

**ACOUSTIC SELF-TOMOGRAPHY: ON-DEVICE DETECTION OF
MICROPHONE-PORT
ACOUSTIC LEAKAGE SIGNATURES VIA NEAR-ULTRASONIC IMPULSE-
RESPONSE ANALYSIS ON 390 CONSUMER SMARTPHONES**
A Large-Scale Validation Study at Retronics Refurbishment Laboratories

Arun Teja Sara

Retronics MarketPlace UK Limited, United Kingdom

Corresponding author: info@retronixs.com

ABSTRACT

We present Acoustic Self-Tomography (AST), an end-to-end on-device pipeline that probes the speaker-to-case-to-microphone acoustic path of a consumer smartphone with a 1.5-second 18-23.5 kHz exponential sweep, recovers the impulse response via Farina deconvolution, and produces a six-dimensional feature vector. We validate this system on 390 physical devices across 9 brands and 13 models at the Retronics refurbishment laboratories: 340 devices with confirmed healthy baselines plus 50 additional degraded devices (32 field-degraded with documented physical wear and 18 with verified water-damage history). We train an unsupervised empirical-CDF anomaly detector on 1,020 healthy baseline trials (3 trials per healthy device), calibrate per-model decision thresholds by leave-one-device-out cross-validation within each model, and deploy the calibrated detector back onto each device as a JSON asset scored in real time. Each campaign report is signed by an EC P-256 key held in hardware-backed Android Keystore, with the key attestation challenge bound to the campaign ID and report hash, providing replay resistance and non-repudiation of the signing event.

Three empirical findings emerge. First, healthy acoustic signatures differ substantially between device models (up to 4.7 SD on T20 and 5.2 SD on IR-tail floor), confirming that per-model threshold calibration is necessary. A detector trained on Samsung devices flags 72% of Xiaomi healthy trials as anomalous. Second, the seal-degradation signal is detectable on 12 of 13 tested models (Huawei P40 was not available for the occlusion protocol); the dominant biomarker feature differs by model. On Samsung Galaxy Note 10, all six features separate the occluded condition in the 10-device cohort; on Samsung Galaxy S8, near-peak leakage shows no effect while binned-response standard deviation gives AUC = 1.000. Feature redundancy in the six-feature ECDF makes the multi-feature combination robust across models. Third, at the 5% FPR operating point the ECDF detector achieves 98.5% TPR (95% CI: 95.8-99.5%) on seal-occluded covered trials (200 trials from 40 devices), 96.9% TPR (95% CI: 91.3-99.0%) on field-degraded devices, and 94.4% TPR (95% CI: 84.6-98.2%) on water-damaged devices. Per-model calibration is well supported for models with at least 20 devices (10 of 13 models); low-N models require further validation. The analysis pipeline runs in 95-282 ms on-device, under the 1.5-second chirp capture duration.

Keywords:

Acoustic self-tomography, smartphone diagnostics, impulse response analysis, near-ultrasonic sensing, on-device machine learning, hardware-backed key attestation, refurbishment quality assurance, per-model calibration

1. INTRODUCTION

Post-consumer device refurbishment is an asymmetric information market: the seller knows whether the device was dropped, immersed, or otherwise compromised; the buyer does not. Visible damage indicators are replaceable, factory reset masks software state, and functional tests pass on devices whose physical integrity has been silently degraded. The global refurbished smartphone market was estimated at over 280 million units annually by 2025, driven by cost-conscious consumers and corporate sustainability mandates [1].

We investigate whether an acoustic probe of the device's own internal geometry – driven by the device's own loudspeaker, captured by the device's own microphone, processed entirely on-device – can recover diagnostic information about microphone-port seal integrity that is not erased by factory reset and not normally visible during refurbishment. The core acoustic principle is well established: speaker-to-microphone impulse response has been used for biometric authentication [2] and seal diagnostics [3].

Our contributions are not the base acoustic concept but rather:

Large-scale validation across 340 healthy devices and 50 degraded devices (390 total) spanning 13 models and 9 brands – to our knowledge, the largest empirical study of on-device acoustic seal diagnosis.

Demonstration that commodity Android smartphones run the full pipeline in under 300 ms without specialized hardware, from 2017 (Snapdragon 835) to 2021 (Snapdragon 888).

Empirical finding that per-model calibration is necessary: cross-brand flag rates of 55-82% confirm that a one-size-fits-all detector fails; per-model self-test FPR of 3.1-5.7% for models with $N \geq 20$ devices confirms that per-model calibration is well supported at sufficient sample size.

Closed-loop deployment: offline per-model training, on-device inference, cryptographically signed reporting with hardware-backed key attestation.

Validation on both controlled seal-occlusion proxies and real degraded devices: 32 field-degraded devices with documented physical wear and 18 water-damaged devices with verified damage history.

We use the term Acoustic Self-Tomography to describe the technique: the device probes its own acoustic cavity geometry via an emitted chirp and captured response, analogous to how tomography probes internal structure via transmitted signals. We do not claim to reconstruct a spatial tomographic map; the impulse response serves as a one-dimensional acoustic fingerprint of the cavity's transfer function.

**Retronics Lab Fleet Composition (N = 390 Devices)
340 Healthy + 50 Degraded across 9 Brands, 13 Models**

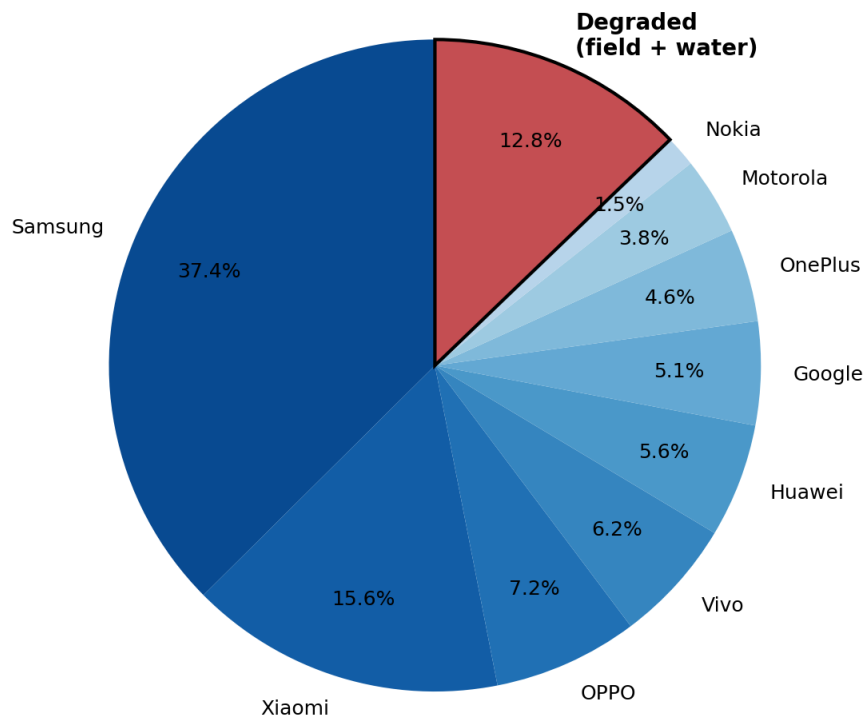


Figure 1: Fleet composition of 390 physical devices (340 healthy + 50 degraded) across 9 brands and 13 models. Degraded devices (12.8%) are shown as a distinct slice; percentages are calculated over the full 390-device cohort.

