi:10.1145/3411764.3445205
- [42] Samsung Electronics. 2016. Samsung EHS64 Wired Stereo Headset with Remote and Mic. Samsung product page. Retrieved August 27, 2026 from <https://www.samsung.com/in/mobile-accessories/headphone-ehs64-ehs64avfbecinu/> Product model EHS64AVFBECINU; product launch date: May 1, 2016.
- [43] Xingzhe Song, Boyuan Yang, Ge Yang, Ruihong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, Article 52, 14 pages. doi:10.1145/3372224.3419209
- [44] Xue Sun, Jie Xiong, Chao Feng, Wenwen Deng, Xudong Wei, Dingyi Fang, and Xiaojiang Chen. 2022. Earmonitor: In-ear Motion-resilient Acoustic Sensing Using Commodity Earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2022), 182:1–182:22. doi:10.1145/3569472
- [45] Titan | Handheld tympanometer. 2024. Instructions for Use – Titan. Accessed: August 30, 2024. Available at: https://www.interacoustics.com/images/files/manuals/en/8100759_D-0100608-M_-2023_12_Instructions_for_Use_Titan_EN.pdf.
- [46] Monica Treviño, Andie Zang, and Edward Lobarinas. 2023. The Middle Ear Muscle Reflex: Current and Future Role in Assessing Noise-Induced Cochlear Damage. *The Journal of the Acoustical Society of America* 153, 1 (2023), 436–466. doi:10.1121/10.0016853
- [47] Luca Turchet, Simone Luiten, Tjebbe Treub, Marloes van der Burgt, Costanza Siani, and Alberto Boem. 2024. Hearing loss prevention at loud music events via real-time visuo-haptic feedback. *Journal on Multimodal User Interfaces* 18, 1 (2024), 43–53.
- [48] Michelle D. Valero, Kelsey E. Hancock, and M. Charles Liberman. 2016. The Middle Ear Muscle Reflex in the Diagnosis of Cochlear Neuropathy. *Hearing Research* 332 (2016), 29–38. doi:10.1016/j.heares.2015.11.005
- [49] Yong Wang and Chongyu Ren. 2012. Effects of Repeated “Benign” Noise Exposures in Young CBA Mice: Shedding Light on Age-Related Hearing Loss. *Journal of the Association for Research in Otolaryngology* 13, 4 (2012), 505–515. doi:10.1007/s10162-012-0329-0
- [50] Robert H Withnell and Lauren E Gowdy. 2013. An analysis of the acoustic input impedance of the ear. *Journal of the Association for Research in Otolaryngology* 14 (2013), 611–622.
- [51] World Health Organization and International Telecommunication Union. 2019. Safe Listening Devices and Systems: A WHO–ITU Standard. <https://www.who.int/publications/i/item/9789241515276>. Mode 1: 80 dB for 40 h/week (dose-based guidance).
- [52] Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2023. Ear-Spiro: Earphone-based Spirometry for Lung Function Assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–27. doi:10.1145/3569480
- [53] Chi Xu, Wentao Xie, Baichen Yang, Yizhen Zhang, Yanbin Gong, Jin Zhang, Wei Li, Shifang Yang, and Qian Zhang. 2025. EasySpiro: Assessing Lung Function via Arbitrary Exhalations on Commodity Earphones. In *Proceedings of the 31st Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 513–528. doi:10.1145/3680207.3723489
- [54] Xiangyu Yin, Kai Huang, Erick Forno, Wei Chen, Heng Huang, and Wei Gao. 2022. Out-Clinic Pulmonary Disease Evaluation via Acoustic Sensing and Multi-Task Learning on Commodity Smartphones. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. Association for Computing Machinery, New York, NY, USA, 1182–1188. doi:10.1145/3560905.3568437
- [55] Xiangyu Yin, Kai Huang, Erick Forno, Wei Chen, Heng Huang, and Wei Gao. 2023. PTEase: Objective Airway Examination for Pulmonary Telemedicine using Commodity Smartphones. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. Association for Computing Machinery, New York, NY, USA, 110–123. doi:10.1145/3581791.3596854

