Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mental Health Effects of a Wearable Heartbeat-Synchronized Haptic Device

Nathan Whitmore

nathanww@media.mit.edu

MIT https://orcid.org/0000-0001-7624-2725

Minsol Kim

MIT

Phoebe Chua

MIT

Yuyang Zhang

MIT

Serena Pei

MIT https://orcid.org/0009-0008-9655-6519

Isabel Wellins

MIT

Anan Afrida

MIT

Pattie Maes

MIT Media Lab

Article

Keywords:

Posted Date: June 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9918164/v1

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

Additional Declarations: There is **NO** Competing Interest.

Mental Health Effects of a Wearable Heartbeat-Synchronized Haptic Device

MINSOL KIM*, Massachusetts Institute of Technology, USA
NATHAN W WHITMORE*†, Massachusetts Institute of Technology, USA
PHOEBE CHUA, Massachusetts Institute of Technology, USA and National University of Singapore, Singapore
YUYANG S ZHANG, Massachusetts Institute of Technology, USA and Wellesley College, USA
ISABEL WELLINS, Massachusetts Institute of Technology, USA
SERENA PEI, Massachusetts Institute of Technology, USA
ANAN AFRIDA, Massachusetts Institute of Technology, USA and Wesleyan University, USA
PATTIE MAES, Massachusetts Institute of Technology, USA

Interoception–the capacity to sense and process internal bodily signals–is strongly linked to emotional experience and mental health, but whether amplifying access to bodily signals causally improves mental well-being remains unknown. We conducted a preregistered, within-subject wearable intervention testing whether everyday cardiac haptic feedback improves psychological outcomes. Participants wore a customized watch delivering either real-time heartbeat-synchronized feedback or temporally shifted sham feedback for two weeks, with repeated measures of mental health, interoceptive accuracy, awareness, confidence, and daily mood. Feedback condition showed no significant effect on mental health, interoceptive accuracy, confidence, daily mood, or mood variability. However, exploratory analyses revealed that baseline anxiety significantly moderated the effects of feedback: individuals with higher baseline anxiety showed significantly greater reductions in anxiety, depressive symptoms, and general distress, with significantly larger effects under sham feedback than real feedback. Baseline general distress also significantly moderated changes in depressive symptoms across conditions. Symptom improvements were not accompanied by significant increases in interoceptive accuracy, confidence, or awareness; These findings provide the first causal, in-the-wild test of interoceptive amplification on mental health; our results suggest it may be helpful for anxiety but appears to work by shifting the interpretation of interoceptive information rather than interoceptive accuracy as initially hypothesized

1 Introduction

Emotional experience and psychological well-being are influenced in part by how individuals perceive and interpret signals arising within their bodies. A growing body of research suggests that disruptions in these internal sensing processes are associated with a range of psychiatric conditions, including anxiety disorders, depression, and alexithymia [17, 23, 34]. These internal sensing processes are collectively referred to as *interoception*–the perception and processing of internal bodily signals, including visceral and somatic sensations [8, 17].

Several theoretical frameworks suggest that interoception may play a foundational role in emotional experience and mental health. Interoceptive signals have been proposed to form a basis for subjective feelings and emotional experience [5, 7]. According to the *somatic marker hypothesis*, bodily signals contribute to emotional evaluation and decision-making by providing affective feedback that informs cognitive processes [9]. Similarly, *predictive processing* accounts of interoception suggest that emotional experiences arise, in part, from actively inferred predictive models of the causes of internal bodily signals [29]. Related accounts further propose that interoceptive experience may largely reflect predictions about the expected state of the body, and that disruptions in these interoceptive predictions may increase vulnerability for mental and physical illness [5]. Within these perspectives, disruptions in the perception or interpretation

*Both authors contributed equally to this research.
†Corresponding author:nathanww@media.mit.edu

of bodily signals could impair emotional awareness and regulation. Alexithymia, for example—characterized by difficulty identifying and describing emotions—has been hypothesized to reflect reduced access to or impaired interpretation of internal bodily cues [15]. Together, these accounts suggest that interoception may be one mechanism linking bodily signals with emotional experience and psychological well-being.

Empirical findings broadly link interoception to mental health, while also suggesting that this relationship is complex. Behavioral studies using tasks such as heartbeat detection have reported associations between interoceptive processing and emotional traits including anxiety, depression, and alexithymia [22, 26, 33]. However, different dimensions of interoception–including perceptual accuracy, subjective bodily awareness, and metacognitive confidence–do not necessarily correlate, suggesting that interoceptive processing is multidimensional and may influence emotional functioning through multiple pathways [12, 21]. Because most existing work is correlational, it remains unclear whether improving interoceptive processing can causally influence psychological well-being.

Several approaches have attempted to strengthen interoception, but they have limitations for isolating its causal role in mental health. Mindfulness-based practices, for example, often involve sustained attention to bodily sensations and have been associated with improvements in emotional regulation and psychological outcomes [11]. However, these interventions typically involve multiple components, including attentional training, cognitive reappraisal, and stress reduction, which makes it difficult to attribute the observed effects specifically to changes in interoception. Prior experiments have also manipulated perceived cardiac signals using artificially accelerated or delayed heartbeat feedback, demonstrating that perceived bodily signals can influence emotional judgments [32, 35]. Yet, these studies manipulate perceived physiological feedback rather than directly improving individuals' awareness of their internal bodily state.

Wearable technology offers a way to study interoception as it unfolds in everyday life. Wearable devices capable of detecting physiological signals and synchronizing feedback to those signals allow internal bodily activity to be rendered perceptually salient through modalities such as vibration. By amplifying real-time feedback aligned with ongoing physiological activity, such systems may increase access to naturally occurring bodily signals. Initial research suggests that heartbeat-synchronized haptic feedback can influence emotional responses under certain conditions, with exploratory evidence indicating that effects may depend on individual differences in interoceptive self-regulatory capacity [18]. However, existing studies have primarily examined short-term emotional responses rather than whether longitudinal amplified feedback in everyday life affects mental health outcomes, for whom, and through what mechanisms.

In sum, to our knowledge, no prior studies have experimentally tested whether enhancing access to interoceptive signals in everyday life through prolonged heartbeat-synchronized haptic feedback can causally influence mental health outcomes. Most existing work examining interoception and psychological well-being has relied on cross-sectional associations or short-term laboratory manipulations. Consequently, it remains unclear whether modulating interoceptive experience in daily life can produce measurable improvements in psychological well-being.

We therefore conducted a preregistered (AsPredicted #248,039), sham-controlled wearable intervention to test whether increasing the perceptual salience of cardiac signals in everyday life influences mental health, interoceptive accuracy, self-reported interoceptive awareness, and mood dynamics. Participants wore a smartwatch delivering either real-time heartbeat-synchronized vibrations (real condition) or temporally shifted sham vibrations (sham condition) derived from their own cardiac activity. The sham condition preserved comparable sensory input while disrupting real-time synchronization, allowing us to isolate the contribution of cardiac synchrony beyond nonspecific vibration exposure or expectancy effects [28, 36]. Across two weeks of everyday use, we assessed validated mental health outcomes, cardiac interoceptive accuracy and confidence, self-reported interoceptive awareness, and ecological momentary assessment of

Table 1. Overview of preregistered research questions, hypotheses, and outcome measures.

RQ	Domain	Hypotheses	Primary measures
RQ1	Mental health outcomes	**H1-a:** Real feedback will lead to significantly greater improvements in one or more mental health metrics than sham feedback. **H1-b:** The effect of feedback condition on mental health outcomes will be significantly moderated by one or more of the following: baseline TAS-20 alexithymia score, baseline MAIA-2 self-regulation subscale score, baseline GHQ-12 score, wear time in the prior week, change in SensHB accuracy, and change in heartbeat-count confidence. **H1-c:** Real–sham difference scores in GHQ-12 will be significantly associated with real–sham difference scores in MAIA-2 overall score, indicating that reduced psychological distress under real feedback co-occurs with increased interoceptive awareness.	CES-D, GAD-7, GHQ-12, SHAPS, TAS-20
RQ2	Cardiac interoceptive accuracy	**H2-a:** Real feedback will lead to significantly greater heart rate interoceptive accuracy than sham feedback. **H2-b:** The effect of feedback condition on interoceptive accuracy will be significantly moderated by one or more of the following: baseline TAS-20 alexithymia score, baseline MAIA-2 self-regulation subscale score, wear time in the prior week, and change in heartbeat-count confidence. **H3:** Confidence in heartbeat counting will be significantly higher in the real condition than in the sham condition.	SensHB slope, confidence ratings
RQ3	Interoceptive awareness development	**H4-a:** MAIA-2 total and/or Noticing and Body Listening subscale scores will be significantly higher at the final lab visit than at the first lab visit, regardless of condition order. **H4-b:** These scores will remain significantly elevated one week after the final lab visit compared to baseline.	MAIA-2 total, Noticing subscale, Body Listening subscale
RQ4	Daily mood dynamics	**H5:** Real feedback will lead to significantly higher average BMIS total scores across the week than sham feedback. **H6:** Real feedback will lead to significantly greater variability in BMIS total scores across the week than sham feedback.	BMIS mean, BMIS SD

Hypotheses are presented using preregistered formulations (AsPredicted #248,039), with minor grammatical edits for readability. Between-condition differences refer to differences between the real and sham feedback conditions.

mood. We also explored whether individual differences predicted the response to the intervention. The preregistered research questions, hypotheses, and outcome measures are summarized in Table 1.