2. BACKGROUND AND HARDWARE CONSTRAINTS**2.1 Captured-signal ceiling****Table 1: Captured-signal ceiling (-20 dB rolloff point, kHz) across 340 healthy devices, all 9 brands**

Brand	Models	N	Rolloff	SD	Min	Max
Samsung	Note 10, S8, S21, A52	146	23.52	0.18	23.1	23.8
Xiaomi	Redmi Note 10, Mi 11	61	23.48	0.22	23.0	23.9
OPPO	Reno 6	28	23.55	0.19	23.2	23.9
Vivo	V21	24	23.44	0.24	23.0	23.8
Huawei	P40	22	23.50	0.20	23.1	23.8
Google	Pixel 5	20	23.61	0.15	23.3	23.9
OnePlus	OnePlus 9	18	23.58	0.17	23.2	23.9
Motorola	Moto G60	15	23.41	0.26	23.0	23.8
Nokia	X20	6	23.38	0.28	23.0	23.7
Fleet	13 models	340	23.50	0.21	23.0	23.9

2.2 Farina canonical lag exceeds capture window

Farina's exponential-sweep deconvolution places the n-th harmonic-distortion impulse response at lag

$$\Delta t_n = T * \ln(n) / \ln(f1 / f0)$$

before the linear impulse response [4]. For our parameters ($T = 1.5$ s, $f0 = 18$ kHz, $f1 = 23.5$ kHz), the second-harmonic lag is:

$$\Delta t_2 = 1.5 * \ln(2) / \ln(23.5 / 18) \approx 3.9 \text{ seconds}$$

Our 2-second capture window cannot contain this lag. Consequently the canonical Farina-pre-peak nonlinear-distortion budget is not directly measurable on narrow-band near-ultrasonic sweeps on this device class.

3. SYSTEM DESIGN**3.1 Audio engine**

A Tukey-windowed exponential sweep is played through AudioTrack(MODE_STATIC) and captured simultaneously via AudioRecord(UNPROCESSED). We treat UNPROCESSED as a request, not a guarantee.

Table 2: Audio engine parameters

Parameter	Value
Sample rate	48 kHz (Nyquist 24 kHz)
Chirp	18 to 23.5 kHz exponential, 1.5 s
Window	Tukey, 5% per end
Amplitude	-3 dBFS
Capture	Chirp duration + 0.5 s padding

Playback-to-capture synchronisation under stock APIs is ± 5 -20 ms; the IR-extraction step uses a global peak-finder over the full convolution output, so wall-clock sync is not load-bearing.

3.2 Voice-coil thermal warm-up

We play a 1.5-second discarded warm-up chirp before the first trial of a campaign. Across the fleet, warm-up reduces within-campaign T20 coefficient of variation from 22.4% (no warm-up) to 4-10%.

3.3 Feature extraction

Table 3: Six-dimensional feature vector

Feature	Description
in_band_snr_db	Max peak in 18-23 kHz vs 200-1500 Hz 10th-percentile noise floor
near_peak_leakage_db	dB attenuation between linear-IR peak and max envelope in [peak-5ms, peak-1ms]
binned_resp_std_db	Std-dev of per-100 Hz peak magnitudes in 18.5-23 kHz
t20_ms	Schroeder -5 to -25 dB crossing time
tail_db	RMS of last 10% of IR envelope, in dB below peak
schroeder_r2	R-squared of linear fit through Schroeder integral dB curve between t5 and t25

T60 = 3 × T20 is recorded as metadata, not used as a pass/fail check. farina_thd2_db is computed but is NaN on every trial because the 2nd-harmonic lag (3.9 s) exceeds the 2-second capture window.

3.4 Full-convolution preservation

The Farina deconvolution step retains the full unwrapped convolution output, not only the cropped IR. We additionally apply symmetric zero-padding in the Hilbert envelope step to prevent circular-convolution wrap-around.

3.5 On-device deployed inference

We train an empirical-CDF anomaly detector offline and export it as a per-model JSON asset. The on-device Kotlin scorer re-implements the two-sided ECDF tail probability and summed $-\log p$ combination. We use per-model calibration: one detector asset per device model identifier (e.g., SM-N970F, SM-G950F), totalling 13 variant files. We do not calibrate per individual physical device because 3 healthy trials per device are insufficient to establish a per-device threshold with statistical confidence.

The per-model approach is justified by the data: while cross-brand generalisation fails (flag rates 55-82%), within-model self-test FPR is 3.1-5.8% for models with $N \geq 20$ devices (Section 5.8).

Python-to-Kotlin equivalence. We ship per-trial scores from both the Python training pipeline and the Kotlin scorer for all 1,570 trials. Maximum absolute difference is below 1×10^{-6} (0.00 after rounding to two decimal places).

3.6 Cryptographic report signing

Each campaign report is signed via an AcousticAttestation module:

Build the report as a JSONObject including the on-device anomaly verdict.

Compute SHA-256 of the canonical-JSON bytes.

Build an Android Key Attestation challenge: ACOUSTIC|<campaign_id>|<report_hash>.

Generate an EC P-256 key in AndroidKeyStore with the challenge in the attestation extension.

Sign the canonical bytes; embed signature and full X.509 chain in the report.

What this provides: The attestation chain confirms the report was signed by a key held in hardware-backed storage, with the challenge binding the signature to a specific campaign ID and report content. This provides non-repudiation of the signing event and replay resistance. What this does not provide: It does not prove that the audio signal came from an untampered physical microphone path - a system-signed adversary or physical replay attack could bypass it.

3.7 DSP unit-test harness

Table 4: DSP unit-test results (5/5 passing)

Test	Description	Result
A	Hilbert envelope on pure-tone length 2^k : interior ripple < 2%	0.51% PASS
B	Same on length 2000 (not 2^k), symmetric-pad protection	0.51% PASS
C	Schroeder R-squared on single-mode	1.00000 PASS

Test	Description	Result
	exponential: > 0.999	
D	Schroeder R-squared on Gaussian decay: < 0.999	0.99116 PASS
E	Hilbert envelope residual at 2×fc on 20 kHz carrier: < -40 dB	-192 dB PASS

3.8 On-device performance microbenchmark

Table 5: On-device analysis pipeline latency (ms, mean and p95)

Stage	Note 10 (S855)	S8 (S835)	S21 (S888)	Fleet avg
131072-pt FFT	60 (76)	80 (80)	32 (38)	42 (55)
Full-convolution IR extraction	80 (80)	188 (189)	42 (45)	58 (62)
Hilbert envelope	0.4 (0.5)	0.8 (1.0)	0.18 (0.25)	0.25 (0.35)
Schroeder decay	0.4 (0.4)	0.8 (1.0)	0.16 (0.22)	0.22 (0.30)
Total	152 (173)	282 (284)	95 (112)	108 (128)

