

When Feeling and Physiology Diverge: Understanding Dual-Indicator Stress Sensemaking and Micro-Interventions in an Emotional-Labor Workplace

JEONGHYUN KIM, KAIST, Republic of Korea

YUGYEONG JUNG, KAIST, Republic of Korea

JUNMO LEE, KAIST, Republic of Korea

KWANGYOUNG LEE, KAIST, Republic of Korea

HWAJUNG HONG, KAIST, Republic of Korea

UICHIN LEE*, KAIST, Republic of Korea

Sensing-based stress management systems integrate self-report, physiological, and contextual data to assess users' perceived stress and physiological indicators related to stress responses and deliver just-in-time interventions. Prior work has largely focused on improving detection accuracy or evaluating intervention effectiveness, with relatively limited attention to how users interpret and make sense of these data in everyday work contexts. Our study addresses this gap by examining how emotional labor workers construct meaning around alignment and divergence between perceived stress and physiological responses to stress. We present a four-week in-the-wild study with 19 call center workers, combining mobile and wearable sensing, a reflection dashboard, and in-depth interviews. Our findings show that participants interpreted stress indicators through work context, bodily conditions, and prior experiences, and treated divergence as an informative cue rather than a simple error. They also recognized that perceived and physiological responses to stress could change differently before and after micro-interventions. Based on these findings, we discuss design implications for sensing-based stress management systems that support stress data literacy and flexible, context-grounded stress sensemaking in practice.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: dual-indicator stress, perceived stress, physiological stress, sensemaking, emotional labor, micro-interventions

ACM Reference Format:

Jeonghyun Kim, Yugyeong Jung, Junmo Lee, Kwangyoung Lee, Hwajung Hong, and Uichin Lee. 2026. When Feeling and Physiology Diverge: Understanding Dual-Indicator Stress Sensemaking and Micro-Interventions in an Emotional-Labor Workplace. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 2, Article 49 (June 2026), 46 pages. <https://doi.org/10.1145/3810198>

1 INTRODUCTION

Managing stress has become central to sustaining mental health in everyday life [79, 121]. Accordingly, researchers in ubiquitous computing and personal informatics have explored sensing-based stress management systems

*Corresponding author.

Authors' Contact Information: [Jeonghyun Kim](mailto:jeonghyun.kim@kaist.ac.kr), KAIST, Daejeon, Republic of Korea, jeonghyun.kim@kaist.ac.kr; [Yugyeong Jung](mailto:yugyeong.jung@kaist.ac.kr), KAIST, Daejeon, Republic of Korea, yugyeong.jung@kaist.ac.kr; [Junmo Lee](mailto:junmo.lee@kaist.ac.kr), KAIST, Daejeon, Republic of Korea, junmo.lee@kaist.ac.kr; [Kwangyoung Lee](mailto:kwangyoung.lee@kaist.ac.kr), KAIST, Daejeon, Republic of Korea, kwangyoung@kaist.ac.kr; [Hwajung Hong](mailto:hwajung.hong@kaist.ac.kr), KAIST, Daejeon, Republic of Korea, hwajung@kaist.ac.kr; [Uichin Lee](mailto:uichin.lee@kaist.ac.kr) (corresponding author), KAIST, Daejeon, Republic of Korea, uilee@kaist.ac.kr.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

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

ACM 2474-9567/2026/6-ART49

<https://doi.org/10.1145/3810198>

that leverage mobile and wearable technologies [45, 60, 133]. These systems estimate users' stress states by capturing diverse physiological and behavioral signals, including heart rate, physical activity, sleep patterns, location, and device usage [33, 118], and support stress reduction in daily life by delivering just-in-time (JIT) micro-interventions [96, 103].

In these studies, two primary types of stress indicators have been used to measure stress states. One is *perceived stress*, measured through self-reports or experience sampling methods (ESM) [20, 146]. The other is *physiological indicator related to stress responses*, often operationalized using biological signals such as heart rate variability (HRV) [125, 140]. In our work, this physiological indicator was quantified using RMSSD (Root Mean Square of Successive Differences), a commonly used HRV metric. We utilize RMSSD as a key physiological signal that reflects stress-related autonomic activity. Perceived stress reflects individuals' subjective experiences grounded in their emotions, cognitions, and appraisals [20, 79], whereas the HRV-based physiological indicator is associated with bodily responses such as arousal, recovery, and fatigue [69, 140]. Prior work has shown that these indicators can exhibit divergent patterns depending on individual differences, environments, and contextual factors [15, 40, 91]. This suggests that stress is inherently multifaceted and cannot be fully captured by a single indicator alone.

Despite this understanding of stress as a multifaceted phenomenon, much of the prior work on sensing-based stress management systems has primarily focused on technical goals, such as improving the performance of stress detection or prediction models [42, 101, 118] and providing feedback through visualization [127, 153]. While these efforts have advanced system accuracy and efficiency, they have paid relatively limited attention to how users actually understand and interpret stress data, or how such interpretations shape their perceptions of their own stress. A similar emphasis appears in prior studies on stress interventions, which typically assess effectiveness by comparing stress indicators before and after an intervention [16, 22, 72, 110]. This outcome-oriented perspective, however, tends to obscure how interventions are experienced differently across individuals, how users make sense of changes in their own states during the intervention process, and what these interpretations mean for their broader understanding of stress [88, 141].

These limitations matter because stress is not merely a physiological reaction but also an experience constructed through individuals' appraisals, interpretations, and meaning-making processes [79]. Even in the same situation, one person may experience high levels of stress by interpreting it as a threat, whereas another may experience it differently by construing it as a challenge. From this perspective, stress data are not inherently meaningful when presented solely as *outcomes* or *metrics*; instead, they become meaningful through interpretation and when situated within the context of users' lived experiences [85, 109]. Accordingly, the process through which users reflect on the causes and effects of their stress based on their data can serve as an important lens into how individuals understand their stress experiences.

Building on this background, our study investigates how people understand and interpret diverse stress-related and intervention data collected through sensing-based stress management systems. We focus on *call center environments* as an emotional labor workplace where stress is frequent and salient. Due to the repetitive and emotionally demanding nature of their work, call center workers commonly experience both perceived stress (e.g., pressure during customer interactions) and physiological stress-related bodily responses captured through HRV-based indicators (e.g., increased heart rate and decreased heart rate variations), making this a representative context for studying stress experiences [44, 50, 120, 144]. In addition, the short intervals between calls and the high density of tasks make it important not only to provide immediate stress relief but also to support workers in understanding when and why they experience stress, which is critical for sustaining mental well-being and performance.

In our study, we conducted a four-week in-the-wild study with 19 call center workers at a call center in South Korea. We collected multiple types of stress-related data through mobile and wearable devices, including perceived stress, an HRV-based physiological indicator, work context, and stress intervention outcomes. After the data collection, we provided participants with a dashboard visualizing their stress levels and post-intervention

changes, and explored how they interpreted this data and constructed meaning around their experiences. For ease of communication in the user-facing system and interviews, participants encountered the label “physiological stress,” but in this paper we refer to it as an HRV-based physiological stress indicator. Based on this study, we formulated two research questions:

- **RQ1:** In the emotional labor workplace, how do participants integrate and interpret **perceived and physiological stress indicators with contextual data** to make sense of their stress?
- **RQ2:** In the emotional labor workplace, how do participants evaluate the provided micro-interventions, and how do these interventions influence their sensemaking of **changes in perceived and physiological stress indicators**?

Our findings indicate that participants perceived the two stress indicators (e.g., perceived and physiological stress) as representing distinct dimensions. They tended to interpret perceived stress as a subjective emotional experience, whereas they understood physiological stress in terms of bodily reactions such as discomfort or tension. However, most participants did not clearly recognize that physiological stress was measured using RMSSD, an HRV-derived metric based on biological signals. The collected data also revealed divergent patterns between the two indicators, and changes in perceived stress did not consistently correspond with changes in physiological stress following interventions. This reflective process appeared to reshape how participants understood their own stress experiences. Based on these findings, we derive design implications for future stress management systems. In summary, our work has the following contributions:

- We conducted a four-week in-the-wild study in a call center setting characterized by emotional labor, systematically examining how people understand and interpret multiple stress indicators (including perceived and physiological stress), contextual information, and intervention outcomes.
- We propose design implications for future sensing-based stress management systems, including support for stress data literacy, sensemaking that treats discrepancy as an informative signal, and value-based intervention and reflection mechanisms that balance measured efficacy, user-perceived value, and feasibility constraints.

2 BACKGROUND AND RELATED WORK

In this section, we review how stress has been defined and measured, examine prior work that leverages stress data to support self-reflection and regulation, and discuss studies on emotional laborers (e.g., call center workers), highlighting opportunities for deploying stress-sensing and intervention technologies.

2.1 Defining, Measuring, and Contextualizing Stress

Stress is not a single construct but an umbrella term that spans subjective appraisal (perceived stress) and bodily response (physiological stress) [126]. Across psychology, Lazarus and Folkman’s *cognitive appraisal theory* has provided a foundational framework for understanding stress as a constructed process, in which individuals’ interpretations of situations shape both subjective experience and subsequent bodily responses [79]. Prior work has therefore conceptualized stress not as a unitary response but as a multifaceted process in which subjective appraisal and physiological responses operate at distinct yet interacting levels [15, 55, 147].

Perceived stress is defined as a subjective state grounded in how individuals appraise and make meaning of a given situation. As such, perceived stress reflects lived experience and is most commonly assessed through in-situ self-report using a standardized instrument such as the Perceived Stress Scale (PSS) using an Ecological

Momentary Assessment (EMA) [61, 64, 65]. Researchers have explored the use of behavioral and expressive cues, such as facial expressions and vocal characteristics, to infer subjective stress states [76, 143].

Physiological stress refers to autonomic and endocrine responses mediated by biological pathways such as the Sympathetic Adrenal Medullary (SAM) and the Hypothalamic–Pituitary–Adrenal (HPA) axes [55, 77, 123]. It is understood as conceptually distinct from subjective experience. In the mobile health domain, it is commonly operationalized using *derived physiological features and patterns*, most prominently heart rate variability (HRV), extracted from wearable sensor signals reflecting heart activity (e.g., reduced HRV under stress-related autonomic arousal) [51, 93, 94, 107, 108, 112]. This approach enables continuous, passive data collection without requiring active user input and has been widely adopted in commercial wearable devices to represent physiological stress, where proprietary stress indicators are typically grounded in HRV-based autonomic features; for example, Garmin’s stress scoring relies largely on HRV [116].

Prior work has treated perceived stress and physiological stress as separate but interacting constructs operating at different levels, with each indicator following different measurement strategies and interpretive logics. Building on this distinction, studies have incorporated contextual information such as location, activity, and smartphone usage patterns to better situate and interpret stress states [23, 70, 73, 78, 92, 101, 132, 149]. Collectively, this line of work has established a foundation for understanding stress as a multidimensional and context-dependent phenomenon in everyday settings.

2.2 Relationships between Perceived and Physiological Stress in Emotional Labor Contexts

Psychology and affective science conceptualize perceived and physiological stress as related yet dissociable processes rather than a single unified construct. Prior work suggests that psychosocial and physiological stressors engage distinct response patterns, underscoring that lived stress experiences and bodily signals do not necessarily map one-to-one [9, 55, 74]. Consistent with this view, ambulatory and workday studies have reported that the correspondence between self-reported stress and cardiovascular measures (e.g., HRV) is often modest and heterogeneous in daily life. This variability reflects differences in measurement conditions or individual characteristics, reinforcing the need to interpret physiological indicators alongside contextual information [36, 90, 105, 142].

In occupational settings, including call centers, work stress is commonly framed through psychosocial job characteristics (e.g., high demands, low control, low support, effort–reward imbalance, and resource constraints) that can influence both subjective strain and physiological regulation [5, 66, 131]. Studies in call center settings have examined links between work-related strain and autonomic indicators, including HRV-related indices, and have reported associations that can vary by task context and individual differences [31]. Despite this accumulated evidence highlighting contextual and individual variability, HCI and ubiquitous sensing research in call center settings has primarily focused on quantifying relationships between indicators, building stress recognition and prediction models [31, 49], or evaluating intervention efficacy [68]. In these paradigms, one indicator is typically treated as a “ground truth” for the other.

As a result, while perceived–physiological relationships are well explored at statistical and experimental levels, prior work offers limited insight into how workers themselves interpret relationships between indicators. To address this gap, our work shifts the focus from estimating association strength or minimizing prediction error to examining how workers make sense of divergence between perceived and physiological stress in everyday practice. We analyze how participants mobilize contextual information, accumulated experiences, and intervention outcomes to explain stress indicators, and how these interpretations shape subsequent reflection and intervention preferences.

2.3 Data-Driven Stress Self-Reflection and Just-in-time (JIT) Intervention

Prior personal informatics research has explored systems that help people self-track personal data [71] and reflect on their stress by feeding back their data through wearables, mobile apps, and visualization systems [64, 65, 100]. These studies aim to increase self-awareness and support personalized coping strategies through data-driven feedback.

Prior work makes stress more interpretable by transforming raw signals into digestible representations. For example, *VetherReflect* uses weather metaphors to surface mismatches between expectations and sensed data [145]. *LifelogExplorer* personalizes lifelog-based monitoring of daily stress [73], and smartwatch-based mHealth tools [54] provide real-time feedbacks. Other work visualizes affect and physiology during commuter driving [35] and combines journaling with data feedback [111]. Momentary logging enables in situ reflection [101]. Beyond individuals, collective displays foster team-level awareness [153] and collaborative reinterpretation in high-stress training [2].

These systems demonstrate that data-driven reflection can make stress *visible* and can motivate coping. While prior systems successfully surface stress data to end users, reflection is usually organized around a single target or an aggregate score. Even when multiple streams exist, interfaces seldom scaffold *comparative* reasoning (e.g., side-by-side juxtaposition, prompts to reconcile divergences between perceived and physiological responses). To address this gap, effective support should help users determine whether perceived and physiological responses *align* or *diverge*, given imperfect interoceptive awareness; if individuals can recognize their own physiological changes, they can apply coping strategies more promptly and effectively [40]. Moreover, greater discordance between perceived and physiological stress has been associated with higher internalizing symptoms and accelerated symptom growth, while alignment between the two is a characteristic of more resilient trajectories [62, 148]. This motivates the design that co-presents perceived and physiological stress with contextual factors and explicitly prompt users to examine (dis)agreement and plausible causes.

2.4 Emotional Labor and Call Center Workers

Workplaces frequently impose stressors outside an individual's control (e.g., customer demands, time pressure, monitoring), creating conditions under which subjective appraisal and bodily response can decouple. In particular, *emotional labor*, which regulates inner feelings to display organizationally required emotions (e.g., friendliness, courtesy), is governed by explicit display rules that promote *surface acting* and suppression [50]. Such regulation produces *emotional dissonance*, the gap between felt and displayed emotions, which is a known contributor to burnout, depression, and anxiety [44]. In these environments, limited autonomy and heavy workloads further constrain attentional resources, making interoceptive awareness imperfect and increasing the likelihood that perceived stress and physiological arousal will *misalign* [40]. Call center workers exemplify high-risk emotional laborers: over half report mental health problems and about 45% are at risk, driven by heavy workloads, time pressure, and surface acting, exacerbated by low autonomy and limited organizational support [138]. These conditions harm well-being and retention, underscoring the need for effective stress-management support.

Prior work has instrumented emotional labor with multimodal sensing of voice, behavior, and physiology (e.g., PPG, EDA, EEG) [104], non-invasive workplace tracking [1], and longitudinal emotion datasets [137]. Intervention studies have widely explored micro- and just-in-time microinterventions delivered via wearables or mobile applications [30, 54, 73, 117]. Automated e-coaching and JITAI receptivity modeling further personalize support [83, 84]. Biofeedback visualizations and group-reflection systems help workers interpret and act on stress signals [2, 152, 153]. For agent-facing support, systems modulate customer vocal negativity in real time or offer response tips [21, 80]. Delivery timing is typically biosignal-triggered or scheduled around anticipated stressors [30, 117].

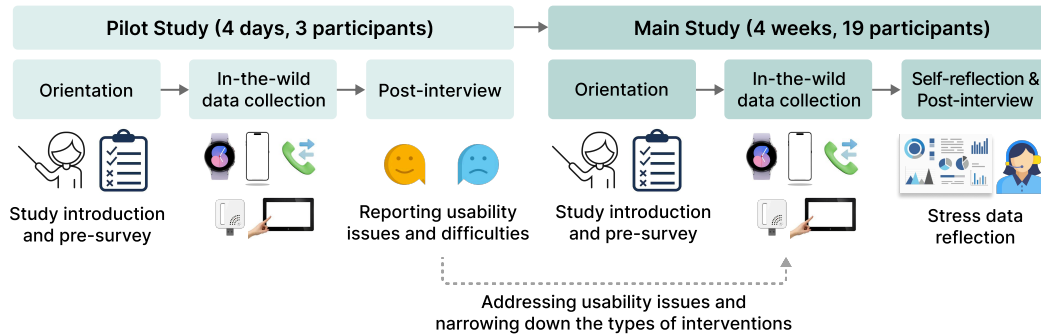


Fig. 1. Overview of the study protocol. A pilot study was conducted to identify usability issues and refine the design of intervention delivery for better work process integration. After addressing these issues and narrowing down the types of interventions, a 4-week main study was conducted, including an orientation session, in-the-wild stress data collection, and stress data self-reflection and post-interviewing.

In call centers, emotional display rules, low autonomy, and high call volume are precisely the conditions under which perceived and physiological indicators may diverge [44, 50]. For example, suppressed expression can lower reported stress while autonomic arousal remains elevated, or habituation can blunt autonomic change despite high perceived strain. At the same time, the transactional, call-by-call structure affords fine-grained self-tracking (pre/post call, intervention windows) and rich contextual labeling (call type/difficulty, environment). This combination makes call centers a suitable testbed for examining how workers interpret *alignment* and *misalignment* between perceived and physiological stress, and for identifying what kinds of interface scaffolds are needed to build dual-indicator literacy in practice.

3 STUDY DESIGN

This section presents the overall study design, covering the study procedure, in-the-wild stress data collection with emotional workers, and stress data reflection after the data collection.

3.1 Overview of Study Procedure

Figure 1 shows the two-phase study procedure, consisting of a pilot and main study. Both followed the same core protocol: (1) orientation, (2) in-the-wild data collection, and (3) post-interview. During orientation, participants were briefed on the study and completed two pre-surveys on work-related stress (KOSS-19 [135], K-BAT [17]). The pilot study (4 days, 3 participants) ensured that real-world deployment would not disrupt workflow and helped refine the procedure. The main study (4 weeks, 19 participants) applied the refined protocol from the pilot study, and post-interviews explored how participants interpreted and reflected on their stress-related data.

All procedures were IRB-approved, and informed consent was obtained. Because many aspects of the pilot and main studies overlapped, we primarily describe the study design of the *main study* and describe the refinements informed by the pilot study where relevant.