2 Results

We preregistered four research questions and hypotheses (Table 10) after the start of data collection but prior to any data analysis (AsPredicted #248,039).

RQ1 focuses on whether real-time heartbeat-synchronized feedback, relative to sham feedback, improves mental health outcomes (H1-a), and whether the effect is moderated by baseline measures, including general distress (GHQ-12), alexithymia (TAS-20), interoceptive accuracy or confidence, change in overall interoceptive awareness (MAIA-2) or interoceptive self-regulation (MAIA-2 subscale), or dosage measured by average everyday wear time (H1-b). We also examined whether the condition difference on GHQ-12 would significantly covary with changes in overall MAIA-2 (H1-c). We focused on the self-regulation subscale, following a prior study [18] indicating self-regulation significantly modulated cardiac-synchronized haptic effects on emotion regulation, though the experiment was done in a one-time session.

RQ2 examines whether the feedback condition affects interoceptive accuracy outcomes. Specifically, we tested whether receiving real feedback would significantly improve cardiac interoceptive accuracy measured by the SensHB metric (4.1) compared to sham feedback (H2-a). Like H1-b, we also tested if this was moderated by the baseline individual differences in alexithymia, MAIA-2, self-regulation, interoceptive confidence, or the wear time (H2-b). In addition, we

Table 2. Baseline psychological measures (week 0; $N = 28$).

Measure	Mean ± SD	Median	Min	Max
GAD-7	12.39 ± 3.78	11.50	7	28
CES-D	13.43 ± 6.80	12.50	2	30
GHQ-12	11.89 ± 5.27	11.00	4	27
SHAPS	23.46 ± 2.17	23.00	19	28
MAIA-2 Total	97.96 ± 22.70	99.00	55	140

Table 3. Correlation between interoceptive accuracy (SensHB slope) and baseline mental health assessments.

Measure	Coefficient	SE	p
SHAPS	-0.34	0.26	0.197
GAD-7	-1.49	0.34	0.0003***
GHQ-12	-1.53	0.59	0.016*
CES-D	-1.54	0.76	0.055
MAIA-2 total	-1.63	2.46	0.515
TAS-20 total	-0.08	1.02	0.940

Values are unstandardized linear regression coefficients. $N = 28$. SE, standard error; SensHB, heartbeat sensitivity; SHAPS, Snaith–Hamilton Pleasure Scale; GAD-7, Generalized Anxiety Disorder scale; GHQ-12, General Health Questionnaire; CES-D, Center for Epidemiologic Studies Depression Scale; MAIA-2, Multidimensional Assessment of Interoceptive Awareness; TAS-20, Toronto Alexithymia Scale.

tested the condition effect on the participants' reported confidence during the SensHB task (H3), to compare with interoceptive accuracy results (H2-a).

RQ3 focused on whether wearing heartbeat-synchronized haptic feedback changed self-reported interoceptive awareness over time. We first hypothesized that self-reported interoceptive awareness (MAIA-2 overall score), and or specifically Body Listening and Noticing subscales, would significantly increase at the end of the intervention week (Week 2) compared to the baseline (H4-a), and remain elevated one week after the device removal at Week 3 (H4-b).

To understand whether any condition effects (RQ1, RQ2) of amplifying real heartbeat signals might be associated with everyday mood improvement, **RQ4** tested whether real, compared to sham, feedback would significantly increase the average BMIS pleasant–unpleasant mood score (H5) or mood variability (H6).

Before reporting preregistered analyses, we summarize baseline psychological and interoceptive characteristics of the recruited participants (Table 2) and examine associations between baseline mental health severity and interoceptive measures to confirm expected patterns from prior literature before examining intervention effects.

Interoceptive accuracy was negatively associated with baseline anxiety (GAD-7) and general psychological distress (GHQ-12). As an exploratory analysis, we examined associations between interoceptive accuracy (SensHB slope) and baseline mental health measures reported prior to any intervention. We separately fit linear regression models using SensHB slope, a measure of interoceptive accuracy, as the outcome variable and each psychological measure as a predictor. Our results (Table 3) show that higher baseline anxiety (GAD-7; $\beta = -1.49$, $p = 0.0003$) and general psychological distress (GHQ-12; $\beta = -1.53$, $p = 0.0163$) were significantly associated with lower SensHB slope values, indicating that individuals with greater anxiety and general distress had lower baseline interoceptive accuracy, consistent with prior literature linking altered interoception to psychopathology.

Table 4. Correlation between interoceptive awareness (MAIA-2 total score) and baseline mental health assessments.

Measure	Coefficient	SE	p
SHAPS	0.044	0.017	0.013*
GAD-7	0.039	0.032	0.226
GHQ-12	-0.024	0.045	0.600
CES-D	-0.047	0.058	0.421
TAS-20 total	-0.038	0.074	0.608

Values are unstandardized linear regression coefficients. $N = 28$. SE, standard error; MAIA-2, Multidimensional Assessment of Interoceptive Awareness; SHAPS, Snaith–Hamilton Pleasure Scale; GAD-7, Generalized Anxiety Disorder scale; GHQ-12, General Health Questionnaire; CES-D, Center for Epidemiologic Studies Depression Scale; TAS-20, Toronto Alexithymia Scale.

Table 5. Real vs. sham feedback effects on preregistered measures.

Variable	Real Mean (SD)	Sham Mean (SD)	Test Used	P-Value
SHAPS	23.64 (2.31)	23.96 (2.28)	Wilcoxon	0.210
GAD-7	12.60 (4.00)	12.04 (3.12)	Wilcoxon	0.651
GHQ-12	11.88 (6.46)	10.88 (6.48)	Wilcoxon	0.692
CES-D	13.84 (8.48)	13.68 (10.53)	Wilcoxon	0.446
TAS	54.36 (8.87)	54.44 (9.23)	Paired t-test	0.897
Interoceptive accuracy	-0.01 (0.78)	0.23 (1.31)	Paired t-test	0.414
BMIS total	44.03 (3.58)	44.90 (4.24)	Paired t-test	0.263
BMIS standard deviation across week	4.04 (2.02)	4.49 (2.48)	Paired t-test	0.469

Table 6. Preregistered moderation of GHQ-12 psychological distress by stimulation condition.

Term	**Estimate**	**SE**	p
Condition (sham−real)	2.66	4.74	0.61
Weekly watch wear time (hours) × condition	-0.02	0.07	0.82
Baseline TAS-20 × condition	0.16	0.14	0.36
Change in SensHB accuracy from prior week × condition	-0.98	3.84	0.81
Baseline MAIA-2 Self-Regulation × condition	-0.38	1.19	0.77
Change in MAIA-2 total score from prior week × condition	0.14	0.49	0.80
Baseline GHQ-12 × condition	0.26	0.30	0.47
Increase in SensHB confidence with real feedback × condition	3.39	4.68	0.52

Estimates are from preregistered interaction models testing whether participant covariates moderated the effect of stimulation condition on GHQ-12 psychological distress. Baseline measures were obtained before intervention. SE, standard error; GHQ-12, General Health Questionnaire; MAIA-2, Multidimensional Assessment of Interoceptive Awareness; SensHB, heartbeat sensitivity; TAS-20, Toronto Alexithymia Scale.

Overall interoceptive awareness (MAIA-2) was significantly associated only with anhedonia (SHAPS).. We also analyzed the correlation between interoceptive awareness (MAIA-2 total score) with baseline psychological measures (Table 4). Higher MAIA-2 scores were significantly associated with lower anhedonia measured by SHAPS ($\beta = 0.044$, $p = 0.013$), but were not significantly related to anxiety (GAD-7), general psychological distress (GHQ-12), depressive symptoms (CES-D), or alexithymia (TAS-20 total), showing a different pattern from interoceptive accuracy (Table 4). These contrasting associations suggest that behavioral interoceptive accuracy and self-reported interoceptive awareness capture distinct aspects of interoceptive functioning.

Table 7. Preregistered moderation of cardiac interoceptive accuracy by stimulation condition.

Term	Estimate	SE	p
Condition (sham–real)	-1.51	1.57	0.35
Weekly watch wear time (hours) × condition	-0.01	0.03	0.75
Baseline TAS-20 × condition	-0.22	0.19	0.26
Baseline MAIA-2 Self-Regulation × condition	-0.25	0.36	0.49
Increase in SensHB confidence with real feedback × condition	-1.28	1.56	0.42

Estimates are from preregistered interaction models testing whether participant covariates moderated the effect of stimulation condition on SensHB cardiac interoceptive accuracy. Predictors follow the preregistered H2-b specification. SE, standard error; MAIA-2, Multidimensional Assessment of Interoceptive Awareness; SensHB, heartbeat sensitivity; TAS-20, Toronto Alexithymia Scale.