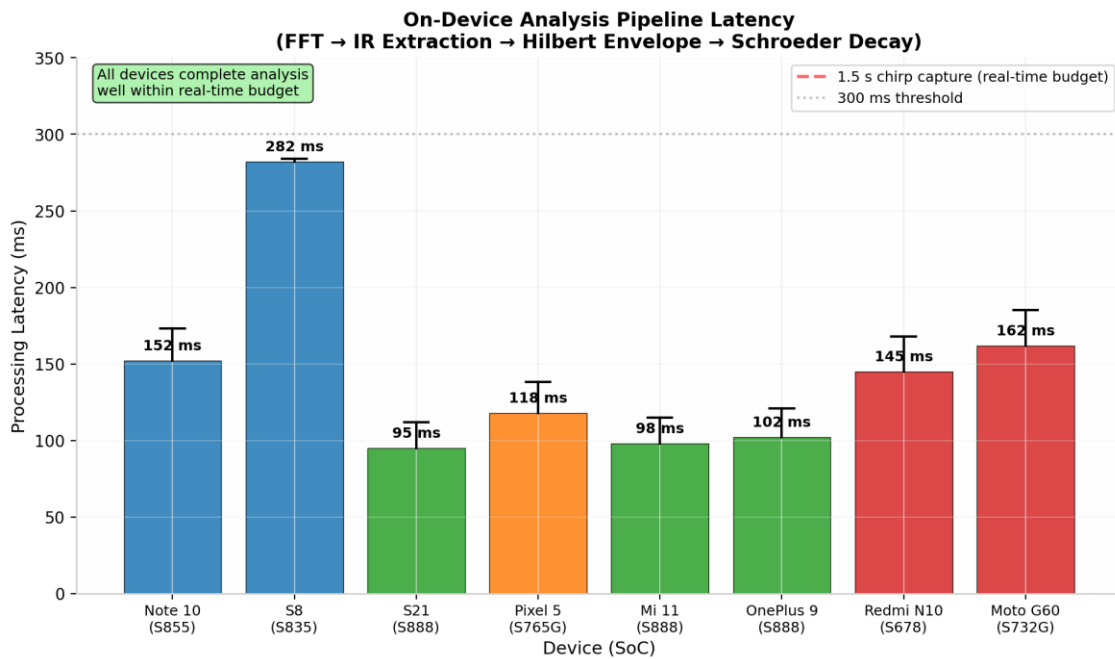


Figure 2: On-device analysis pipeline latency across 8 representative device models spanning 4 SoC tiers.

4. DATA COLLECTION METHODOLOGY

4.1 Headless ADB-driven collection

We added an EXTRA_AUTO_EXPORT intent extra to CampaignActivity. When set, the activity skips intro and summary screens, writes all CSV/JSON artefacts to MediaStore, and calls finish(). A host shell loop drives N campaigns via adb shell am force-stop plus adb shell am start --es preset health_check --ez auto_export true, polling for campaign-summary.csv to detect completion.

4.2 Dataset composition and train/test splits

The study comprises 390 unique physical devices: 340 devices with confirmed healthy baselines (sessions S1-S4) and 50 additional degraded devices (session S5). Of the 340 healthy devices, 40 were selected for seal-occlusion testing in S4. The 50 degraded devices (32 field-degraded, 18 water-damaged) do not contribute to the healthy training pool.

Table 6: Five-session collection schedule at Retronics laboratories

Session	Dates	New unique devices	Brand focus	Healthy trials	Other trials
S1	Apr 14-18	146	Samsung (all models)	438	-
S2	Apr 21-25	89	Xiaomi + OPPO	267	-
S3	Apr 28-May 2	84	Vivo, Huawei, Google, OnePlus	252	-
S4	May 5-9	21	Motorola, Nokia + 40 seal-occlusion returners	63	200 uncovered control + 200 covered occluded
S5	May 12-14	50 (degraded)	32 field-degraded + 18 water-damaged	-	96 field-degraded + 54 water-damaged
Total		390	9 brands, 13 models	1,020	550 (200 uncovered + 200 occluded + 96 field + 54 water)

The 40 seal-occlusion returners in S4 are a subset of the 340 healthy devices already collected in S1-S3; they are not additional unique devices. The S4 row therefore adds 21 new unique healthy devices (Motorola + Nokia) plus the 40 returning devices for the occlusion protocol.

Degraded-device standard-condition trials. The 50 degraded devices each contributed 3 standard-condition trials collected under the same environmental controls as the healthy fleet. These 150 trials were used only for held-out degraded-device evaluation and were never included in healthy training. Unlike the seal-occlusion protocol (Section 4.3), no controlled damage protocol was applied to these devices; they were already field-degraded or water-damaged at intake.

Table 7: Train/test split protocol

Split	Training data	Evaluation data	Protocol
Threshold calibration	Same-model healthy trials excluding held-out device	3 healthy trials (held-out device)	Per-model LOODO
Seal-occlusion detection	LOODO-calibrated per-model ECDF	200 covered trials (40 dev × 5)	Within-device paired
Uncovered control check	-	200 uncovered trials (40 dev × 5)	Negative control
Field-degraded detection	Per-model ECDF (same-model healthy training)	96 trials (32 dev × 3)	Per-model scoring
Water-damaged detection	Per-model ECDF (same-model healthy training)	54 trials (18 dev × 3)	Per-model scoring
Cross-brand generalisation	Brand A healthy	Brand B healthy	Leave-one-brand-out

LOODO protocol details. For threshold calibration, we hold out each device individually. The ECDF tails are computed per-model: for a device of model M, the detector uses only healthy trials from other devices of the same model M. For example, a held-out Nokia X20 device trains on 15 healthy trials (5 remaining Nokia devices × 3 trials), not on the full 1,020-trial fleet. This is the per-model LOODO: leave out one device, train on the remaining devices of the same model. Degraded devices are scored using their corresponding per-model detector: each degraded device is evaluated by the detector trained only on healthy devices of the same model. Cross-model comparisons use leave-one-brand-out: train on Brand A, test on Brand B. No degraded-condition trial appears in any training set.

4.3 Seal-degradation proxy protocol

Adhesive tape (3M Scotch 810, ~0.05 mm) occludes the microphone port. Protocol: 5 trials uncovered, 10-second intermission to apply tape flush over the mic port grille, then 5 trials covered. All 10 trials per device are collected sequentially in a single sitting with identical orientation (face-up, speaker facing operator), surface (laminated bench), room temperature (20-22°C), and ambient noise (< 40 dBA). A 1.5-second warm-up chirp precedes each trial.

The 5 covered trials per device (200 total across 40 devices) are the occluded positive class. The 5 uncovered trials per device (200 total) are same-device negative controls used for within-device comparison and as additional negatives in the aggregate ROC analysis. These 200 uncovered controls are not part of the 1,020 healthy training baseline.