A Clinical Background and Illustrative Content Confounding

A.1 Scope of This Appendix

The argument in Sec. 3 rests on two observations. First, a commodity in-ear path can capture small reflection changes, but it does not inherit the stable absolute criterion of a clinical system. Second, arbitrary playback changes which ear-canal paths are excited, so the observed reflection changes with content even when the ear does not. The following sections explain these two effects. Appendix B then derives the residual cues used for AR detection.

A.2 Clinical Acoustic-Admittance Background

Clinical AR assessment [35, 39] commonly describes the middle-ear response through acoustic admittance, the reciprocal of acoustic impedance:

$$Y = \frac{1}{Z}.$$

With controlled single-frequency stimulation, a low-level probe tone gives a reference reflection coefficient R_{ref} . A higher-level elicited state gives a detection coefficient R_{det} . Their acoustic-admittance change is

$$\Delta Y = \frac{1 - R_{\text{ref}}}{Z_1(1 + R_{\text{ref}})} - \frac{1 - R_{\text{det}}}{Z_2(1 + R_{\text{det}})},$$

where Z_1 and Z_2 are calibrated impedances for the same ear.

Controlled Titan–COTS comparison. The comparison in Sec. 3.1 uses an Interacoustics Titan clinical system [32, 45] and a matched COTS path built from wired Samsung EHS64 headphones [42] and an ICS-40720 analog differential-output MEMS microphone. Titan presents a 500 Hz elicitor and monitors a 226 Hz probe tone. The COTS path uses SPL-matched playback and records the reflected signal under the corresponding stimulus condition.

Figure 2 reports pairwise percentage contrasts relative to the original playback. The pink curve compares the human ear with the silicone artificial ear. The blue curve compares the silicone ear with the IEC 60318-4 ear simulator. The orange line is the Titan admittance-change criterion, the green bands are Titan-detected AR ranges, and the dashed green curve fits the Titan measurements. The experiment includes five volunteers, a silicone artificial ear [25], and an IEC 60318-4-compliant ear simulator [24]. This comparison is separate from the contralateral music-validation protocol in Sec. 6.

With dedicated probe and elicitor tones and stable instrumentation, Titan attributes small changes in reflected sound power to AR onset. It begins to detect small responses near 65 dB SPL, and the response grows markedly above 75 dB SPL. The two COTS contrasts also rise overall, but they vary irregularly with SPL. Neither passive fixture can generate a physiological AR. Their different geometries and acoustic impedances can still create a nonzero offset, and that offset changes with level. The passive-fixture contrast even crosses the Titan criterion. An absolute value calibrated at one level therefore cannot serve as a global AR threshold on the COTS path.

A.3 An Illustrative Multipath-Content Model under Arbitrary Playback

Absolute reflection criteria face a second problem under arbitrary playback. Unlike a clinical probe tone, music is nonstationary and ranges from sparse tonal passages to broadband mixtures. Each spectrum excites a different combination of ear-canal paths. We use a factorized model to show how temporal overlap and frequency-dependent absorption enter the observed reflection.

The reverberant field in the canal decays over time. We denote its envelope by

$$L_p(t) = L_{p0} - \beta t / T_{\text{floor}},$$

where L_{p0} is the initial sound-pressure level, β controls the decay rate, and T_{floor} is the time required to approach the ambient-noise floor. The envelope $g_{\text{mix}}(t)$ summarizes the cumulative effect of overlapping reflections.

The canal wall, surrounding tissues, and canal morphology also shape frequency-dependent absorption and scattering. We collect these effects in $g_{\text{abs}}(f, x)$, where f is frequency and x is propagation distance along the canal.

