On-device fingertip HRV against a chest strap: first 30 participants

Report date 2026-09-19 · registered estimator/gate system v5 (frozen 2026-09-18) · 303 paired readings recorded, 265 accepted · prepared from the study pipeline in ppgcore/benchmarks/polar_study

PARTICIPANTS	HEART RATE	RMSSD	READINGS ≥ 20 MS OFF	SDNN	WITHHELD BY GATE
30	**0.47**	**4.03**	**0**	**4.13**	**3.3%**
21 iOS · 9 Android · 20 phone models	bpm MAE · r 0.999	ms MAE · r 0.975 · 92% within 10 ms	of 265 accepted (5 of 274 with the gate off)	ms MAE · r 0.984	9 of 274 usable readings

Summary. Thirty volunteers each took ten 60-second fingertip readings with the NEUROFIT app, five at rest and five during paced breathing at 6 breaths/min, while wearing a Polar H10 chest strap that recorded beat-to-beat intervals. Of the 303 readings, 277 passed the app's on-device signal-quality check, 3 were excluded because the strap record itself contained too many ectopic beats to define RMSSD, and the consistency gate withheld 9 more, leaving 265 accepted readings. On those, heart rate agreed with the strap to 0.47 bpm, RMSSD to 4.03 ms with 95% limits of agreement of -9.2 to +11.7 ms and no reading off by 20 ms or more, and SDNN to 4.13 ms. Agreement was consistent across platform, condition, skin tone, gender and age, with the caveat that the sample is concentrated at Monk tones 2–6 and ages 25–34. The iOS path has been tested on an independent holdout of 5 participants; the Android-specific parts of v5 have not yet been tested out of sample.

1. Methods

Protocol

Participants sat with the phone flat and the index finger covering the rear camera and flash. Each session comprised five rest readings and five paced-breathing readings at 6 breaths/min following the in-app pacer, 60 s each. The Polar H10 was worn per manufacturer instructions and streamed RR intervals to the phone over Bluetooth; the app uploaded the camera signal, the strap intervals and per-session demographics (gender, age range, Monk skin tone) anonymously. Participants were recruited from 2026-09-10 onward; the two investigators are excluded from all participant statistics because the estimator was developed on their readings.

Reference

Strap RR intervals were cleaned of isolated premature beats and their compensatory pauses before computing reference RMSSD/SDNN, using a fixed detector (short interval < 0.78× local median with normal neighbours, or short + pause pair). Records in which ≥ 5% of beats were flagged were excluded as invalid RMSSD references (HRV Task Force convention; RMSSD is a sinus-rhythm measure). This removed 3 readings.

Camera pipeline and gate (v5)

Beat timing from the red channel at 30 fps (Android) or 60 fps (iOS) with the production ppgcore engine. Reported RMSSD is the production estimate plus an isolated merged-interval correction; on Android it is the geometric mean of the parabolic- and Hilbert-refined estimates when they agree within 1.5×, else the Hilbert estimate, with a re-timing repair for deep band compression that applies only above 12 ms. A consistency gate withholds readings whose beat series shows a jump > 26% of the median interval together with high-frequency spectral energy not corroborated by an independent beat-timing method. Withheld readings are reported to the user as an unsuccessful reading.

Statistics

Mean absolute error (MAE), mean absolute percentage error, bias (camera − strap), Pearson r, Bland–Altman 95% limits of agreement (bias ± 1.96 SD), share of readings within 10 ms, and count of readings ≥ 20 ms off ("catastrophic"). Strata are reported without inferential tests given the small cell sizes.

2. Overall agreement, accepted readings (n = 265)

METRIC	N	MAE	MAPE	BIAS	R	95% LOA	NOTES
Heart rate (bpm)	265	0.47	0.6%	-0.24	0.999	-1.4 to +0.9	90% within 1 bpm, 99% within 2 bpm
RMSSD (ms)	265	4.03	10.7%	+1.25	0.975	-9.2 to +11.7	92.5% within 10 ms; 0 readings ≥ 20 ms
SDNN (ms)	265	4.13	7.0%	+1.10	0.984	-9.8 to +12.0	p90 |error| 9.2 ms
Baevsky stress index (ln)	265	0.18	4.1%	-0.04	0.970	-0.51 to +0.42	76% of readings within ±25% of strap SI, 92% within 1.5×
Poincaré SD2:SD1	265	0.36	12.3%	-0.05	0.852	-1.05 to +0.94	ratio units; error tracks RMSSD error
Respiratory rate (bpm), paced 6 bpm, confidence-gated	95	0.46	—	+0.14	—	—	89% within 1 bpm, 97% within 2 bpm, 0 gross misses; ungated n=104 MAE 0.73, 3 misses