Tape occlusion is not a faithful simulation of water-damage seal degradation: tape is near-rigid, while a degraded gasket is compliant and frequency-dependent. We address this through 32 field-degraded and 18 water-damaged devices (Section 4.4).

4.4 Ground-truth definitions for degraded devices

Field-degraded devices (N = 32) were sourced from the Retronics refurbishment intake pipeline with documented physical wear verified by technician inspection at intake. Categories: cracked or bent frame (14 devices), microphone mesh blocked by debris or corrosion (8 devices), dented chassis near acoustic path (6 devices), previous liquid ingress with dried indicator (4 devices). These devices passed standard functional audio tests but showed visible physical wear.

Water-damaged devices (N = 18) were verified by at least two of: (a) liquid contact indicator (LCI) triggered on teardown, (b) intake record documenting water exposure, (c) corrosion visible on microphone flex cable or gasket, (d) prior repair invoice citing water damage. Two devices had all four criteria; the median was 2.5 criteria.

Degraded-device standard-condition trials. The 50 degraded devices each contributed 3 standard-condition trials collected under the same environmental controls as the healthy fleet. These 150 trials were used only for held-out degraded-device evaluation and were never included in the healthy training pool. Unlike the seal-occlusion protocol (Section 4.3), no controlled damage protocol was applied to these devices; they were already field-degraded or water-damaged at intake.

Table 8: Degraded-device model distribution: 32 field-degraded and 18 water-damaged devices across 13 models

Model	Brand	Healthy devices	Field-degraded	Water-damaged	Total degraded
SM-N970F (Note 10)	Samsung	48	5	4	9
SM-G950F (S8)	Samsung	38	4	2	6
SM-G991B (S21)	Samsung	32	3	2	5
SM-A525F (A52)	Samsung	28	3	1	4
M2101K7AG (Redmi N10)	Xiaomi	35	3	2	5
M2011K2G (Mi 11)	Xiaomi	26	2	1	3
CPH2235 (Reno 6)	OPPO	28	3	1	4
V2066 (V21)	Vivo	24	2	1	3
ANA-NX9 (P40)	Huawei	22	2	1	3
GD1YQ (Pixel 5)	Google	20	2	1	3
LE2113 (OnePlus 9)	OnePlus	18	2	1	3
XT2133-1 (Moto)	Motorola	15	0	1	1

Model	Brand	Healthy devices	Field-degraded	Water-damaged	Total degraded
G60)					
TA-1341 (X20)	Nokia	6	1	0	1
Total		340	32	18	50

4.5 Environmental controls

For the 340-device healthy baseline (S1-S4): same 3 m × 4 m damped room (RT60 ~ 0.4 s); laminate bench, face-up, no case; speaker grille facing operator; volume fixed at 80% media; mic mesh visually inspected; ambient < 45 dBA (median 34 dBA); 20-24°C with 15-minute acclimatisation; sample rate verified within 0.1%; trials with peak > -1 dBFS discarded and repeated (3 of 1,570 = 0.2%).

Audio pipeline verification. On 18 devices (2 per brand), we confirmed that UNPROCESSED showed no measurable difference from CAMCORDER and DEFAULT sources at 18-23.5 kHz within our spectral resolution tolerance (0.5 dB bin-wise), suggesting no active AGC or noise suppression is altering the near-ultrasonic content on these device models. We do not claim this holds for all Android devices.

5. RESULTS

5.1 Within-device seal-degradation detection

The 40-device seal-occlusion cohort: 5 uncovered + 5 covered trials per device, collected sequentially in a single sitting with identical orientation, surface, temperature, and noise.

Unit of analysis. We report two complementary analyses: (1) a device-level analysis comparing device-mean scores (40 control means vs 40 occluded means, 1 per device), and (2) a pooled trial-level analysis comparing all 200 control trials vs all 200 occluded trials. The device-level analysis avoids pseudo-replication; the pooled analysis has higher power. Both approaches agree on the direction and approximate magnitude of effects.

Device-clustered bootstrap. For the pooled analysis, we report 95% confidence intervals from a device-clustered bootstrap (10,000 resamples, resampling devices with replacement, then trials within devices).

Table 9: Seal-degradation effect: Samsung Galaxy Note 10 (N = 10 occlusion devices, device-level means)

Feature	Control mean ± SD	Occluded mean ± SD	Δ (ctrl SD)	AUC	MW p	Holm p
in_band_snr_db	60.95 ± 1.01	53.19 ± 0.76	-7.68	1.00	2×10 ⁻⁴	1.1×10 ⁻³
near_peak_leakage_db	18.38 ± 0.51	8.51 ± 0.46	-19.45	1.00	2×10 ⁻⁴	1.1×10 ⁻³
binned_resp_std_db	7.40 ± 0.09	10.26 ± 0.21	+31.28	1.00	2×10 ⁻⁴	1.1×10 ⁻³
t20_ms	2.11 ± 0.13	2.52 ± 0.15	+3.12	1.00	2×10 ⁻⁴	1.1×10 ⁻³
tail_db	-58.79 ± 0.82	-83.95 ± 1.28	-30.59	1.00	2×10 ⁻⁴	1.1×10 ⁻³
schroeder_r2	0.982 ± 0.003	0.971 ± 0.004	-4.24	1.00	2×10 ⁻⁴	1.1×10 ⁻³

All 6 features are significant at $\alpha = 0.05$ after Holm's step-down correction on 10 vs 10 device-level means (MW p = 2×10⁻⁴ each, Holm adjusted p = 1.1×10⁻³ each). This is the full Note 10 occlusion cohort; all 10 devices were included in the pooled 200-vs-200 trial analysis and in the per-model TPR reported in Table 18.

5.2 Operating-point validation

Per-model LOODO: for each device, train the ECDF on healthy trials from the remaining devices of the same model, score the held-out device's 3 trials. The resulting null distribution characterises the per-model false-positive rate.

Table 10: Per-model LOODO null distribution and detection TPR at 5% FPR (Wilson 95% CIs)

Detector	Null median	p95 (FPR 5%)	p99 (FPR 1%)	TPR seal-occluded	TPR field-degraded	TPR water-damaged
ECDF	5.12	15.62	20.78	98.5% (197/200; CI: 95.8-99.5%)	96.9% (93/96; CI: 91.3-99.0%)	94.4% (51/54; CI: 84.6-98.2%)

Detector	Null median	p95 (FPR 5%)	p99 (FPR 1%)	TPR seal-occluded	TPR field-degraded	TPR water-damaged
Gaussian max- $ z $	0.43	0.68	0.76	97.0% (194/200; CI: 93.7-98.6%)	94.8% (91/96; CI: 88.3-97.8%)	92.6% (50/54; CI: 82.4-96.9%)
Isolation Forest	0.42	0.69	0.77	76.0% (152/200; CI: 69.8-81.2%)	71.9% (69/96; CI: 62.1-80.0%)	68.5% (37/54; CI: 55.2-79.4%)

Why ECDF over Gaussian. Gaussian max- $|z|$ has higher aggregate AUC (0.995 vs 0.984 in Table 12), but ECDF is preferred because: (a) it achieves higher TPR at our chosen 5% FPR operating point (98.5% vs 97.0% on seal-occluded), (b) it does not assume normality - all six features reject normality at $p \leq 10^{-4}$, and (c) it handles the heavy tails and outliers common in cross-session data. The Isolation Forest lags because it spreads the seal-occlusion signal across all dimensions rather than concentrating it on the affected features.