3.2 In-the-Wild Stress Data Collection

3.2.1 Participants. We recruited 19 call center workers (17 female, age: $M = 38.90$, $SD = 5.81$) from a Metropolitan City Call Center in South Korea (see Appendix A for demographics and presurvey results). This female-skewed sample reflects the broader South Korea call center workforce, where prior studies report that over 90% of workers

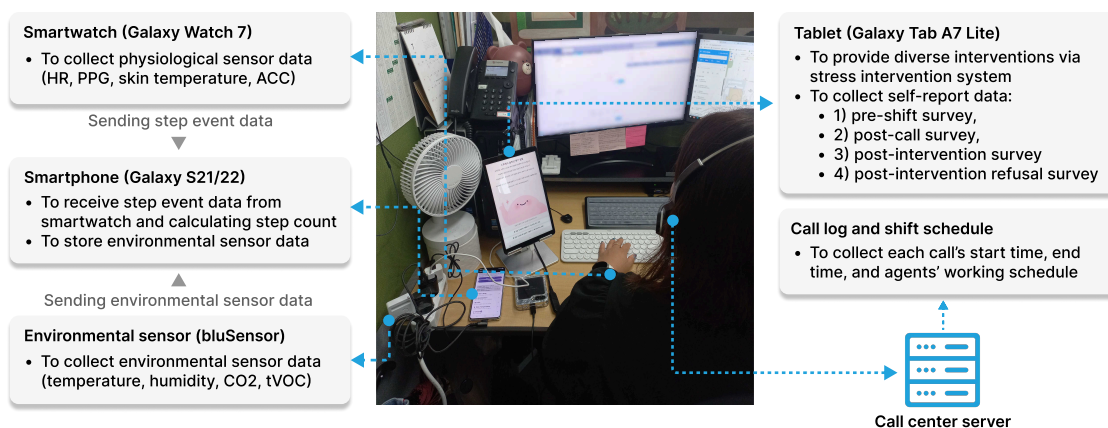


Fig. 2. Overview of the in-the-wild stress data collection setup. Physiological signals were collected using a smartwatch, while environmental data were captured using an environmental sensor. A smartphone relayed step and environmental sensor data, and a tablet application delivered stress interventions and collected self-reports. Call logs and shift schedules from the call center server were used to contextualize stress data in relation to working conditions.

are women [10, 18, 58]. The average job tenure was 3.58 years ($SD = 1.35$). Participation was voluntary with written informed consent, and each participant received \$284.39 compensation. The call center operates with staggered 8-hour shifts that ensure continuous service coverage across weekdays and weekends. Inbound calls enter a shared queue and are routed to the next available worker, creating social pressure to minimize breaks because downtime increases colleagues' workload. Performance evaluations emphasize call volume, and workers frequently handle complaints and emotionally demanding inquiries, resulting in high work-related stress.

3.2.2 Data Collection Setup. We designed the data collection environment to allow call center workers to perform regular call-handling tasks while receiving intervention and logging multimodal stress data. In line with prior in-the-wild stress sensing studies [12, 124, 134], we collected ecologically valid data to later support participant reflection. Figure 2 shows the deployed sensing configuration. A tablet (Galaxy A7 Lite) ran our stress intervention application to deliver intervention prompts and capture self-reports (Section 3.2.3 details the stress intervention system). A smartwatch continuously recorded physiological signals, and an environmental sensor was placed at each seat. A gateway phone (Galaxy S21 or S22) relayed smartwatch data and stored environmental readings to ensure reliable data capture during work tasks; details on the collected data and measures are presented in Section 3.2.5 and implementation details can be found in Section 3.4.

3.2.3 Stress Intervention System. To examine how interventions during real work tasks influence perceived and physiological stress indicators, we developed a tablet-based system integrated into workers' call-handling workflows (Fig. 8). The system delivers brief coping interventions during inter-call periods and collects short self-reports. To minimize workload interference, we selected eight micro-interventions completable within a few minutes. Based on prior studies on intervention delivery [48, 52, 53], we designed an initial set of interventions commonly adopted in occupational stress contexts. These interventions were then tested in the pilot study, and those that workers found difficult to engage with or ineffective for stress relief were excluded; the main study used the remaining eight interventions (Table 1). Following Lazarus and Folkman's stress coping theory [79],

Table 1. Intervention options evaluated in the study.

Coping Strategy	Category	Intervention Type	Intervention Name	
Emotion-focused	Exercise	Stretching	Stretching	
		Breathing	Deep Breathing	
	Mindfulness	Journaling	Gratitude diary	I Did Well, Right?
			Cheering	Words I Need Right Now
		Emotional diary	Anger-Eating Fairy	
	Fidgeting	Pop	Protect Me	
Eating	Eating snacks	Sugar Boost		
Problem-focused	Reframing	Emotion detachment strategy	Is It Because of Me?	

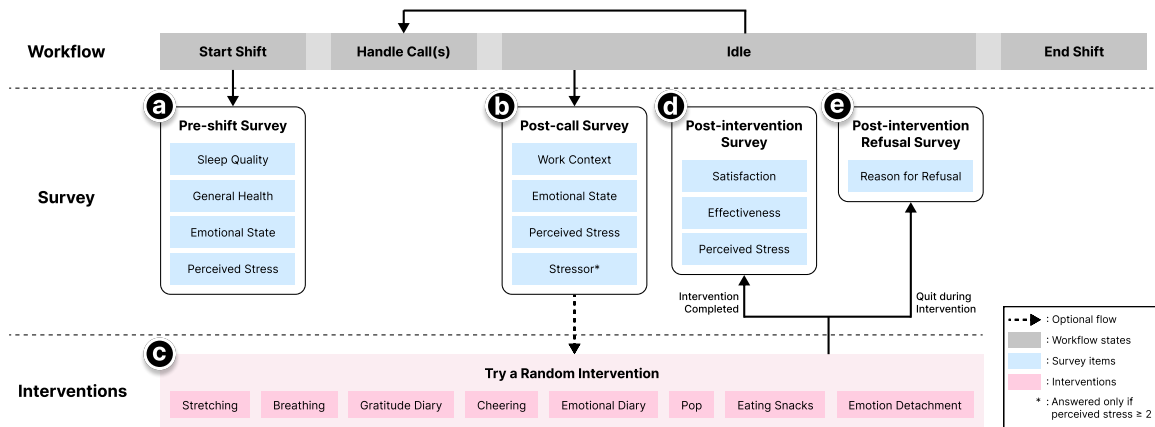


Fig. 3. Overview of the daily task of call center workers. (a) At the beginning of the shift, they completed the pre-shift survey. (b) In the idle state (between the calls), they completed the post-call survey. (c) After the survey, they tried an intervention randomly recommended by the app. (d) If they completed the intervention, they responded to the post-intervention survey. (e) If they quit during the intervention, they completed a post-intervention refusal survey.

we categorized interventions as emotion-focused (regulating emotional responses) or problem-focused (modifying the stressor or appraisal). *Emotion-focused interventions* included stretching [95], guided breathing [39], micro-journaling (gratitude [24], affirmation [129], emotional labeling [39]), fidgeting for discreet tactile engagement [155], and a quick snack option [151]. The *problem-focused intervention* was cognitive reframing to encourage brief emotional detachment [114], which briefly helps call center workers reinterpret uncontrollable caller behavior and adopt a neutral, task-focused stance before the next call. We initially designed 12 interventions but removed four after the pilot due to disruption or increased workload. The details of the final intervention sets are presented in Appendix B.

3.2.4 User Tasks. At the start of each shift, workers completed a *pre-shift survey* (Fig. 3(a)) assessing their overall condition before work, including sleep quality [14], general health [118], perceived stress [20], arousal [156], valence [156], and fatigue [118] on 5-point Likert scales. After finishing a call, they could complete a *post-call*

Table 2. Overview of collected data with source, device, and frequency. Steps were aggregated with a 10-minute window, with the sampling rate of 0.2 Hz. IBI data are collected irregularly, with each measurement providing a batch of interbeat intervals for the entire segment. Abbreviations: HR = heart rate, IBI = inter-beat interval, PPG = photoplethysmogram, ACC = accelerometer, tVOC = total volatile organic compounds.

Method	Source	Device	Data	Frequency
Passive sensing	Call log server	Call center server	Call start time, Call end time, Shift schedule	Per call
			HR	1 Hz
			IBI	-
	Wrist band	Galaxy Watch 7	PPG	25 Hz
			Skin temperature	$\frac{1}{60}$ Hz
			ACC	25 Hz
			Phone	Galaxy S21–22
Environmental sensor	bluSensor (BSP02AIR)	Temperature, Humidity, CO ₂ , tVOC	0.1 Hz	
Self-report	Tablet	Galaxy Tab (A7 Lite)	<i>Pre-shift survey</i> : Sleep quality, General health, Perceived stress, Arousal, Valence, Fatigue	Per day
			<i>Post-call survey</i> : Previous call type, Workload, Perceived stress, Stressor, Arousal, Valence, Fatigue, Surface acting	≥ 15 per day
			<i>Post-intervention survey</i> : Intervention satisfaction, Stress relief effect, Perceived stress	≥ 9 per day
			<i>Post-intervention refusal</i> : Reason for refusal	-

survey (Fig. 3(b)) assessing stress levels and capturing the preceding call context: call type (e.g., general inquiry or complaint), perceived workload, perceived stress, arousal, valence, fatigue, and the extent of surface acting [13]. Stressor details were collected only when perceived stress was 2 or higher, indicating at least some level of stress. A **randomly** chosen intervention followed the post-call survey (Fig. 3(d)) to balance exposure across the eight intervention types and minimize bias. To minimize disruption to their workflow, we supported intervention skipping by allowing them to choose between two options freely: responding only to the post-call survey for the call or responding to the *post-call survey* and completing the intervention. If a call arrived during an intervention, they could quit and indicate the reason through *post-intervention refusal survey* (Fig. 3(e)). After an intervention, they completed a *post-intervention survey* (Fig. 3(d)) assessing satisfaction, perceived stress relief, and perceived stress after the intervention. Full survey items are listed in Appendix C.

To balance ecological validity with burden, we set soft per-shift targets of nine post-call surveys and six post-call surveys paired with an intervention; the application displayed progress toward these targets and cumulative completions to help workers pace participation.

Interface details and screenshots are in Appendix G.

3.2.5 Collected Data and Measures. We collected self-report and passive sensing data to capture detailed contextual information about call center workers' stress experiences during calls and interventions (Tables 2 and 8).

Call log data. We obtained timestamps for all calls and shift schedules to align sensing and self-reports with call events. Because calls can trigger stress, these records enabled analysis of physiological and perceived changes around stress-inducing moments. Shift schedules were used to verify that surveys were completed during their working hours.

Wearable and smartphone sensing data. Participants wore a Galaxy Watch 7 to collect physiological and behavioral signals. We recorded heart rate (HR), inter-beat intervals (IBI), and photoplethysmogram (PPG), which are common inputs for stress estimation [87], and derived our physiological stress indicator from IBI (see Section 3.4 for details). We also collected accelerometer (ACC) data and step counts to provide behavioral context relevant to stress and mental health [43, 128].

Environmental sensing data. A bluSensor captured environmental conditions related to stress, including temperature, humidity, CO₂ concentration, and air quality (tVOC).

Self-reported data. We administered *pre-shift*, *post-call*, *post-intervention*, and *post-intervention refusal* surveys.

3.3 Stress Data Reflection

After data collection, we supported participants' reflection on their stress-related data to help understanding their experiences. We designed a visualization dashboard, extracted salient stress-related features, and used it during post-interviews. The following subsections detail our dashboard design, feature extraction, and reflection procedures.

3.3.1 Data Visualization: The Dual-Indicator Reflection Dashboard. We developed a *dual-indicator reflection dashboard* as the primary tool for our post-interview sensemaking probes.

To address our RQs, the dashboard supports intuitive reflection on the complex and often divergent relationships between stress indicators and context, through two key visualization targets, which became the dashboard's two modes:

- (1) **Stress states** integrating perceived, physiological, and contextual data (Fig. 4)
- (2) **Stress changes** before and after intervention delivery (Fig. 5)

A brief primer explains perceived and physiological stress to aid interpretations of each facet of stress (Fig. 4(a) and Fig. 5(a)). Because presenting this multifaceted data in an interpretable way that effectively surfaced potential divergences was a core design challenge, we iteratively refined the visualization: three authors built rapid prototypes, five HCI researchers reviewed visual encodings and task support, and the lead author implemented the final design.

Exploring multimodal stress states. In the state mode, users first view a calendar summarizing daily stress trends (Fig. 4(b)), where color saturation reflects average stress score. Selecting a date (e.g., July 25th) reveals a daily timeline of perceived and physiological stress indicators (Fig. 4(c)): purple bars indicate perceived stress from surveys, and yellow bars represent normalized RMSSD computed at the same timestamps, black dots mark call events, and green dashed lines indicate performed interventions. This supports observing how the two stress states fluctuate over time and identify moments where they align or diverge.

To explore "why" stress occurs, users scroll to a treemap of factors associated with perceived and physiological stress indicators (Fig. 4(d)). Rectangles represent factors; size and saturation encode association strength, and hue indicates category (stressor, environment, situation, pre-shift). Stressor category refers to direct triggers of stress originating from work or peers, reported in the post-call survey. The environment category reflects sensor-measured workplace conditions. Situation category refers to contextual and physiological factors representing the participant's state during stressful events, including call context, affect, and physiological states (e.g., skin temperature, step count). The pre-shift category captures baseline daily condition from the pre-shift survey. Hovering reveals a tooltip with the factor name and correlation score, and a fixed-template summary highlights top associations for each stress type (Appendix F.1).



Fig. 4. Overview of data visualization dashboard in state mode. (a) Brief explanation of perceived and physiological stress indicators. (b) Calendar that summarizes daily stress trends. (c) Clicking on a specific day of a calendar reveals a timeline of perceived and physiological stress indicators, alongside intervention and call records. (d) Treemap of correlation score between the two stress types and related factors.



Fig. 5. Overview of data visualization dashboard in change mode. (a) Brief explanation of perceived and physiological stress indicators. (b) Calendar that summarizes stress reduction trends. (c) Clicking on a specific day of a calendar reveals a timeline of perceived and physiological stress indicators, alongside intervention and call records. Hovering over a dashed line shows how the intervention changed stress, along with its name. (d) Bar chart showing average stress reduction for each intervention.

Exploring stress changes according to intervention delivery. In the change mode, users begin with a monthly calendar showing whether stress *increased*, *decreased*, or *no-changed* across both weeks and weekdays (Fig. 5(b)); darker colors indicate greater average stress reduction, and days without data remain white. Selecting a date (e.g., July 29th) opens a daily timeline showing how stress changed after each intervention, enabling them to examine “how stress changed” for both perceived and physiological data (Fig. 5(c)). Hovering over a dashed line shows the direction of change for perceived and physiological stress indicators and the completed activity, enabling comparison across instances and stress types.

To explore which interventions were helpful, users scroll to a bar chart ranking interventions by observed stress-reduction effects for each stress type (Fig. 5(d)). Fixed-template summaries highlight key findings to help users compare outcomes with expectations (Appendix F.2).

3.3.2 Data Preprocessing. We derived stress-related features across psychological, physiological, and contextual dimensions for visualization and reflection.

Stress Indicators. Perceived stress scores from surveys served as the psychological indicator. We used root mean square of the successive differences (RMSSD) from inter-beat intervals (IBI) as the physiological indicator, as RMSSD is a standard short-term heart rate variability (HRV) metric reflecting vagally mediated regulation and is widely used in wearable-based stress research [29, 125]. Higher RMSSD indicates greater recovery capacity and reduced stress. While HRV is a widely used indicator of psychological stress [57, 106, 119, 158], it is prone to confounders in free-living settings. Thus, we interpret RMSSD as an *interpretable* cue rather than a direct or clinical proxy of “stress level” [47]. Although log-transformed RMSSD (e.g., lnRMSSD) is often used to reduce skewness and improve statistical comparability across individuals [34, 136], our goal was not to claim an absolute or population-level interpretation. Instead, we focused on a *relative, within-person* representation to facilitate sensemaking by presenting an *interpretable, personalized index* alongside the 0–4 Likert scale for perceived stress. Therefore, we applied *per-participant* min–max scaling [4, 99] over all valid windows to obtain a relative index in the range of 0–4:

$$\text{normalized RMSSD index} = 4 \times \left(1 - \frac{\text{RMSSD} - \min(\text{RMSSD})}{\max(\text{RMSSD}) - \min(\text{RMSSD})} \right)$$

A higher index value indicates higher physiological stress indicator (i.e., lower RMSSD).

This linear scaling preserves equal visual steps in the dashboard and is used as a participant-specific, visualization-oriented index for sensemaking rather than a population-wide measure.

Changes in perceived stress were calculated as the pre-post differences around each intervention. For the physiological feature, we computed RMSSD in pre- and post-intervention windows and measured change as the difference between the corresponding normalized RMSSD indices, yielding a -4 to 4 range.

Data cleansing. We removed missing values, modality-specific outliers, and invalid intervals before feature computation.

Environmental sensors: readings were normalized per device by subtracting each sensor’s full-period mean, aggregated into 10-second windows, and windows without timestamps were discarded.

IBI: Invalid-status samples (e.g., non-wear, low accuracy) were excluded; intervals <300 ms or >2000 ms and 3×IQR outliers [59] were removed; windows without valid IBIs were discarded.

HR: Invalid-status samples (e.g., non-wear, low accuracy) and HR<30 bpm or >200 bpm were removed.

Others: skin temperature, step count, and accelerometer magnitude were validated by timestamp only. Survey responses were joined only when valid timestamps were available across data types.

From 5,699 *post-call* survey-anchored samples, 15.02% (n=856) were excluded, leaving 4,843 valid samples (mean=254.89 per person, SD=30.76). Most exclusions were due to missing valid timestamps in at least one

modality (e.g., loose wristbands yielding only invalid data); when multiple surveys occurred between two calls, we retained only the first.

Windowing. *Post-call* surveys were prompted immediately after call completion. Between call end and the next call start, the gap had a median of 5.6 min (Q1=3.7, Q3=8.3; mean=8.9, SD=12.8; min=17s, max=133.8min), indicating typically short intervals with occasional long delays (e.g., lunch breaks). Between consecutive *post-call* surveys, the gap had a median of 15.7 min (Q1=8.4, Q3=38.4; mean=31.1, SD=38.2; min=0 s, max=325.2 min), with occasional long gaps due to the daily target (fifteen *post-call* surveys per shift).

After cleansing, we computed features in windows anchored to each *post-call* timestamp t_s . By default, we used a 5-minute window ($[t_s - 5\text{min}, t_s]$) to capture the immediate post-call period, consistent with short-term HRV feature extraction and operationalizing early post-stressor recovery [115]. We used fixed-length windows rather than full call duration because call lengths were highly variable (median=105s; Q1=64s, Q3=172s; mean=136.0s, SD=111.6s; min=1s, max=28.05min), reducing comparability across calls. All modalities used the 5-minute window except step count, which used 60 minutes because 5-minute windows yielded near-zero values under back-to-back calls while a longer window provided meaningful variation.

Given typically short between-call intervals, this window is a practical approximation of post-call state in our setting. In a small number of cases with long gaps, it may include break-related activity, so we interpret windowed features as a practical approximation of the post-call state in an in-the-wild workplace setting, rather than a precise measure of recovery timing for every instance.

Feature extraction. We aggregated sensor features within each window using the mean (and accelerometer standard deviation) and aligned survey responses by timestamp.

For analysis, we constructed sessions anchored at each *post-call survey*, containing the pre-intervention state (survey and concurrent sensor features), the recommended/performed intervention (if any), and the *post-intervention survey*. We report the gap between the preceding call and intervention *completion* (median=2.6min; Q1=1.8, Q3=3.9; mean=3.2, SD=2.6; min=23s, max=38.4min). The long tail reflects rare postponements in real-world operations, such as workers taking a brief rest after emotionally taxing calls or delaying completion during high workload and tightly packed call schedules. Nonetheless, the majority of sessions were completed promptly, suggesting that delayed cases are outliers rather than the dominant between-call pattern. Notably, this provides an *upper bound* on the true call-to-intervention-start gap, since the intervention necessarily started earlier than its completion time. Accordingly, pre-post physiological changes around interventions may partially reflect ongoing post-call recovery dynamics in addition to intervention engagement, and we interpret them as in-the-wild approximations rather than precise causal estimates.

