



# HHS Public Access

Author manuscript

*Proc ACM Interact Mob Wearable Ubiquitous Technol.* Author manuscript; available in PMC 2021 June 06.

Published in final edited form as:

*Proc ACM Interact Mob Wearable Ubiquitous Technol.* 2020 December ; 4(4): . doi:10.1145/3432210.

## Automated Detection of Stressful Conversations Using Wearable Physiological and Inertial Sensors

**RUMMANA BARI,**

University of Memphis, Electrical and Computer Engineering, Memphis, TN, 38152, USA

**MD. MAHBUBUR RAHMAN\***,

University of Memphis, Computer Science, Memphis, TN, USA

**NAZIR SALEHEEN,**

University of Memphis, Computer Science, Memphis, TN, USA

**MEGAN BATTLES PARSONS,**

University of Memphis, Communication Science and Disorder, Memphis, TN, USA

**EUGENE H. BUDER,**

University of Memphis, Communication Science and Disorder, Memphis, TN, USA

**SANTOSH KUMAR**

University of Memphis, Computer Science, Memphis, TN, USA

### Abstract

Stressful conversation is a frequently occurring stressor in our daily life. Stressors not only adversely affect our physical and mental health but also our relationships with family, friends, and coworkers. In this paper, we present a model to automatically detect stressful conversations using wearable physiological and inertial sensors. We conducted a lab and a field study with cohabiting couples to collect ecologically valid sensor data with temporally-precise labels of stressors. We introduce the concept of stress cycles, i.e., the physiological arousal and recovery, within a stress event. We identify several novel features from stress cycles and show that they exhibit distinguishing patterns during stressful conversations when compared to physiological response due to other stressors. We observe that hand gestures also show a distinct pattern when stress occurs due to stressful conversations. We train and test our model using field data collected from 38 participants. Our model can determine whether a detected stress event is due to a stressful conversation with an F1-score of 0.83, using features obtained from only one stress cycle, facilitating intervention delivery within 3.9 minutes since the start of a stressful conversation.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

[rummanabari@gmail.com](mailto:rummanabari@gmail.com).

\*The author contributed while he was at the University of Memphis. He is currently working at Samsung Research America

## 1 INTRODUCTION

Stress is unavoidable in our everyday life. There are several reasons for stress (i.e., stressors), such as long-standing pressures at work, deadlines, test-taking, conflict in conversations, financial difficulties, health issues, care-giving, etc. Prior work has investigated and organized different types of stressors. For example, 1,031 participants were studied in [4]. They observed 4,000 stressful events from the daily life of these participants and organized the stressors in seven broad categories — interpersonal argument and tensions, work, home related stress, finances, health, networking, and miscellaneous. Among them, interpersonal argument and tensions occur most frequently (50% of the time) as people interact with partners, friends, family members, colleagues, and supervisors regularly. This is followed by work-related stress (13.4% of the time) including work demand, overload, technical issues, and job security. Another study with 225 graduate students found that academic or professional demands, interpersonal demands, financial strains, and commuting were the most common stressors [22].

As interactions with partner, family, friends, and colleagues are a fundamental aspect of our daily life, stressful interaction is a major daily stressor for a large population. Healthy interactions can provide happiness, social support, and cause fewer health issues [25, 51]. But, stressful interactions such as conflicts may lead to deleterious consequences to physical and psychological health (e.g., depression, anxiety, and substance abuse) and may affect the relationship quality, happiness, and overall life satisfaction [12, 17, 23, 49]. Moreover, stressful conversations at work can adversely impact productivity, job performance, and job satisfaction [53].

Therefore, it is important to understand the timing, frequency, and duration of stressful conversations to reduce their harmful effect in daily life. Sensor-based automated detection of stressful conversations from the natural environment can be used by researchers to investigate the antecedents, dynamics, and consequents of stressful conversations, potentially leading to novel therapies and interventions. Moreover, real-time detection of such conversations can be used to trigger just-in-time mobile interventions for deescalating a tense situation and for pacifying the users so that they can recover and cope better with the situation.

In this paper, we demonstrate the feasibility of detecting stressful conversations from stress time-series data. In particular, we show that by analyzing the dynamics of stress time series, we can detect whether the current stress event is due to interpersonal interaction or other stressors such as commuting or work related stress. To develop a model for automatic detection of stressful conversations using wearable sensors, we need carefully labeled sensor data with unambiguous and temporally-precise labels (i.e., tight time-synchronization between event labels and sensor data) that mark the timing, duration, and the reason for the stress, all collected from real-life.