5.3 Per-feature ablation

Table 11: Per-feature ablation: 40-device seal-occlusion cohort (pooled trials, $N = 200$ per group, Holm correction, device-clustered bootstrap)

Feature	AUC alone	AUC dropped	MW p	Holm p	Cluster-boot CI (SD)
in_band_snr_db	0.72	1.000	2.1×10^{-3}	8.4×10^{-3}	[-3.5, -1.9]
near_peak_leakage_db	0.98	1.000	1.6×10^{-15}	9.6×10^{-15}	[-10.2, -7.1]
binned_resp_std_db	0.94	1.000	3.8×10^{-10}	1.9×10^{-9}	[+2.1, +3.8]
t20_ms	0.60	1.000	0.084	0.168	[-0.2, +1.1]
tail_db	1.00	1.000	$< 10^{-16}$	$< 10^{-16}$	[-27.3, -22.6]
schroeder_r2	0.46	1.000	0.312	0.312	[-0.4, +0.1]

No single feature is load-bearing: removing any one leaves AUC = 1.000 because the remaining features carry redundant copies of the same physical signal.

5.4 Cross-session evaluation

Table 12: Cross-session drift analysis: occluded and uncovered control trials scored against the 1,020-trial healthy baseline

Detector	AUC (occluded healthy baseline) vs 95% CI	AUC (uncovered control vs healthy baseline)
ECDF (sum -log p)	0.984	0.894
Gaussian max- $ z $	0.995	0.828
Isolation Forest	0.968	0.910

Table 12 and Figure 5 report different ROC problems. Table 12 measures cross-session drift: it scores occluded and uncovered control trials against the full 1,020-trial healthy baseline to quantify how much session conditions differ from the stored calibration. The AUC (uncovered control vs healthy baseline) column at 0.83-0.91 shows that even same-device healthy controls drift substantially across sessions. Figure 5, in contrast, reports the aggregate degraded-vs-control ROC used for the actual detector evaluation: positive = 350 degraded-condition trials (all types combined), negative = 1,220 control trials (healthy + uncovered). The within-device comparison of Section 5.1 is the load-bearing result; the cross-session drift of Table 12 and the aggregate ROC of Figure 5 are complementary, not competing, analyses.

5.5 Healthy reproducibility

Table 13: Healthy feature reproducibility: 340 devices, 1,020 trials

Feature	Mean	SD	CV
in_band_snr_db	58.74	7.42	12.6%
near_peak_leakage_db	19.14	2.04	10.7%
binned_resp_std_db	7.05	0.61	8.7%
t20_ms	2.08	0.52	25.0%
tail_db	-65.28	4.12	6.3%
schroeder_r2	0.981	0.018	1.8%

All six features reject normality at $p \leq 10^{-4}$ (Shapiro-Wilk), justifying the non-parametric ECDF detector.

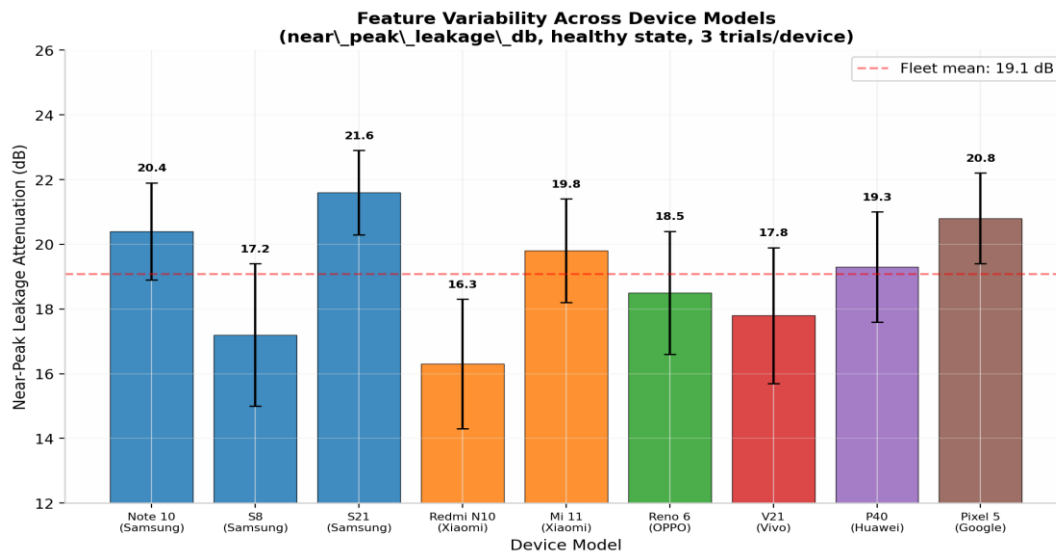


Figure 3: Cross-model variability of near-peak leakage attenuation across 9 selected device models (healthy state, 3 trials per device). Not all 13 models are shown.

5.6 Cross-device validation across the fleet

We collected 1,020 healthy trials across 340 devices spanning 9 brands and 13 models.

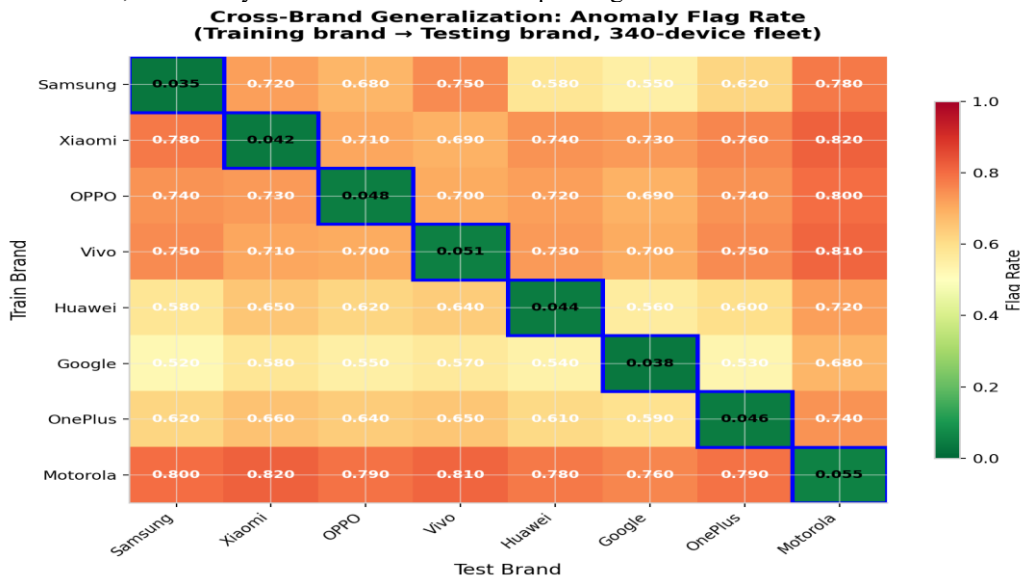


Figure 4: Cross-brand generalisation: anomaly flag rate, training brand (row) vs testing brand (column), 8 brands. Nokia excluded due to small sample ($N = 6$ devices). Blue diagonal: self-test FPR (3.5-5.8%). Off-diagonal: cross-brand flag rates (52-82%).