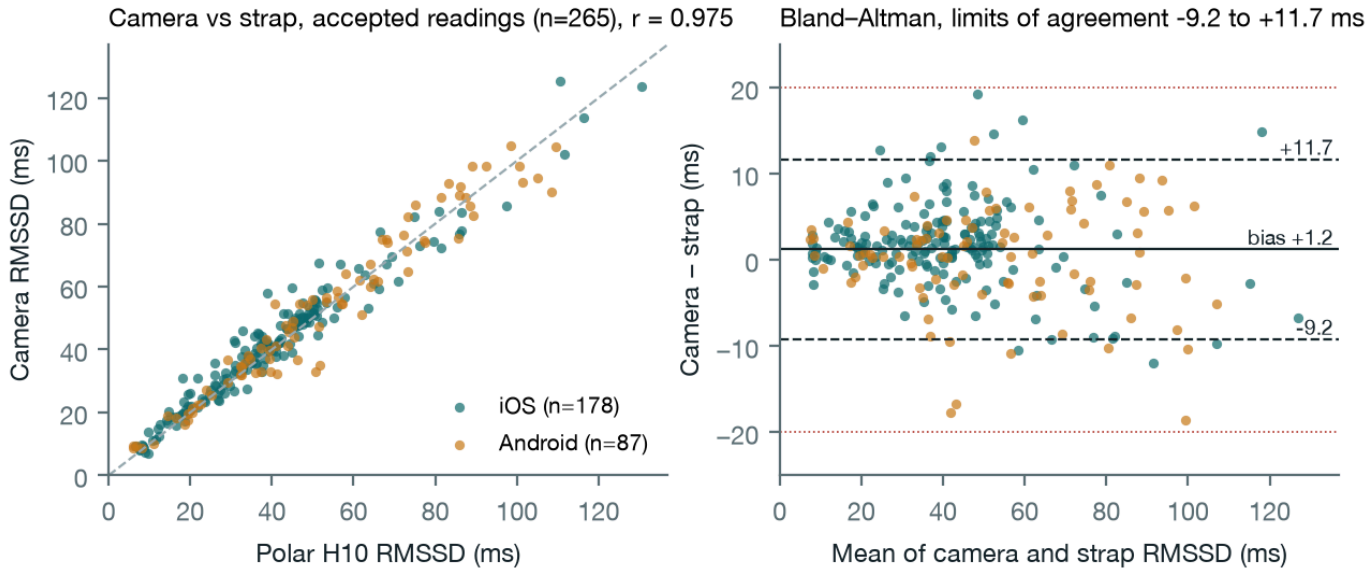

Figure 1. Left: camera RMSSD against strap RMSSD for every accepted reading; dashed line is identity. Right: Bland–Altman plot of the same readings; solid line bias, dashed lines 95% limits of agreement, dotted lines the ±20 ms catastrophe threshold.

3. Stratified agreement

Each row reports the accepted readings in that stratum: number of participants and readings, heart-rate MAE, and RMSSD MAE, bias, r, 95% limits of agreement, share within 10 ms and count ≥ 20 ms off. Cells with one or two participants describe those individuals, not the stratum.