The traditional approach is to request users to proactively provide labels by manually keeping a diary [3], retrospectively via an interview [4], or ecological momentary self-reports [20, 22]. However, these methods lack the temporal resolution and reliability needed

to develop a sensor-based model successfully [35]. Alternatively, an observer can be assigned to follow each participant in their daily life. However, this approach involves significant expense, burden, and may still not capture several real-life scenarios in order to respect participants' privacy.

We designed and conducted a lab study and a field study to collect ecologically valid data (i.e., data representing events expected in the daily life) with unambiguous and temporally-precise labels. Stressful conversations usually involve two (or more) parties, all of whose consent is needed, especially for capturing sensor data during stressful conversations. As cohabiting couples typically spend a lot of time together, we recruited couples to wear sensors and collect data concurrently. We adapted the model presented in [45] to find the start and end times of stress events from the sensor data. We developed an automated stress visualization system utilizing Day Reconstruction Method (DRM) [24] to present the detected stress events to users with surrounding contexts (i.e., conversation, location, and physical activity). The goal was to help users confirm or refute a detected stress event and recall the reason for the confirmed stress events, providing us labels of stressors for each identified and confirmed stress event. Finally, as automated detection of conversations from audio or respiration data is limited to an F1 score of around 0.7 [6], we collected high-quality raw audio to verify the presence of conversations via human annotation. As collection of raw audio poses privacy concern and burden because the participants need consent from anyone they talk to, we limited the data collection with each couple to one full day, similar to other studies that also recruited couples and collected wearable sensor and audio data from them [18, 55]. To increase between-person and between-situation diversity in the data, we recruited 38 participants (19 cohabiting couples) in the field.

To understand the nature of physiological response during stressful conversations, we conducted a lab study with 12 participants (6 cohabiting couples) that was structured to trigger stressful conversations among couples. The lab study ensures control of other potentially confounding events in the field that may affect physiology, allowing us to discover the unique patterns of stress response in sensor data during stressful conversations.

In the lab data, we observe that the stress time-series follows a cyclical pattern that likely results from the interplay between the sympathetic and parasympathetic nervous system during a stress response, similar to that found in physiological response during stress [26, 57]. Next, we develop a method to automatically identify this cyclical pattern or cycles in the stress time-series data. We use these cycles as a dynamic, natural window to segment the stress time series during a stress event. We then identify discriminative features from each stress cycle and train a machine learning model to determine whether a stress event is due to stressful conversations.

We show that using features from one stress cycle, the model can identify whether a stress event is due to stressful conversation with an F1 score of 0.74. We also observe distinct patterns in hand gestures during stressful conversations. By augmenting the model with hand gesture features (derived from wrist-worn inertial sensors) within each stress cycle, the F1 score improves to 0.83. A stressful conversation usually consists of multiple stress cycles.

Using all cycles improves the F1 score to 0.89, providing a trade-off between accuracy and how rapidly from the start of a stressful conversation an intervention can be delivered.

## 2 BACKGROUND ON PHYSIOLOGICAL RESPONSE TO STRESSORS

A stressor presents a challenge, opportunity, or threat to users. To help users prepare for stress response, their autonomic nervous system (ANS) activates their physiology, including the cardio-respiratory system (i.e., heart and lungs), endocrine system (e.g., hormone secretion), and the thermoregulatory system (e.g., temperature and sweating). ANS comprises the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS) [47]. The SNS elevates the physiology, preparing the body for a ‘fight-or-flight’ response. To provide the needed energy, SNS stimulates several physiological parameters (e.g., heart rate, respiration rate, blood volume, body temperature, etc.). To limit any damage to the end organs, PNS acts as a counterbalance mechanism to restore calm and thus maintain homeostasis. Its strength is usually proportional to the increase caused by SNS, and it eventually brings the physiology back to a resting state.

The interplay of SNS and PNS systems can be illustrated by considering their impact on the cardiovascular system. In response to a stressor, the SNS increases the heart rate (HR). Once the threat is over, the PNS reduces HR, bringing it back to a resting state [11]. Heart rate variability (HRV) is a measure commonly used to quantify the interaction of SNS and PNS. The HRV is defined as the variation in the beat-to-beat intervals. An increased/decreased HRV indicates increased activity of the PNS/SNS, respectively. Therefore, HRV is a simple measure to quantify the contributions of the PNS/SNS and has traditionally been used to estimate stress response. Heart rate variabilities have been found to follow cyclical patterns in lab settings [26, 57]. De Geus et. al., showed that the heart rate increases when users face stressors [13]. For stressors, they used a tone avoidance task, a memory search task, and a cold pressor test. They found that the heart rate remains high as long as the stressor is present and it goes back to the pre-stress level with the removal of stressors, resulting in a cyclical pattern.