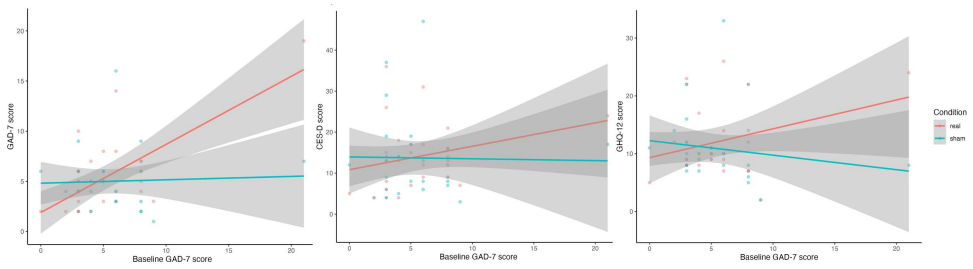


Fig. 1. **Interaction between baseline anxiety (GAD-7) and feedback condition on symptom outcomes.** Panels show model-estimated symptom scores as a function of baseline anxiety separately for real and sham stimulation for anxiety symptoms (GAD-7; left), depressive symptoms (CES-D; middle), and general psychological distress (GHQ-12; right). Lines represent model-predicted values with 95% confidence intervals. Higher baseline anxiety was associated with greater symptom improvement under sham stimulation, with smaller effects observed under real stimulation.

Feedback condition did not have any significant effects on pre-registered measures (RQ1, RQ2, RQ4). Our main registered research questions were whether real feedback had significant effects on mental health outcomes (RQ1; H1a), interoceptive accuracy and confidence (RQ2; H2a, H3), and overall affect measures (RQ4; H5, H6). We did not observe any significant main effects of feedback condition (real versus sham) on these measures (Table 5). To examine potential individual differences in responsiveness, we tested whether preregistered baseline factors moderated the effect of feedback (H1b, H2b). None of the preregistered covariates–baseline alexithymia (TAS-20), interoceptive self-regulation (MAIA), baseline distress (GHQ-12), wear time, or changes in interoceptive confidence–significantly moderated mental health outcomes (RQ1; H1-b; Table 6) or interoceptive accuracy (H2-b; Table 7). Because these preregistered analyses did not reveal average effects across the full sample, we next examined whether feedback effects varied as a function of baseline symptom severity.

Baseline anxiety moderated symptom outcomes, with the strongest effects under sham feedback (Exploratory; RQ1). Given prior evidence that interoceptive processing may relate differentially to anxiety and depressive symptom dimensions, we explored whether baseline psychological traits moderated the effects of feedback condition on symptom outcomes. Baseline GHQ-12, GAD-7, CES-D, SHAPS, TAS-20, and MAIA-2 total scores were included as candidate moderators. Baseline anxiety (GAD-7) showed the strongest interactions with feedback condition after false discovery rate (FDR) correction (Table 8). Specifically, baseline anxiety significantly interacted with feedback condition in predicting changes in depressive symptoms (CES-D; $\beta = 1.42$, $p_{\text{FDR}} = 0.003$), anxiety symptoms (GAD-7; $\beta = 0.84$, $p_{\text{FDR}} = 0.001$), and

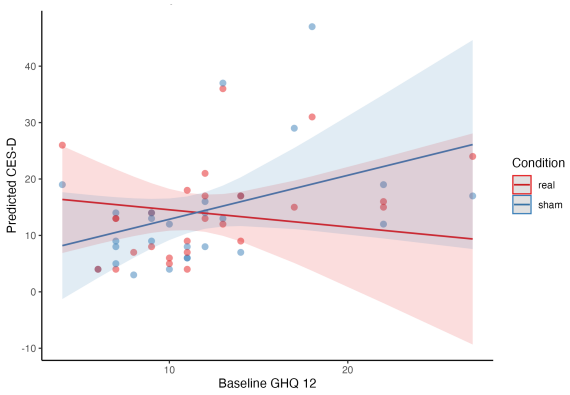


Fig. 2. **Interaction between baseline GHQ-12 and feedback condition.** Panels show estimated symptom change in depression severity (CES-D) as a function of baseline GHQ-12 for real versus sham stimulation. Lines represent model-predicted values with 95% confidence intervals. Higher baseline GHQ-12 was associated with a relative advantage of real over sham stimulation, although condition-specific slopes were not significant when tested separately.

general psychological distress (GHQ-12; $\beta = 1.06$, $p_{\text{FDR}} = 0.036$). Interaction patterns are visualized in Fig. 1. We also examined whether baseline anxiety predicted the effects of sham and real stimulation relative to baseline (reported in Table 9). Baseline anxiety significantly interacted with both sham and real stimulation in predicting changes in anxiety symptoms (GAD-7), with effects in the same direction but larger in magnitude under sham stimulation (sham: $\beta = -1.45$, 95% CI [-1.80, -1.10], $p < 0.001$; real: $\beta = -0.62$, 95% CI [-0.97, -0.27], $p = 0.001$). Baseline anxiety predicted the effect of sham stimulation, but not of real stimulation, on general psychological distress and depressive symptoms. These moderation effects were specific to mental health outcomes: baseline anxiety did not significantly moderate affect, behavioral interoceptive accuracy, perceived accuracy, or confidence.

Baseline general psychological distress (GHQ-12) significantly moderated the relative condition effect on depressive symptoms (CES-D) (Exploratory; RQ1). In contrast to the pattern observed for baseline anxiety, baseline general psychological distress (GHQ-12) also interacted with stimulation condition in predicting changes in depressive symptoms (CES-D) ($\beta = -1.08$, $p = 0.036$); Figure 2. Specifically, participants with poorer baseline mental health exhibited greater reductions in CES-D scores under real stimulation relative to sham stimulation. However, baseline psychological distress did not predict the effects of real or sham stimulation relative to baseline, suggesting that this effect should be interpreted as a relative condition difference rather than as evidence for a robust real-stimulation effect alone (Table 9). No significant moderation effects were observed for anhedonia (SHAPS), alexithymia (TAS-20), affect (BMIS), behavioral interoceptive accuracy (SensHB slope), or interoceptive confidence.

We next examined whether participation in the study was associated with changes in self-reported interoceptive awareness (RQ3). In H4a and H4b, we hypothesized that interoceptive awareness would increase by Week 2 regardless of condition order (H4a), and that these gains would persist one week after participants stopped using the watch (H4b). Interoceptive awareness was assessed using the MAIA-2 total score and the subscales Noticing and Body Listening.

Device use, regardless of feedback condition, produced temporary gains in Body Listening and Noticing interoceptive awareness subscales (RQ3; H4a-b). At week 2, the MAIA-2 total score did not differ significantly from baseline. However, participants showed significant increases in Noticing ($t(27) = 2.61$, $p = 0.015$) and Body Listening ($t(27) = 2.67$, $p =$

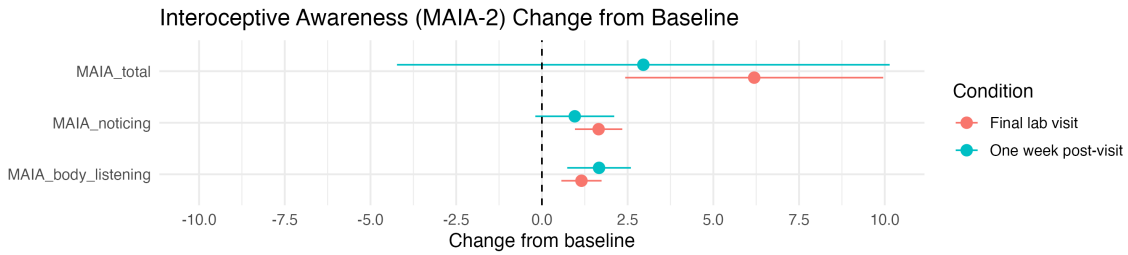


Fig. 3. Changes in MAIA-2 overall score and subscales at the final lab visit (left) and one week after final lab visit (right) compared to the baseline measurement.

0.032). At week 3, one week after discontinuing the watch, none of the interoceptive measures remained significantly elevated relative to baseline (Figure 3). These findings indicate short-term increases in specific dimensions of subjective interoceptive awareness during active participation, without evidence of sustained effects after discontinuation.

Mental health improvements did not significantly covary with gains in interoceptive awareness (H1-c). We hypothesized that if conscious awareness of bodily states directly supports mental health, improvements in interoceptive awareness (MAIA-2) from the sham week to the real week would significantly correlate with improvements in general distress. This hypothesis was not supported; as there was no significant correlation between change in GHQ-12 and change in interoceptive awareness across weeks.

The intervention had lasting effects on mental health and self-reported interoceptive awareness among anxious participants (Exploratory; related to RQ1, RQ3). Our observation that both real and sham feedback benefited mental health in anxious participants suggests that symptom improvements were not driven by improved access to interoceptive information, but rather by changes in the interpretation or processing of this information. One prediction of this hypothesis is that changes in interpretation would persist after the deice was removed. We therefore conducted exploratory follow-up analyses to examine whether participants with higher baseline anxiety showed lasting symptom changes after device removal, and whether these changes co-occurred with changes in self-reported interoceptive awareness.