The resulting illustrative effective reflection coefficient is

$$R_{\text{eff}}(f, t, x) = R_{\text{ideal}}(f) g_{\text{mix}}(t) g_{\text{abs}}(f, x),$$

where $R_{\text{ideal}}(f)$ is the idealized reflection coefficient under simplified conditions.

This factorization exposes why a fixed reflectance threshold fails. The overlap and absorption terms are weighted by the current playback spectrum and gain, so the microphone observes a different composite response as the content changes. *EarLog* therefore predicts the no-AR response for the current playback and tests the residual, rather than comparing the raw reflection with one absolute reference.

B Residual-Cue Derivations and Online AR Decision

Section 4.2.2 defines the estimator, and Secs. 4.2.3 and 4.2.4 define the residual descriptors and online decision. This appendix derives the residual relation and describes pLDS window construction, regularization details, training-time feature normalization, and the hysteresis parameters.

B.1 From Middle-Ear Load Change to LDS

AR changes the effective middle-ear load seen from the ear canal [1, 20, 50]. This change modifies the acoustic boundary condition and becomes observable through the ear-canal standing-wave response in a subset of frequency bands.

In analysis band b , let $S_b(t)$ denote the short-time root-mean-square energy of the decoded playback signal and let $R_b(t)$ denote the corresponding inward-microphone energy.

We define the source-side log loudness and the band-wise log transmission ratio as

$$L_{s,b}(t) = \log S_b(t), \quad \tau_b(t) = \log R_b(t) - \log S_b(t).$$

Within a short window, we approximate the headphone gain in band b by G_b . We represent ear-canal reverberation by H_b^{rev} and the additional middle-ear transfer change caused by AR by H_b^{AR} . The transmission ratio can then be written as

$$\tau_b(t) \approx \log |G_b H_b^{\text{rev}}(f_b, t)| + \log |H_b^{\text{AR}}(L_{s,b}(t), t)|.$$

The first term describes the device and ear-canal response that the no-AR baseline seeks to predict. During AR, we write

$$H_b^{\text{AR}}(L_{s,b}) = 1 + \varepsilon_b(L_{s,b}),$$

where ε_b denotes the loudness-dependent transfer perturbation associated with the effective middle-ear load change $\Delta Z_L(f_b, L_{s,b})$. Applying the chain rule gives

$$\frac{\partial \tau_b}{\partial L_{s,b}} \approx \frac{\partial \log |H_b^{\text{AR}}|}{\partial Z_L} \frac{\partial Z_L}{\partial L_{s,b}}.$$

Without AR, the no-AR baseline accounts for the ordinary level-dependent response and the remaining loudness sensitivity stays close to zero. When AR engages, the effective load becomes loudness-dependent and the first factor above can be amplified near ear-canal resonance bands.

The main pipeline computes this quantity after baseline subtraction. Let $\tau_b^{\text{base}}(t)$ denote the predicted no-AR transmission ratio for the same playback condition. The residual is

$$r_b(t) = \tau_b(t) - \tau_b^{\text{base}}(t).$$

The LDS used in the main text is therefore

$$\text{LDS}_b \approx \frac{\partial r_b(t)}{\partial L_{s,b}(t)}.$$

This expression links the observed residual slope to the loudness-dependent middle-ear transfer change while suppressing the dominant device and playback response.

B.2 pLDS Window Construction and Regularization Details

A single-band LDS can absorb variation transferred from neighboring bands. For the short estimation window $t = t_1, \dots, t_W$, *EarLog* arranges the residual and source-side loudness observations as

$$\mathbf{r}(t) = [r_1(t), \dots, r_B(t)]^\top, \\ \mathbf{L}_s(t) = [L_{s,1}(t), \dots, L_{s,B}(t)]^\top,$$

which supply the per-frame observations for the multiband regression in Eq. 2. Dynamic audio often produces strongly correlated loudness trajectories across bands, so ordinary least squares can yield an unstable coefficient matrix. The ridge term governed by λ_A regularizes that collinearity. The

estimator still relies on the excitation present in the current window, which is why Sec. 4.2.2 excludes low-energy bands.