3.3.3 Data Reflection and Post-interview. In the post-interview session, participants engaged in data reflection tasks by exploring their data via the dashboard for 10-15 minutes. Then they participated in a semi-structured interview about their observations. We structured this session as a *reflection-on-action* activity commonly used in personal informatics to support trace-grounded sensemaking, helping participants articulate interpretations and reconcile prior expectations with observed patterns [65, 75, 97]. For each mode (state and change), we assigned two reflection tasks.

Data exploration task in state mode. Participants inspected calendar views to obtain an overview of weekly/weekday variation in stress states. For a chosen date, they identified *aligned or misaligned intervals* between perceived and physiological stress indicators on the timeline and jointly reviewed *situation, environment, stressor, and environment* data at those moments to derive *candidate factors* related to each indicator. During the interview, participants reported newly recognized patterns and the data points referenced to interpret mismatches (e.g., felt experience, call type/difficulty, environmental conditions), including factors judged most associated with each indicator.

Data exploration task in change mode. Participants used the calendar to observe weekly/weekday variation in stress reduction. For a selected date, participants inspected whether perceived or physiological stress indicator *increased, decreased, or no-changed* after an intervention and explored which *intervention types* appeared effective using *data-driven rankings and patterns*. In the interview, they described their criteria for judging an intervention’s effectiveness (e.g., duration of the effect and (dis)agreement of pre- and post-intervention changes across indicators), the intervention they felt was most effective, its (mis)alignment with data-derived rankings, and any noticed bodily responses.

3.3.4 Analysis. We excluded P15 and P16 from all analyses due to survey response quality issues (survey straight-lining); results therefore report $N=17$ (P01–P14, P17–P19).

We conducted an inductive thematic analysis [11] of interview data to understand participants’ perceptions of their stress data. Korean speech-to-text service¹ transcribed the recorded interviews. Three authors independently open-coded the first nine interviews, then discussed and consolidated themes into an initial codebook. Using this initial codebook, we coded the remaining interviews and finalized the codebook through team discussion.

For quantitative analysis, we computed within-person Spearman correlations (ρ) between each stress indicator (perceived, physiological) and contextual features spanning *stressor, environment, situation*, and *pre-shift* categories, controlling false discovery rate with Benjamini–Hochberg (BH–FDR) at $q < .05$. We also correlated (Spearman) satisfaction with perceived effectiveness, and satisfaction with realized reductions in perceived and physiological stress indicators across completed intervention sessions within person (BH–FDR per participant, $q < .05$). To assess correspondence between perceived and physiological reductions, we computed within-person Spearman correlations between the two reduction measures across sessions (BH–FDR, $q < .05$).

3.4 Implementation

We developed an end-to-end system for sensing, data transport, intervention delivery, and post-hoc visualization. Using the Samsung Health Sensor SDK [28], Galaxy Watch 7 continuously sampled onboard sensors and buffered data in 10-minute batches, transferred to a paired phone via Bluetooth Low Energy (BLE). HR and IBI were obtained as device-provided outputs via the SDK (derived on-device from PPG), rather than recomputed from raw PPG, following prior work using device-provided HR/IBI [19, 67, 157]. The phone time-stamped wearable data, collected step counts through the Samsung Health Data SDK [27]², and received environmental readings from bluSensor. A tablet app (TypeScript, React Native) delivered interventions and surveys, uploading data independently. Both devices used data-enabled USIMs; 10-minute batches were uploaded with a retry-enabled queue for reliability. All streams (wearable, mobile, environmental, intervention, and survey data) were stored in Supabase (PostgreSQL) with linked participant and device IDs.

For visualization-assisted interviews, we implemented a web dashboard in TypeScript and React.js. To support reproducibility, we release the source code of the dashboard and visualization pipeline (excluding the sensing/data-collection) with a subset of de-identified processed data and input-schema documentation. The repository is available at: <https://github.com/Jeonghyun109/stress-data-dashboard>.

4 RESULTS

In this section, we present our findings from the interviews, highlighting the sensemaking patterns participants interpret in an emotional labor workplace setting. We organize results to answer our research questions. We first explore how participants interpret perceived and physiological stress indicators with contextual data (RQ1). We

¹<https://clovanote.naver.com>

²We retrieved step counts on the phone via the Samsung Health Data SDK [27] because Samsung Health aggregates step data in the phone’s Samsung Health data store, providing a single consistent step stream for analysis.

then examine how participants evaluate micro-interventions and make sense of the subsequent, often divergent, stress changes (RQ2).

4.1 Understanding of Perceived and Physiological Stress Indicators with Contextual Data (RQ1)

4.1.1 Initial Perception of Perceived and Physiological Stress Indicators. Before dashboard exploration, we asked how participants differentiated perceived and physiological stress indicators. Participants described both under stress but experienced in distinct dimensions. Thirteen participants described perceived stress as a subjectively perceived or mentally experienced form of stress: “*Perceived stress is just the stress I feel in my mind, something I sense in my head.*” (P09) Fifteen participants understood physiological stress indicator as bodily manifestations of internal physiological reactions, citing symptoms such as headaches (n=4), increased heart rate (n=4), muscle tension (n=2), chest tightness (n=2), indigestion (n=1), warmth/heat sensation (n=1), and shortness of breath (n=1). For example, P01 stated, “*When my body reacts, like grabbing the back of my neck or feeling heat rise, that feels like physiological stress.*” However, most were unaware that physiological stress indicator could be measured through biosignals (e.g., pulse/heart rate) despite the written explanation of indicators on the user interface; only three mentioned such signals. As P04 noted, “*When people get nervous, their heart beats faster. I think it’s checked through heart rate, just like when your heart races not only from stress but also when someone you like is in front of you.*”

In summary, participants framed perceived stress as subjective emotion and physiological stress indicator as bodily sensation, but largely lacked awareness of stress measurement via physiological signals.

4.1.2 Contextualizing Perceived and Physiological Stress Indicators. Interestingly, participants’ interpretations of the data did not align with the aforementioned correlation analysis or the correlation patterns in the dashboard. Instead, they reconstructed their understanding of data based on personal experiences and intuitions, articulating subjective interpretations of their stress. While exploring the treemap on the dashboard, participants identified factors they believed were most closely related to perceived and physiological stress indicators. Eleven participants primarily pointed to self-reported stressor items such as *communication issues with customers and rude customers* as major causes of perceived stress. P10 explained, “*When communication with customers doesn’t go well, it’s exhausting, and since I have to maintain politeness, I end up suppressing my emotions, which itself becomes stressful.*” Similarly, three participants cited *complaint calls* and *workload* as recurring challenges that demanded emotional regulation, thereby amplifying perceived stress. P13 remarked, “*Complaint calls cause the most stress... those are the hardest because I have to start by hiding my emotions.*”

Regarding HRV-based indicator of physiological stress, participants mentioned similar stressors. Eleven participants identified *communication issues with customers, rude customers, and time pressure* as key triggers, explaining that they elicited immediate bodily reactions. P07 said, “*When I hear profanity during a call, my heart races and my hands tremble.*” Compared to the cases of perceived stress, interestingly, five participants pointed to *fatigue* as a core factor contributing to physiological stress indicator, attributing heightened stress levels to accumulated exhaustion. P11 stated, “*Since I’m at the menopausal age, I often can’t sleep well, so when I come to work tired, that situation likely increased my physiological stress.*” Notably, these interpretations did not align with the actual correlation analysis results (Figure 11).

Overall, these participants reinterpreted the data through the lens of their lived experiences, identifying everyday fatigue, emotional regulation efforts, and customer interactions as primary contributors to physiological stress indicator. Most participants did not mention environmental data when interpreting their perceived or physiological stress indicators, largely due to their limited understanding of such indicators. As P02 noted, “*I’m not really sure about the environmental data, and it’s hard to tell how much it actually reflects stress.*” However, some participants began to recognize potential links between physiological stress indicator and environmental

factors: “*Since physiological stress reflects changes in my body, I think stressors and environmental factors might have had a big influence. Looking back now, the temperature data really stands out to me.*” (P04).

In summary, rather than passively accepting data-driven results, participants sought to make sense of their stress by reconciling the data with their personal experiences, intuitions, and levels of understanding, thus constructing their own interpretations of perceived and physiological stress indicators.

4.1.3 Interpreting Discrepancies between Perceived and Physiological Stress Indicators. Participants compared perceived and physiological stress indicators side by side on the dashboard to make sense of their stress experiences. Eight participants tended to trust objectively measured indicators, such as physiological stress indicator, even when the data differed from their subjective judgment. P02 noted, “*Perceived stress reflects my subjective judgment, while physiological stress reflects bodily responses with subjective elements. So I think physiological stress is accurate.*” In contrast, four participants interpreted the discrepancy based on situational factors, such as P07 explained, “*When I had COVID-19, I was coughing so much that I couldn’t handle calls properly. The calls themselves were easy, but since my body was exhausted, I think my physiological stress appeared higher.*” Two participants attributed their higher perceived stress compared to physiological stress indicator to *emotional suppression and accumulated tension*. P06 described, “*As I keep doing work, stress builds up, and when emotions are suppressed like that, I get upset over even small things. So my perceived stress feels much higher.*”

After recognizing the discrepancy, participants differed in which indicator they considered a more accurate reflection of their stress. Similar to earlier findings, seven participants regarded physiological stress indicator as more trustworthy. Four of them believed that bodily sensations are more accurate than subjective feelings. P17 said, “*First of all, I feel that the physical stress my body sensed seems more accurate. It literally tracks my bodily reactions, and since those reactions are used to indicate stress, it even picks up moments when I wasn’t consciously aware that I was feeling stressed.*” The remaining three participants found the physiological stress indicator trends consistent with their daily stress experiences, thus deeming it more reliable: “*The [contextual] pattern of physiological stress was almost what I expected. When I usually feel stressed during work, I could see those moments reflected in slightly higher physiological stress values, which matched my expectations.*” (P09) Conversely, five participants felt that perceived stress better captured their actual condition since it was directly experienced and consciously assessed. P05 remarked, “*Perceived stress aligns with what I think, but physiological stress made me realize that it could be high even when I don’t feel it that way.*” Furthermore, several participants viewed the two indicators not as separate but as complementary. P10 explained, “*I feel stress when both the cognitive and bodily parts act together. It’s hard to think of them as independent.*” Similarly, P09 added, “*When perceived stress is high, I believe the body also reacts, so I don’t see them as unrelated.*”

Thus, participants did not perceive the discrepancy between perceived and physiological stress indicators as merely a measurement issue. Rather, they used these differences as opportunities to reflect on their cognitive and bodily responses, reinterpreting and making sense of their overall stress experiences through the comparison of both indicators, along with contextual factors.

4.2 Preference of Intervention and Understanding of Post-intervention Changes (RQ2)

4.2.1 Intervention Preference and Effectiveness. We first examined participants’ preferences and perceived effectiveness for the eight stress-relief interventions. Participants demonstrated diverse preferences for each intervention, and their reasons varied across individuals (see Table 3). The most preferred interventions were *Stretching*, *Protect Myself*, and *Breathing* (each n=14), followed by *Sugar Recharge* (n=13), *Anger Fairy* (n=10), *I Did Well, Right?* (n=9), *Words I Want to Hear Now* (n=9), and *Is It Because of Me Now?* (n=7).

Figure 6 illustrates the distributions of satisfaction, perceived effectiveness, and actual reduction rates in both perceived stress and HRV-based physiological indicator for each intervention. Satisfaction and perceived effectiveness generally ranged from moderate to high levels (3-4 points), though variation existed across interventions. In

Table 3. Reasons for preferring and not preferring each interventions. *N* indicates the number of participants who preferred the intervention.

Intervention (<i>N</i>)	Preference reasons	Dislike reasons
Stretching (14)	Refreshing, Following instructions is helpful	Lack of time, Lack of space, Feeling self-conscious (noticed by others)
Protect Me (14)	Relieve anger, Stress releases through fidgeting (like a game), Refreshing	Cannot relate to the idea of a character protecting me
Deep Breathing (14)	Refreshing, Calms the mind, Allows time to prepare for the next call	Lack of time, Difficult to follow instructions as directed
Sugar Boost (13)	Eating snacks relieves stress	Health issue or is not preferred, Effect is not immediate
Anger-Eating Fairy (10)	Emotional dumping ground, Can calmly reflect on and organize the stressful situation	Makes one reflect on stress, Cannot think up a word
I Did Well, Right? (9)	Can reflect on one's own feelings, Self-reassurance, Receives comfort	Don't want to recall stressful situation, Cannot think up a word as time passes, Lack of time
Words I Need Right Now (9)	Can reflect on one's own feelings, Enjoy the feeling of being talked to	Vague feeling, Cannot think up a word to say as time passes
Is It Because of Me? (7)	Separate emotion from work which helps to hold it in	Cannot relate as the content does not apply to their situation

contrast, actual stress reduction rates showed substantial individual differences for both stress types, and in some cases, stress even increased (negative values). The median reduction rate of perceived stress was approximately 10–20% across most interventions, with large interindividual variability. The physiological feature (RMSSD) exhibited even greater variability, ranging from -50% to +100%.

To further explore the relationships among intervention satisfaction, perceived effectiveness, and actual reduction in perceived and physiological stress indicators, we conducted correlation analyses (Table 4). Across all participants, satisfaction and perceived effectiveness showed a strong positive correlation, indicating that participants consistently regarded satisfying interventions as effective. However, correlations between satisfaction and actual perceived stress reduction varied among individuals. For instance, P09 exhibited a strong positive correlation ($r=0.937$, $q<.01$), suggesting that satisfying interventions effectively reduced perceived stress for this participant, whereas mostly others showed weak to moderate positive correlations that were not statistically significant. Similarly, correlations between satisfaction and changes in the physiological indicator varied substantially across participants and were not significant.

The discrepancy between perceived effectiveness and actual stress reduction was also reflected in interviews. P07 noted, “*I thought breathing would help, but it actually didn't, and on the other hand, interventions that required me to write something felt stressful but turned out fine in the data.*” Fourteen participants mentioned that when intervention outcomes differed from their expectations, they were prompted to reassess their prior beliefs or judgments. For example, P08 recalled, “*I was surprised that a solution I thought was good increased my stress. Seeing the data made me realize it might have been more about the call itself than the intervention.*” Some participants also discovered previously overlooked effects through the data, particularly noticing benefits from interventions they had initially perceived as less effective. P12 reflected, “*I didn't feel much of a difference, but my physiological stress actually decreased. It seems like 'Words I Want to Hear Now' was more effective than I thought.*”

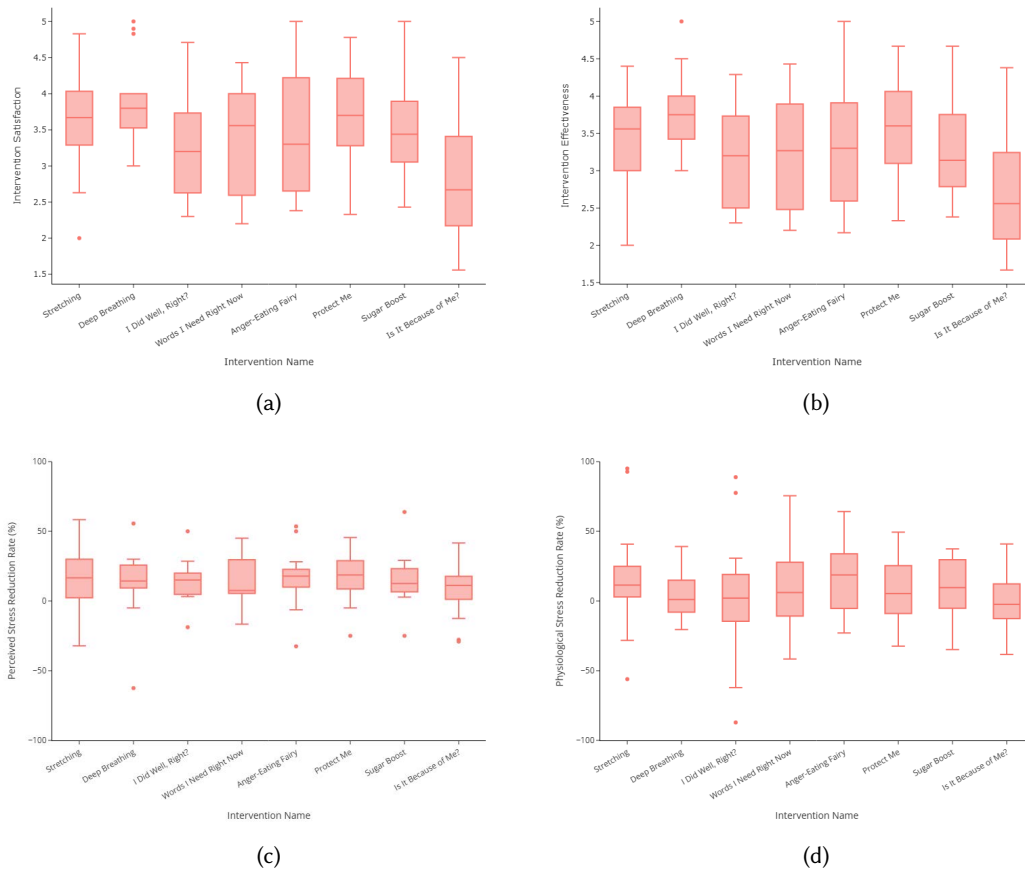


Fig. 6. Distribution of intervention evaluations and outcomes. (a) intervention satisfaction (5-point Likert scale), (b) perceived effectiveness (5-point Likert scale), (c) the reduction rate of a perceived stress indicator (%), and (d) the reduction rate of a physiological stress indicator (%). For the stress reduction rate, positive values indicate a decrease in stress, and negative values indicate an increase.

4.2.2 Interpreting Discrepancies between Changes in Perceived and Physiological Stress Indicators. In addition to the mismatch between satisfaction and intervention effects, no consistent relationship was observed between changes in perceived and physiological stress indicator. Participant-wise correlation analyses revealed no statistically significant associations between the two stress reduction rates for any participant (see Table 5). This finding suggests that the same intervention may influence perceived stress and physiological stress indicator in distinct ways.

Figure 7 compares the reduction rates of perceived stress and physiological stress indicator for four participants (P04, P09, P14, P18) across interventions. For P04, perceived stress decreased by approximately 20% after engaging with *I Did Well, Right?*, while the physiological stress indicator increased by 90%. P09 exhibited generally similar changes across both stress types, but in *Protect Myself*, perceived stress decreased while a physiological stress indicator slightly increased. P14 showed pronounced discrepancies, with perceived stress increasing in *Is It*

Table 4. Per-participant Spearman’s correlation between intervention satisfaction and other measures related to interventions. Asterisks indicate significance levels after FDR correction for each metric: * $q < 0.05$, ** $q < 0.01$, *** $q < 0.001$.

PID	Correlation		
	Perceived effectiveness	Perceived reduction	Physiological reduction
P01	0.945***	0.525	0.195
P02	1***	0.279	-0.031
P03	0.989***	-0.03	0.028
P04	0.811*	0.611	-0.028
P05	1***	0.554	0.515
P06	1***	0.676	0.399
P07	0.991***	0.361	0.299
P08	0.989***	0.59	0.576
P09	0.977***	0.937**	0.856
P10	0.988***	0.476	0.648
P11	0.975***	0.494	0.101
P12	0.968***	0.668	-0.522
P13	0.86**	0.357	0.414
P14	0.787*	0.375	-0.233
P17	0.969***	0.874	0.745
P18	0.957***	0.687	0.153
P19	0.941***	0.463	0.034

Table 5. Per-participant Spearman’s correlation between the reduction of perceived and physiological stress indicators.