STRATUM	PARTICIPANTS	READINGS	HR MAE (BPM)	RMSSD MAE (MS)	BIAS (MS)	R	LOA (MS)	WITHIN 10 MS	≥ 20 MS
PLATFORM									
iOS	21	178	0.51	3.76	+1.77	0.974	-7.7 to +11.3	93.3%	0
Android	9	87	0.37	4.58	+0.17	0.974	-11.7 to +12.1	90.8%	0
CONDITION									
rest	30	134	0.37	3.79	+1.40	0.961	-8.4 to +11.2	94.0%	0
paced breathing	30	131	0.57	4.27	+1.09	0.976	-10.0 to +12.2	90.8%	0
QUALITY TIER									
SQ3	30	244	0.43	3.92	+0.95	0.976	-9.4 to +11.3	93.0%	0
SQ2	12	21	0.95	5.32	+4.68	0.974	-4.2 to +13.5	85.7%	0
MONK SKIN TONE									
Monk 2	4	33	0.43	2.83	+1.88	0.970	-4.2 to +8.0	97.0%	0
Monk 3	3	28	0.46	1.60	+0.70	0.974	-3.2 to +4.6	100.0%	0
Monk 4	5	46	0.49	4.56	-0.55	0.975	-13.2 to +12.1	89.1%	0
Monk 5	9	79	0.51	4.23	+1.65	0.968	-9.0 to +12.3	88.6%	0
Monk 6	6	53	0.41	5.25	+1.53	0.965	-10.9 to +14.0	92.5%	0
Monk 8	3	26	0.45	4.10	+2.40	0.943	-7.0 to +11.8	96.2%	0
GENDER									
female	12	105	0.45	4.31	+1.35	0.971	-9.6 to +12.3	90.5%	0
male	18	160	0.48	3.85	+1.18	0.977	-8.9 to +11.3	93.8%	0
AGE RANGE									
18-24	1	8	1.02	2.07	+1.72	0.783	-2.2 to +5.6	100.0%	0
25-34	21	181	0.46	4.03	+1.13	0.971	-9.3 to +11.6	93.9%	0
35-44	5	47	0.44	4.77	+0.66	0.974	-11.3 to +12.7	89.4%	0
45-54	1	10	0.17	1.28	+0.43	0.108	-2.8 to +3.6	100.0%	0
55-64	2	19	0.52	4.48	+4.04	0.984	-5.4 to +13.4	78.9%	0

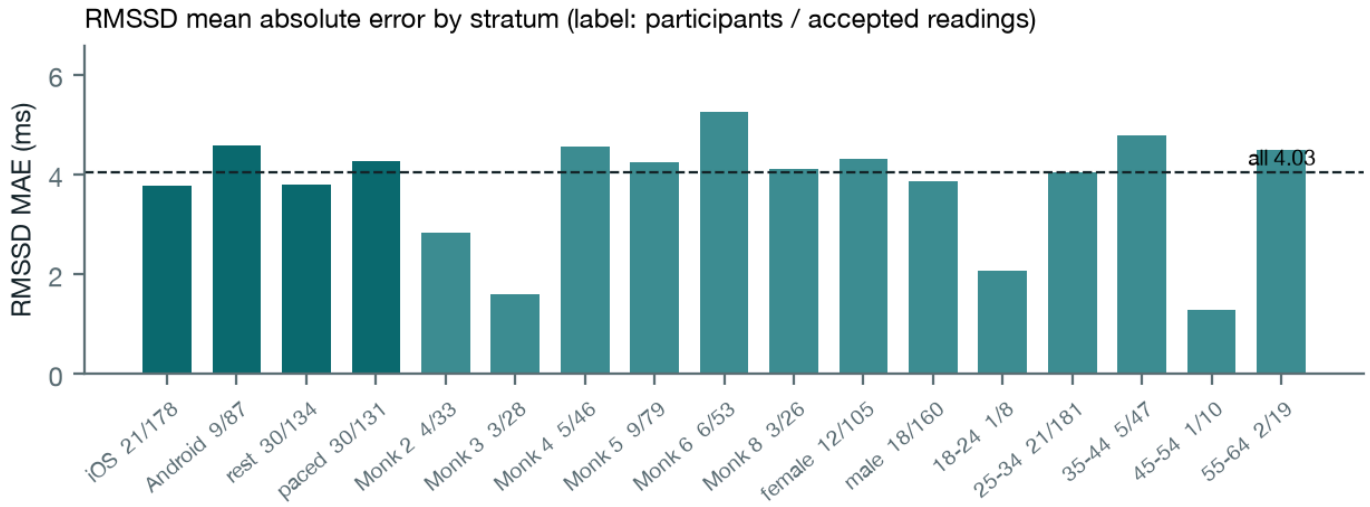

Figure 2. RMSSD MAE by stratum; dashed line is the all-readings value. Labels give participants and accepted readings per stratum.

4. Independent test (holdout) status

System constants were frozen in two steps and every participant enrolled after a freeze is scored blind against the frozen system. The iOS path was frozen on P1–P23 (2026-09-16) and is unchanged in v5; its holdout is every iOS participant from P24 onward. The Android-specific components of v5 were frozen on P1–P29 (2026-09-18) after the first two Android holdout participants exposed two failure modes in the previous system (a missed over-read on a Galaxy S25 Ultra and an estimator blow-up on a moto g play); their Android holdout begins at P30 and is currently empty.