B.3 Verifier Training and Hysteresis Parameters

Equations 5 and 6 define the feature vector and online score. The z-score mean and variance are computed from the training participants and then fixed for evaluation. Covariance shrinkage is applied when fitting the LDA discriminant to reduce instability under the available labels.

The runtime parameters are the smoothing coefficient λ_s , entry and exit thresholds γ_{hi} and γ_{lo} , and consecutive-window counts K_+ and K_- . The two thresholds prevent rapid switching near the decision boundary, while the consecutive-window counts require sustained evidence from valid windows. Their state semantics and event aggregation are defined in Sec. 4.2.4.

C Details of FIR Identification and Volume-Dependent Extrapolation

The three subsections describe bucketed identification, low-order baseline fitting, and runtime construction used by Sec. 4.1. The coefficient update itself is defined in Sec. 4.1.2.

C.1 Bucketed LMS Identification

For each band-energy bucket V , Sec. 4.1.1 defines a local no-AR FIR on the tap support selected by Sec. 4.1.2. Equation 1 is the authoritative coefficient update. After convergence, the supported coefficient vector provides the local no-AR impulse response $\mathbf{h}^{(V)}$ used by the bucket-level baseline model.

C.2 Band-Wise Polynomial Extrapolation Across Loudness Buckets

The gammatone filterbank maps each bucket-specific response into the B analysis bands used throughout the main pipeline. Section 4.1.3 defines the band-wise polynomial model for $\theta_b(V)$. Here, V is the bucket-level playback condition, and the model coefficients are fitted from the corresponding low-volume calibration buckets. The low-order fit captures the smooth variation observed across those buckets and provides a baseline for playback states without an exact stored match.

C.3 Online Construction of the No-AR Baseline

For each analysis window, *EarLog* computes the current playback state V_t , evaluates the band-wise models from Sec. C.2,

and reconstructs $H_{\text{baseline}}(f, V_t)$. Section 4.1.3 defines the resulting log-magnitude residual. This stage predicts the no-AR baseline but does not make an AR decision; the residual is passed to the LDS, pLDS, and roughness calculations in Sec. 4.2.

D Calibration Transfer and Baseline Refresh

This appendix expands the calibration-transfer and refresh mechanisms in Sec. 4.4. The main text defines the anchor-matching and online transfer objectives. Here, we describe paired-difference and drift-basis construction, anchor-set construction, support-window formation, frequency-domain transport, and the consistency failure that requires a new low-volume baseline.

D.1 Building the Drift Basis

Let $\theta_i^{(a)} \in \mathbb{R}^B$ and $\theta_i^{(b)} \in \mathbb{R}^B$ denote band-wise no-AR baseline parameters from the same playback bucket under two fit or environmental conditions. Both observations come from valid low-volume windows used to form the no-AR reference. Their difference is

$$\Delta\theta_i = \theta_i^{(b)} - \theta_i^{(a)}.$$

Stacking N paired differences gives

$$\Delta = [\Delta\theta_1, \dots, \Delta\theta_N] \in \mathbb{R}^{B \times N}.$$

EarLog applies singular value decomposition:

$$\Delta = U\Sigma R^\top.$$

The leading r left singular vectors form the drift basis

$$\Psi = U_{[:,1:r]} \in \mathbb{R}^{B \times r}.$$

Its columns capture the dominant band-wise response changes observed across refitting, seal, humidity, and related non-thermal conditions.

D.2 Anchor-Band Matching

Bands that are weakly affected by the drift basis provide a stable reference for matching the current condition. *EarLog* defines the anchor set

$$\mathcal{A} = \{b \in \{1, \dots, B\} : \|\Psi_{b,:}\|_2 \leq \tau_A\},$$

where $\Psi_{b,:}$ is row b of the drift basis. Let $P_{\mathcal{A}}$ denote the diagonal selector that preserves these bands.