PID	Correlation	PID	Correlation	PID	Correlation
P01	0.038	P07	0.124	P13	0.058
P02	-0.092	P08	-0.111	P14	-0.148
P03	0.058	P09	0.152	P17	0.301
P04	-0.194	P10	0.081	P18	0.112
P05	0.066	P11	0.053	P19	0.084
P06	0.095	P12	-0.025		

Because of Me Now? and *Stretching*, whereas the physiological stress indicator either decreased or increased in the opposite direction. P18’s reduction rates of the physiological stress indicator were generally higher than perceived stress reductions, yet in *Sugar Recharge*, the physiological stress indicator increased while perceived stress decreased. These examples illustrate that even within an individual, perceived stress and physiological stress indicator can respond differently depending on the intervention.

After using the dashboard, a majority of participants ($n=12$) reported newly recognizing that a decrease in perceived stress does not necessarily accompany a decrease in a physiological stress indicator. Participants did not regard this inconsistency as a mere measurement error but instead sought to interpret it in relation to their personal context, habits, or bodily characteristics, rather than assuming that the interventions directly modulated their physiological signals. Five participants attributed the discrepancy to their lifestyle or behavioral factors.

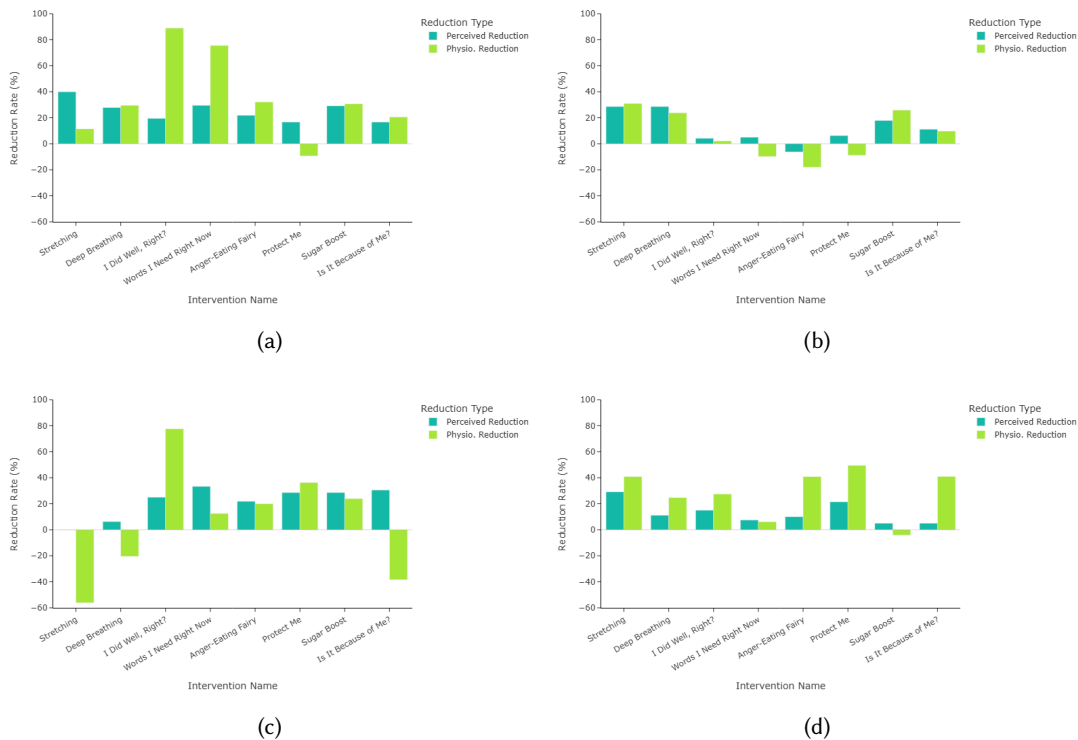


Fig. 7. Per-intervention stress-reduction rates (%) for four participants: (a) P04, (b) P09, (c) P14, and (d) P18. For each intervention, paired bars show the reduction of perceived and physiological stress indicators relative to the pre-intervention baseline; positive values indicate decreases in a stress indicator and negative values indicate increases after the intervention. *Perceived Reduction* indicates the perceived stress reduction rate and *Physio. Reduction* indicates the reduction rate of a physiological stress indicator.

For instance, P03, who experienced a decrease in perceived stress but an increase in the physiological stress indicator, linked the pattern to physical activity: “*The perceived stress reduction from stretching was the highest, around 20.8%, but here [physiological stress] it shows -28.2%, completely opposite results. Maybe it’s related to exercise intensity or heart rate; I stretched quite actively.*” Three participants focused on the temporal gap between cognitive and physiological recovery. P04 explained, “*The data show perceived stress going down but physiological stress going up. Maybe my mood improved while stretching, but the stress embedded in my body, like dizziness, headaches from complaint calls, or chest tightness, didn’t subside immediately.*” Likewise, P05 commented that physiological recovery might not be immediate under prolonged stress: “*I think breathing or stretching helped me mentally, but when stress has been accumulating for a while, physiological stress levels don’t go down right away.*”

Therefore, participants did not interpret discrepancies between changes in perceived stress and physiological stress indicators as errors. Rather, they contextualized these differences within their behaviors, conditions, and recovery pace, constructing personalized explanations that integrated both perceived and bodily dimensions of stress experience.

5 DISCUSSION

Our study explored how emotional laborers who frequently experience stress [138] understand and interpret stress indicators, contextual data, and stress changes following interventions. Through four weeks of data collection and in-depth interviews, we analyzed how participants perceived and made sense of these data. Participants interpreted moments of alignment and misalignment between the two indicators, as well as post-intervention changes, by drawing on their personal experiences and contextual environment, with related data. These findings distinguish our work from prior studies that have primarily focused on improving stress-modeling [42, 118] or evaluating intervention effectiveness [16, 22]. We foreground how people actually understand and interpret complex stress and intervention data in everyday contexts. In the following, we discuss these findings in depth and present several implications for future sensing-based stress management systems.

5.1 Supporting Stress Data Literacy in Call Center Context

In our study, despite providing a dashboard with explanatory information, some participants had difficulties in developing clear mental models for interpreting their stress-related data. For example, they tended to interpret physiological stress indicator primarily in terms of subjective sensations such as ‘fatigue’, rather than in relation to HRV signals and underlying physiological processes. Even with basic explanations of multiple indicators, participants often struggled to relate these data to their concrete stress experiences and work contexts, reflecting limitations in their overall stress data literacy.

Importantly, data literacy is not acquired through one-time information delivery, but is gradually developed through an iterative process of connecting their data with everyday experience and reflecting on it over time [150]. For non-expert users who are unfamiliar with data, this process often requires continuous interpretive scaffolding to support the connection between diverse data sources and their everyday work and personal experiences [63]. However, in high-intensity and emotionally demanding environments such as call centers [138], participants have limited opportunities to engage in sustained reflection. Under these conditions, conventional reflection-support approaches that assume prolonged interaction are unlikely to align with users’ actual work practices.

In response to these constraints, supporting reflection-in-action [122] through brief, in-situ reflective practices embedded in everyday work may offer a more practical alternative [7]. For example, systems can leverage AI-based guidance to provide lightweight and situationally relevant support. Recent studies have proposed approaches that use large language models to scaffold users’ interpretation of sensor data through structured explanations and contextualized feedback [19, 86, 154]. In call center contexts, such guidance is better implemented through short, structured prompts delivered at low-burden moments, such as between calls or after work, rather than through open-ended conversations [25, 100]. By encouraging users to relate their work experiences to stress-related signals, these reflection-in-action practices can gradually strengthen users’ interpretive capabilities, even under constrained time and cognitive capacity.

5.2 Understanding Discrepancy between Perceived and Physiological Stress Indicators

Participants experienced situations in which perceived stress and physiological stress indicators showed discrepancy, or in which the two indicators exhibited different patterns of change following an intervention. Participants interpreted discrepancies in relation to their own bodily conditions, experiences, and work contexts. Some participants explained the mismatches by referring to factors such as emotional suppression, accumulated fatigue, work-related burdens, and lifestyle habits. Even when the same intervention affected the two indicators differently, participants tended to interpret these outcomes not as failures of the intervention, but as temporary relief, delayed recovery, or individual differences. Through this process, they came to use both indicators as complementary sources of information for understanding their cognitive and bodily states. These findings suggest

that discrepancies between perceived and physiological stress indicators are difficult to explain through a single cause or definitive interpretation.

These discrepancies can be interpreted in light of prior studies reporting differences between stress reactivity and recovery [41, 81]. Previous studies have shown that physiological responses to stressors tend to emerge relatively quickly, whereas recovery may take longer and exhibit individual differences [6]. As a result, physiological indicators measured at a specific point in time may not fully reflect individuals' psychological states due to variations in recovery speed [81]. Conversely, individuals may cognitively perceive themselves as stressed while exhibiting relatively weak or indistinct bodily responses.

This temporal mismatch is particularly relevant in the context of micro-interventions, which, while brief in duration, are often sufficient to elicit partial physiological recovery. In our study, the HRV-based physiological stress indicator after intervention may capture early recovery processes rather than complete recovery since we use short-term HRV features over a 5-minute window anchored to each post-intervention. Within this time frame, physiological signals can reasonably reflect initial down-regulation following a stressful call, yet may not fully capture longer-term recovery trajectories that unfold over extended periods. We posit that the degree and timing of such recovery could vary substantially across individuals, reflecting differences in baseline autonomic regulation, accumulated fatigue, and ongoing work demands. Moreover, participants were embedded in work contexts characterized by repeated and closely spaced stressors, such that physiological regulation often unfolded within overlapping cycles of stress and recovery rather than discrete, isolated episodes. Under these conditions, physiological measures computed in short post-call or post-intervention windows may capture a mixture of partial recovery and residual activation from prior calls, even when individuals subjectively perceive relief or improvement following an intervention.

Taken together, discrepancies between perceived and physiological stress indicators in our findings should not be interpreted as failures of measurement or intervention. Rather, they reflect the temporal, individual, and contextual complexity of stress regulation in everyday work settings, thus underscoring the value of supporting within-person sensemaking through multiple, complementary indicators rather than relying on a single, absolute measure of stress.

5.3 Misalignment between Measure-Derived Efficacy and User-Perceived Value

One of our findings is the misalignment between the measure-derived efficacy of interventions (changes in perceived and physiological stress indicators) and user-perceived value (e.g., satisfaction and perceived effectiveness). For instance, in the case of the *breathing* intervention, despite minimal reduction in the physiological stress indicator in the data, one participant (P07) stated that the ability to forget the frustrating emotion and focus on oneself was the greatest reward. This highlights a tension in stress management systems; optimizing measured efficacy risks low acceptance when users are engaged with non-preferred options, while prioritizing preference may favor enjoyable interventions with limited measurable impact.

To address this dilemma, we can utilize the concept of *value-based intervention*, which emphasizes aligning system goals with users' intrinsic priorities and tolerability. In healthcare, value-based approaches respect patients' lived experiences rather than focusing solely on measurable outcomes [139]. Adapting this to our context, we define *user-perceived value* as the benefits users prioritize and are willing to act on in the moment (e.g., emotional relief, satisfaction, and tolerability). Accordingly, systems should not seek a single optimal intervention, but instead frame intervention recommendation as a *multi-objective decision* problem that considers measure-derived efficacy, user-perceived value, and feasibility constraints together [56, 113]. For example, a system can present a small set of Pareto-efficient options and briefly explain their trade-offs, thereby reducing users' reliance on satisfaction-only heuristics when making choices. This approach is consistent with our findings showing that workplace conditions, such as time pressure and limited privacy, decisively shape what users can tolerate and

value in the moment. In this context, feasibility should be treated as an important design consideration [98], motivating the need for low-burden micro-interventions that can be completed within brief and unpredictable gaps between tasks [3, 8].

While value-sensitive design [37, 38] is important at the intervention stage, it is equally critical in *reflection* to reduce the misalignment over time. We propose *value-based reflection* as a learning loop that continuously links and adjusts user-perceived value with measure-derived efficacy. Specifically, the system can (i) summarize which intervention users selected under a given context and what they considered valuable in that moment, (ii) present these choices alongside subsequent trajectories of perceived and physiological stress indicators, and (iii) surface recurring mismatch patterns (e.g., “high relief but limited recovery”). In this process, measure-derived efficacy functions not as a criterion to evaluate or invalidate users’ choices, but as a *calibration signal* that incrementally adjusts future recommendations. Through this reflection-driven calibration loop, systems can respect what users value in the moment while progressively personalizing intervention recommendations toward options that better align with longer-term stress recovery.

5.4 Limitations and Future Work

Our study has several limitations. First, although HRV has been widely used as an indicator for estimating physiological stress in prior stress and health research [46, 57, 106, 119, 158], there are limitations in interpreting HRV as a direct representation of participants’ physiological stress. In in-the-wild settings, HRV can be influenced by various confounding factors such as physical activity, posture, respiration, caffeine, illness, and sensor artifacts, which may prevent it from fully capturing physiological stress [47, 130]. Although we applied careful data cleansing procedures to reduce the impact of such confounders, physiological stress can still be influenced by other factors. For example, activity-related signals such as subtle wrist-motion changes (e.g., fidgeting) may co-occur with stress for some individuals [26]. Prior work suggests that stress can manifest through subtle behaviors in workplace settings (e.g., keystroke dynamics) [32, 82, 89]. However, we could not disentangle stress-related behaviors from measurement confounds.

Building on this limitation, future research could incorporate additional signals, such as physical responses (e.g., headache, bodily tension, and fidgeting), as well as other physiological or contextual data, to better reflect physiological stress. In addition, the interface should avoid presenting physiological indicators in overly definitive ways and instead adopt representations that make uncertainty explicit. For example, labeling such values as “estimates” or “reference indicator” may help users treat them as interpretive cues rather than as ground truths.

Second, data-driven reflection was conducted after the four-week data collection period, which may have made it difficult for participants to accurately recall their stress experiences and the underlying causes at the time. This corresponds to *reflection-on-action*, which refers to retrospective reflection that takes place after an activity has occurred [122]. When the timing of reflection is temporally distant from the lived experience, as in our study, participants may face limitations in reconstructing and interpreting their stress experiences. Future work should therefore be designed to provide more frequent and timely opportunities for reflection in everyday life. For example, providing brief summary feedback at moments temporally close to the experience, such as immediately after the end of the workday or following a break, may help users connect their data with their lived experiences in context and engage in more situated reflection.

Another limitation is that the intervention set was refined after the pilot primarily based on usability and workflow fit. For example, the “Hydration Break” was removed because pilot participants found it less helpful or practical in the tightly constrained between-call context, potentially excluding interventions with physiological benefits. Future work should more clearly separate usability-driven refinement from efficacy evaluation.

Additionally, because interventions were executed between calls, pre-post physiological changes around interventions may be partially confounded by ongoing post-call recovery rather than reflecting intervention

effects alone. Future work should model recovery to decouple intervention effects from call-related physiological residuals.

Finally, this study focused on stress within the specific call center workplace setting and the sensemaking processes of workers within that context. However, the implications of sensemaking with physiological data extend to other high-stress domains beyond emotional labor. Future research should investigate whether similar patterns of discrepancy appear in professions characterized by extreme physiological stress, such as soldiers in training or delivery workers under time constraints. For instance, in physically demanding jobs (e.g., delivery, construction), physiological spikes might be misinterpreted as physical exertion rather than stress, presenting a different challenge for stress tracking systems compared to the sedentary setting of call centers. Extending this research to a broader set of target populations will provide a better understanding of how different categories of work stressors, such as emotional and physical demands, impact users' interpretation of bio-data.

6 CONCLUSION

This work investigated how emotional laborers make sense of dual stress indicators (perceived and physiological stress) in a call center setting. In a four-week in-the-wild deployment with a reflection dashboard and interviews, participants compared both indicators alongside contextual factors. Participants grounded their interpretations in lived experience, often referencing call context and fatigue. When indicators diverged, many treated mismatches as informative, with several trusting the physiological indicator over subjective judgment; however, limited literacy about biosignals and environmental measures constrained interpretation depth. Intervention satisfaction aligned with perceived effectiveness but it did not reliably predict measured stress reduction: reduction rates exhibited substantial individual variability, and perceived and physiological changes often diverged for the same intervention even within participants. Participants attributed divergence to behavior, lifestyle, and recovery lags rather than dismissing it as noise. From these findings, we suggest systems should visualize divergence and potential recovery lags, scaffold physiological data literacy so biosignals are interpreted as uncertain cues rather than definitive ground truth, and personalize interventions via multi-objective trade-offs among efficacy, value, and feasibility.

Acknowledgments

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2022-II220064) and by the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (RS-2022-NR068758)

References

- [1] Fatema Akbar. 2021. *Stress and Human-Computer Interaction at the Workplace: Unobtrusive Tracking With Wearable Sensors and Computer Logs*. Ph. D. Dissertation. University of California, Irvine.
- [2] Surely Akiri, Vasundhara Misal, Sanaz Taherzadeh, J Lee Jenkins, Gary Williams, Helena Mentis, and Andrea Kleinsmith. 2024. Enhancing stress understanding through team reflection: technology-driven insights in high-stress training scenarios. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–18.
- [3] Patricia Albulescu, Irina Macsinga, Andrei Rusu, Coralia Sulea, Alexandra Bodnaru, and Bogdan Tudor Tulbure. 2022. "Give me a break!" A systematic review and meta-analysis on the efficacy of micro-breaks for increasing well-being and performance. *PLoS one* 17, 8 (2022), e0272460.
- [4] Jana Bahrova, Martin Augustynek, Tereza Hrnčířová, Eliska Szalbotova, Lukas Tomaszek, Martin Malcik, and Jan Kubicek. 2025. Unsupervised profiling of meditation-induced autonomic responses using electrodermal and heart rate variability features. *Results in Engineering* 27 (2025), 106481.
- [5] Arnold B. Bakker and Evangelia Demerouti. 2007. The Job Demands-Resources model: state of the art. *Journal of Managerial Psychology* 22, 3 (04 2007), 309–328.
- [6] Melanie Bamert and Jennifer Inauen. 2022. Physiological stress reactivity and recovery: Some laboratory results transfer to daily life. *Frontiers in Psychology* 13 (2022), 943065.