The stress response can also be explained in terms of endocrine response to stress, i.e., salivary cortisol levels. In [38], authors investigated the cortisol level in 124 heterosexual dating couples during a conflict negotiation task. Cortisol was assessed at 7 points before and after the task, creating a trajectory of stress reactivity and recovery for each participant, resulting in a cyclical pattern.

The interplay of SNS and PNS can be distinct when presented with different stressors because the persistence of stress stimuli can differ. For example, during a cold pressor test, the initial stress response can be high due to shock from cold temperature, but physiology can gradually recover as the body gets used to the temperature difference. But in a stressful conversation, there can be highly stressful moments that may be followed by either further escalation or de-escalation, dynamics which can drive the activation of SNS and PNS differently than from a cold pressor test. In fact [7] showed that the stress responses to three different stressors (i.e., cognitive, emotional, and physical) are sufficiently distinct that they can be detected using a machine learning model. In another recent work, [27] showed that

respiration pattern during stressful conversation is different than that during a stressor not involving conversation (i.e., cognitive). Both of these works used controlled lab experiments to show the distinction in stress response due to stressful conversations when compared with that due to other stressors. We build upon these works to observe the physiological responses to various naturally-occurring stressors in real-life and develop a model that can successfully identify when a stress response is due to a stressful conversation.

### 3 RELATED WORKS

Prior works on the detection and characterization of conversations has largely used audio, with recent works using microphones in smartphones. They show that audio data collected by smartphones can be used to automatically detect the occurrence of a conversation [28], measure the frequency and duration of conversations [14], identify the speaker [42], count the number of parties involved in a conversation [59], and quantify the contribution of conversation partners and their turn-taking behaviors [21, 28]. In addition, prior work has also developed algorithms to detect stress from audio data captured using smartphones [30], assessed the correlation between speech characteristics and stressful situations such as job interview or public speaking, and investigated correlations between vocal characteristics and social stress for adolescents with autism spectrum disorder [9]. As detection of stress from audio data can usually be done only when the user is speaking, works on detecting stress from audio data do not explore how to distinguish stressful conversation from other stressors. Further, despite the richness of audio data in detecting and characterization conversations, audio applications have been limited in real-life due to privacy concerns, despite development of algorithms that make inferences from audio data with privacy-preserving features [58]. Therefore, models have been developed to detect conversation from physiological signals (e.g., breathing) [6, 15, 39, 43], to obviate the need for collecting audio data. But none of these works explore the feasibility of detecting stressful conversations from physiological or inertial sensor data.

Separately, physiological sensors such as ECG and respiration have been extensively used in detecting stress, first in the lab settings [2], gradually moving to the field environment via ambulatory Holter Monitors in backpacks [34], then to selected tasks in the field environment with wired wearable sensors [19], and finally to the free-living environment with wireless sensors [20, 37]. Recent works use pulse plethysmograph (PPG) sensor in conveniently-worn wrist devices [10, 36]. They use diverse lab stressors, e.g., using a cold pressor as a physical stressor, mental arithmetic as a cognitive stressor, and public speaking as a social stressor. The focus of the machine learning models in these works has been to detect stress irrespective of the type of stressor by extracting commonality in stress response captured by sensor data so that a single trained model can detect all stress events. Our goal instead is to discover uniqueness in the stress responses due to stressful conversations.

A recent work [27] collected respiration data in the lab settings, where they included a non-verbal relaxer (watch 10 minutes neutral movie), a verbal relaxer (talk in mother tongue for 5 minutes on a chosen topic), a verbal stressor (prepare for and participate in an interview), and a non-verbal stressor (take part in a cognitive task). In order to improve the accuracy of stress detection, they developed a two-stage model. In the first stage, they detect whether a

conversation is taking place, and depending on the outcome, they apply different stress models to detect whether the signals exhibit a stress response. They show that using a two-stage classifier, they achieve 83% accuracy compared to 76% when using a one-layer classifier that does not detect conversations, demonstrating that stress response in respiration is different during stress events with or without a conversation. As their goal was to improve the detection of stress model similar to other works in stress detection, they did not address the issue of distinguishing verbal stressors from non-verbal stressors on their lab dataset. In Section 6.4, we construct a baseline model motivated by this work that uses an automated detection of conversation and automated detection of stress and combines both to detect stressful conversations. We find the best performance from such a model is limited to an F1 score of 0.6.