Baseline anxiety significantly interacted with timepoint, such that participants with higher baseline anxiety showed reductions in alexithymia, anxiety, and general psychological distress one week after the intervention period (Fig. 4; all FDR-corrected $p < 0.05$). Notably, the interaction between baseline anxiety and timepoint was associated with decreased self-reported interoceptive awareness, including reductions in MAIA-2 total score, Body Listening, Emotional Awareness, Trusting, and Self-Regulation subscales (all FDR-corrected $p < 0.05$). No significant main effects of timepoint survived FDR correction.

We next tested whether the decreases in interoceptive awareness we observed among participants with higher baseline anxiety emerged during the active watch-wear period (Fig. 5). During active wear, sham stimulation was associated with reductions in anxiety and general distress among participants with higher baseline anxiety, but did not significantly change interoceptive measures. This suggests that the decrease in self-reported interoceptive awareness emerged after the watch-wear period ended. We further provide a potential explanation for these findings in the following section.

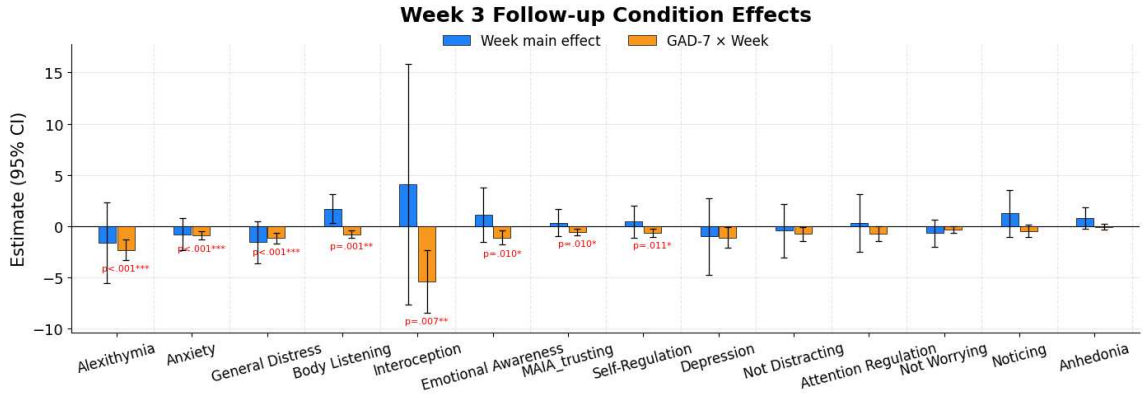


Fig. 4. **Follow-up effects after device removal.** Bars indicate model estimates for the main effect of timepoint and the timepoint × baseline anxiety interaction at Week 3 relative to baseline.Interoception refers to the overall MAIA-2 score. Anhedonia refers to the SHAPS total; higher values reflect less anhedonia. Error bars indicate 95% confidence intervals. Significant FDR-corrected effects are shown in red.

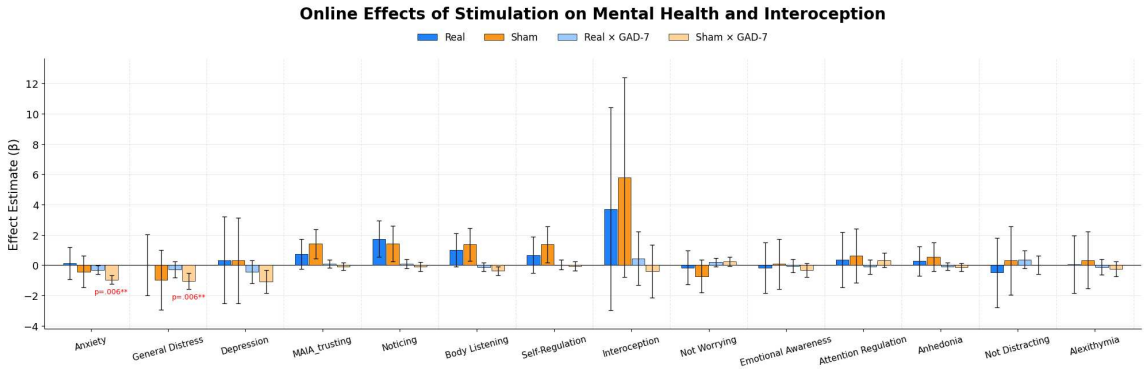


Fig. 5. **Effects during watch wear.** Bars indicate model estimates for real stimulation, sham stimulation, real stimulation × baseline anxiety, and sham stimulation × baseline anxiety during weeks 1 and 2 relative to baseline. Interoception refers to the overall MAIA-2 score. Anhedonia refers to the SHAPS total; higher values reflect less anhedonia. Error bars indicate 95% confidence intervals. Significant FDR-corrected effects are shown in red.

3 Discussion

This study investigated whether wearing a smartwatch delivering heartbeat-synchronized haptic feedback over two weeks of everyday use would improve mental health, interoceptive accuracy, or daily mood in individuals with elevated alexithymic traits. Across preregistered confirmatory tests, we found no significant main effects of feedback condition on any primary outcome (all $p > .20$; Tables 1, 10, 5). Device use, regardless of condition, produced temporary increases in MAIA-2 Noticing and Body Listening, but these gains did not persist after device removal. Exploratory analysis showed that baseline anxiety significantly moderated condition effects on anxiety, depressive symptoms, and general distress (all $p_{\text{FDR}} < .05$; Tables 8), with stronger condition-specific effects under the sham feedback. Condition-specific analyses showed that higher baseline anxiety predicted reductions in anxiety symptoms under both real and sham feedback, with larger effects under sham feedback, and predicted reductions in general psychological distress under

sham feedback only (Table 9). Below, we discuss these findings and propose a recalibration mechanism that may explain why symptom reductions were concentrated among participants with higher baseline anxiety, particularly under sham feedback.

Symptom reductions among high-anxiety participants were strongest under sham feedback. The most robust finding in this study was a significant interaction between feedback condition and baseline anxiety. Higher initial GAD-7 scores moderated condition effects on anxiety, depressive symptoms, and general psychological distress (Table 8). Condition-specific analyses indicated that this pattern was strongest under sham feedback: baseline anxiety predicted reductions in anxiety symptoms under both conditions, but predicted reductions in general psychological distress only under sham feedback (Table 9). These exploratory interaction effects survived correction for multiple comparisons and were not observed for alexithymia, anhedonia, affect measures, interoceptive accuracy or confidence.

Symptom reductions were not explained by improved interoceptive accuracy or awareness

While previous work has shown that real-time haptic heartbeat feedback can acutely enhance interoceptive accuracy and confidence [10], we found that real heartbeat-synchronized haptic feedback did not produce significant gains in interoceptive accuracy, self-reported interoceptive awareness, or confidence (Table 5). Furthermore, the strongest symptom-related effects among participants with higher baseline anxiety occurred under sham feedback (Table 9), which is difficult to reconcile with a mechanism based solely on improved heartbeat-detection accuracy. Furthermore, symptom improvements did not correlate with improvements in interoceptive accuracy and lasting improvements were associated with *decreases* in self-reported interoceptive awareness at follow-up. Thus, the observed mental health symptom reductions among participants with higher baseline anxiety do not rely on improving interoceptive accuracy or awarenss.

Instead, we propose a *recalibration mechanism*: heartbeat-synchronized feedback may have changed how participants attend to and interpret body signals. This is a plausible explanation, as awareness and interpretation of interoceptive signals is recognized as a facet of interoception not necessarily coupled with interoceptive accuracy [13, 17]. *Predictive processing* theory further suggests that individuals with mental disorders may involve maladaptiva interpretation about their internal bodily states, such that ordinary bodily fluctuations are interpreted as signals of threat or harm [17], and emotional disorder may arise in part from the miscalibrated predictions about their internal bodily states [5, 29]. High anxiety symptoms may have been linked to maladaptive predictive associations between bodily arousal and danger [17, 29], and repeated exposure to salient cardiac feedback, whether or not it is synchronized with the actual heartbeat, may have provided participants with repeated opportunities to observe that bodily fluctuations are not followed by harm or threat. Thus, we suggest that rather than enhancing accuracy, heart rate feedback may have altered how highly-anxious participants interpret their interoceptive signals.

A prediction of this "recalibration" hypothesis is that the effects should be durable. We observed that one week after discontinuing watch use, participants with higher baseline anxiety had decreases in symptoms alongside persistent decreases in interoceptive awareness across multiple subscales (Fig. 4). This observation is consistent with the possibility that heart rate feedback reduced pathological interpretation, potentially by training participants to learn that small fluctuations in their heart rate may not be always associated with negative everyday experiences.

This recalibration hypothesis also provides an explanation for why sham feedback was more effective than real feedback. Sham vibrations provide a delayed feedback, which makes them particularly well-suited for breaking the learned association between cardiac sensations and perceived threat, as participants receive a strongly plausible, but not truly synchronized signal that is unrelated to any actual physiological event. In contrast, real heartbeat-synchronized feedback may preserve the correspondence between bodily fluctuations and vibratory input, which may at times reinforce,

rather than recalibrate existing associations between arousal and threat in anxious individuals compared to the sham feedback. We also propose that the differential response to sham versus real feedback in participants with higher baseline general distress–where real feedback appeared advantageous for depression specifically–may have been due to a distinct mechanism in which real cardiac information is beneficial when the primary difficulty is motivational disengagement (as in depression) rather than interoceptive hypervigilance (as in anxiety).