- [7] Matthew Barker-Canler, Daniel Gooch, Janet Van Der Linden, and Marian Petre. 2024. Flexible minimalist self-tracking to support individual reflection. *ACM Transactions on Computer-Human Interaction* 31, 3 (2024), 1–35.
- [8] Amit Baumele, Theresa Fleming, and Stephen M Schueller. 2020. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. *Journal of medical Internet research* 22, 10 (2020), e20631.
- [9] Sara Becker, Birgit Spinath, Beate Ditzen, and Tobias Dörfler. 2023. Psychological stress= Physiological stress? An experimental study with prospective teachers. *Journal of Psychophysiology* 37, 1 (2023), 12.
- [10] Vicki Belt, Ranald Richardson, and Juliet Webster. 2002. Women, social skill and interactive service work in telephone call centres. *New technology, work and employment* 17, 1 (2002), 20–34.
- [11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [12] Amelie Edith Brinkmann, Sophia Antonia Press, Eduard Helmert, Martin Hautzinger, Inna Khazan, and Jan Vagedes. 2020. Comparing effectiveness of HRV-biofeedback and mindfulness for workplace stress reduction: A randomized controlled trial. *Applied psychophysiology and biofeedback* 45, 4 (2020), 307–322.
- [13] Céleste M Brotheridge and Raymond T Lee. 2003. Development and validation of the emotional labour scale. *Journal of occupational and Organizational Psychology* 76, 3 (2003), 365–379.
- [14] Daniel J Buysse, Charles F Reynolds III, Timothy H Monk, Susan R Berman, and David J Kupfer. 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research* 28, 2 (1989), 193–213.
- [15] Jana Campbell and Ulrike Ehlert. 2012. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology* 37, 8 (2012), 1111–1134.
- [16] Markus Canazei, Luis Glenzer, Siegmund Staggl, Verena Dresen, Johannes Weninger, and Elisabeth Margarete Weiss. 2025. Light-guided resonant breathing enhances psychophysiological stress recovery in a simulated office environment. *Scientific Reports* 15, 1 (2025), 40953.
- [17] Soohyun Cho. 2020. A preliminary validation study for the Korean version of the burnout assessment tool (K-BAT). *Korean Journal of Industrial and Organizational Psychology* 33, 4 (2020), 461–499.
- [18] Seong-Sik Cho, Hyunjoo Kim, JinWoo Lee, Sinye Lim, and Woo Chul Jeong. 2019. Combined exposure of emotional labor and job insecurity on depressive symptoms among female call-center workers: A cross-sectional study. *Medicine* 98, 12 (2019), e14894.
- [19] Akshat Choube, Ha Le, Jiachen Li, Kaixin Ji, Vedant Das Swain, and Varun Mishra. 2025. GLOSS: Group of LLMs for Open-Ended Sensemaking of Passive Sensing Data for Health and Wellbeing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–32.
- [20] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A Global Measure of Perceived Stress. *Journal of Health and Social Behavior* 24, 4 (1983), 385–396.
- [21] Vedant Das Swain, Qiuyue" Joy" Zhong, Jash Rajesh Parekh, Yechan Jeon, Roy Zimmermann, Mary P Czerwinski, Jina Suh, Varun Mishra, Koustuv Saha, and Javier Hernandez. 2025. AI on my shoulder: Supporting emotional labor in front-office roles with an llm-based empathetic coworker. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–29.
- [22] Marijke De Couck, Ralf Caers, Liza Musch, Johanna Fliegau, Antonio Giangreco, and Yori Gidron. 2019. How breathing can help you make better decisions: Two studies on the effects of breathing patterns on heart rate variability and decision-making in business cases. *International Journal of Psychophysiology* 139 (2019), 1–9.
- [23] Glen Debar, Nele De Witte, Romy Sels, Marc Mertens, Tom Van Daele, and Bert Bonroy. 2020. Making wearable technology available for mental healthcare through an online platform with stress detection algorithms: the Carewear project. *Journal of Sensors* 2020, 1 (2020), 8846077.
- [24] Geyze Diniz, Ligia Korkes, Luca Schiliró Tristão, Rosangela Pelegrini, Patrícia Lacerda Bellodi, and Wanderley Marques Bernardo. 2023. The effects of gratitude interventions: a systematic review and meta-analysis. *Einstein (São Paulo)* 21 (2023), eRW0371.
- [25] Poorvesh Dongre. 2024. Physiology-driven empathic large language models (EmLLMs) for mental health support. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.
- [26] Heather Elahi, Ha Min Son, Catrina A Calub, Nahid Nasiri, Daniel Shapiro, Katherine Isbister, Jared Borden, Paul D Hastings, Xin Liu, and Julie B Schweitzer. 2025. Impact of fidget devices on anxiety and physiological responses in adults with ADHD. *Research in Developmental Disabilities* 158 (2025), 104944.
- [27] Samsung Electronics. 2025. *Samsung Health Data SDK*. Samsung Electronics Co., Ltd. <https://developer.samsung.com/health/data/overview.html> Version 1.0.0.
- [28] Samsung Electronics. 2025. *Samsung Health Sensor SDK*. Samsung Electronics Co., Ltd. <https://developer.samsung.com/health/sensor/overview.html> Version 1.0.0.
- [29] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. 1996. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 5 (1996), 1043–1065.
- [30] Don Samitha Elvitigala, Philipp M Scholl, Hussel Suriyaarachchi, Vipula Dissanayake, and Suranga Nanayakkara. 2021. StressShoe: a DIY toolkit for just-in-time personalised stress interventions for office workers performing sedentary tasks. In *Proceedings of the 23rd international conference on mobile human-computer interaction*. 1–14.

- [31] Mamiko Enoki, Eri Maeda, Toyoto Iwata, and Katsuyuki Murata. 2017. The association between work-related stress and autonomic imbalance among call center employees in Japan. *The Tohoku Journal of Experimental Medicine* 243, 4 (2017), 321–328.
- [32] Clayton Epp, Michael Lippold, and Regan L. Mandryk. 2011. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 715–724.
- [33] Emre Ertin, Nathan Stohs, Santosh Kumar, Andrew Raij, Mustafa al'Absi, and Siddharth Shah. 2011. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems* (Seattle, Washington) (SenSys '11). Association for Computing Machinery, New York, NY, USA, 274–287.
- [34] Michael R Esco, Andrew D Fields, Matthew A Mohammadnabi, and Brian M Kliszczewicz. 2025. Monitoring Training Adaptation and Recovery Status in Athletes Using Heart Rate Variability via Mobile Devices: A Narrative Review. *Sensors* 26, 1 (2025), 3.
- [35] Stephen H Fairclough and Chelsea Dobbins. 2020. Personal informatics and negative emotions during commuter driving: Effects of data visualization on cardiovascular reactivity & mood. *International Journal of Human-Computer Studies* 144 (2020), 102499.
- [36] Tiina Föhr, Asko Tolvanen, Tero Myllymäki, Elina Järvelä-Reijonen, Sanni Rantala, Riitta Korpela, Katri Peuhkuri, Marjukka Kolehmainen, Sampsa Puttonen, Raimo Lappalainen, et al. 2015. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. *Journal of Occupational Medicine and Toxicology* 10, 1 (2015), 39.
- [37] Batya Friedman and David G Hendry. 2019. *Value sensitive design: Shaping technology with moral imagination*. MIT Press.
- [38] Batya Friedman, David G Hendry, Alan Borning, et al. 2017. A survey of value sensitive design methods. *Foundations and Trends® in Human-Computer Interaction* 11, 2 (2017), 63–125.
- [39] Éva Gál, Simona Ștefan, and Ioana A Cristea. 2021. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. *Journal of affective disorders* 279 (2021), 131–142.
- [40] Sarah N Garfinkel, Anil K Seth, Adam B Barrett, Keisuke Suzuki, and Hugo D Critchley. 2015. Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. *Biological psychology* 104 (2015), 65–74.
- [41] Sabine AE Geurts and Sabine Sonnentag. 2006. Recovery as an explanatory mechanism in the relation between acute stress reactions and chronic health impairment. *Scandinavian Journal of Work, Environment & Health* 32, 6 (2006), 482–492.
- [42] Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. 2016. Continuous stress detection using a wrist device: in laboratory and real life. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) (UbiComp '16). Association for Computing Machinery, New York, NY, USA, 1185–1193.
- [43] Nuno Gomes, Matilde Pato, Andre Ribeiro Lourenco, and Nuno Datia. 2023. A survey on wearable sensors for mental health monitoring. *Sensors* 23, 3 (2023), 1330.
- [44] Alicia A Grandey. 2000. Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of occupational health psychology* 5, 1 (2000), 95.
- [45] Yunjo Han, Hyemin Lee, Kobiljon E Toshnazarov, Youngtae Noh, and Uichin Lee. 2022. Stressbal: personalized just-in-time stress intervention with wearable and phone sensing. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*. 41–43.
- [46] Yeaminul Haque, Rahat Shahriar Zawad, Chowdhury Saleh Ahmed Rony, Hasan Al Banna, Tapotosh Ghosh, M Shamim Kaiser, and Mufti Mahmud. 2024. State-of-the-art of stress prediction from heart rate variability using artificial intelligence. *Cognitive Computation* 16, 2 (2024), 455–481.
- [47] Junichiro Hayano and Emi Yuda. 2019. Pitfalls of assessment of autonomic function by heart rate variability. *Journal of physiological anthropology* 38, 1 (2019), 3.
- [48] Elena Heber, Dirk Lehr, David Daniel Ebert, Matthias Berking, and Heleen Riper. 2016. Web-based and mobile stress management intervention for employees: a randomized controlled trial. *Journal of medical Internet research* 18, 1 (2016), e5112.
- [49] Javier Hernandez, Rob R. Morris, and Rosalind W. Picard. 2011. Call Center Stress Recognition with Person-Specific Models. In *Affective Computing and Intelligent Interaction*, Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 125–134.
- [50] Arlie Russell Hochschild. 2022. The managed heart. In *Working in America*. Routledge, 40–48.
- [51] Annie Høgh, Åse M Hansen, Eva G Mikkelsen, and Roger Persson. 2012. Exposure to negative acts at work, psychological stress reactions and physiological stress response. *Journal of psychosomatic research* 73, 1 (2012), 47–52.
- [52] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–16.
- [53] Hyunchan Hwang, Sun Mi Kim, Bo Netterstrøm, and Doug Hyun Han. 2022. The efficacy of a smartphone-based app on stress reduction: randomized controlled trial. *Journal of medical Internet research* 24, 2 (2022), e28703.

- [54] Mahmudul Islam, Sami Rashid, Lishan Rafid, Tasnuba Badrul, Ashraful Islam, and Beenish Moalla Chaudhry. 2024. Design and Evaluation of a Smartwatch-Based Physiological Signal-Driven Workplace Stress Management mHealth Tool for Bangladeshi Healthcare Professionals. In *2024 Advances in Science and Engineering Technology International Conferences (ASET)*. 1–10.
- [55] Katharine Ann James, Juliet Ilena Stromin, Nina Steenkamp, and Marc Irwin Combrinck. 2023. Understanding the relationships between physiological and psychosocial stress, cortisol and cognition. *Frontiers in endocrinology* 14 (2023), 1085950.
- [56] Dietmar Jannach and Himan Abdollahpouri. 2023. A survey on multi-objective recommender systems. *Frontiers in big Data* 6 (2023), 1157899.
- [57] Susanna Järvelin-Pasanen, Sanna Sinikallio, and Mika P Tarvainen. 2018. Heart rate variability and occupational stress—systematic review. *Industrial health* 56, 6 (2018), 500–511.
- [58] Kyung-Sook Jeong, Su-Jeong Choi, Myeong-Ok Park, and Yan Li. 2015. The effects of customer service representatives' emotional labor by emotional display rules on emotional dissonance, emotional exhaustion and turnover intention in the context of call centers. *Korean J Bus Adm* 28, 2 (2015), 529–51.
- [59] L Y Ji, Y J Yang, A G Li, S F Wang, and J K Wu. 2011. Robust time series processing for Heart Rate Variability analysis in daily life. In *2011 Computing in Cardiology*. 301–304.
- [60] Alvaro A. Jiménez-Ocaña, Andrés Pantoja, Mario Andrés Valderrama, and Luis Felipe Giraldo. 2023. A Systematic Review of Technology-Aided Stress Management Systems: Automatic Measurement, Detection and Control. *IEEE Access* 11 (2023), 116109–116126.
- [61] Eunkyung Jo, Austin L. Toombs, Colin M. Gray, and Hwajung Hong. 2020. Understanding Parenting Stress through Co-designed Self-Trackers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13.
- [62] Ellen Jopling, Katerina Rnic, Taylyn Jameson, Alison Tracy, and Joelle LeMoult. 2023. Discordance indices of stress sensitivity and trajectories of internalizing symptoms in adolescence. *Research on Child and Adolescent Psychopathology* 51, 10 (2023), 1521–1533.
- [63] Matthew Jörke, Yasaman S Sefidgar, Talie Massachi, Jina Suh, and Gonzalo Ramos. 2023. Pearl: A technology probe for machine-assisted reflection on personal data. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 902–918.
- [64] Gyuwon Jung and Uichin Lee. 2025. CounterStress: Enhancing Stress Coping Planning through Counterfactual Explanations in Personal Informatics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 714, 20 pages.
- [65] Gyuwon Jung, Sangjun Park, and Uichin Lee. 2024. DeepStress: Supporting Stressful Context Sensemaking in Personal Informatics Systems Using a Quasi-experimental Approach. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1000, 18 pages.
- [66] Robert A. Karasek. 1979. Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. *Administrative Science Quarterly* 24, 2 (1979), 285–308.
- [67] Kianoosh Kazemi, Iman Azimi, Pasi Liljeberg, and Amir M Rahmani. 2025. Respiration Rate Estimation via Smartwatch-based Photoplethysmography and Accelerometer Data: A Transfer Learning Approach. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 1 (2025), 1–24.
- [68] Justin J Kennedy and Miranda Pretorius. 2008. Integrating a portable biofeedback device into call centre environments to reduce employee stress: Results from two pilot studies. *Journal of Workplace Behavioral Health* 23, 3 (2008), 295–307.
- [69] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235.
- [70] Taewan Kim, Haesoo Kim, Ha Yeon Lee, Hwarang Goh, Shakhboz Abdigapurov, Mingon Jeong, Hyunsung Cho, Kyungsik Han, Youngtae Noh, Sung-Ju Lee, et al. 2022. Prediction for retrospection: Integrating algorithmic stress prediction into personal informatics systems for college students' mental health. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–20.
- [71] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–28.
- [72] Ulrich Kirk and Johanne L Axelsen. 2020. Heart rate variability is enhanced during mindfulness practice: A randomized controlled trial involving a 10-day online-based mindfulness intervention. *PloS one* 15, 12 (2020), e0243488.
- [73] Rafal Kocielnik and Natalia Sidorova. 2015. Personalized stress management: enabling stress monitoring with lifelogexplorer. *KI-Künstliche Intelligenz* 29, 2 (2015), 115–122.
- [74] Lydia Kogler, Veronika I Müller, Amy Chang, Simon B Eickhoff, Peter T Fox, Ruben C Gur, and Birgit Derntl. 2015. Psychosocial versus physiological stress—Meta-analyses on deactivations and activations of the neural correlates of stress reactions. *Neuroimage* 119 (2015), 235–251.
- [75] Youngji Koh, Chanhee Lee, Eunki Joung, Hyunsoo Lee, and Uichin Lee. 2025. Harnessing Home IoT for Self-tracking Emotional Wellbeing: Behavioral Patterns, Self-reflection, and Privacy Concerns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–36.

- [76] Agata Kolakowska, Wioleta Szwoch, and Mariusz Szwoch. 2020. A review of emotion recognition methods based on data acquired via smartphone sensors. *Sensors* 20, 21 (2020), 6367.
- [77] Jaap M Koolhaas, Alessandro Bartolomucci, Bauke Buwalda, Seitse F de Boer, Gabriele Flügge, S Mechiel Korte, Peter Meerlo, Robert Murison, Berend Olivier, Paola Palanza, et al. 2011. Stress revisited: a critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews* 35, 5 (2011), 1291–1301.
- [78] Saewon Kye, Junhyung Moon, Juneil Lee, Inho Choi, Dongmi Cheon, and Kyoungwoo Lee. 2017. Multimodal data collection framework for mental stress monitoring. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) (*UbiComp '17*). Association for Computing Machinery, New York, NY, USA, 822–829.
- [79] Richard S Lazarus. 1984. *Stress, appraisal, and coping*. Vol. 445. Springer.
- [80] Duri Lee, Kyungmin Nam, and Uichin Lee. 2024. Supporting Interpersonal Emotion Regulation of Call Center Workers via Customer Voice Modulation. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI EA '24*). Association for Computing Machinery, New York, NY, USA, Article 306, 8 pages.
- [81] Jinhak Lee, Ho Bin Hwang, Seungjae Lee, Jayon Kim, Jeyeon Lee, Sanghag Kim, Jung Hee Ha, Yoojin Jang, Sejin Hwang, Hoon-Ki Park, Jongshill Lee, and In Young Kim. 2024. Analysis of Acute Stress Reactivity and Recovery in Autonomic Nervous System Considering Individual Characteristics of Stress Using HRV and EDA. *IEEE Access* 12 (2024), 115400–115410.
- [82] Uichin Lee, Duri Lee, Eunji Park, and Heejeong Lim. 2025. 21: Artificial intelligence for emotion regulation at work. In *Handbook of Emotion Regulation at Work*. Edward Elgar Publishing, Cheltenham, UK, 409 – 435.
- [83] Aniek Lentferink, Matthijs L Noordzij, Anouk Burgler, Randy Klaassen, Youri Derks, Hilbrand Oldenhuis, Hugo Velthuisen, and Lisette van Gemert-Pijnen. 2022. On the receptivity of employees to just-in-time self-tracking and eCoaching for stress management: a mixed-methods approach. *Behaviour & information technology* 41, 7 (2022), 1398–1424.
- [84] Aniek Lentferink, Hilbrand Oldenhuis, Hugo Velthuisen, and Lisette van Gemert-Pijnen. 2023. How Reflective Automated e-Coaching Can Help Employees Improve Their Capacity for Resilience: Mixed Methods Study. *JMIR human factors* 10 (2023), e34331.
- [85] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies (*UbiComp '11*). Association for Computing Machinery, New York, NY, USA, 405–414.
- [86] Jiachen Li, Xiwen Li, Justin Steinberg, Akshat Choube, Bingsheng Yao, Xuhai Xu, Dakuo Wang, Elizabeth Mynatt, and Varun Mishra. 2025. Vital Insight: Assisting Experts' Context-Driven Sensemaking of Multi-modal Personal Tracking Data Using Visualization and Human-in-the-Loop LLM. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–37.
- [87] Kun Liang, Anfu Zhou, Zhan Zhang, Hao Zhou, Huadong Ma, and Chenshu Wu. 2023. mmStress: Distilling human stress from daily activities via contact-less millimeter-wave sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–36.
- [88] Zilu Liang, Daewon Hwang, Samantha Chen, Nhung Huyen Hoang, Kingkarn Khotchasing, and Edward F Melcer. 2025. User Preferences for Interaction Timing in Smartwatch Sleep Hygiene Games. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [89] Aicha Maalej and Ilhem Kallel. 2020. Does Keystroke Dynamics tell us about Emotions? A Systematic Literature Review and Dataset Construction. In *2020 16th International Conference on Intelligent Environments (IE)*. IEEE, 60–67.
- [90] Maria Marin-Farrona, Manuel Leon-Jimenez, Jorge Garcia-Unanue, Leonor Gallardo, Carmen Crespo-Ruiz, and Beatriz Crespo-Ruiz. 2020. Transtheoretical model is better predictor of physiological stress than perceived stress scale and work ability index among office workers. *International journal of environmental research and public health* 17, 12 (2020), 4410.
- [91] Iris B Mauss, Robert W Levenson, Loren McCarter, Frank H Wilhelm, and James J Gross. 2005. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* 5, 2 (2005), 175.
- [92] Varun Mishra, Tian Hao, Si Sun, Kimberly N Walter, Marion J Ball, Ching-Hua Chen, and Xinxin Zhu. 2018. Investigating the role of context in perceived stress detection in the wild. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 1708–1716.
- [93] Varun Mishra, Gunnar Pope, Sarah Lord, Stephanie Lewia, Byron Lowens, Kelly Caine, Sougata Sen, Ryan Halter, and David Kotz. 2020. Continuous detection of physiological stress with commodity hardware. *ACM transactions on computing for healthcare* 1, 2 (2020), 1–30.
- [94] Varun Mishra, Sougata Sen, Grace Chen, Tian Hao, Jeffrey Rogers, Ching-Hua Chen, and David Kotz. 2020. Evaluating the reproducibility of physiological stress detection models. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 4 (2020), 1–29.
- [95] Jesus Montero-Marin, Sonia Asun, Nerea Estrada-Marcen, Rosario Romero, and Roberto Asun. 2013. Effectiveness of a stretching program on anxiety levels of workers in a logistic platform: a randomized controlled study. *Atención primaria* 45, 7 (2013), 376–383.
- [96] Margaret E Morris, Qusai Kathawala, Todd K Leen, Ethan E Gorenstein, Farzin Guilak, Michael Labhard, and William Deleeuw. 2010. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *Journal of medical Internet research* 12, 2 (2010), e10.