Finally, [55] showed the feasibility of detecting whether an interpersonal conflict occurred in each hour (reporting an accuracy of 69.2%) using wearable sensor and audio data for that hour from romantic couples who wore sensors for a day in field. As the focus of this work was to detect for each hour whether any conflict occurred or not, they did not present any model to distinguish among different stressors. In summary, to the best of our knowledge, our work is the first attempt at demonstrating that stressful conversations can be detected automatically from wearable physiological sensors in daily life, without the need for audio data.

## 4 STUDIES TO COLLECT ECOLOGICALLY VALID SENSOR DATA WITH PRECISE LABELS

We designed and conducted a lab and a field study to understand the nature of stress patterns during stressful conversations and collect ecologically valid sensor data with precise labels for model development. Both studies were approved by the Institutional Review Board (IRB), and all participants provided written consent. We first describe the study requirements before describing the details of both studies.

### 4.1 Study Requirements

We sought a study design that satisfies the following requirements to produce the necessary sensor data and associated labels for our model development.

1. **Ecologically Valid Sensor Data:** The study should capture physiological sensor data from the field environment during real-life stressors of different types. (Section 4.2)
2. **Stress Event Localization:** The start and end times of each stress event should be located precisely in the sensor data stream. (Section 4.3)
3. **Stressor Labels:** Each stress event should have an assigned label of reason, i.e., stressor. (Section 4.4)
4. **Resolving Ambiguity in Stressor Labels:** Each detected stress event, especially stressful conversations, should be independently confirmed so as to remove any



ambiguity due to machine learning models or recall errors by the participants. (Section 4.6)

5. **Coverage of Stressful Conversations:** The study should have appropriate consent and sensor data available from both the conversing partners, including during stressful conversations. (Section 4.5)
6. **Confounder-free Data:** As field-collected data can be affected by confounding events that can affect physiological signals (e.g., due to physical activity), clean data should also be collected during stressful conversations that are largely free from other confounders, to seed the model development. (Section 4.7)

In the following, we describe how our study design satisfied each of these requirements.

## 4.2 Wearable Devices for Ecologically Valid Sensor Data

Participants wore the AutoSense chest band with Electrocardiogram (ECG) and respiration sensors [16] to capture physiological data in their daily life. To capture physical activity that can confound the inference of stress from physiological sensors and to provide physical activity context surrounding stress events, the chestband included 3-axis accelerometer signals. The participants also wore a wristband consisting of a 3-axis accelerometer and a 3-axis gyroscope on their dominant hand to capture hand gestures during conversations. They wore a LENA audio recorder [1] to capture high-quality audio that could be used to unambiguously verify the occurrence and timing of stressful conversations. They were instructed to carry the recorder in a pouch placed around the waist to reduce occlusion of the microphone and to increase the likelihood of capturing high quality audio.

Each participant was provided with an Android smartphone that collected GPS-traces from which location cues can be inferred. For time synchronization among all sensor signals, the smartphone also received and stored data from all wearable sensors. Participants were asked to carry all the devices during their waking hours except during showers and contact sports, to maximize the opportunity to capture sensor data during stress events.

## 4.3 Stress Detection and Stress Event Localization

We employed previously validated algorithms on the collected sensor data to meet the requirements of precisely locating the start and end times of stress events. We first use the cStress model [20] to obtain a stress state from each minute of ECG and respiration signals that represent the physiological response to a stressor. The model outputs a probability measure of stress scaled between 0 and 1, termed ‘stress likelihood’ as shown in Figure 1. From ECG, the model computes the mean, median, 20<sup>th</sup>, and 80<sup>th</sup> percentiles of heart rate, variance, and quartile deviation of HRV and energy of HRV in different frequency bands (0.1–0.2Hz, 0.2–0.3Hz, 0.3–0.4Hz). From the respiration signal, it computes mean, median, 80<sup>th</sup> percentile, and quartile deviation from inhalation (I), exhalation (E) duration, ratio between I/E, stretch, and inspiration volume, computed in each breath cycle within a minute. In cross-subject validation using SVM on lab data, the cStress model classified stress and non-stress minutes with an F1 score of 0.81 in ( $n = 21$ ) participants who were subjected to three validated stressors — public speaking, mental arithmetic, and cold-pressor tasks. When tested on a dataset from another group of participants ( $n = 26$ ) subjected to the same lab

stress protocol, the model was able to classify stress and non-stress minutes with an F1 score of 0.9. The model was also evaluated against self-reports collected in the field. In the first study of ( $n = 20$ ) healthy adults who provided 1,060 self-reports in a 7-day study, the model reported an F1-score of 0.71 for the median participant. On a second field study with ( $n = 38$ ) polydrug users who wore the sensors for four weeks, the model reported a median F1 score of 0.72 [45]. In a third field study of ( $n = 53$ ) newly-abstinent smokers who wore the sensors for 4 days, the model reported a median F1 score of 0.65 [44].