3.1 Device Use Transiently Increased Interoceptive Awareness Regardless of Feedback Condition.

Another noteworthy finding is that regardless of the feedback condition, wearing the device for two weeks produced temporary significant increases in MAIA-2 Noticing ($t(27) = 2.61$, $p = .015$) and Body Listening ($t(27) = 2.67$, $p = .032$) subscales at Week 2 (Fig. 3; H4-a), without persistent gains after the device removal (H4-b). The absence of a condition difference in these effects suggests the temporary improvement in interoceptive awareness might be due to engagement with the device or the study rather than specifically due to cardiac signal amplification. For example, attention or expectancy effects may be accompanied by device use. Whether these transient increases in interoceptive awareness contributed to the symptom changes observed during the active period cannot be determined from the present data and requires future work. *SensHB Captures Baseline Individual Differences but Not Intervention-Related Change.* While feedback did not significantly alter interoceptive accuracy, our exploratory analysis showed that baseline interoceptive accuracy was significantly associated with anxiety and general distress at baseline. These findings are consistent with results from previous literature showing mental health conditions are associated with poorer interoceptive accuracy [22]. However, we found a dissociation after intervention. In highly-anxious participants, mental health improved without a change in interoceptive accuracy. These findings imply that changes in interoceptive perception and interpretation, rather than accuracy, may be the basis for the improvement in mental health in highly-anxious people.

3.2 Limitations and Future Directions

There are several limitations to this study. The modest sample size limits the precision of interaction estimates and the robustness of MAIA-2 subscale analyses. Replication in larger samples with complementary behavioral and physiological measures of interoception will be important. The MAIA-2 total score may also not capture short-term or context-dependent fluctuations in interoceptive experience. The study was conducted in a non-clinical sample with elevated alexithymia levels, which might limit generalizability. Given our finding that individuals with high baseline anxiety showed significant symptom improvements from the sham haptic feedback, we encourage future work to replicate these effects in clinical population with clinical anxiety.

Despite these limitations, the empirical findings presented in this paper provide preliminary evidence for uncovering possible mechanism that can inform the future research on interoception and mental health and the design of novel interventions. Cardiac feedback did not appear to improve mental health by enhancing the accuracy of heartbeat detection (SensHB slope). Instead, symptom reductions were contingent on baseline anxiety and occurred without measurable changes in interoceptive accuracy, suggesting that attentional, inferential, or interpretive processes are more important for mental health than accuracy alone. Future work should directly assess how heart rate feedback influences interoceptive inference, attentional allocation to bodily signals, expectancy formation, and learning processes. A testable prediction of our recalibration hypothesis is that participants may be less emotionally reactive to changes in body signals after receiving heart rate feedback, empirically using methods such as a CO_2 challenge [25]. More broadly, these findings suggest a dissociation among interoceptive accuracy, awareness, and interpretation, indicating that their contributions to mental health may not be equivalent.

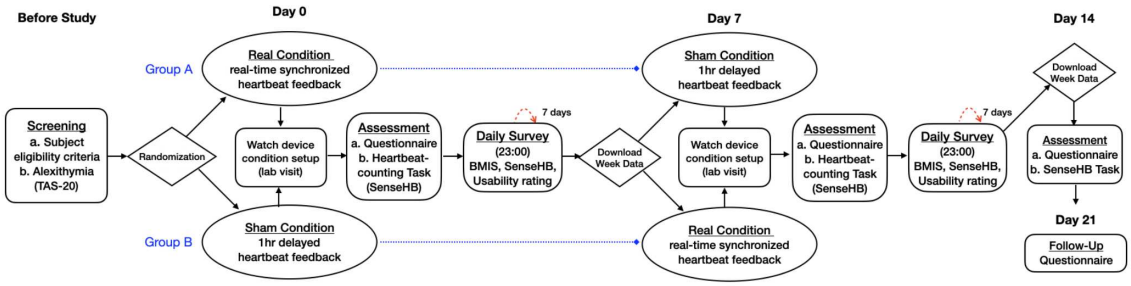


Fig. 6. **Overview of study procedures.** Participants were randomized to condition order, completed baseline and weekly assessments, wore the smartwatch during two intervention weeks, switched feedback condition after Week 1, and completed a one-week follow-up after device removal.

4 Methodology

The study employed a two-week, within-subject, counterbalanced, sham-controlled wearable intervention design within a three-week study timeline, followed by a Week 3 online follow-up. Each participant completed one week of real-time heartbeat-synchronized feedback and one week of sham, temporally delayed feedback in randomized order. Condition order was assigned using computer-generated randomization without stratification prior to enrollment. Participants were blinded to condition assignment during data collection and were told only that feedback timing might vary across weeks. Debriefing occurred after the follow-up survey. No adverse events were reported.

Participants. Twenty-nine adults (18–46 years; average 26.8 ± 5.9 years; 19 female, 10 male) were recruited through university mailing lists and community advertisements. Sample size was determined *a priori* based on feasibility and prior within-subject heartbeat-feedback studies reporting medium effect sizes ($d \approx 0.6$–0.8), indicating that 24 participants would provide > 80% power to detect within-subject effects at $\alpha = 0.05$. Eligibility criteria included age 18–65 years, English fluency, and absence of self-reported cardiac arrhythmia or other cardiovascular conditions that could interfere with photoplethysmographic sensing. To enrich the sample for elevated alexithymic traits, participants were required to score > 51 on the 20-item Toronto Alexithymia Scale (TAS-20) during online screening [1, 3]. TAS-20 scores in the final sample averaged 56.97 ± 4.32 (range 51–66). Based on established cutoffs (≤51 non-alexithymic; 52–60 borderline; ≥61 alexithymic), we recruited 7 participants (24.1%) who were classified as alexithymic and 22 (75.9%) as borderline. Table 12 summarizes the recruited participants. Individuals with academic or supervisory relationships with the investigators were excluded. The study was approved by the Institutional Review Board, and all participants provided written informed consent. Participants received $80–$120, prorated by study completion. One participant withdrew after Week 2, yielding 27 final participants.

Apparatus and Feedback Manipulation. Participants wore a programmable Bangle.js 2 smartwatch equipped with a photoplethysmography sensor sampling heart rate at 25–30 Hz. Heartbeat detection and vibration triggering were processed locally on-device to minimize latency. In the real condition, vibration pulses were triggered within 100 ms of detected heartbeat peaks derived from the photoplethysmographic waveform. In the sham condition, vibration timing was derived from the participant's own heart rate recorded one hour earlier, preserving individual-specific variability while eliminating real-time synchrony. Device-reported heart rate confidence values were used to exclude unreliable trials. Device logs, including heart rate and wear-time indicators, were downloaded at weekly visits.

Study Procedure. Figure 6 provides an overview of the study procedure. At baseline (Week 0), participants completed questionnaire measures and three 30-second heartbeat-counting trials separated by rest intervals. Participants were instructed not to manually check their pulse or use external timing cues. After each trial, they reported perceived heartbeat count and rated confidence on a 7-point Likert scale. No vibratory feedback was delivered during baseline assessment. During Weeks 1 and 2, participants wore the smartwatch during waking hours, removing it only for charging or water exposure. Daily surveys were distributed at 23:00 local time and included the Brief Mood Introspection Scale (BMIS), a one-minute heartbeat estimation task, and ratings of perceived synchrony and comfort. At the end of Week 1, participants returned for reassessment, repeated questionnaire and heartbeat-counting measures, and switched feedback condition. At the end of Week 2, participants completed post-intervention assessments and returned the device. One week after device removal (Week 3), participants completed an online follow-up survey.

4.1 Measures

Psychological Scales for Mental Health. Mental health outcomes were assessed at baseline (Week 0), midpoint (Week 1), post-intervention (Week 2), and follow-up (Week 3). Depression symptoms were measured using the Center for Epidemiologic Studies Depression Scale (CES-D), anxiety symptoms using the Generalized Anxiety Disorder scale (GAD-7), hedonic capacity using the Snaith–Hamilton Pleasure Scale (SHAPS), and general psychological distress using the General Health Questionnaire (GHQ-12). Alexithymic traits were assessed using the Toronto Alexithymia Scale (TAS-20). A summary of these measures, their constructs, and corresponding clinical cutoffs is provided in Table 13. We scored the SHAPS with higher scores indicating greater hedonic capacity (less anhedonia) with a score range of 0-3 on each item for a total score range of 0 to 42.