- [97] Hugh Munby. 1989. Reflection-in-action and reflection-on-action. *Current issues in education* 9, 1 (1989), 31–42.
- [98] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2016. Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of behavioral medicine* (2016), 1–17.
- [99] Talha Nazir, Zahid Hussain, and Petr Panfilov. 2025. A Real-Time Multi-modal Machine Learning Framework for Tri-Level Stress Classification Using Physiological and Behavioral Data. Available at SSRN 5526498 (2025).
- [100] Sameer Neupane, Poorvesh Dongre, Denis Gracanin, and Santosh Kumar. 2025. Wearable Meets LLM for Stress Management: A Duoethnographic Study Integrating Wearable-Triggered Stressors and LLM Chatbots for Personalized Interventions. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 588, 8 pages.
- [101] Sameer Neupane, Mithun Saha, Nasir Ali, Timothy Hnat, Shahin Alan Samiei, Anandathirtha Nandugudi, David M. Almeida, and Santosh Kumar. 2024. Momentary Stressor Logging and Reflective Visualizations: Implications for Stress Management with Wearables. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 809, 19 pages.
- [102] Christina Orphanidou, Timothy Bonnici, Peter Charlton, David Clifton, David Vallance, and Lionel Tarassenko. 2014. Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE journal of biomedical and health informatics* 19, 3 (2014), 832–838.
- [103] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014. PopTherapy: coping with stress through pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (Oldenburg, Germany) (PervasiveHealth '14)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 109–117.
- [104] Eunji Park, Duri Lee, Yunjo Han, James Diefendorff, and Uichin Lee. 2024. Hide-and-seek: Detecting Workers' Emotional Workload in Emotional Labor Contexts Using Multimodal Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–28.
- [105] Sangjun Park, Dejiang Zheng, and Uichin Lee. 2025. A PPG Signal Dataset Collected in Semi-Naturalistic Settings Using Galaxy Watch. *Scientific Data* 12, 1 (2025), 892.
- [106] Jeremy E Peabody, Rebecca Ryznar, Markus T Ziesmann, Lawrence Gillman, Rebecca J Ryznar, and Lawrence M Gillman. 2023. A systematic review of heart rate variability as a measure of stress in medical professionals. *Cureus* 15, 1 (2023).
- [107] Stephen W Porges. 1995. Cardiac vagal tone: a physiological index of stress. *Neuroscience & Biobehavioral Reviews* 19, 2 (1995), 225–233.
- [108] Pooja Prajod, Bhargavi Mahesh, and Elisabeth André. 2024. Stressor Type Matters! — Exploring Factors Influencing Cross-Dataset Generalizability of Physiological Stress Detection. In *Proceedings of the 26th International Conference on Multimodal Interaction (San Jose, Costa Rica) (ICMI '24)*. Association for Computing Machinery, New York, NY, USA, 508–517.
- [109] Amon Rapp and Federica Cena. 2016. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies* 94 (2016), 1–17.
- [110] Xipei Ren, Bin Yu, Yuan Lu, Biyong Zhang, Jun Hu, and Aarnout Brombacher. 2019. LightSit: An unobtrusive health-promoting system for relaxation and fitness microbreaks at work. *Sensors* 19, 9 (2019), 2162.
- [111] Xipei Ren, Xiaoyu Zhang, Renyao Zou, Ran Yan, and Bin Yu. 2025. EmoVis: exploring data-enabled analogue journaling to promote self-reflection for mental wellness among college students. *Behaviour & Information Technology* 44, 4 (2025), 859–881.
- [112] Abdelmounaam Rezgui. 2025. On the Causality between Cognitive Stress and Physiological Stress: The Stroop Test as a Case Study. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP Adjunct '25)*. Association for Computing Machinery, New York, NY, USA, 495–502.
- [113] Marco Tulio Ribeiro, Nivio Ziviani, Edleno Silva De Moura, Itamar Hata, Anisio Lacerda, and Adriano Veloso. 2014. Multiobjective pareto-efficient approaches for recommender systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 4 (2014), 1–20.
- [114] James P Robson Jr and Meredith Troutman-Jordan. 2014. A concept analysis of cognitive reframing. *Journal of Theory Construction & Testing* 18, 2 (2014).
- [115] Charlotte M Roddick, Yeon Soo Seo, Shan-Li Barkovich, Luke Forrester, and Frances S Chen. 2025. Cardiac vagal recovery following acute psychological stress in human adults: A scoping review. *Neuroscience & Biobehavioral Reviews* 176 (2025), 106268.
- [116] Hadar Rosenbach, Alon Itzkovitch, Yori Gidron, and Tom Schonberg. 2025. Assessing Stress Level Scores Against Wearables-Driven Physiological Measurements. *Stress and Health* 41, 6 (2025), e70125.
- [117] Akane Sano, Paul Johns, and Mary Czerwinski. 2017. Designing opportune stress intervention delivery timing using multi-modal data. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 346–353.
- [118] Akane Sano and Rosalind W. Picard. 2013. Stress Recognition Using Wearable Sensors and Mobile Phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 671–676.
- [119] Zhanna Sarsenbayeva, Niels van Berkel, Danula Hettiachchi, Weiwei Jiang, Tilman Dingler, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2019. Measuring the effects of stress on mobile interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable*

- and *Ubiquitous Technologies* 3, 1 (2019), 1–18.
- [120] Regina Franziska Schmid, Joachim Thomas, and Katrin Rentzsch. 2024. Individual differences in parasympathetic nervous system reactivity in response to everyday stress are associated with momentary emotional exhaustion. *Scientific Reports* 14, 1 (2024), 26662.
- [121] Neil Schneiderman, Gail Ironson, and Scott D Siegel. 2005. Stress and health: psychological, behavioral, and biological determinants. *Annu. Rev. Clin. Psychol.* 1, 1 (2005), 607–628.
- [122] Donald A Schon and Vincent DeSanctis. 1986. The reflective practitioner: How professionals think in action.
- [123] André Schulz, Dana Schultchen, and Claus Vögele. 2020. Interoception, Stress, and Physical Symptoms in Stress-Associated Diseases. *European Journal of Health Psychology* 27, 4 (2020), 132–153.
- [124] Andreas Richard Schwerdtfeger, Josef Martin Tatschl, and Christian Rominger. 2025. Effectiveness of 2 Just-in-Time Adaptive Interventions for Reducing Stress and Stabilizing Cardiac Autonomic Function: Microrandomized Trials. *Journal of Medical Internet Research* 27 (2025), e69582.
- [125] Fred Shaffer and Jay P Ginsberg. 2017. An overview of heart rate variability metrics and norms. *Frontiers in public health* 5 (2017), 258.
- [126] Amir Mohammad Shahsavarani, Esfandiar Azad Marz Abadi, and Maryam Hakimi Kalkhoran. 2015. Stress: Facts and theories through literature review. *International Journal of Medical Reviews* 2, 2 (2015), 230–241.
- [127] Moushumi Sharmin, Andrew Raji, David Epstien, Inbal Nahum-Shani, J. Gayle Beck, Sudip Vhaduri, Kenzie Preston, and Santosh Kumar. 2015. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Osaka, Japan) (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 505–516.
- [128] Mahsa Sheikh, Meha Qassem, and Panicos A Kyriacou. 2021. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Frontiers in digital health* 3 (2021), 662811.
- [129] David K Sherman. 2013. Self-affirmation: Understanding the effects. *Social and Personality Psychology Compass* 7, 11 (2013), 834–845.
- [130] Hangsik Shin. 2016. Ambient temperature effect on pulse rate variability as an alternative to heart rate variability in young adult. *Journal of clinical monitoring and computing* 30, 6 (2016), 939–948.
- [131] Johannes Siegrist. 1996. Adverse health effects of high-effort/low-reward conditions. *Journal of occupational health psychology* 1, 1 (1996), 27.
- [132] Myriam Sillevs Smitt, Mehdi Montakhabi, Jessica Morton, Cora van Leeuwen, Klaas Bombeke, and An Jacobs. 2022. Users' Perceptions of a Digital Stress Self-monitoring Application: Research Insights to Design a Practical Innovation. In *Universal Access in Human-Computer Interaction. Novel Design Approaches and Technologies*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 325–341.
- [133] Amisha Singh, Akshita Agarwal, Ajeet Kumar, Umang Rastogi, and Sushil Kumar. 2024. A Comprehensive Approach to Stress Detection and Management Using Machine Learning, Deep Learning, and Chatbot Integration. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 1–5.
- [134] Changwon Son, Sudeep Hegde, Carl Markert, Karim Zahed, and Farzan Sasangohar. 2023. Use of a mobile biofeedback app to provide health coaching for stress self-management: Pilot quasi-experiment. *JMIR Formative Research* 7, 1 (2023), e41018.
- [135] Hansoo Song, Hyoung Ryoul Kim, Inah Kim, Jin-Ha Yoon, Sang-Baek Koh, Sung-Soo Oh, Hee-Tae Kang, Da-Yee Jeung, Dae-Sung Hyun, Chunhui Suh, et al. 2025. Occupational stress (KOSS® 19): scale development and validation in the Korean context. *Annals of Occupational and Environmental Medicine* 37 (2025), e12.
- [136] Derek P Spangler, Emily J Dunn, Amelia Aldao, Nicole R Feeling, Matthew L Free, Brandon L Gillie, Michael W Vasey, DeWayne P Williams, Julian Koenig, and Julian F Thayer. 2021. Gender matters: Nonlinear relationships between heart rate variability and depression and positive affect. *Frontiers in Neuroscience* 15 (2021), 612566.
- [137] Xiao Sun. 2025. WELD: A Large-Scale Longitudinal Dataset of Emotional Dynamics for Ubiquitous Affective Computing. [arXiv:2510.15221](https://arxiv.org/abs/2510.15221) [cs.AI]
- [138] Sushma Suri and Saba Rizvi. 2008. Mental health and stress among call center employees. *Journal of the Indian academy of applied psychology* 34, 2 (2008), 215–220.
- [139] Elizabeth Teisberg, Scott Wallace, and Sarah O'Hara. 2020. Defining and implementing value-based health care: a strategic framework. *Academic Medicine* 95, 5 (2020), 682–685.
- [140] Julian F Thayer, Fredrik Åhs, Mats Fredrikson, John J Sollers III, and Tor D Wager. 2012. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews* 36, 2 (2012), 747–756.
- [141] Xin Tong, Matthew Louis Mauriello, Marco Antonio Mora-Mendoza, Nina Prabhu, Jane Paik Kim, and Pablo E Paredes Castro. 2023. Just do something: Comparing self-proposed and machine-recommended stress interventions among online workers with home sweet office. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [142] Thomas Vaessen, Aki Rintala, Natalya Otsabryk, Wolfgang Viechtbauer, Martien Wampers, Stephan Claes, and Inez Myin-Germeys. 2021. The association between self-reported stress and cardiovascular measures in daily life: A systematic review. *PLoS one* 16, 11 (2021), e0259557.

- [143] Rebecca Valerio and Marwa Mahmoud. 2025. *A multimodal Framework for exploring behavioural cues for automatic Stress Detection*. Association for Computing Machinery, New York, NY, USA, 535–539.
- [144] Melanie M van der Ploeg, Jos F Brosschot, Bart Verkuil, Brandon L Gillie, DeWayne P Williams, Julian Koenig, Michael W Vasey, and Julian F Thayer. 2017. Inducing unconscious stress: Cardiovascular activity in response to subliminal presentation of threatening and neutral words. *Psychophysiology* 54, 10 (2017), 1498–1511.
- [145] Nadine Wagener, Marit Bentvelzen, Bastian Dänekas, Paweł W. Woźniak, and Jasmin Niess. 2023. VeatherReflect: Employing Weather as Qualitative Representation of Stress Data in Virtual Reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (Pittsburgh, PA, USA) (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 446–458.
- [146] Jeannette Weber, Peter Angerer, and Jennifer Apolinário-Hagen. 2022. Physiological reactions to acute stressors and subjective stress during daily life: A systematic review on ecological momentary assessment (EMA) studies. *PLoS one* 17, 7 (2022), e0271996.
- [147] Jeannette Weber, Meike Heming, Jennifer Apolinário-Hagen, Stefan Liszto, and Peter Angerer. 2024. Comparison of the Perceived Stress Reactivity Scale with physiological and self-reported stress responses during ecological momentary assessment and during participation in a virtual reality version of the Trier Social Stress Test. *Biological Psychology* 186 (2024), 108762.
- [148] Andrea Wiglesworth, Jessica Butts, Katherine A Carosella, Salahudeen Mirza, Victoria Papke, Jason José BendeZú, Bonnie Klimes-Dougan, and Kathryn R Cullen. 2023. Stress system concordance as a predictor of longitudinal patterns of resilience in adolescence. *Development and psychopathology* 35, 5 (2023), 2384–2401.
- [149] Brent D Winslow, Rebecca Kwasinski, Jeffrey Hullfish, Mitchell Ruble, Adam Lynch, Timothy Rogers, Debra Nofziger, William Brim, and Craig Woodworth. 2022. Automated stress detection using mobile application and wearable sensors improves symptoms of mental health disorders in military personnel. *Frontiers in Digital Health* 4 (2022), 919626.
- [150] Annika Wolff, Daniel Gooch, Jose J Cavero Montaner, Umar Rashid, and Gerd Kortuem. 2016. Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics* 12, 3 (2016).
- [151] Saskia Wouters, Nele Jacobs, Mira Duif, Lilian Lechner, and Viviane Thewissen. 2018. Negative affective stress reactivity: The dampening effect of snacking. *Stress and Health* 34, 2 (2018), 286–295.
- [152] Mengru Xue, Rong-Hao Liang, Jun Hu, Bin Yu, and Loe Feijs. 2022. Understanding How Group Workers Reflect on Organizational Stress with a Shared, Anonymous Heart Rate Variability Data Visualization. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 27, 7 pages.
- [153] Mengru Xue, Rong-Hao Liang, Bin Yu, Mathias Funk, Jun Hu, and Loe Feijs. 2019. AffectiveWall: designing collective stress-related physiological data visualization for reflection. *IEEE Access* 7 (2019), 131289–131303.
- [154] Xiaofan Yu, Lanxiang Hu, Benjamin Reichman, Dylan Chu, Rushil Chandrupatla, Xiyuan Zhang, Larry Heck, and Tajana S Rosing. 2025. Sensorchat: Answering qualitative and quantitative questions during long-term multimodal sensor interactions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–35.
- [155] Dilan Yüksel and Dilek Güneş. 2024. The effect of stress ball on anxiety and pain levels in angiography: A Randomized Controlled Trial. *Cardiovascular Therapeutics* 2024, 1 (2024), 5049092.
- [156] Young Ho Yun, Xin Shelley Wang, Jung Suk Lee, Ju Won Roh, Chang Geol Lee, Won Sup Lee, Keun Seok Lee, Soo-Mee Bang, Tito R Mendoza, and Charles S Cleeland. 2005. Validation study of the korean version of the brief fatigue inventory. *Journal of pain and symptom management* 29, 2 (2005), 165–172.
- [157] Yiran Zhao and Tanzeem Choudhury. 2024. Evaluate Closed-Loop, Mindless Intervention in-the-Wild: A Micro-Randomized Trial on Offset Heart Rate Biofeedback. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing (Melbourne VIC, Australia) (UbiComp '24)*. Association for Computing Machinery, New York, NY, USA, 307–312.
- [158] Yiran Zhao, Yujie Tao, Grace Le, Rui Maki, Alexander Adams, Pedro Lopes, and Tanzeem Choudhury. 2023. Affective touch as immediate and passive wearable intervention. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–23.

A Demographic Information

Table 6. Overview of participants' demographic characteristics and pre-survey responses.

PID	Gender	Age	Years of Experience	Pilot	PID	Gender	Age	Years of Experience	Pilot
P01	F	38	≥ 2 years	Y	P11	F	34	≥ 2 years	N
P02	F	43	≥ 5 years	Y	P12	F	50	≥ 5 years	N
P03	F	35	≥ 5 years	Y	P13	F	50	≥ 5 years	N
P04	M	37	≥ 3 years	N	P14	F	31	≥ 5 years	N
P05	F	38	≥ 3 years	N	P15	F	42	≥ 3 years	N
P06	F	38	≥ 1 year	N	P16	F	44	≥ 5 years	N
P07	M	42	≥ 3 years	N	P17	F	31	≥ 3 years	N
P08	F	34	≥ 5 years	N	P18	F	44	≥ 3 years	N
P09	F	42	≥ 5 years	N	P19	F	33	≥ 2 years	N
P10	F	33	≥ 3 years	N					

Table 7. Participants' results on pre-survey. Score values are *mean ± standard deviation* across participants (n=19). KOSS-19 (Korean Occupational Stress Scale) is a job stress questionnaire (emotion labor-related), and K-BAT (Korean Burnout Assessment Tool) is a Korean version of the job burnout assessment scale.

Survey type	Category	Score	Max score
KOSS-19	Physical environment	4.21±1.18	10
	Job demands	5.37±1.16	15
	Job autonomy	5.00±1.05	10
	Relationship conflict	6.47±1.12	10
	Job insecurity	5.05±1.08	10
	Organizational system	11.47±1.54	20
	Unfair compensation	5.84±0.83	10
	Work-life balance	5.52±0.70	10
K-BAT	Exhaustion	24.63±4.80	40
	Mental distance	11.11±2.82	20
	Impaired cognitive control	11.84±2.91	25
	Impaired emotional control	12.63±3.15	25

In the pre-survey (see Table 7), participants showed moderate levels of stress across most categories, with lower stress in job demands and less burden from impaired cognitive control. However, they reported higher stress in relationship conflict and greater burden from exhaustion.

B Intervention Details

- **Stretching:** Participants were given three types of stretching exercises: shoulder and hip, rounded shoulder and back, and lower back and shoulder. Participants selected one type and followed video instructions. The entire procedure took less than five minutes.
- **Deep Breathing:** Participants were instructed to breathe in for 4 seconds, hold their breath for 7 seconds, and breathe out for 8 seconds. This procedure was repeated three times.
- **I Did Well, Right?:** Participants wrote at least one thing they did well in the previous call, including even minor accomplishments.
- **Words I Need Right Now:** Participants were asked to write short words or phrases they would like to hear at the moment, such as words of comfort or determination.
- **Anger-Eating Fairy:** Participants were asked to write about frustrating moments they experienced during the previous call. Then, an animation played in which a character consumed (or “ate up”) the written text.
- **Protect Me:** Participants tapped on the screen to create a shield over a character, which expanded with each tap. They were required to tap 30 times to complete the intervention.
- **Sugar Boost:** Participants were asked to eat some sweets.
- **Is It Because of Me?:** Participants answered three questions, each accompanied by three suggested responses. Afterward, they were shown a predetermined answer script for each question, described in Appendix B.1.

B.1 Script of “Is It Because of Me?” Intervention

- Q1. Some words come and go, but some linger in your heart. Are you feeling that way now?
- Options: Yes, kind of / I’m not sure / I’m okay
 - Answer: Whatever you’re feeling is okay. Emotions often stay quietly inside us longer than we think, and sometimes they just rise to the surface unexpectedly.
- Q2. Do you think they really meant those words? Perhaps they said it not because they dislike you, but because they were frustrated with the situation. What do you think?
- Options: It felt like they meant it to me / It seemed because of the situation / I’m not sure
 - Answer: Whatever you’re feeling is okay. Emotions often stay quietly inside us longer than we think, and sometimes they just rise to the surface unexpectedly.
- Q3. Do you feel like you need to keep holding on to this feeling? What would you like to do with it right now?
- Options: I want to let it go for a while / I’ll hold on to it a bit longer / I’m not sure
 - Answer: That’s okay. Whatever you choose is okay. Just remember—this feeling is not your fault. If you can let it go, that’s a way of protecting yourself.