The online support set $\mathcal{S}_t$ contains the first valid low-volume support windows available under the current condition. Their band-wise average forms $\mathbf{q}_t$ in the same log-transfer representation as the stored baseline parameters. Equation 8 matches this support to the stored bucket that agrees most closely on the anchor bands.

D.3 Online Transfer Estimation

Equation 9 estimates the current drift coefficients after anchor matching. At the transfer-function level,

$$\log |\tilde{H}_{\text{baseline}}(f, V_t, u_t)| = \log |H_{\text{baseline}}(f, V_t)| + \sum_{j=1}^r \hat{\alpha}_{t,j} \psi_j(f),$$

where $\psi_j(f)$ is the frequency-domain profile of drift component j . The residual passed to Sec. 4.2 is

$$\Delta \log \tilde{H}(f, t) = \log |H_{\text{obs}}(f, t)| - \log |\tilde{H}_{\text{baseline}}(f, V_t, u_t)|.$$

D.4 Consistency Failure and Baseline Refresh

The transferred baseline is checked on valid low-volume support windows. *EarLog* monitors the residual energy and the low-volume LDS, pLDS, and roughness values produced by the transported baseline. After a transient failure, the validity gate in Sec. 4.3 rejects the affected window. Persistent failure across consecutive support windows suspends AR detection and triggers a low-volume refresh of the band-energy buckets and baseline parameters.

Listening exposure and signal-quality status remain active during refresh. The refreshed baseline must pass the same consistency test before AR detection resumes. This fallback covers shifts that are not represented by the learned drift basis without introducing a second AR classifier or recursively transporting a previous estimate.

E Ear-Canal Length Estimation and Sparse FIR Tap Placement

This appendix gives the resonance-based propagation prior used to select the sparse FIR support in Sec. 4.1.2. We first obtain a coarse effective length L from the standing-wave condition

$$2\pi fL/c = (2n+1)\pi/2,$$

where c is the speed of sound and n is a nonnegative integer. For the fundamental mode,

$$L = c/(4f).$$

In our implementation, L is constrained to 15–45 mm. The lower bound accounts for the shorter effective propagation distance introduced by headphone insertion relative to the theoretical 25–45 mm ear-canal range. The upper bound corresponds to a minimum candidate resonance frequency of approximately 1906 Hz.

A single resonance may be shifted by multipath. *EarLog* therefore forms a relative-gain curve by dividing the measured frequency-response magnitude by the source magnitude and identifies prominent resonance peaks within the

allowed frequency range. For a detected resonance f_n , the corresponding effective length is

$$L = \frac{c}{4f_n}(2n + 1).$$

The resulting one-way propagation delay is $t_0 = L/c$. EarLog places nonzero FIR taps near the direct arrival and the round-trip delays

$$k_i \approx 2it_0f_s, \quad i = 0, 1, \dots, N,$$

where f_s is the sampling rate and i indexes successive reflections. This propagation prior reduces the number of free FIR coefficients and supports stable response estimation within the short analysis window.

F Details of Observation Screening

This appendix complements Sec. 4.3 by defining the validity indicators and window-routing rules. It specifies how windows are assigned to valid, invalid, or overload-only states and how persistent baseline inconsistency is handed to baseline maintenance.

F.1 Window Validity and Routing

Let $y_{\text{raw}}(n)$ denote the inward-microphone samples in an analysis window of length N . EarLog identifies samples that approach the calibrated saturation boundary of the analog front end:

$$C = \left\{ n \in [1, N] : |y_{\text{raw}}(n)| \geq \eta_{\text{sat}} \right\},$$

where η_{sat} is obtained during front-end calibration. The clipped-sample ratio is

$$r_{\text{clip}} = \frac{|C|}{N}.$$

A calibrated operating range separates hard saturation from mild clipping. Hard-saturated windows are logged as *overload-only*; mildly clipped windows retain enough response structure to remain eligible for the constrained repair path summarized in Sec. 4.3.