The cStress model only provides a stress likelihood for each minute, which does not indicate the start and end time of a stress event. To obtain stress events from the noisy and largely discontinuous (due to missing data or confounding from physical activity) time series of stress likelihoods, we apply the stress event detection model presented in [45]. This model first generates stress likelihood in minute-windows using the cStress model, but sliding every 5 seconds, to reduce the noise in the stress likelihood time series. Second, it excludes any data when participant may be recovering from physical activity (after accelerometer signals show no activity). Third, it uses  $k$ -nearest neighbor approach to impute any missing values of stress likelihood that is ‘missing at random’. Fourth, it applies a moving average convergence divergence (MACD) method to find the cross over points that partition the continuous stress likelihood time-series into stress events, clearly marking the start and end times, as shown in Figure 1. Fifth, it excludes any windows that have more than 50% of stress likelihoods imputed. Finally, it applies a density threshold (to the area under the stress likelihood curve) to decide which windows are stressful events. In the field-collected data, between 2 and 4 stress events per day were detected [45].

#### 4.4 Context Inferences and Visualization for Stressor Labeling

To aid the participants in recalling the stressor for each detected stress event, we detected several cues such as location from GPS, conversation from respiration, and activity from accelerometers. This information surrounding each stress event was presented to the participants so they could reconstruct those moments to confirm or refute the detection of these stress events and to recall the stressor responsible for that stress event. We first describe how we process the sensor data to obtain the surrounding contexts and then present the visualization.

##### 4.4.1 Inferring Significant Locations Using Historical Map-Based

**Visualization:** Location is an important memory cue. When it is annotated with a time range, this information can help users to reconstruct their day and facilitate self-reflection [46]. Locations of interest are places where a user spends a significant amount of time. We adopted the spatio-temporal clustering algorithm from [31] to infer significant locations, arrival time, departure time, duration of stay, and sequence and frequency of location visits throughout the day, all from GPS traces. A distance threshold of 100 meters and a time threshold of 10 minutes were used to find the spatio-temporal clusters.

We utilized a map-based visualization technique (as shown in Figure 2) developed in [54] to observe the location clusters on Google Earth. Labeling of the location clusters was semi-automated. The two most common location clusters, *home* and *work*, were automatically



labeled based on the address provided by the participants at the beginning of the study. To label the remaining location clusters, the participants were asked to provide the semantic labels during the data review session. This helped resolve ambiguities for co-located places (e.g., grocery store and a restaurant). Distinct semantic locations thus obtained included: *own home*, *parent's home*, *others home*, *work*, *restaurant*, *store*, *grocery*, *religious place* (e.g., church, mosque), and *recreation center* (e.g., gymnasium).

**4.4.2 Inferring Commute:** Driving episodes are detected from GPS-derived speed by applying a threshold for maximum gait speed of 2.533 meters/second [8]. A driving session is composed of driving segments separated by stops, e.g., due to a traffic light. The in-between stops usually are of short duration unless there is traffic congestion. The end of a driving session is defined as a stop (i.e., zero speed) for more than two minutes. Driving segments, separated by less than two-minute stop, are considered to be part of the same driving episode [56].

**4.4.3 Inferring Physical Activity:** For activity inference, we use the on-body accelerometer based activity detection approach presented in [41]. The pre-processing steps include filtering of raw data and removal of gravitational acceleration and drift from the filtered data. Finally, we compute the standard deviation of the magnitude of acceleration ( $a_{mag} = \sqrt{a_x^2 + a_y^2 + a_z^2}$ ), which is independent of the orientation of the accelerometers.

**4.4.4 Inferring Conversation Episodes:** For detecting conversations from respiration data, we used the method proposed in [40]. This model extracts features in respiration cycles in each 30 second window, trains a machine learning model to produce speaking, listening, or quiet states, and then applies a Hidden Markov Model (HMM) to construct the conversation status for each 30 seconds window of respiration data. It achieves 87% accuracy in distinguishing conversation from non-conversation.

**4.4.5 Contextualized Timeline Visualization to Assist in the Recall of Stressors.**—We developed a contextualized timeline visualization by building upon stress visualizations presented in [48]. In order to help the participants reconstruct the moments surrounding a stress event, we made several adaptations in the visualization, guided by the *Day Reconstruction Method (DRM)* [24].