HOLDOUT SET	PARTICIPANTS	USABLE	ACCEPTED	HR MAE	RMSSD MAE	BIAS	R	LOA	WITHIN 10 MS	≥ 20 MS	WITHHELD	SDNN MAE	RR (GATED)
iOS, P24–P30 (P24, P25, P26, P28, P30)	5	41	40	0.38	3.13	+2.33	0.978	-4.5 to +9.2	97.5%	0	1	2.55	n=15, MAE 0.39, 100% within 1
Android, P30 onward	0	no Android participant has enrolled since the v5 freeze											

> Every headline figure in sections 2 and 3 is therefore a development-set result for the Android branch and a mixed development/holdout result for iOS. The iOS holdout alone (40 accepted readings, 0 readings ≥ 20 ms) bounds the iOS catastrophe rate below about 7% at 95% confidence; ten holdout participants per platform would bring both bounds near 3% and give the Android gate four to five genuine out-of-sample stress tests.

5. Coverage and withheld readings

STAGE	READINGS	SHARE
Recorded, 30 participants	303	100%
Passed on-device signal quality (usable)	277	91.4%
Strap reference valid (< 5% ectopic beats)	274	90.4%
Accepted by consistency gate	265	87.5%

The gate withheld 9 of 274 reference-valid readings (3.3%; iOS 4.3%, Android 1.1%). Applied to the production reading mix and the existing quality-check failure rate, the projected share of production readings ending without a number is about 8.8%.

Effect of the gate

With the v5 estimator but no gate, 5 of the 274 reference-valid readings were off by 20 ms or more (RMSSD MAE 4.84 ms, r 0.945, LoA -13.3 to +17.6 ms); with today's production estimator and no gate the count is 14. The gate withheld 9 readings and no accepted reading is off by 20 ms or more (MAE 4.03 ms, r 0.975, LoA -9.2 to +11.7 ms). None of the 9 withheld readings had an error below 10 ms.

Reference exclusions

3 readings were excluded because ≥ 5% of strap beats were ectopic; in one of them the strap RMSSD was 148 ms against 14–28 ms on the same participant's other rest readings, with the camera at 38.5 ms. This is a limitation of RMSSD as a measure during frequent ectopy rather than of either device.

6. Sample composition

GENDER	PARTICIPANTS	AGE RANGE	PARTICIPANTS	MONK TONE	PARTICIPANTS
male	18	**18-24**	1	**Monk 2**	4
female	12	**25-34**	21	**Monk 3**	3
		35-44	5	**Monk 4**	5
		45-54	1	**Monk 5**	9
		55-64	2	**Monk 6**	6
				Monk 8	3

PLATFORM	PHONE MODEL	PARTICIPANTS
Android	24090RA29G	1
Android	Pixel 8	1
Android	SM-A156U	1
Android	SM-A235F	1
Android	SM-G991U1	1
Android	SM-S906U1	1
Android	SM-S911U	1
Android	SM-S938U	1
Android	moto g play - 2024	1
iOS	iPhone18,2	5
iOS	iPhone17,1	3
iOS	iPhone17,2	3
iOS	iPhone16,1	2
iOS	iPhone18,1	2
iOS	iPhone13,4	1
iOS	iPhone14,2	1
iOS	iPhone14,5	1
iOS	iPhone15,5	1
iOS	iPhone17,4	1
iOS	iPhone18,3	1

7. Limitations and next steps

- **Holdout.** The Android branch of v5 has no independent test yet; the iOS holdout is five participants. Target: ten holdout participants per platform, scored blind, with no constant changed until they are in.
- **Sample.** No participant at Monk tones 7, 9 or 10; 21 of 30 participants aged 25–34 and 3 aged 45 or over. Strata with one or two participants describe individuals. Recruitment should prioritise Monk 7+, ages 45+, high-HRV Android users and budget Android phones.
- **Bias at low HRV.** iOS reads 2–3 ms high below 60 ms of reference RMSSD; for participants under 25 ms this is the dominant error (MAPE ≈ 17%). No dataset-fitted correction has been applied.
- **Respiratory rate** is validated only at paced 6 breaths/min, where the strap-derived reference is reliable; resting rate has no usable reference in this design. The confidence gate that removes harmonic-lock errors also withholds about 10% of paced readings.
- **Ectopy.** Frequent ectopic beats invalidate RMSSD on both devices. A camera-side irregular-rhythm detector based on beat-interval statistics was evaluated and does not work at this reading length; pulse-morphology approaches remain open.
- **Single site, seated protocol, 60 s readings, one breathing rate.** Results should not be extrapolated to motion, ambulatory use or other reading durations.

Data and code: paired uploads in the validation bucket; analysis scripts and frozen-system definitions in ppgcore/benchmarks/polar_study (README documents the pipeline, registered systems v2–v5 and the reference-quality rule).