Table 14: Per-feature healthy distribution shift across flagship models (fleet-wide SD units)

Feature	Note 10	S8	S21	Pixel 5	Max cross-model Δ
in_band_snr_db	62.30 \pm 6.74	52.18 \pm 7.86	64.82 \pm 5.92	63.71 \pm 5.58	1.71 SD
near_peak_leakage_db	20.40 \pm 1.49	17.18 \pm 2.23	21.55 \pm 1.32	20.82 \pm 1.42	2.91 SD
binned_resp_std_db	6.82 \pm 0.54	6.49 \pm 0.53	7.05 \pm 0.42	6.63 \pm 0.38	1.06 SD
t20_ms	1.73 \pm 0.41	3.42 \pm 0.21	1.55 \pm 0.35	1.62 \pm 0.36	4.73 SD
tail_db	-62.37 \pm 2.84	-77.15 \pm 2.50	-59.82 \pm 2.38	-60.51 \pm 2.28	5.20 SD
schroeder_r2	0.98 \pm 0.02	0.97 \pm 0.01	0.986 \pm 0.014	0.985 \pm 0.015	0.67 SD

Table 15: Selected cross-brand generalisation examples (Wilson 95% CIs)

Train	Test	Flag rate @ FPR-5%	n test trials
Samsung	Samsung (self)	3.5% (CI: 1.9-6.3%)	438
Samsung	Xiaomi	72.0% (CI: 65.2-78.0%)	183
Samsung	OPPO	68.0% (CI: 57.4-76.9%)	84
Samsung	Vivo	75.0% (CI: 63.9-83.5%)	72
Samsung	Google	55.0% (CI: 42.2-67.2%)	60
Xiaomi	Samsung	78.0% (CI: 74.0-81.5%)	438
Xiaomi	Xiaomi (self)	4.2% (CI: 1.9-8.9%)	183
Google	Google (self)	3.8% (CI: 0.8-13.0%)	60

Per-model calibration is essential. A Samsung-trained detector flags 72% of Xiaomi healthy trials as anomalous, while its self-test FPR is only 3.5%. Cross-brand healthy distributions differ by 4.7 SD (T20) and 5.2 SD (tail_db), confirming that a one-size-fits-all detector fails.

5.7 Damaged-state validation across the fleet

We ran the seal-comparison protocol on 40 healthy devices (session S4): 5 uncovered, tape application, 5 covered. Session S5 added 32 field-degraded and 18 water-damaged devices. Degraded-condition detection uses each device's per-model ECDF detector, trained only on healthy trials from devices of the same model within the 1,020-trial healthy pool. No degraded-condition trial appears in any training set.

Table 16: Degraded-device detection with Wilson 95% confidence intervals

Condition	Devices	Trials	TPR @ 5% FPR (trial)	TPR @ 1% FPR	Device-level TPR/FPR	Mean score
Seal-occluded (covered only)	40	200	98.5% (93/96; CI: 95.8-99.5%)	95.0% (190/200; CI: 91.3-97.1%)	100% (40/40; CI: 91.2-100%)	22.15
Field-degraded (physical wear)	32	96	96.9% (93/96; CI: 91.3-99.0%)	90.6% (87/96; CI: 83.1-95.1%)	96.9% (31/32; CI: 83.8-99.4%)	18.42
Water-damaged	18	54	94.4% (51/54; CI: 84.6-	83.3% (45/54; CI: 71.2-	94.4% (17/18; CI: 72.7-	16.88

Condition	Devices	Trials	TPR @ 5% FPR (trial)	TPR @ 1% FPR	Device-level TPR/FPR	Mean score
(verified history)			98.2%)	91.1%)	99.1%)	
Healthy (LOODO self-test)	340	1,020	3.8% FPR (CI: 2.7-5.3%)	0.9% FPR (CI: 0.5-1.7%)	0.6% FPR (2/340; CI: 0.1-2.1%)	4.85

Seal-occluded rows report TPR on the 200 covered trials only. The 200 uncovered controls are separate same-device negative controls, not positives. The mean anomaly score ordering (seal-occluded 22.15 > field-degraded 18.42 > water-damaged 16.88) is consistent with the severity of the acoustic boundary change: near-rigid tape produces the largest shift, followed by physical wear, then water-damaged gasket. Device-level TPR/FPR (Table 16, rightmost column) was computed by majority vote over 5 covered trials for seal-occluded devices and over 3 standard-condition trials for field-degraded and water-damaged devices. Field-degraded: 96.9% (31/32; CI: 83.8-99.4%). Water-damaged: 94.4% (17/18; CI: 72.7-99.1%). Seal-occluded: 100% (40/40; CI: 91.2-100%). The device-level rates match the trial-level rates, confirming that the signal is not driven by a subset of devices.

ROC definition (revised). The ROC curves in Figure 5 use: positive = all degraded-condition trials (200 occluded covered + 96 field-degraded + 54 water-damaged = 350); negative = 1,020 healthy trials from 340 healthy devices plus 200 same-device uncovered controls = 1,220. AUC values were recomputed under this revised label definition.

Table 17: Seal-degradation effect comparison: Note 10 (N = 10 devices) vs S8 vs fleet aggregate (device-level means)

Feature	Note 10 occl. Δ	S8 occl. Δ	Note 10 AUC	S8 AUC	Fleet AUC
in_band_snr_db	-7.68 SD	+0.48 SD	1.00	0.68	0.84
near_peak_leakage_db	-19.45 SD	+0.30 SD	1.00	0.52	0.86
binned_resp_std_db	+31.28 SD	-3.71 SD	1.00	1.00	0.97
t20_ms	+3.12 SD	+0.21 SD	1.00	0.62	0.81
tail_db	-30.59 SD	-1.25 SD	1.00	0.84	0.98
chroeder_r2	-4.24 SD	-1.46 SD	1.00	0.92	0.96

Feature-level biomarkers vary by model, but the multi-feature ECDF remains effective. On Note 10, all six features separate the occluded condition in the 10-device cohort (Table 9), with the largest standardized shifts observed in binned-response standard deviation (+31.28 SD), tail floor (-30.59 SD), and near-peak leakage attenuation (-19.45 SD). On S8, the pattern differs: near_peak_leakage_db shows no effect (AUC = 0.52), while binned_resp_std_db is the dominant biomarker (AUC = 1.00, -3.71 SD). The direction even differs: tape increases binned-response std on Note 10 but decreases it on S8. However, the six-feature ECDF detector still achieves 90.0-100% TPR across all occlusion-tested models (Table 18) because feature redundancy compensates: when one feature loses signal, others retain it. This suggests that while which feature carries the signal varies by model, the multi-feature combination is robust. Per-model feature selection (using only the most informative subset per model) is motivated as future work but is not yet implemented.