These measures were selected to examine mental health domains that have been linked to interoceptive processing in prior literature. Anxiety and depression have been associated with alterations in interoceptive processing and bodily signal perception; however, prior work reports mixed findings, with different dimensions of interoception (e.g., interoceptive accuracy, sensibility, and awareness) showing variable and sometimes inconsistent relationships with affective symptoms [6, 17]. Alexithymia, characterized by difficulties in identifying and describing internal states, has been consistently associated with reduced interoceptive awareness, although the strength and specificity of this relationship vary across measures and tasks [16]. Anhedonia, indexed by the SHAPS, has been discussed in relation to altered interoceptive and reward processing in affective disorders, though direct empirical links between interoceptive measures and anhedonia remain limited [24]. General psychological distress, measured by the GHQ-12, serves as a broad screening measure of mental health and does not isolate interoceptive mechanisms, but provides a complementary index of overall affective functioning [14]. Together, these measures enable us to examine whether—and to what extent—individual differences in interoceptive processes are associated with affective symptoms and related cognitive–emotional processes, providing in-the-wild empirical evidence in the context of prior mixed findings [17, 24].

Multidimensional Assessment of Interoceptive Awareness (MAIA-2). Interoceptive awareness was assessed using MAIA-2, a validated self-report instrument designed to capture multiple dimensions of self-reported interoceptive awareness. The MAIA-2 is a 37-item questionnaire comprising eight subscales that reflect distinct facets of how individuals perceive, attend to, and interpret internal bodily signals [20]. Each MAIA-2 subscale measures a distinct component of interoceptive awareness. In this study, we focused on the Noticing and Body Listening subscales, which we hypothesized to be directly relevant to the heartbeat-synchronized haptic feedback provided in the intervention. The *Noticing* subscale indexes awareness of basic bodily sensations, including uncomfortable, comfortable, and neutral sensations such as changes in heart rate, tension, or physical discomfort. It reflects a perceptual aspect of interoception

(whether individuals are aware that bodily sensations are present). The *Body Listening* subscale captures the tendency to actively attend to bodily signals as a source of information for guiding behavior and decision-making. It reflects a more integrative aspect of interoception, in which individuals use bodily sensations to inform emotional understanding or action. These complementary measures allow us to distinguish between more basic perceptual awareness of bodily sensations (Noticing) and higher-order engagement with and use of these signals (Body Listening), consistent with multidimensional models of interoception [13, 17].

SensHB for Cardiac interoceptive accuracy. Interoceptive accuracy was assessed using the SensHB metric [19]; participants were asked to estimate their heart rate at three points during the lab visit and SensHB was defined as the slope of the regression line between participants' estimated heartbeat counts and measured heartbeats. The SensHB method is considered less biased by prior knowledge of resting heart rate than traditional heartbeat counting measures, because it is not biased by the participant's knowledge of their resting heart rate. In this experiment, SensHB was calculated based on three measurements during the lab visit with approximately 2 minutes between each measurement. We also collected participants estimates of their confidence in their heart rate count and their estimate of the accuracy of the watch feedback; these values were averaged across trials to yield one value of each for each lab visit.

Brief Mood Introspection Scale for Daily affect change. To measure everyday mood, we used the average Brief Mood Introspection Scale (BMIS) score, as well as the within-week standard deviation to capture mood variability.

4.2 Preregistration and Analysis

The study hypotheses, outcome measures, exclusion criteria, and statistical analysis plan were preregistered on AsPredicted (Registration #248,039; https://aspredicted.org/hnyw-cw85.pdf). The preregistration was submitted after data collection. but prior to any data analysis. All confirmatory analyses reported in this manuscript follow the preregistered protocol unless explicitly noted otherwise. Analyses involving demographic variables (e.g., age, sex) and usability-related factors were exploratory.

Exclusion criteria. Exclusion criteria were preregistered and applied prior to hypothesis testing. Intervention weeks were excluded if cumulative wear time was less than 48 hours between laboratory visits, if more than 11 days elapsed between weekly assessments, or if fewer than two daily surveys were completed. Follow-up responses completed more than 10 days after Week 2 were excluded. Heartbeat-counting trials were excluded if device-reported heart rate sensing confidence was below 80%. We excluded incomplete surveys.

Statistical Analysis. All analyses were preregistered prior to data inspection. Normality was assessed using the Shapiro–Wilk test. For within-subject comparisons between real and sham conditions, two-tailed paired t-tests were used for normally distributed variables and Wilcoxon signed-rank tests for non-normal distributions. False discovery rate (FDR) correction was applied separately within predefined outcome families (mental health, interoceptive measures, and daily affect metrics). All statistical tests were two-sided with $\alpha = 0.05$. Effect sizes and 95% confidence intervals are reported where appropriate. Longitudinal changes in interoceptive awareness were assessed using repeated-measures ANOVA with Time (Week 0, Week 2, Week 3) as a within-subject factor. Moderator analyses were conducted using regression models testing whether baseline or intervention-related covariates moderated the effect of feedback condition, such as linear regression models predicting per-participant difference scores (Real − Sham) from baseline alexithymia (TAS-20), baseline interoceptive self-regulation (MAIA-2 Self-Regulation), baseline general mental health (GHQ-12), wear duration, and within-subject changes in interoceptive accuracy and confidence.

4.3 Ethics statement

The study was approved by the MIT Committee on the Use of Humans as Experimental Subjects (Protocol 2502001554)

4.4 Generative AI use

We used generative AI to generate figures from raw data and for AI-assisted copy editing. All content generated using AI was verified for accuracy by the authors.

5 Author Contributions

MK: Conceptionalization, Methodology, Software, Investigation,Formal analysis, Data curation, Writing (original draft and review), Visualization, Validation, Supervision. NWW: Conceptionalization, Methodology, Software, Investigation, Formal analysis Data curation, Writing (original draft and review), Visualization, Validation,Supervision. PC: Conceptionalization, Methodology, Writing (review). YSZ:Investigation, Data curation, Formal analysis, Writing (review), Visualization. IW: Investigation, Data curation, Formal analysis, Writing (review), Visualization. SP: Investigation, Writing (original draft and review). AA: Software, Investigation. PM: Project administration, funding acquisition, Writing (review).

6 Data availability

Data and analysis code for the analyses in this paper are available at https://osf.io/wfexj

References

[1] R. Michael Bagby, James D. A. Parker, and Graeme J. Taylor. 1994. The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research* 38, 1 (1994), 23–32.

[2] R. Michael Bagby, James D. A. Parker, and Graeme J. Taylor. 1994. The Twenty-Item Toronto Alexithymia Scale—I. Item Selection and Cross-Validation of the Factor Structure. *Journal of Psychosomatic Research* 38, 1 (1994), 23–32. doi:10.1016/0022-3999(94)90005-1

[3] R. Michael Bagby, Graeme J. Taylor, and James D. A. Parker. 1994. The twenty-item Toronto Alexithymia Scale—II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research* 38, 1 (1994), 33–40.

[4] R. Michael Bagby, Graeme J. Taylor, and James D. A. Parker. 1994. The Twenty-Item Toronto Alexithymia Scale—II. Convergent, Discriminant, and Concurrent Validity. *Journal of Psychosomatic Research* 38, 1 (1994), 33–40. doi:10.1016/0022-3999(94)90006-X

[5] Lisa Feldman Barrett and W. Kyle Simmons. 2015. Interoceptive predictions in the brain. *Nature Reviews Neuroscience* 16, 7 (2015), 419–429. doi:10.1038/nrn3950

[6] Wei Chen et al. 2023. Interoception and psychopathology: dissociable dimensions across anxiety and depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2023).

[7] A.D. Craig. 2009. How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience* 10, 1 (2009), 59–70.

[8] A. D. Craig. 2002. How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience* 3, 8 (2002), 655–666. doi:10.1038/nrn894

[9] Antonio R. Damasio. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences* 351, 1346 (1996), 1413–1420. doi:10.1098/rstb.1996.0125

[10] Olga Dobrushina et al. 2024. Interoceptive training with real-time haptic versus visual heartbeat feedback. *Psychophysiology* (2024). doi:10.1111/psyp.14648 Advance online publication.

[11] Dorothee Fischer, Markus Messner, and Olga Pollatos. 2017. Improvement of interoceptive processes after mindfulness training. *Psychophysiology* 54 (2017), 1659–1667.

[12] Sarah N. Garfinkel, Anil K. Seth, Adam B. Barrett, Kei Suzuki, and Hugo D. Critchley. 2015. Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology* 104 (2015), 65–74. doi:10.1016/j.biopsycho.2014.11.004

[13] Sarah N. Garfinkel, Anil K. Seth, Lisa Feldman Barrett, Kei Suzuki, and Hugo D. Critchley. 2015. Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology* 104 (2015), 65–74. doi:10.1016/j.biopsycho.2014.11.004

[14] David P. Goldberg. 1972. *The Detection of Psychiatric Illness by Questionnaire.* Oxford University Press.