The other three indicators examine different parts of the playback–response relation. The leakage ratio $\lambda_{\text{leak}}(t)$ compares outer- and inner-microphone energy to reveal seal loss or abnormal external-sound leakage. The source-coherence score $c_{\text{src}}(t)$ tests whether the inward observation remains aligned with the known playback. The baseline-consistency score $m_{\text{fit}}(t)$ measures whether the current response remains compatible with the recent no-AR baseline.

A window remains valid only while all four indicators stay within their calibrated ranges. The gate changes state after either a violation or a recovery persists across consecutive windows, avoiding rapid switching around a boundary. Saturated windows are routed to *overload-only*; other failures are routed to *invalid*. Rejected windows do not update the



Figure 13: Representative headphone models used in the robustness study.

AR state or event count, but they remain in the listening-exposure and signal-quality records. Persistent baseline inconsistency is passed to Sec. 4.4 for baseline adaptation or refresh.

G Implementation Details

This appendix expands Sec. 5 with the analog front end, thermal compensation, electromagnetic shielding, deployment geometry, runtime, and power measurements.

G.1 Analog Front End and Thermal Compensation

The inner and outer ICS-40720 microphones are conditioned through separate analog paths and digitized on independent ADC channels. A resistor-capacitor anti-alias filter is followed by active high-pass and low-pass stages that confine the analog response to 50 Hz–20 kHz. Final band limiting and equalization are performed digitally.

Let $G_0(T)$ denote the uncompensated microphone-to-ADC gain at temperature T , and let $C_{\text{NTC}}(T)$ denote the gain correction introduced by the NTC feedback network. The feedback topology is selected so that

$$\begin{aligned} \partial_T \log C_{\text{NTC}}(T) &\approx -\partial_T \log G_0(T), \\ \partial_T \log [G_0(T)C_{\text{NTC}}(T)] &\approx 0 \end{aligned}$$

over the calibrated operating range. The analog network removes the dominant first-order thermal gain drift before the ADC, and digital normalization absorbs the smaller residual offset.

G.2 Shielding and Deployment

The printed circuit board sits close to the headphone driver, where it is exposed to low-frequency magnetic fields

from coil current and high-frequency electromagnetic interference from driver wiring. The circuit is mounted in a 0.5 mm polylactic-acid carrier. An inner 0.5 mm 1J85 permalloy shell diverts low-frequency magnetic fields, while an outer grounded 200 μm copper mesh attenuates high-frequency interference.

Figure 6b shows the in-ear and over-ear placement. In an in-ear assembly, the module lies between the driver and the canal entrance. In an over-ear assembly, the electronics remain at least 1 cm from the driver magnet to reduce magnetic coupling to the MEMS front end. Figure 13 shows representative devices from the robustness study.

G.3 Runtime and Power

The fixed-point implementation runs on an Arduino Portenta H7 with an STM32H747 processor. The Cortex-M7 core runs at 480 MHz and the Cortex-M4 core at 240 MHz. The complete pipeline processes the two 48-kHz microphone streams at $1.05\times$ real time.

The prototype development board draws 0.45 W during continuous inference and 0.40 W during a fallback low-volume refresh. These values include the general-purpose MCU board and its peripheral circuitry. Relative to an audio-only baseline on the same platform, the *EarLog*-specific incremental overhead is 16 mW, corresponding to 0.016 Wh per listening hour. This is approximately 8% of a representative 0.20 Wh earbud battery. The whole-board and incremental measurements describe different quantities and should not be compared as product-level power figures. Clock alignment, SPL matching, and the evaluation platform are summarized in Appendix H.3.

H Evaluation Details

This appendix provides participant statistics, stimulus schedules, hardware profiling, and the detailed robustness protocols supporting Secs. 6 and 7. Section 6 is the authoritative description of the contralateral Titan reference, independent label generation, window states, event aggregation, and metric denominators.