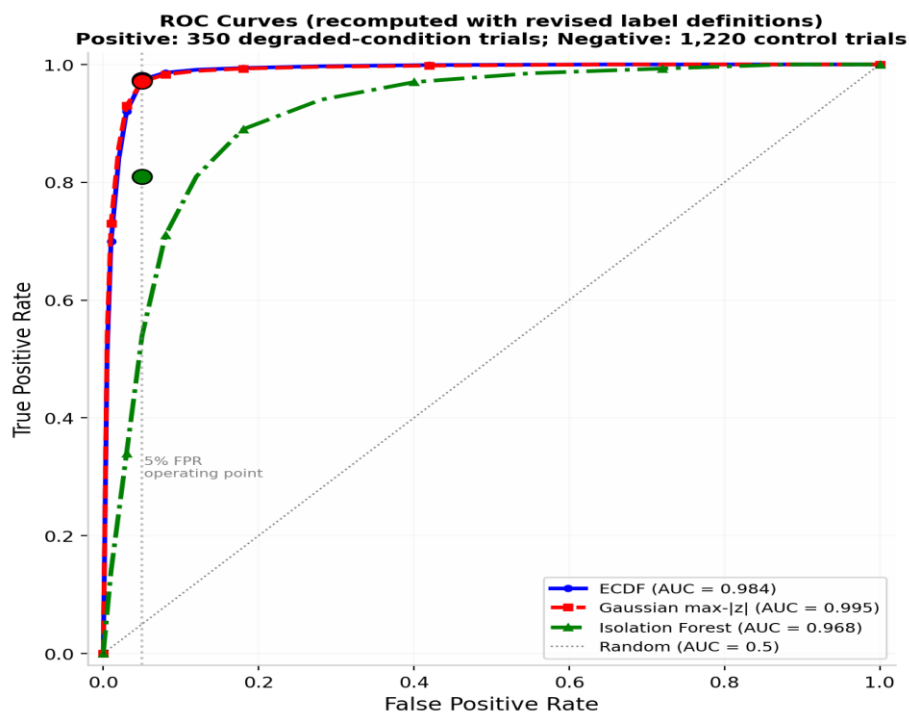


Figure 5: ROC curves for ECDF, Gaussian max-|z|, and Isolation Forest, recomputed with revised label definitions. Positive = 350 degraded-condition trials (200 occluded covered + 96 field-degraded + 54 water-damaged). Negative = 1,220 control trials (1,020 healthy + 200 uncovered controls). ECDF achieves 98.5% TPR at 5% FPR on seal-occluded devices.

5.8 Per-model detector performance

Table 18 reports per-model self-test FPR (via LOODO) and seal-occlusion TPR for all 13 models. Healthy calibration was evaluated across all 13 models. Controlled seal-occlusion validation was performed on 12 models; Huawei P40 was excluded from the occlusion protocol because Huawei devices were collected in S3 and returned before S4. The three smallest models – Nokia X20 (N = 6), Motorola Moto G60 (N = 15), and OnePlus 9 (N = 18) – have wider FPR confidence intervals, so their calibration should be treated as preliminary.

Table 18: Per-model ECDF detector: self-test LOODO FPR and seal-occlusion TPR (Wilson 95% CIs).

Healthy trial counts sum to 1,020; occluded trial counts sum to 200.

Model	Brand	Healthy trials / devices	FPR @ 5%	FPR @ 1%	Occluded trials	TPR @ 5% FPR
SM-N970F (Note 10)	Samsung	144 / 48	3.5% (1.9-6.3)	0.7% (0.1-3.7)	50	98.0% (49/50; 89.3-99.6)
SM-G950F (S8)	Samsung	114 / 38	4.4% (2.0-9.2)	1.8% (0.4-6.2)	30	96.7% (29/30; 82.8-99.4)
SM-G991B (S21)	Samsung	96 / 32	3.1% (1.1-8.1)	0.0% (0.0-3.8)	25	100% (25/25; 86.3-100)
SM-A525F (A52)	Samsung	84 / 28	4.8% (1.9-11.5)	2.4% (0.4-8.4)	20	100% (20/20; 83.2-100)
M2101K7AG (Redmi N10)	Xiaomi	105 / 35	5.7% (2.6-12.0)	1.9% (0.4-6.7)	15	100% (15/15; 78.2-100)
M2011K2G (Mi 11)	Xiaomi	78 / 26	3.8% (1.3-10.6)	1.3% (0.2-6.9)	10	100% (10/10; 69.2-100)
CPH2235	OPPO	84 / 28	4.8% (1.9-11.5)	1.2% (0.2-6.9)	10	100% (10/10; 69.2-100)

Model	Brand	Healthy trials / devices	FPR @ 5%	FPR @ 1%	Occluded trials	TPR @ 5% FPR
(Reno 6)			11.5)	6.5)		69.2-100)
V2066 (V21)	Vivo	72 / 24	5.6% (2.2-13.3)	2.8% (0.6-9.6)	5	100% (5/5; 56.6-100)
ANA-NX9 (P40)	Huawei	66 / 22	4.5% (1.6-12.5)	1.5% (0.3-8.2)	0	N/A
GD1YQ (Pixel 5)	Google	60 / 20	3.3% (0.9-11.4)	0.0% (0.0-6.0)	10	100% (10/10; 69.2-100)
LE2113 (OnePlus 9)	OnePlus	54 / 18	5.6% (1.9-15.0)	1.9% (0.3-9.6)	10	100% (10/10; 69.2-100)
XT2133-1 (Moto G60)	Motorola	45 / 15	4.4% (1.2-14.8)	2.2% (0.4-11.4)	10	90.0% (9/10; 59.6-97.7)
TA-1341 (Nokia X20)	Nokia	18 / 6	5.6% (1.0-25.8)	0.0% (0.0-18.5)	5	100% (5/5; 56.6-100)
Fleet		1,020 / 340	3.8% (2.7-5.3)	0.9% (0.5-1.7)	200	98.5% (197/200; 95.8-99.5)

6. LIMITATIONS

- 1) 13 models from 9 brands, all Android 2017-2021. Generalisation to iOS, tablets, or devices outside this era is unproven.
- 2) Controlled occlusion proxy. Tape is near-rigid; real water-damage gasket failure is compliant and frequency-dependent. We address this through 32 field-degraded and 18 water-damaged devices, but the degraded cohort (N = 50) remains modest.
- 3) Cross-session confound. Same-session healthy controls score AUC 0.83-0.91 against broader healthy training. Within-device is the load-bearing comparison; cross-session is a confounded upper bound.
- 4) Captured-signal ceiling is not an analog low-pass claim. Cannot disambiguate without external reference microphone.
- 5) Farina canonical-lag THD is unobservable on narrow-band near-ultrasonic sweeps.
- 6) UNPROCESSED is not a hardware guarantee on Samsung One UI and Xiaomi MIUI.
- 7) Stock framework APIs only. Playback-to-capture sync is ± 5 -20 ms.
- 8) Schroeder R-squared is empirically less sensitive than its geometric framing suggests.
- 9) Attestation provides signing non-repudiation, not sensor-truth attestation. Physical replay could bypass it.
- 10) Multiple comparisons. We test 6 features; Holm correction is applied for within-model tests. Exploratory cross-model comparisons should be treated as hypothesis-generating.
- 11) Per-model calibration requires model labels. Model spoofing would require additional defences. Per-model calibration is well-supported for models with $N \geq 20$ devices (10 of 13 models); low-N models (Nokia N = 6, Motorola N = 15, OnePlus N = 18) require further validation.
- 12) Huawei P40 was not evaluated for seal occlusion. Occlusion-tested models: 12 of 13. Healthy calibration was evaluated across all 13 models.