[15] Beate M. Herbert, Cornelia Herbert, and Olga Pollatos. 2011. On the relationship between interoceptive awareness and alexithymia. *Biological Psychology* 87, 3 (2011), 345–352. doi:10.1016/j.biopsycho.2011.05.006

[16] Beate M. Herbert, Cornelia Herbert, and Olga Pollatos. 2011. On the Relationship Between Interoceptive Awareness and Alexithymia: Is Interoceptive Awareness Related to Emotional Awareness? *Journal of Personality* 79, 5 (2011), 1149–1175. doi:10.1111/j.1467-6494.2011.00717.x

[17] Sahib S. Khalsa and Rachel C. Lapidus. 2018. Interoception and Mental Health: A Roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3, 6 (2018), 501–513. doi:10.1016/j.bpsc.2017.12.004

[18] Minsol Michelle Kim, Nathan W. Whitmore, Phoebe Chua, Serena Pei, Malak Abdalla, and Pattie Maes. 2025. Exploring the Emotional Effects of Enhanced Interoception via Heartbeat-Synchronized Haptic Feedback. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 4, Article 183 (Dec. 2025), 21 pages. doi:10.1145/3770663

[19] Dennis EO Larsson, Giulia Esposito, Hugo D Critchley, Zoltan Dienes, and Sarah N Garfinkel. 2021. Sensitivity to changes in rate of heartbeats as a measure of interoceptive ability. *Journal of Neurophysiology* 126, 5 (2021), 1799–1813.

[20] Wolf E. Mehling, Michael Acree, Anita Stewart, Jonathan Silas, and Alexander Jones. 2018. The Multidimensional Assessment of Interoceptive Awareness, Version 2 (MAIA-2). *PLOS ONE* 13, 12 (2018), e0208034. doi:10.1371/journal.pone.0208034

[21] Jennifer Murphy, Caroline Catmur, and Geoffrey Bird. 2019. Classifying individual differences in interoception: Implications for affective experience. *Psychological Science* 30, 4 (2019), 597–611. doi:10.1177/0956797619831610

[22] Camilla L Nord and Sarah N Garfinkel. 2022. Interoceptive pathways to understand and treat mental health conditions. *Trends in cognitive sciences* 26, 6 (2022), 499–513.

[23] Martin P. Paulus and Murray B. Stein. 2010. Interoception in anxiety and depression. *Brain Structure and Function* 214, 5-6 (2010), 451–463. doi:10.1007/s00429-010-0258-9

[24] Martin P Paulus and Murray B Stein. 2010. Interoception in anxiety and depression. *Brain structure and Function* 214, 5 (2010), 451–463.

[25] Giampaolo Perna, Barbara Barbini, Silvia Cocchi, Angelo Bertani, and Mariangela Gasperini. 1995. 35% CO2 challenge in panic and mood disorders. *Journal of Affective Disorders* 33, 3 (1995), 189–194.

[26] Olga Pollatos, Eva Traut-Mattausch, and Rainer Schandry. 2009. Differential effects of anxiety and depression on interoceptive accuracy. *Depression and Anxiety* 26, 2 (2009), 167–173. doi:10.1002/da.20504

[27] Lenore S. Radloff. 1977. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* 1, 3 (1977), 385–401. doi:10.1177/014662167700100306

[28] Christopher Ring, Jules Brener, Kent Knapp, and John Mailloux. 2015. Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: a cautionary tale about interoceptive awareness. *Biological Psychology* 104 (2015), 193–198. doi:10.1016/j.biopsycho.2014.12.010

[29] Anil K. Seth. 2013. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences* 17, 11 (2013), 565–573. doi:10.1016/j.tics.2013.09.007

[30] R. P. Snaith, M. Hamilton, S. Morley, A. Humayan, D. Hargreaves, and P. Trigwell. 1995. A Scale for the Assessment of Hedonic Tone: The Snaith–Hamilton Pleasure Scale. *British Journal of Psychiatry* 167, 1 (1995), 99–103. doi:10.1192/bjp.167.1.99

[31] Robert L. Spitzer, Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe. 2006. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* 166, 10 (2006), 1092–1097. doi:10.1001/archinte.166.10.1092

[32] Ana Tajadura-Jiménez, Aleksander Väljamäe, and Daniel Västfjäll. 2008. Self-representation in mediated environments: the experience of emotions modulated by auditory-vibrotactile heartbeat. *Cyberpsychology & Behavior* 11, 1 (2008), 33–38. doi:10.1089/cpb.2007.0002

[33] Dominic A. Trevisan, Matthew R. Altschuler, Anna Bagdasarov, Cristina Carlos, Shuqiao Duan, Efrat Hamo, Shreya Kala, Madeline L. McNair, Tyler Parker, Daniel Stahl, et al. 2019. A meta-analysis on the relationship between interoceptive awareness and alexithymia: Distinguishing interoceptive accuracy and sensibility. *Journal of Abnormal Psychology* 128, 8 (2019), 765–776. doi:10.1037/abn0000454

[34] Daniela A. Trevisan, Wolf E. Mehling, and James C. McPartland. 2019. Interoception and alexithymia: A systematic review and meta-analysis of the relationship between interoception and alexithymia. *Neuroscience and Biobehavioral Reviews* 98 (2019), 12–29. doi:10.1016/j.neubiorev.2018.12.004

[35] Stuart Valins. 1966. Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology* 4, 4 (1966), 400–408. doi:10.1037/h0023791

[36] Yiran Zhao, Jatin Arora, Yugie Tao, David Miller, Alexander T. Adams, and Tanzeem Choudhury. 2024. Translation Effectiveness of Offset Heart Rate Biofeedback as a Mindless Intervention for Alcohol Craving Among Risky Drinkers: Controlled Experiment. *JMIR Formative Research* 8 (2024), e54438.

Table 8. Exploratory interactions between baseline psychological traits and stimulation condition across outcome measures.

Outcome	Estimate	s.e.	P value	FDR P
Baseline CES-D				
Perceived accuracy	-0.03	0.07	0.6550	0.8238
BMIS	-0.10	0.23	0.6744	0.8238
BMIS s.d.	0.05	0.17	0.7685	0.8238
CES-D	0.29	0.24	0.2417	0.7093
Confidence	0.06	0.06	0.3796	0.7093
GAD-7	-0.04	0.12	0.7784	0.8238
GHQ-12	0.20	0.25	0.4249	0.7498
MAIA total	-0.16	0.72	0.8275	0.8451
SHAPS	-0.14	0.12	0.2648	0.7093
SensHB	-0.70	0.49	0.1776	0.7093
Baseline GAD-7				
Perceived accuracy	-0.02	0.08	0.7793	0.8238
BMIS	-0.27	0.28	0.3482	0.7093
BMIS s.d.	0.34	0.21	0.1182	0.7093
CES-D	1.42	0.29	0.0001	0.0031**
Confidence	-0.10	0.08	0.2053	0.7093
GAD-7	0.84	0.15	< 0.0001	0.0012**
GHQ-12	1.06	0.30	0.0022	0.0361*
MAIA total	0.44	0.86	0.6136	0.8238
SHAPS	0.09	0.14	0.5224	0.8036
SensHB	-0.25	0.71	0.7251	0.8238
Baseline GHQ-12				
Perceived accuracy	0.10	0.09	0.2925	0.7093
BMIS	0.30	0.30	0.3244	0.7093
BMIS s.d.	-0.31	0.22	0.1675	0.7093
CES-D	-1.08	0.31	0.0024	0.0361*
Confidence	-0.05	0.08	0.5566	0.8145
GAD-7	-0.17	0.16	0.2961	0.7093
GHQ-12	-0.58	0.32	0.0846	0.7093
MAIA total	0.63	0.91	0.4996	0.8036
SHAPS	0.19	0.15	0.2152	0.7093
SensHB	0.86	0.72	0.2471	0.7093
Baseline MAIA total				
Perceived accuracy	0.00	0.01	0.7126	0.8238
BMIS	-0.01	0.04	0.7510	0.8238
BMIS s.d.	0.01	0.03	0.7540	0.8238
CES-D	-0.04	0.05	0.3901	0.7093
Confidence	0.01	0.01	0.3724	0.7093
GAD-7	-0.01	0.02	0.7776	0.8238
GHQ-12	-0.02	0.05	0.7132	0.8238
MAIA total	0.18	0.14	0.2024	0.7093
SHAPS	0.03	0.02	0.1410	0.7093
SensHB	-0.05	0.09	0.6027	0.8238
Baseline SHAPS				
Perceived accuracy	-0.17	0.14	0.2169	0.7093
BMIS	0.52	0.46	0.2685	0.7093
BMIS s.d.	-0.42	0.34	0.2302	0.7093
CES-D	-0.48	0.48	0.3296	0.7093
Confidence	0.05	0.12	0.7115	0.8238
GAD-7	0.01	0.24	0.9560	0.9560
GHQ-12	0.14	0.49	0.7826	0.8238
MAIA total	-1.00	1.41	0.4881	0.8036
SHAPS	-0.62	0.23	0.0150	0.1794
SensHB	0.41	1.01	0.6936	0.8238
Baseline TAS total				
Perceived accuracy	-0.02	0.03	0.5187	0.8036
BMIS	0.09	0.10	0.3424	0.7093
BMIS s.d.	-0.11	0.07	0.1410	0.7093
CES-D	-0.24	0.10	0.0268	0.2676
Confidence	0.02	0.03	0.3574	0.7093
GAD-7	-0.01	0.05	0.8310	0.8451
GHQ-12	-0.06	0.10	0.5498	0.8145
MAIA total	-0.43	0.30	0.1640	0.7093
SHAPS	-0.03	0.05	0.4836	0.8036
SensHB	0.23	0.20	0.2760	0.7093

Exploratory interaction models tested whether baseline psychological traits moderated the effect of stimulation condition across mental health, affective, and interoceptive outcomes. p-values were corrected using false discovery rate (FDR). Perceived accuracy refers to participants' ratings of how accurately the device reflected their heart rate. BMIS indicates mean Brief Mood Introspection Scale valence across the week; BMIS s.d. indicates intra-week variability in BMIS valence. SE, standard error; CES-D, Center for Epidemiologic Studies Depression Scale; GAD-7, Generalized Anxiety Disorder scale; GHQ-12, General Health Questionnaire; MAIA, Multidimensional Assessment of Interoceptive Awareness; SHAPS, Snaith–Hamilton Pleasure Scale; SensHB, heartbeat sensitivity; TAS, Toronto Alexithymia Scale.