H.1 Participant Statistics and WB-ART Stratification

The cohort contains 81 participants, with one evaluation ear from each participant. For subgroup analysis, the evaluation ears are stratified by WB-ART into ART-1 (65–75 dB, $n = 16$), ART-2 (75–85 dB, $n = 29$), and ART-3 (above 85 dB, $n = 36$). Figure 14 shows the cohort distribution.

H.2 Stimulus Schedule, Control Conditions, and Held-Out Content

The complete corpus spans controlled measurement, robustness testing, and ordinary-listening deployment. A 35-mm straight tube provides a non-physiological control. Playback includes broadband white noise over 20 Hz–20 kHz, piano tones, instrument sounds, and music excerpts at 50–90 dB SPL. The total recording duration exceeds 400 hours. The clinically referenced 145,800-window benchmark and unilateral music validation are defined separately in Sec. 6.

The robustness study spans 16 headphone models and ambient-noise conditions at 45, 55, 65, and 70 dB. These conditions represent a quiet dormitory, classroom discussion, office conversation, and roadside noise. The headphone set includes over-ear, in-ear, and semi-in-ear form factors, together with wired and wireless devices. Figure 13 shows representative models.

Unless stated otherwise, supervised training uses wideband stimuli. Piano notes and instrument timbres are held out for cross-content evaluation without fine-tuning. The unilateral music experiment keeps the same wideband-trained classifier fixed and uses the synchronized contralateral Titan reference defined in Sec. 6.

H.3 Hardware, Clocking, and Deployment Profiling

Training and offline processing use a computer with an Intel Core i7-9750H processor and 8 GB RAM. The model is trained in MATLAB R2023b and deployed on the Arduino Portenta H7. The inner and outer microphone streams are captured on separate ADC channels with 16-bit resolution at 48 kHz. The microphone streams are aligned to the digital playback by cross-correlation, and playback levels are matched to the clinical setup with a calibrated ear simulator. The fixed-point implementation runs at $1.05\times$ real time.

H.4 Confidence Intervals

Confidence intervals, when reported, are computed by participant-level bootstrapping. Window-level resampling is not used because adjacent windows from the same participant are correlated.

H.5 Recognition Across WB-ART Groups Under Cross-Content Playback

Figure 15 complements Sec. 7.1 by combining wideband, piano, and instrument playback in a noisy environment. Each evaluation ear is tested repeatedly at 50–90 dB SPL. Accuracy reaches 0.959 for ART-1 ears around 65–70 dB WB-ART, ranges from 0.883 to 0.925 in the ART-2 transition range around 75 dB, and reaches 0.877 for ART-3 ears

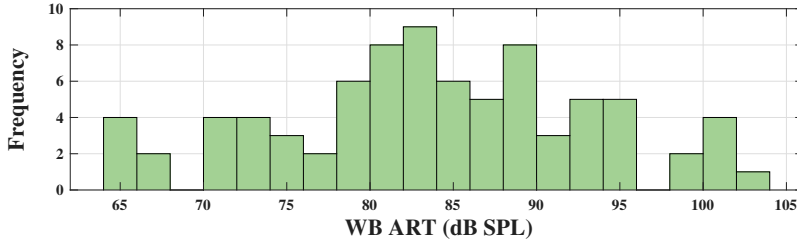


Figure 14: Distribution of evaluation ears across WB-ART values.

around 95–100 dB. The broader content set preserves the cross-threshold operating range established by the clinically referenced benchmark.

H.6 Refitting and Everyday-Behavior Protocol

For the refitting benchmark, each participant repeatedly removes and re-wears the same headphone model. After every re-wear, we record a short low-volume support segment and then play content at 10 dB below the participant’s WB-ART. Synchronized Titan-inactive intervals define the AR-negative reference condition, so every positive *EarLog* output is counted as a false alarm.