We incorporated three design qualities for effective health data representation [52]. (1) the design must feel familiar to users, mirroring their own experience, (2) creating designs that leave space for users' own interpretation of their bodily data, and (3) the modalities used in the design do not contradict one another, but instead harmonize, helping users to make sense of the representation.

We created a stacked timeline visualization shown in Figure 3 for each user. We used horizontal and vertical placement along with color coding as our visual encoding channel as these channels are most effective in supporting the comparison of multiple data streams [33]. In the timeline, the horizontal axis shows the time of day, and vertical axes is divided into four channels that represent location, conversation, activity, and stress likelihood. We use hue as the color component to code different levels of stress — green represents no stress,

yellow stands for medium, and red indicates high levels of stress likelihood (based on perceived stress categories reported in [32]). Deeper shades of color for conversation and activity time series show the occurrence of conversation and physical movement, respectively, and grey color indicates the absence of conversation or absence of movement. Significant locations are marked with corresponding labels. If a transition between locations takes place using a motorized vehicle, then the transition is labeled as commuting. For all the four data streams, the presence of a gap implies missing data for that time period. Aligning all data streams using the same timeline facilitates understanding of the role of different contexts such as location or conversation on stress events.

It is difficult to pinpoint a stressful event when the data is on the scale of several hours (e.g., over 12 hours of data was collected per day). Therefore, we provided users the ability to zoom in and out at different temporal resolutions. By providing details-on-demand, we allowed users to view precise stress likelihood levels and associated contexts (e.g., location, conversation, and physical activity status). To help them in recalling a specific event, we used tool-tip texts displayed at the time of occurrence of each event.

#### 4.5 Participant Selection and Protocol To Capture Real-Life Stress Events

We recruited couples to wear sensors and collect data concurrently to maximize the coverage of stressful conversations. The field study included 38 individuals (19 pairs of cohabiting couples). Field study participants included 20 women (mean age:  $28.53 \pm 4.89$  years) and 18 men (mean age:  $28.92 \pm 2.10$  years). Eighteen participants were Caucasian and the rest were Asian. Twenty participants (10 pairs) participated during weekdays and the rest participated during weekends.

The field study consisted of three phases — (1) an enrollment session, (2) free-living data collection, and (3) a data review session to label detected stress events using the visualization. During the enrollment session, participants gave consent and completed a demographic questionnaire, a dyadic adjustment scale [50], and a pre-study questionnaire. Participants were shown an example visualization generated from previously collected sample data. This was designed to help them understand how the field data collected would help them understand their own stress patterns and identify daily stressors for potential stress management in daily life. This orientation was also designed to motivate the participants for careful data collection when they were in free-living condition.

Afterward, participants were shown how to wear the sensors and monitor the status of sensor data collection. They then proceeded to collect sensor data in the field. After completing at least 24 hours with the sensors since the start of the data collection, both partners came back to the lab next day to review stress visualizations generated from their own data and annotate the automatically detected stress events captured in the field.

Because the field study involved collection of continuous audio, location, and physiological data from the participants, they were given an option to pause data collection during their private moments. They could proactively pause data collection using the “Stop” button in the smartphone software during data collection in the field. Also, they were given the option to retroactively delete data during private moments during the data review session. The data

collection was limited to 24 hours to reduce privacy concerns associated with the raw recording of audio data in the natural environment; participants were instructed to get verbal consent from conversation partner(s) other than their romantic partner before recording audio conversation involving them. If any partner(s) declined the request, participants were instructed to stop recording the audio.

#### 4.6 Stressor Labels Collected and Confirmed

The participants were asked to confirm each stress event in the visualization of their data. This was done to resolve any ambiguity in stress event detection due to the usage of machine learning models from sensor data, including the elimination of any false detection. To further confirm and contextualize the stress events, several follow up questions were asked such as “*what’s going on?*”, “*where were you?*”, “*who were you with?*”.

Participants were asked to rate the usability of the visualization interface on a 5-point Likert scale. All the participants either strongly agreed (32 out of 38) or agreed (6 out of 38) when asked if the interface was “*Easy to understand*”. Nineteen participants strongly agreed (out of 38) and 14 agreed (out of 38) when asked if “*they thought that most people would learn to use the visualization quickly*”. When asked an open ended question: “*What things did you Like and Dislike in the study*”, 20 participants (out of 27 who responded to this question) mentioned that they liked the stress visualization system. For example, C4F commented, “[I] Liked visualization of the day, disliked wearing all the sensors”.