C Survey Items

Table 8. Summary of collected *Pre-shift* survey items.

Survey	Category	Question	Answers
Pre-shift	Sleep quality	How would you rate the overall quality of your sleep last night?	Very poor (1) - Very good (5)
	General health	How would you rate your overall health at this moment?	Very poor (1) - Very good (5)
	Perceived stress	How stressed do you feel right now?	Not at all (1) - Very much (5)
	Arousal	How would you describe your current emotional state?	Very calm (1) - Very energetic (5)
	Valence	How would you describe your current emotional state?	Very negative (1) - Very positive (5)
	Fatigue	How fatigued do you feel right now?	Not at all (1) - Very much (5)

Table 9. Summary of collected *Post-call* survey items.

Survey	Category	Question	Answers
Post-call	Previous call type	Please select the type of the previous call you handled.	General inquiry / Complaint
	Workload	Do you feel that your current workload is too low or too high?	Very low (1) - Very high (5)
	Perceived stress	How stressed did you feel during the previous call?	Not at all (1) - Very much (5)
	Stressor	Please select all factors that caused you stress. (Multiple selections allowed)	Felt that my abilities were insufficient / Difficulty in understanding the tasks / Pressure from performance evaluation / Dissatisfaction with work procedures / Communication issues with customers / Rude customers / Time pressure / Noise from surrounding people / Conflicts or issues with colleagues / Other
	Arousal	During the previous call, how would you describe your emotional state?	Very calm (1) - Very energetic (5)
	Valence	During the previous call, how would you describe your emotional state?	Very negative (1) - Very positive (5)
	Fatigue	How fatigued did you feel during the previous call?	Not at all (1) - Very much (5)
	Surface acting	During the previous call, to what extent did you try to hide your true emotions?	Not at all (1) - Very much (5)

Table 10. Summary of collected *Post-intervention* survey items.

Survey	Category	Question	Answers
Post-intervention	Intervention satisfaction	How satisfied were you with the intervention you completed?	Not at all (1) - Very much (5)
	Perceived effectiveness	To what extent did you feel that this intervention helped relieve your stress?	Not at all (1) - Very much (5)
	Perceived stress	How stressed do you feel right now after completing the intervention?	Not at all (1) - Very much (5)

Table 11. Summary of collected *Post-intervention refusal* survey items.

Survey	Category	Question	Answers
Post-intervention refusal	Reason for Refusal	Please select all reasons for declining the intervention. (Multiple selections allowed)	Felt that the activity would not be helpful / Did not feel like doing it at the moment / Too much work or other tasks to attend to / Concerned about others' opinions / Felt okay as I am right now / Other

D Data Monitoring

During the data collection period, participants' data were monitored in real time. We conducted daily server-side monitoring and weekly call-center monitoring throughout data collection.

For daily checks, we reviewed app usage days, per-day and cumulative response counts versus target, and input integrity (option misuse, free-text anomalies such as meaningless entries and blanks) to monitor self-reported data. Also, we reviewed passive-sensing summary statistics (record counts, mean, and standard deviation per sensor). We identified non-wear episodes, including charging, and used these indicators to screen for insufficient sampling (e.g., network transmission failures) and incorrect wearing of the watch.

For weekly checks, we visited the call center to retrieve locally stored data. We used timeline plots to align self-report timestamps with call logs and passive sensing. Also, we inspected data-quality metrics (missing rate, anomaly rate, and score based on Signal Quality Index (SQI) proposed by Orphanidou et al. [102]) to localize and diagnose problematic intervals.

We contacted participants via a study-specific mobile messaging channel³ for brief reminders and technical support, such as reminders to ensure proper watch wear. Participants with poor data yield or low ESM responses were asked to follow an escalation protocol: 1) a gentle reminder about proper wearing, charging, or Bluetooth Low Energy (BLE) synchronization, 2) stepwise troubleshooting regarding app relaunch, watch reboot, and permission checks, 3) remote or on-site assistance if needed, and 4) extension of the collection window when appropriate.

For six participants whose data quality was suboptimal or who had provided an insufficient number of Experience Sampling Method (ESM) responses, an additional data collection period was carried out from August 18 to August 22.

³KakaoTalk Channel is the official business account feature of KakaoTalk, enabling 1:1 chat and broadcast messages to subscribers; we operated a study-specific Channel for reminders and technical support only.

E Intervention Satisfaction and Effectiveness

Table 12. Participants' perception of each intervention and effect on stress. Satisfaction and perceived effectiveness are answered on a 5-point Likert scale. Values are *mean ± standard deviation* across participants (n=17).

Intervention	Satisfaction	Perceived effectiveness	Stress reduction (%)	
			Perceived	Physiological
Stretching	3.62 ± 0.73	3.4 ± 0.67	16.82 ± 20.81	16.3 ± 36.32
Deep Breathing	3.87 ± 0.59	3.74 ± 0.51	14.04 ± 23.74	4.59 ± 17.32
I Did Well, Right?	3.22 ± 0.77	3.14 ± 0.67	14.28 ± 14.39	2.54 ± 42.18
Words I Need Right Now	3.37 ± 0.75	3.22 ± 0.73	14.6 ± 16.36	10.89 ± 32.95
Anger-Eating Fairy	3.47 ± 0.87	3.35 ± 0.79	16.06 ± 19.75	16.3 ± 23.67
Protect Me	3.7 ± 0.7	3.56 ± 0.68	17.49 ± 16.4	8.33 ± 21.61
Sugar Boost	3.5 ± 0.67	3.28 ± 0.64	15.65 ± 17.92	10.73 ± 21.13
Is it Because of Me?	2.83 ± 0.81	2.75 ± 0.77	8.33 ± 18.82	-0.21 ± 20.75

F Summary Template for Data Visualization

F.1 Treemap in State Mode

The top N factors correlated with stress were identified, with N varying by category. The contents inside square brackets were adapted according to the stress type and data collected from participants: *Your [perceived/physiological] stress showed the strongest associations with [3-4 factors]. Within stressor category, the top correlates were [2-3 stressors]. In environment category, [1-2 environment factors] had the highest association. In situation category, [2-3 situation factors] were most related. In pre-shift category, [1-2 pre-shift factors] showed the strongest associations.*

If some factors exhibited a correlation in the opposite direction from the general direction of correlation with stress, the following template was added. *Interestingly, for some factors, they tended to increase([All factors that should show negative correlation]) or decrease([All factors that should show positive correlation]) as [perceived/physiological] stress increases.*

F.2 Barchart in Change Mode

The top two to three interventions showing the greatest stress reduction were identified, along with the corresponding amount of reduction. If any interventions were associated with increased stress, one to two interventions with the largest stress increase were also presented, together with their increase rates. The contents inside the square brackets were adapted according to the stress type and data collected from participants: *Your [perceived/physiological] stress decreased the most with [2-3 interventions]. In contrast [1-2 interventions] was associated with higher [perceived/physiological] stress.*

G Interface of the Tablet Application

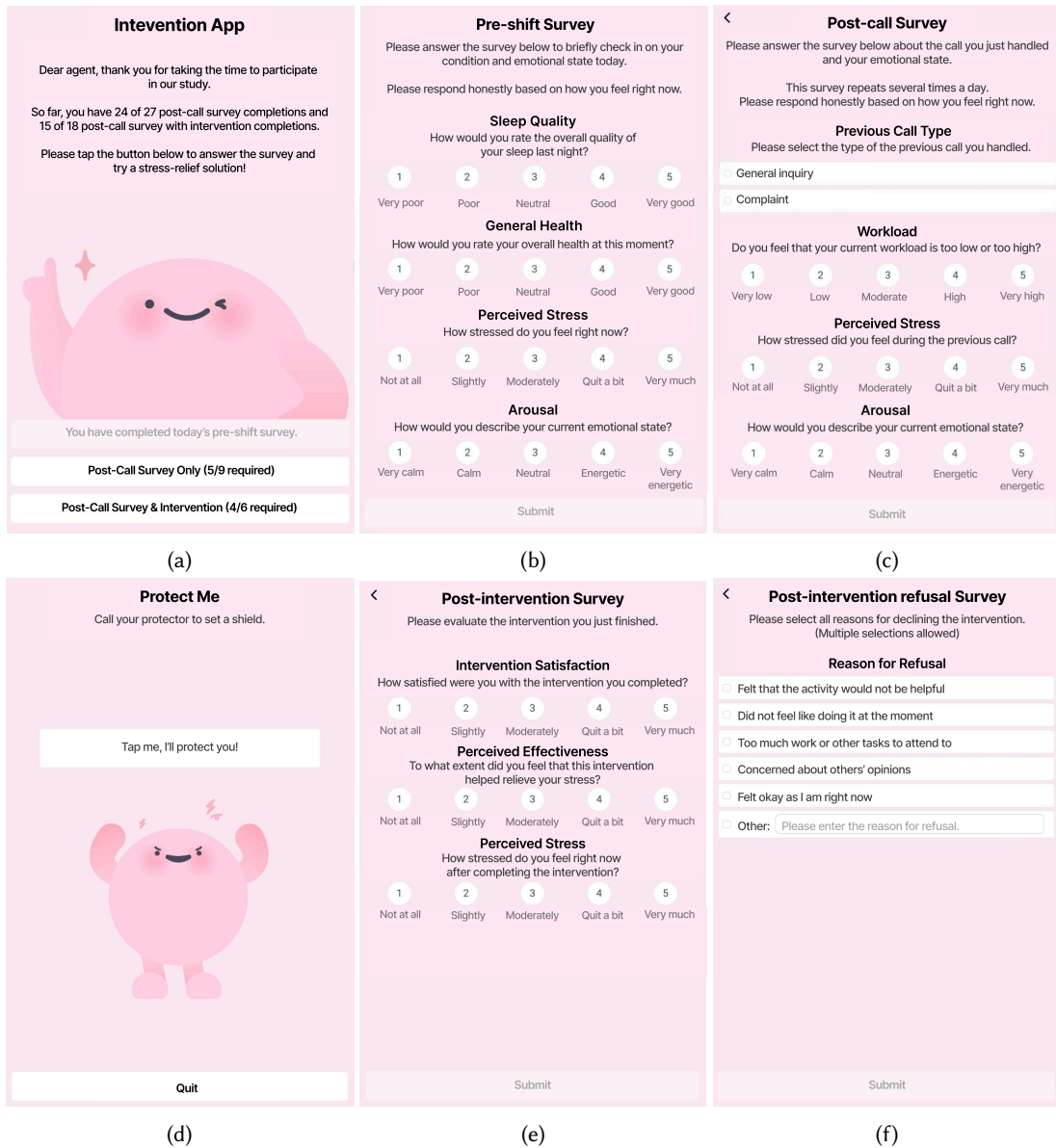


Fig. 8. Screenshots of the tablet application: (a) introduction page, (b) *pre-shift* survey, (c) *post-call* survey, (d) intervention (example: “*Protect Me*”), (e) *post-intervention* survey, and (f) *post-intervention refusal* survey.

Agents could access the pre-shift survey (Fig. 8(b)) using the first button shown in Fig. 8(a). They could complete the post-call survey (Fig. 8(c)) using either the second or the third button in Fig. 8(a). When the third button was used, a random intervention was recommended following the post-call survey (Fig. 8(d)). If participants terminated

the intervention using the “Quit” button (Fig. 8(d)), they were prompted to complete the post-intervention refusal survey (Fig. 8(f)). Otherwise, they proceeded to the post-intervention survey (Fig. 8(e)).

The progress toward the target number of post-call surveys and interventions could be viewed in Fig. 8(a), displayed on the corresponding buttons.

H Interview Protocol

H.1 Pilot Study

Before the interview, we informed the participants as following: We’ll now begin an interview about the experiment you participated in over the past four days. If there are any questions you’d prefer not to answer, please feel free to decline. Before we begin, the audio recording of this interview will be started now.

I. Overall Experience

- (1) How was your overall experience participating in the experiment?
- (2) Was there anything particularly inconvenient or difficult during the experiment?

II. Device Usage

- (1) Did you have any problems using the tablet, smartphone, or smartwatch?
 - (i) Did any of the devices (especially the watch) run out of battery during the experiment?
- (2) Was there anything uncomfortable about wearing or charging the watch? (e.g., feeling the watch was distracting, the watch getting warm, etc.)

III. App Interface and Task Performance

- (1) Was the provided intervention app intuitive to use? (Was there anything confusing?)
 - (a) Please tell us if there were any inconveniences while using the app.
- (2) Was completing 5–10 solution tasks per day realistically manageable while working?

IV. Survey Comprehension

- (1) Were the survey questions easy to understand?
- (2) Were there any questions that were confusing or difficult to answer?

V. Intervention Content and Effectiveness

- (1) Which interventions were helpful, if they exist?
- (2) Regardless of their effectiveness, were there any interventions you particularly liked or found enjoyable?
- (3) Conversely, were there any interventions that were hard to engage with or that you felt weren’t helpful?

VI. Suggestions for Improvement

- (1) The main experiment will begin next Monday. Do you have any suggestions for improvement?
- (2) Please share any ideas on improving data collection or the experimental process, as well as the interventions and app features
- (3) (If the participant previously reported an error and they didn’t mention about it) You mentioned an error occurred before. Could you describe the situation in as much detail as you remember?

H.2 Main Study

H.2.1 Warm-up. While there was no strict protocol for the introduction, participants were informed that their participation was voluntary, that they could skip any questions they did not wish to answer, and that the interview would be audio-recorded.

I. Overall Data Collection

- (1) During the experiment, what positive aspects did you find about responding to the tablet app surveys, performing interventions, or wearing the smartwatch? Why?
- (2) During the experiment, what aspects did you find difficult or inconvenient about responding to the tablet app surveys, performing interventions, or wearing the smartwatch? Why?

II. Mental Model on Perceived Stress and Physiological Stress

- (1) In your own words, how do you understand the difference between perceived stress and physiological stress?
- (2) While participating in this experiment, did you understand how the smartwatch measured your stress? (e.g., heart rate, heart rate variability)

III. Experience with Interventions

- (1) Among the randomly recommended interventions, which ones did you prefer and which ones did you not? Please explain the reasons for your preferences.
- (2) (If the participant answers “because it was quick to complete”) Besides being quick to complete, were there any other reasons that you liked about the intervention?
- (3) For example,
 - (a) Were there any interventions that you initially didn’t expect to be helpful but turned out to be beneficial once you tried them?
 - (b) Among the interventions, which ones were easy to perform during work hours, and which ones were difficult to perform?

H.2.2 Data Exploration on State Mode. The participants explored the visualization dashboard in State Mode for 10 to 15 minutes. They performed below 3 tasks to explore stress states and related factors. Because it may be difficult to explore all data, we allowed them to freely focus on the sections they find most interesting.

- T1. Using the calendar view, observe how your stress levels have changed by week and by day of the week.
- T2. In the timeline view, select a date you wish to explore and examine your cognitive and physiological stress states together with contextual data such as situation, environment, and causes.
- T3. Explore the contextual data (stressor, environment, situation, pre-shift) related to your perceived and physiological stress. Using the data, reflect on what kinds of factors seem to be related to your stress.

After the exploration, we asked them questions based on the following interview protocol. During the interview, participants were allowed to freely look at the dashboard.

I. Overall Experience

- (1) Did viewing your stress data through the dashboard help you better understand your own stress? Please feel free to explain why or why not.
- (2) While using the dashboard today, did you discover any new patterns or insights about your stress that you hadn’t been aware of before? Or, did your way of understanding your own stress change compared to how you usually perceive it?

II. Perception of Perceived vs. Physiological Stress

- (1) Perceived stress and physiological stress represent two different aspects of your stress, captured from different data sources. As you viewed your perceived stress records alongside your physiological stress data, what stood out to you?
 - (a) For example, if your perceived stress appeared high but physiological stress was low, what additional data did you look at to judge whether you were truly under stress? (e.g., your subjective feelings, the nature of the calls, environmental factors, etc.)
- (2) When the two stress types differed, what data did you refer to in order to interpret the situation?
 - (a) For example, there may have been cases where you felt fine mentally, but your physiological stress appeared high, or times when perceived stress was high but physiological stress was low. How did you interpret those differences? Which parts of the data did you focus on to make sense of them?

III. Exploring Factors Affecting Stress States

- (1) While viewing the dashboard, which factors did you feel were most closely related to your perceived stress? Why?
- (2) While viewing the dashboard, which factors did you feel were most closely related to your physiological stress? Why?
- (3) If the factors associated with perceived and physiological stress appeared different or similar, how did you interpret those results?
 - (a) For example, perceived stress might seem more related to call difficulty, while physiological stress might seem more connected to lack of sleep. When such differences appeared between the two types of stress, how did you make sense of them?

H.2.3 Data Exploration on Change Mode. The participants explored the visualization dashboard in Change Mode for 10 to 15 minutes. They performed below 3 tasks to explore stress changes due to interventions and type of interventions that affected stress level. Because it may be difficult to explore all data, we allowed them to freely focus on the sections they find most interesting.

- T1. Using the calendar view, observe how your stress levels changed by week and day of the week following the introduction of interventions.
- T2. Select a date you wish to explore, and in the timeline view, examine whether your perceived and physiological stress levels increased or decreased after performing the interventions.
- T3. Explore which types of interventions were associated with changes in your perceived and physiological stress levels.

After the exploration, we asked them questions based on the following interview protocol. Again, during the interview, participants were allowed to freely look at the dashboard.

I. Perception of Changes in Perceived and Physiological Stress After Interventions

- (1) As you reviewed the changes in your perceived stress and physiological stress together, what did you notice?
 - (a) For example, there may have been times when your physiological stress decreased but your perceived stress increased. How did you interpret such differences, and what do you think caused them?
- (2) When the changes in the two stress types appeared differently, what criteria did you use to decide whether your stress had truly increased or decreased?
 - (a) For example, What additional data did you refer to in judging whether your stress was actually relieved? (e.g., your subjective feelings, the nature of the calls, environmental factors, etc.)

II. Exploring Interventions Influencing Stress Change

- (1) Among the various interventions, which ones did you feel were particularly effective in reducing your stress? Did your personal perception align with what was shown in the data rankings?
 - (a) For example, you may have thought that stretching was the most effective intervention, but the data indicated that having a sugary snack was the most effective.
- (2) After performing a stress-relief activity, have you ever physically felt a change not only in your perceived stress but also in your bodily responses? Please describe the situation in detail.
 - (a) For example, before performing the intervention, you might have felt your heart pounding with anger, but after doing a randomly recommended breathing exercise, you felt your heartbeat calm down. Have you experienced such change in your physiological response?

H.2.4 Closing. At the end, we ask the participants if there is any data or functionality that should be added to the stress management system like the one used in the experiment to help understand their stress.

Table 13. Category-level distribution of correlation magnitudes for perceived and physiological stress indicators. We summarize the distribution of $|\rho|$ by contextual category (Stressor, Pre-shift, Situation, Environment), reporting median $|\rho|$, IQR of $|\rho|$, and the fraction of correlations exceeding $|\rho| \geq 0.30$. This provides a compact overview of which categories most systematically relate to each stress indicator.