Table 9. Condition-specific interactions between baseline mental health factors and stimulation condition.

Outcome	Term	Estimate	SE	p	95% CI lower	95% CI upper
GHQ-12	Real × GAD-7 baseline	-0.025	0.376	0.947	-0.786	0.736
GHQ-12	Sham × GAD-7 baseline	-1.068	0.376	0.007**	-1.828	-0.307
GAD-7	Real × GAD-7 baseline	-0.620	0.173	0.001***	-0.971	-0.270
GAD-7	Sham × GAD-7 baseline	-1.451	0.173	< 0.0001***	-1.801	-1.100
CES-D	Real × GAD-7 baseline	-0.022	0.460	0.961	-0.950	0.906
CES-D	Sham × GAD-7 baseline	-1.329	0.460	0.006	-2.258	-0.340
CES-D	Real × GHQ-12 baseline	-0.255	0.522	0.626	-1.296	0.786
CES-D	Sham × GHQ-12 baseline	0.807	0.512	0.120	-0.216	1.829

Terms are shown separately for real and sham stimulation for outcomes that showed significant baseline mental health × condition interactions. SE, standard error; CI, confidence interval; GHQ-12, General Health Questionnaire; GAD-7, Generalized Anxiety Disorder scale; CES-D, Center for Epidemiologic Studies Depression Scale.

Table 10. Summary of preregistered analyses and findings.

H	Analysis	Key finding	Interpretation
H1-a	Main effect of condition on mental health	No significant main effect of feedback condition on any mental health outcome (all $p > .20$; CES-D, GAD-7, GHQ-12, SHAPS, TAS-20; Table 5)	Real heartbeat-synchronized feedback did not improve mental health outcomes relative to sham across the full sample
H1-b	Moderation of mental health effects by preregistered baseline traits	None of the preregistered moderators significantly moderated GHQ-12 psychological distress (all $p > .36$; Table 6); moderation of other mental health outcomes was similarly non-significant	Preregistered trait and compliance measures did not globally determine responsiveness to feedback condition
H1-c	Covariation of GHQ-12 and MAIA-2 change	Positive but non-significant association between real–sham difference scores in GHQ-12 and MAIA-2 ($\beta = 0.63$, $p = .21$)	Improvements in mental health did not reliably co-occur with increases in self-reported interoceptive awareness
H2-a	Main effect of condition on interoceptive accuracy	No significant main effect of condition on SensHB slope (real: $M = -0.01$; sham: $M = 0.23$; $p = .414$; Table 5)	Heartbeat-synchronized feedback did not enhance behavioral interoceptive accuracy over one week of everyday use
H2-b	Moderation of interoceptive accuracy by preregistered baseline traits	None of the preregistered moderators significantly moderated SensHB outcomes (all $p > .26$; Table 7)	Alexithymia and self-regulation did not shape interoceptive accuracy responsiveness to feedback
H3	Main effect of condition on interoceptive confidence	No significant Real–Sham difference in mean confidence ratings ($t(24) = 0.68$, $p = .502$); the linear mixed model also showed no condition effect ($p = .502$)	Real feedback did not increase subjective confidence in heartbeat estimation relative to sham
H4-a	Change in MAIA-2 from baseline to Week 2	MAIA-2 total score did not change significantly; Noticing ($t(27) = 2.61$, $p = .015$) and Body Listening ($t(27) = 2.67$, $p = .032$) increased significantly	Study participation produced transient gains in specific dimensions of interoceptive awareness
H4-b	Persistence of MAIA-2 gains at Week 3 follow-up	No interoceptive measure remained significantly elevated one week after device removal (Figure 3)	Gains in interoceptive awareness were not sustained beyond the active intervention period
H5	Main effect of condition on mean daily mood	No significant condition effect on BMIS mean (real: $M = 44.03$; sham: $M = 44.90$; $p = .263$; Table 5)	Real feedback did not improve average daily affective tone relative to sham
H6	Main effect of condition on mood variability	No significant condition effect on BMIS SD (real: $M = 4.04$; sham: $M = 4.49$; $p = .469$; Table 5)	Feedback condition did not alter intra-week variability in daily mood

All analyses followed the preregistered statistical plan (AsPredicted #248,039). No preregistered hypotheses were supported at $\alpha = .05$ except H4-a, where Noticing and Body Listening increased significantly at Week 2 but MAIA-2 total score did not. FDR correction was applied within predefined outcome families.

Table 11. Summary of exploratory analyses and findings.

Related RQ	Analysis	Key finding	Interpretation
RQ1	Baseline anxiety moderation of mental health	Baseline GAD-7 significantly moderated condition effects on CES-D ($\beta = 1.42$, $p_{FDR} = .003$), GAD-7 ($\beta = 0.84$, $p_{FDR} = .001$), and GHQ-12 ($\beta = 1.06$, $p_{FDR} = .036$). Baseline CES-D, SHAPS, MAIA-2 total, and TAS-20 did not reach significance after FDR correction.	Baseline anxiety emerged as the strongest and most consistent moderator of symptom change; effects were specific to affective symptoms and did not extend to interoceptive or mood variability outcomes
RQ1	Condition-specific decomposition of anxiety moderation	For GAD-7, interactions were significant under both sham ($\beta = -1.45$, $p < .001$) and real ($\beta = -0.62$, $p = .001$), with a larger effect under sham. For GHQ-12, the interaction was significant under sham only ($\beta = -1.07$, $p = .007$).	High-anxiety participants showed larger anxiety symptom reductions under sham than under real feedback; real feedback showed smaller condition-specific effects on anxiety and no significant condition-specific effect on general distress
RQ1	Baseline general distress moderation of depression	Baseline GHQ-12 significantly moderated condition effects on CES-D ($\beta = -1.08$, $p_{FDR} = .036$); higher baseline distress predicted greater depression reduction under real versus sham.	Higher baseline general distress was associated with a relative advantage of real over sham feedback for depressive symptoms
RQ1, RQ2	Interoception–mental health dissociation	Symptom improvements were not accompanied by changes in SensHB slope, interoceptive confidence, or MAIA-2 total score across conditions.	Mental health changes were not reliably associated with measurable changes in behavioral interoceptive accuracy or awareness
RQ1, RQ3	Long-term follow-up in high-anxiety participants	High-anxiety participants showed significant reductions in GAD-7, GHQ-12, and TAS-20 at Week 3 (all $p < .05$, FDR-corrected); several MAIA-2 subscales also decreased significantly.	Follow-up symptom reductions co-occurred with reduced self-reported interoceptive awareness, supporting proposed recalibration mechanism
RQ2, RQ3	Baseline interoception–mental health correlations	SensHB slope was significantly associated with GAD-7 ($\beta = -1.49$, $p = .0003$) and GHQ-12 ($\beta = -1.53$, $p = .016$), but not MAIA-2, SHAPS, or TAS-20. MAIA-2 total was associated with SHAPS ($\beta = 0.044$, $p = .013$)	Interoceptive accuracy (SensHB) and self-reported awareness showed dissociated baseline correlates, supporting a multidimensional account of interoception

All analyses reported in this table were exploratory unless noted as extensions of preregistered tests. Table 8 presents the comprehensive moderation table. p-values were corrected using false discovery rate (FDR) where applicable.

Table 12. Participant characteristics ($N = 29$).

Characteristic	Value
Sex	Female: 19 (65.5%); Male: 10 (34.5%)
TAS-20 total score	56.97 ± 4.32 (51–66)
TAS-20 category	Borderline: 22 (75.9%); Alexithymic: 7 (24.1%); Non-alexithymic: 0 (0%)

Values are mean ± s.d. (range) or n (%).

Table 13. Summary of psychological measures and clinical cutoffs.

Measure	**Construct**	**Clinical Interpretation**
CES-D [27]	Depressive symptoms	≥16 indicates clinically relevant depressive symptoms
GAD-7 [31]	Anxiety severity	5 = mild, 10 = moderate, 15 = severe
SHAPS [30]	Hedonic capacity	
GHQ-12 [14]	General psychological distress	≥13 indicates psychological distress (Likert scoring)
TAS-20 [2, 4]	Alexithymia traits	≥61 = alexithymia; 52–60 = borderline