After reviewing the visualization, participants were able to recall 125 (out of 137) detected stressful events. Sensor data was fully available and not confounded by physical activity for the 97 confirmed events that were used for our modeling (see Table 1). For 12 events, they either disagreed with the visualization output or could not confirm whether the stress event occurred. In addition, we asked all the participants whether they recalled any stress event that happened during the study that was not identified by the system (false negative). Two participants (out of 38) reported three such false negative events (over 38 person days of data collection). The stress events missed by the sensors were not included in our model training or testing as the start and end times of these events could not be determined precisely because they were missed by the sensors.

Participants recalled several reasons for stress events (i.e., stressors) such as meeting with a supervisor, having deadlines at work, job interviews, conflict with their partner, driving on a busy road, assignment deadlines, etc. We find that the detected and confirmed stress events belong to three major categories — stressful conversations, commute, and work-related stress. Table 1 shows the number of stress events in each category with their average duration and what was happening at the time. In our data set, we find that 53 stressful events (i.e., 54% of all confirmed stressful events) were due to conversations with partner, friends, parents, colleagues, supervisors, etc.

We resolved any ambiguity in the start and end times of stressful conversations by listening to the raw audio before, during, and after each event. We find that each stress event attributed to stressful conversations were correctly labeled. It may be because of our contextualized visualization that showed the participants whether they were having

conversations at the time of a detected stress event and where they were, e.g., at home or office.

We also found 30 stressful events during commute and 14 events due to work. Any stress event that involved a conversation whether at home, work, or anywhere else, is included in the category of stressful conversation. The same would be the case for work-related stressor, unless it involved a conversation, in which case it belongs to the stressful conversation category. We note that the percentage of stress events in each category matches with that reported in [4]. The distribution of stress events in our dataset in these three categories is shown in Figure 4.

Out of 125 confirmed stress events, 28 were only partially observed by sensors (due to missing data and overlapping physical activity) and hence their start and end times could not be precisely determined. Therefore, they were excluded from our modeling. These events included several stressors that did not belong to the above three categories. They included household chores (8), stress during shopping or grocery (5), and miscellaneous (15) stress events that included feeling sick, a sick family member, worrying about the partner, water leak in the house, cleaning the house, etc.

#### 4.7. Lab Study to Collect Confounder-free Data

We designed a lab study to collect clean data during stressful conversations that could be used to find any distinguishing patterns in the stress signals. The lab tasks were designed to create difficult communication situations and thus induce interpersonal conflicts. We recruited 12 individuals (6 pairs of cohabiting couples) from students and employees (both full-time and part-time) at a university. Participants included 7 women (mean age:  $29.9 \pm 7.4$  years) and 5 men (mean age:  $27.2 \pm 2.9$  years). They wore the AutoSense chest band [16] to collect ECG and respiration signals and wore a headset microphone and a throat microphone to capture audio. Each couple took part in several interaction tasks in a sitting position with limited or no movement. To produce baseline measures, participants remained seated face-to-face in a comfortable chair silently for five minutes. Next, they took part in a ‘Scripted Dialogue’ task and then recreated a map [5] to elicit goal-oriented conversation, and finally, they engaged in spontaneous conversation for approximately 15 minutes. During Map Tasks [5], both participants were given maps that have been used in prior literature, one presenting a pre-printed route with a starting and finishing point for the Instruction Giver and the other presenting a map with only a starting point for the Instruction Receiver. The Instruction Receiver attempted to recreate the Instruction Giver’s pre-printed route based on verbal directions from the Instruction Giver. In the maps, several mismatches in the route between the two partner’s map were intentionally included to induce conflict between them. A (blocking) screen was placed between them for visual separation. They then switched roles and were given another set of maps to generate another conversation to complete the Map Task 2. After that, participants took part in a five minute debriefing conversation, as the nature of the map tasks tended to induce some informational conflict between partners that they needed to resolve during this session.

## 5 MODEL DEVELOPMENT

In this section, we describe how we extract distinguishing patterns from the stress event along with the wrist motion sensor data to detect stressful conversations. First, we describe our proposed method to identify cyclical pattern in a stress event followed by wrist motion patterns. Second, we describe feature extraction from the stress and wrist motion data to train a machine learning model for detecting stressful conversations. Finally, we evaluate our models and discuss implications of the models.

### 5.1 Key Ideas and Overall Approach

Input to the model is a continuous stress likelihood time-series that is annotated with the start and end times of stress events. The goal of the model is to determine which stress events are due to stressful conversations.