For the everyday-behavior benchmark, participants walk, speak, chew, and smile during the same Titan-referenced below-threshold playback. Figure 10a reports false-alarm rate among valid windows on the left axis and valid-window ratio on the right axis. The joint measurement verifies that *EarLog* controls erroneous event logging while retaining sensing coverage. Invalid and overload-only intervals remain in the exposure and signal-quality records.

H.7 Cross-Day Drift and Reference Maintenance

The cross-day study follows multiple participants for seven days. Each participant uses the same evaluation ear and headphone model throughout the study. Every daily session contains a prespecified below-WB-ART support segment, a seated controlled-playback segment, and an above-threshold block for performance verification.

We compare two policies. The stale-reference policy reuses the initial baseline throughout the study. The full-system policy enables the calibration transfer, consistency check, and low-volume refresh described in Sec. 4.4. Each thin trajectory in Fig. 10b follows one participant under one policy. The aggregate curves report the participant-level median and interquartile range.

The refresh analysis records trigger count and accumulated low-volume playback time for every participant and

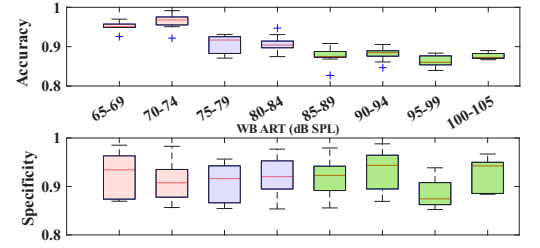


Figure 15: *EarLog* maintains broad cross-threshold performance under wideband, piano, and instrument playback in a noisy environment.

day. These measurements distinguish the continuous contribution of calibration transfer from the occasional cost of rebuilding the no-AR baseline.

I Discussion

EarLog establishes a new paradigm for ear-health monitoring by recording AR recruitment during everyday listening over time. Using *EarLog*, we also found that listeners with normal hearing may still show roughly 20–30 AR activations per hour even when the volume is limited. These event logs may support future studies of how repeated high-load listening contributes to hearing damage and provide listener-specific evidence for long-term risk management.

We did not build *EarLog* around an end-to-end classifier because reliable AR labels are scarce. Like many physiological-sensing problems, repeatedly eliciting AR to collect a large positive dataset would be difficult to justify under safety and ethics constraints. Arbitrary dynamic audio also produces very different signal patterns, making it unrealistic to annotate every type of content window by window. Individual ear anatomy, headphone response, and fit further change how the same physiological event appears acoustically. With limited labels, a model trained directly on raw signals can easily learn playback content or fit instead of physiology. *EarLog* therefore avoids asking the model to learn every source of variation by itself.

Our experiments cover refitting, walking, speaking, chewing, smiling, ambient noise, and multi-day drift. When interference cannot be explained, *EarLog* leaves the affected windows out of the AR event log. More conditions still need to be evaluated. These disturbances are typically random and brief, while AR events occur independently and do not systematically coincide with them. Removing interference therefore takes priority over retaining every window.

The *EarLog* prototype recreates the inner- and outer-microphone arrangement commonly found in active-noise-cancelling headphones, although its implementation remains different from an actual ANC signal chain. It still requires synchronized access to playback and microphone data, together with a stable analog front end. The NTC network compensates for

temperature-dependent front-end gain and reduces baseline drift and low-volume refreshes. This component is optional. Software calibration can handle the same drift, but it requires additional computation and energy.

Using ultrasound instead of everyday playback as the probe is also feasible. Ultrasound offers a more stable excitation, but it requires wider-band transducers and acquisition front ends, an ultrasonic microphone, and software support for synchronizing the probe signal. These additions increase power consumption, hardware compatibility requirements,

and the risk of intermodulation distortion, making direct deployment on ordinary headphones less practical.

EarLog may also extend beyond AR to other ear-health states and conditions. Voluntary middle-ear muscle activity, small pressure transients, and slowly changing ear states may also alter the in-ear response. Distinguishing these processes will require longer temporal context, binaural evidence, or clinical information, opening an important direction for future work.