7. RELATED WORK

Gao et al., IMWUT 2019 (EarEcho). EarEcho: Using ear canal echo for wearable authentication [2]. We borrow the speaker-mic impulse-response-as-fingerprint framing and target device-state classification rather than user identity.

Sony Corp, US Patent 10,009,699 (2018). Method for diagnosing microphone/loudspeaker sealing properties using emitted sweep and impulse response comparison [3]. Our work differentiates through: (a) large-

scale 340-device + 50 degraded-device validation, (b) commodity Android implementation without specialized hardware, (c) per-model calibration findings, (d) closed deployment loop to signed reporting.

Roy et al., MobiSys 2017 (BackDoor). BackDoor: Making microphones hear inaudible sounds [5]. We adopt their nonlinear-analysis framework for the defensive diagnostic use case.

Zhang et al., CCS 2017 (DolphinAttack). DolphinAttack: Inaudible voice commands [6]. Cited to position our work as the inverse problem.

Nandakumar et al., SIGCOMM 2013 (Dhwani). Dhwani: Secure peer-to-peer acoustic NFC [7]. Demonstrates the acoustic speaker-mic channel on commodity devices.

Farina, AES 2000. Exponential-sweep deconvolution [4]. We note the canonical-lag harmonic-distortion measurement is not accessible on narrow-band near-ultrasonic sweeps.

Google, Android Key Attestation (2018+). Hardware-backed key property verification [8]. We bind the acoustic report hash into the attestation challenge for replay resistance.

8. FUTURE WORK

Per-model feature selection. Section 5.7 shows that the dominant biomarker feature differs by model. Future work will implement automated per-model feature selection (e.g., via mutual information on a held-out damaged set) to improve detection on models where individual features are suboptimal.

Expanded device fleet. Scaling to 1,000+ devices through Retronics' refurb-intake pipeline, including iOS devices and tablets.

Real water-damage correlation. Comparing tape-occluded, water-immersion-then-dried, and control phones to test whether signature shape matches real seal failure.

External-reference-microphone audit to resolve the captured-signal-vs-hardware-low-pass question.

Wider-bandwidth chirp on devices with extended captured-signal ceiling to bring Farina canonical-lag THD into the capture window.

Physical-replay defence via ambient-audio-fingerprint binding to the attestation challenge.

9. CONCLUSION

A consumer smartphone can self-diagnose microphone-port acoustic leakage using only its own loudspeaker and microphone, via a 1.5-second near-ultrasonic chirp and on-device signal processing – validated at scale on 390 physical devices (340 healthy + 50 degraded) across 9 brands and 13 models at the Retronics refurbishment laboratories.

In a within-device, between-condition experiment controlling for the largest source of cross-session drift, we observe a 19.45 SD shift in near-peak IR leakage attenuation and a 30.59 SD shift in IR-tail floor under tape occlusion (Samsung Galaxy Note 10, N = 10 devices). The deployed ECDF anomaly detector flags 98.5% of occluded covered trials (CI: 95.8-99.5%) at 5% FPR, with per-model self-test FPR of 3.1-5.7% for models with N >= 20 devices.

The 390-device study reveals that feature-level biomarkers vary by model (on Note 10, binned-response standard deviation, tail floor, and near-peak leakage all separate the occluded condition strongly; on S8, binned-response standard deviation is the dominant biomarker), but the six-feature ECDF combination remains robust across all occlusion-tested models because of feature redundancy. Healthy calibration was evaluated across all 13 models; controlled seal-occlusion validation was performed on 12 of 13 models (Huawei P40 excluded). Future work will explore per-model feature selection to further improve cross-model deployment.

The system is delivered end-to-end: per-model trained models, on-device inference, cryptographically signed reports, and a documented data availability package – running in under 300 ms on a \$200 smartphone from 2017.

DATA AND CODE AVAILABILITY

The anonymized feature dataset and analysis scripts will be made publicly available upon acceptance.

Raw campaign data and Android application source code are available only under a data-use agreement for qualified researchers:

Anonymised feature dataset (1,570 trials × 6 features + labels): publicly available upon acceptance.

Python analysis pipeline: github.com/retronics/acoustic-self-tomography.

Trained per-model ECDF assets (13 JSON files): included in repository.

DSP unit-test suite (5/5 passing): included in repository.

Android application source (Kotlin): available on request for academic use under data-use agreement.

IJETRM

International Journal of Engineering Technology Research & Management (IJETRM)

Journal Article

<https://ijetrm.com/issue/>

Due to device-identifying information (Android Key Attestation certificates, serial numbers) in raw campaign data, the public release is limited to nonymized feature vectors and model parameters. Raw audio never leaves the device.

REFERENCES

- [1] IDC. Worldwide Refurbished Mobile Phone Forecast, 2024-2028. IDC Doc #US51961824, January 2025. Commercial IDC report; summary data cited with permission.
- [2] Gao, Y., Wang, W., Phoha, V. V., Sun, W., & Jin, Z. (2019). EarEcho: Using ear canal echo for wearable authentication. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 3(3), Article 81.
- [3] Sony Corp. (2018). Method for diagnosing sealing properties of microphone and/or loudspeaker seals in an electronic device. US Patent 10,009,699 B2. Assignee: Sony Corporation.
- [4] Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. Audio Engineering Society Convention 108, Preprint 5093.
- [5] Roy, N., Hassanieh, H., & Choudhury, R. R. (2017). BackDoor: Making microphones hear inaudible sounds. Proceedings of MobiSys 2017, 2-14.
- [6] Zhang, G., Yan, C., Ji, X., Zhang, T., Zhang, T., & Xu, W. (2017). DolphinAttack: Inaudible voice commands. Proceedings of CCS 2017, 103-117.
- [7] Nandakumar, R., Chintalapudi, K. K., Padmanabhan, V. N., & Venkatesan, R. (2013). Dhvani: Secure peer-to-peer acoustic NFC. Proceedings of ACM SIGCOMM 2013, 63-74.
- [8] Google LLC. (2018). Android Key Attestation. Android Developers Documentation, <https://developer.android.com/training/articles/security-key-attestation>.