Category	Perceived Stress Indicator			Physiological Stress Indicator		
	Median $ \rho $	$ \rho $ _IQR	$\%(\rho \geq 0.30)$	Median $ \rho $	$ \rho $ _IQR	$\%(\rho \geq 0.30)$
Stressor	0.183143	0.272444	0.352941	0.047252	0.064261	0
Pre-shift	0.067109	0.070259	0.01	0.080737	0.092762	0.02
Situation	0.270918	0.560856	0.485207	0.058637	0.10146	0.029586
Environment	0.076612	0.066074	0	0.073588	0.096232	0.014706

Table 14. Feature-level summary of within-participant Spearman correlations with perceived and physiological stress indicator. For each feature, we report the median correlation across participants with interquartile range (IQR), the sign-agreement rate (SA; proportion of participants sharing the majority sign), and the proportion of participants with moderate-to-strong associations ($|\rho| \geq 0.30$), enabling identification of consistent between-participant patterns beyond visually dense heatmaps.

Category	Feature	Perceived Stress Indicator			Physiological Stress Indicator			Δ median		
		Median ρ	IQR	SA	$\%(\rho \geq 0.30)$	Median ρ	IQR		SA	$\%(\rho \geq 0.30)$
Stressor	lack_ability	0.10	[0.05, 0.13]	0.83	0.000	-0.00	[-0.07, 0.02]	0.58	0.00	0.10
	difficult_work	0.19	[0.12, 0.31]	0.93	0.286	0.03	[-0.02, 0.04]	0.64	0.00	0.16
	eval_pressure	0.05	[0.03, 0.07]	1.00	0.000	0.05	[0.03, 0.09]	1.00	0.00	-0.01
	work_bad	0.22	[0.11, 0.31]	0.86	0.357	0.02	[-0.02, 0.07]	0.57	0.00	0.19
	hard_communication	0.48	[0.27, 0.54]	1.00	0.706	-0.05	[-0.10, -0.01]	0.82	0.00	0.52
	rude_customer	0.32	[0.14, 0.40]	1.00	0.500	-0.01	[-0.07, 0.02]	0.50	0.00	0.33
	time_pressure	0.18	[0.06, 0.19]	1.00	0.231	0.04	[-0.01, 0.10]	0.62	0.00	0.14
	noise	0.09	[0.06, 0.16]	1.00	0.000	0.02	[-0.01, 0.07]	0.60	0.00	0.07
	peer_conflict	0.08	[0.06, 0.08]	1.00	0.000	-0.07	[-0.09, -0.01]	0.67	0.00	0.15
	other	0.37	[0.11, 0.57]	0.88	0.625	0.02	[-0.03, 0.06]	0.63	0.00	0.35
Pre-shift	daily_stress	0.06	[0.04, 0.10]	0.76	0.000	-0.05	[-0.11, 0.06]	0.65	0.00	0.11
	daily_arousal	0.02	[-0.02, 0.07]	0.67	0.000	-0.03	[-0.07, 0.00]	0.73	0.00	0.05
	daily_valence	-0.04	[-0.09, 0.03]	0.65	0.000	0.05	[-0.04, 0.09]	0.65	0.00	-0.09
	daily_tiredness	0.01	[-0.04, 0.07]	0.59	0.059	0.00	[-0.14, 0.12]	0.59	0.00	0.00
	daily_general_health	0.03	[-0.01, 0.10]	0.71	0.000	0.01	[-0.12, 0.02]	0.53	0.06	0.03
	daily_sleep_quality	0.04	[-0.03, 0.10]	0.59	0.000	-0.02	[-0.10, 0.05]	0.59	0.06	0.06
Situation	acc_mean	-0.07	[-0.15, -0.01]	0.76	0.000	0.02	[-0.04, 0.08]	0.53	0.00	-0.09
	acc_std	0.10	[-0.06, 0.12]	0.65	0.000	0.26	[0.17, 0.29]	1.00	0.24	-0.17
	workload	0.32	[0.10, 0.50]	0.94	0.500	0.03	[-0.05, 0.08]	0.56	0.00	0.29
	arousal	0.55	[0.03, 0.74]	0.76	0.647	0.01	[-0.04, 0.02]	0.53	0.00	0.54
	valence	-0.70	[-0.77, -0.60]	0.94	0.941	0.03	[-0.01, 0.07]	0.65	0.00	-0.72
	fatigue	0.82	[0.77, 0.87]	1.00	1.000	0.00	[-0.06, 0.05]	0.53	0.00	0.81
	surface_acting	0.78	[0.65, 0.87]	1.00	1.000	-0.02	[-0.08, 0.04]	0.53	0.00	0.81
	call_type_angry	0.35	[0.30, 0.43]	1.00	0.765	-0.02	[-0.05, -0.00]	0.76	0.00	0.38
	steps	-0.02	[-0.08, 0.02]	0.71	0.000	0.02	[-0.01, 0.08]	0.71	0.00	-0.04
	skintemp	0.02	[0.01, 0.14]	0.82	0.000	-0.15	[-0.20, -0.11]	0.94	0.06	0.17
Environment	humidity_mean	-0.00	[-0.07, 0.06]	0.59	0.000	-0.01	[-0.05, 0.06]	0.59	0.00	0.01
	co2_mean	-0.06	[-0.10, -0.01]	0.76	0.000	0.02	[-0.03, 0.05]	0.65	0.06	-0.08
	tvoc_mean	-0.07	[-0.16, -0.04]	0.76	0.000	0.01	[-0.14, 0.06]	0.53	0.00	-0.07
	temperature_mean	0.08	[-0.02, 0.10]	0.65	0.000	-0.01	[-0.13, 0.04]	0.53	0.00	0.09

1 Correlation between Stress Indicators and Contextual Data

We report category-level distributions of correlation magnitudes, feature-level summary statistics across participants, and frequency-based summaries of the most prominent correlations per participant.

Table 15. Highest-frequency contextual features among the top-K absolute correlations per participant. For each participant, we select the top-K features ranked by $|\rho|$ (K=5) separately for perceived and physiological indicators, and count how often each feature appears across participants (with counts of positive vs. negative associations). This frequency-based view highlights cross-participant regularities.

Category	Feature	Perceived Stress Indicator				Physiological Stress Indicator			
		Top-K Count	Pos	Neg	Top-K Ratio	Top-K Count	Pos	Neg	Top-K Ratio
Stressor	difficult_work	1	1	0	0.058824	0	0	0	0
	work_bad	0	0	0	0	1	1	0	0.058824
	hard_communication	10	10	0	0.588235	2	0	2	0.117647
	rude_customer	1	1	0	0.058824	2	0	2	0.117647
	time_pressure	0	0	0	0	1	1	0	0.058824
	other	7	7	0	0.411765	1	1	0	0.058824
Pre-shift	daily_stress	1	1	0	0.058824	6	2	4	0.352941
	daily_arousal	0	0	0	0	1	0	1	0.058824
	daily_valence	0	0	0	0	3	2	1	0.176471
	daily_tiredness	1	1	0	0.058824	8	3	5	0.470588
	daily_general_health	0	0	0	0	4	0	4	0.235294
	daily_sleep_quality	0	0	0	0	3	1	2	0.176471
Situation	acc_mean	0	0	0	0	5	4	1	0.294118
	acc_std	0	0	0	0	14	14	0	0.823529
	workload	5	5	0	0.294118	3	0	3	0.176471
	arousal	10	9	1	0.588235	0	0	0	0
	valence	15	0	15	0.882353	2	2	0	0.117647
	fatigue	16	16	0	0.941176	0	0	0	0
	surface_acting	15	15	0	0.882353	1	0	1	0.058824
	call_type_angry	3	3	0	0.176471	0	0	0	0
	steps	0	0	0	0	1	1	0	0.058824
skintemp	0	0	0	0	10	0	10	0.588235	
Environment	humidity_mean	0	0	0	0	2	2	0	0.117647
	co2_mean	0	0	0	0	2	1	1	0.117647
	tvoc_mean	0	0	0	0	7	2	5	0.411765
	temperature_mean	0	0	0	0	6	3	3	0.352941

Table 13 summarizes correlation magnitudes by contextual category. Overall, perceived stress shows stronger alignment with Situation and Stressor features, whereas the physiological stress indicator exhibits smaller magnitudes across categories. Table 14 highlights which features show strong and directionally consistent associations across participants (median ρ [IQR], sign agreement, and the fraction with $|\rho| \geq 0.30$). This summary helps identify features that generalize across participants versus those that are heterogeneous. Table 15 complements Table 14 by listing the highest-frequency features among each participant's top-K absolute correlations, making cross-participant regularities immediately visible.

Finally, the category-specific heatmaps visualize participant-level correlation patterns in full detail and provide a reference for the summarized trends above.

Fig. 9-10 presents the Spearman correlation coefficients between features belonging to four categories of context data (stressor, environment, situation, and pre-shift) and each stress indicator for individual participants (these results are also displayed in the dashboard, Fig. 4.(d)). For perceived stress, features in the *situation* and *stressor* categories, obtained from self-reports, generally exhibited higher correlations. In particular, *emotional arousal*, *fatigue*, *surface acting*, and *complaint calls* showed strong positive correlations for many participants, whereas

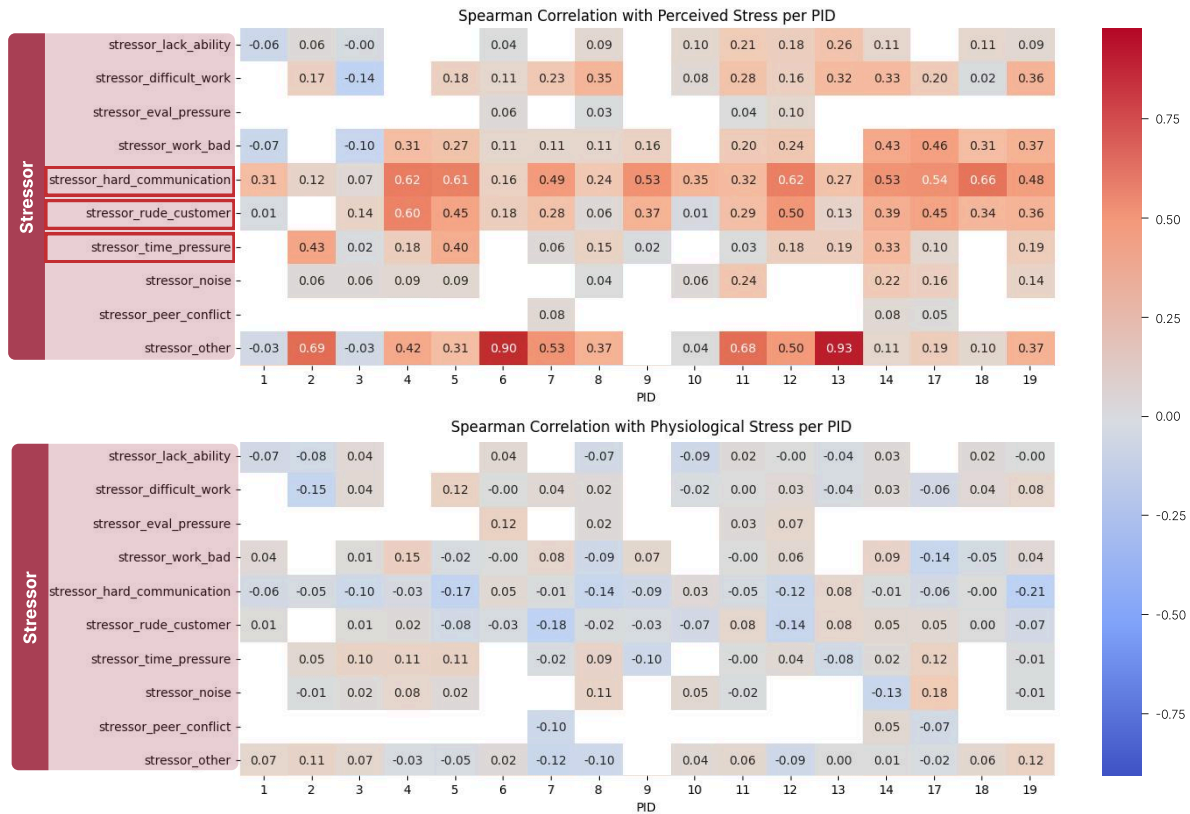


Fig. 9. Spearman correlation heatmaps across participants (x-axis; IDs) and features (y-axis) for stressor data

emotional valence revealed a negative correlation (see purple box in Figure 11). Stressor-related features such as *communication issues*, *rude customers*, and *time pressure* also demonstrated weak to moderate positive correlations (see red boxes in Figure 9). Correlations for the physiological stress indicator were smaller in magnitude and more dispersed. Among relatively consistent patterns, the activity-related feature *acc_std* exhibited positive correlations across multiple participants (see green box in Figure 11). Given the desk-based workflow, step counts in 5-minute windows were near-zero (median=0; Q1=0, Q3=0; mean=0.03, SD=0.86; min=0, max=60), suggesting limited locomotion. However, *acc_std* can still capture fine-grained wrist motion (e.g., typing/mouse use, posture adjustments), and HRV estimates may remain sensitive to such motion. Therefore, we interpret the association with *acc_std* as a potential confound for HRV interpretation, rather than evidence that the physiological indicator purely reflects physical movement. *Skin temperature* and *indoor temperature* showed weak negative correlations in some cases (see blue boxes in Figure 11), while environmental factors (e.g., CO₂ concentration, air quality) exhibited mixed directions across participants, showing no consistent trends (see Figure 12).

In summary, the perceived stress indicator was strongly modulated by emotional and social stimuli such as call situations, whereas the physiological stress indicator exhibited a multidimensional pattern influenced by a complex interplay of environmental factors and physiological responses to stressors.

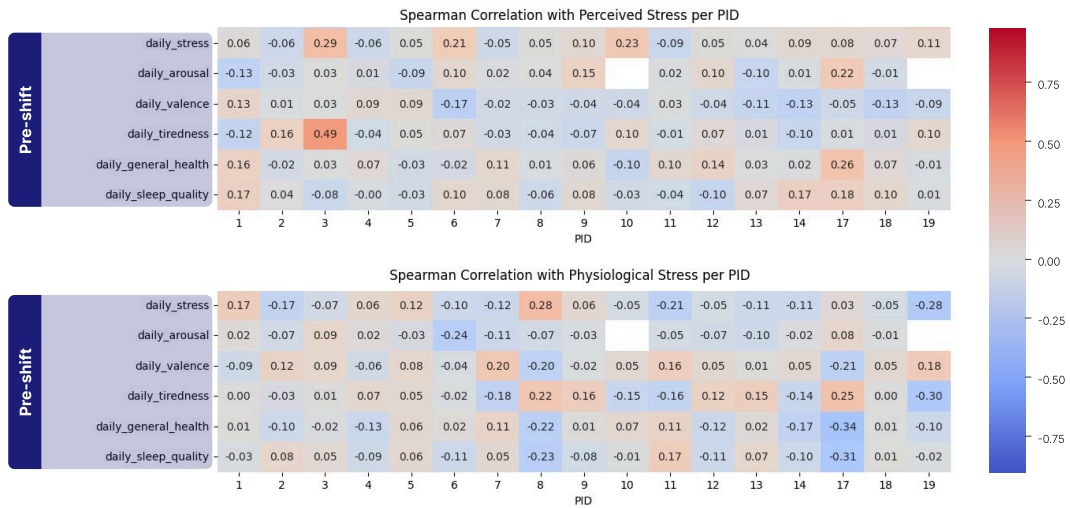


Fig. 10. Spearman correlation heatmaps across participants (x-axis; IDs) and features (y-axis) for pre-shift data

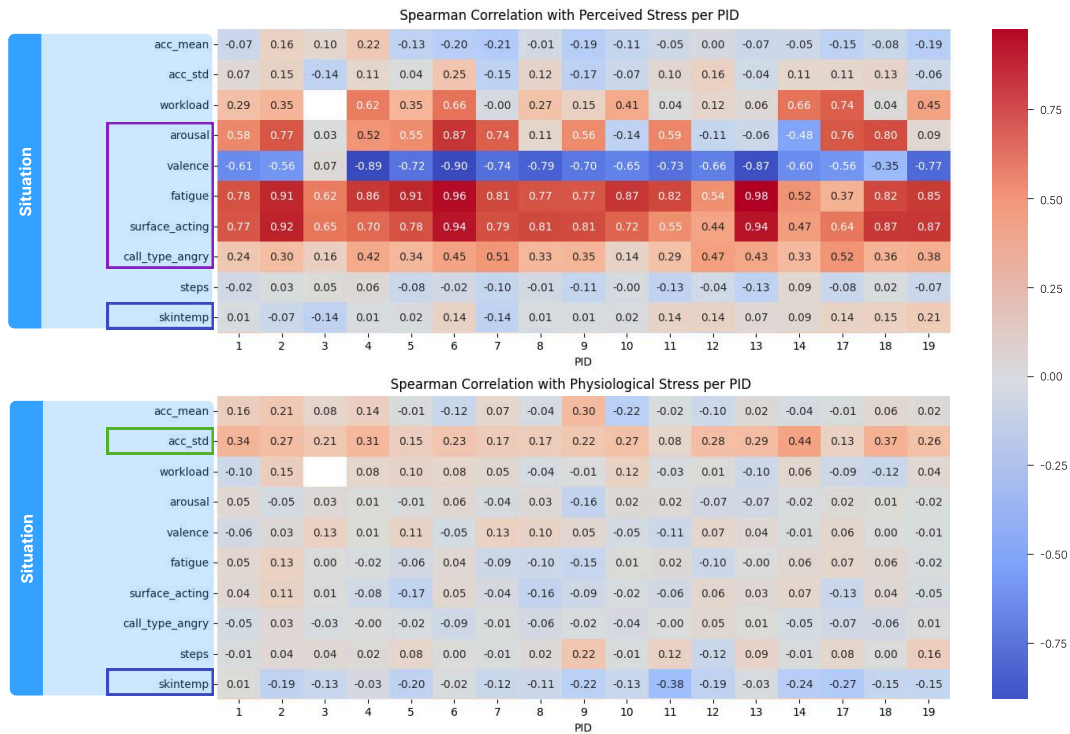


Fig. 11. Spearman correlation heatmaps across participants (x-axis; IDs) and features (y-axis) for situation data

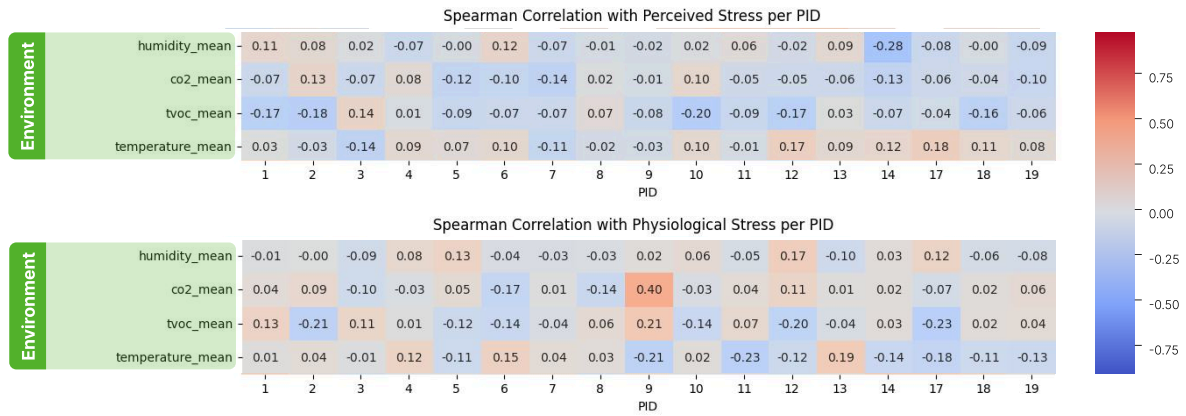


Fig. 12. Spearman correlation heatmaps across participants (x-axis; IDs) and features (y-axis) for environmental data