**5.1.1 Key Ideas.**—Our model development is based on three key ideas. First, we notice that stress time-series signal during stress events is episodic and often periodic, exhibiting peaks and troughs that can be used to naturally segment the stress data. Second, we identify several novel features from these cycles. Third, we observe that the pattern of hand gestures when stress occurs due to stressful conversations is distinct in nature, as compared to when stress is due to work or commute. With the increasing adoption of smartwatches and fitness trackers, it is increasingly feasible to capture hand movement patterns continuously. We also note that with recent improvements in optical sensing in smartwatches, stress may also be detected from smartwatches [10, 36], making for a ubiquitous device on which our model can be implemented.

**5.1.2 Overall Approach.**—Our model development consists of the following major steps.

- 1. Cyclical Pattern Identification:** Cyclical patterns in stress events are different than those in regular physiological signals such as respiration cycles. Respiration cycle is well defined by inhalation and exhalation phases associated with each breath, but the cyclical pattern in stress events does not have any such naturally defined phases. They are generated by the interplay of the SNS/PNS system. Therefore, existing methods for detecting peaks and troughs are not directly applicable to stress event cyclical pattern identification. We propose a new method to detect cycles in the stress likelihood timeseries and characterize portions of interest from which distinguishable features can be computed.
- 2. Intra-cycle Feature Extraction:** Unlike respiration, there is no natural phenomenon of inspiratory and expiratory time. Therefore, we discover new features that can characterize and interpret each stress cycle.
- 3. Inter-cycle Feature Extraction:** To capture any patterns that span multiple stress cycles within a stress event, potentially covering all stress cycles within a stressful event, we compute features spanning multiple stress cycles.
- 4. Wrist Motion Features:** Wrist motion sensors data have been researched extensively for activity and posture detection. We compute these features within

each stress cycle to determine their utility in capturing the distinct signatures of hand gestures observed during stress events, to improve the accuracy of detecting stressful conversations.

## 5.2 Observation and Characterization of Cyclical Patterns in Stress Likelihoods within Stress Events

As described in Section 2, we expect the physiological response during a stress event to exhibit a cyclical pattern. To investigate whether we observe a cyclical pattern during stressful conversations, we analyzed the physiological data collected during the lab study (see Section 4.7), where stressful conversations took place and the physiological data was mostly free of any confounders. As described in Section 4.3, we apply the cStress model on physiological data to convert the physiological sensor data into stress likelihoods (in sliding minute-windows, starting every 5 seconds) as shown in Figure 1. We also mark the start and end of stress events.

We observe that the cyclical patterns previously observed in the physiological response (see Section 2) during stress tasks (due to the interplay between SNS and PNS) is also observed in the stress likelihood time series within a stress event. The activation of SNS results in the elevation of physiological arousal which is captured by an increase in the stress likelihood produced by the cStress model. We define this point as stress ‘Rising point’ where stress arousal starts to elevate from its pre-stress condition, i.e., an average of daily stress likelihood as shown in Figure 1. Concurrently, each time SNS activates, the PNS gets activated as well to provide the corresponding counterbalance so as to keep the physiology in homeostasis balance. When the influence of PNS exceeds that of SNS, then it reaches a ‘Saturation point’, after which the stress arousal starts to decay, indicated by the ‘Decay point’ when the effect of stressor starts to mitigate. Finally, it reaches the pre-stress value or below the daily average of stress likelihood denoted by the ‘Recovery point’. We define this structure as a ‘stress cycle’, where stress cycle begins at a ‘Rising point’ and ends at a ‘Recovery point’.

The cycle repeats if the current episode continues to produce new stress triggers (e.g., conflicting words spoken by the conversation partner). A stress event may consist of one or more stress cycles depending on the repetition of stress triggers within a stress event. In Figure 1, the depicted stress event consists of three stress cycles.

We illustrate the cyclical patterns in the stress likelihood time-series data during lab tasks in Figure 5a. It shows that stress likelihood was low during the baseline session. Stress likelihood rises during the scripted dialogue task as the individual was waiting for his/her turns, and they were focusing on their performance to make the dialogues look more natural. As the nature of the map tasks tended to induce some informational conflict between partners, we see high arousal stress cycles during Map Task 2 and during the debrief session when they were trying to resolve their conflict. Stress arousal in Map Task 1 is not as visible due to missing data.

We observe a similar cyclical pattern during stress events in the field data. Figure 5b depicts the stress arousal of a participant in the field during two separate conversational interactions



at two different times. The first interaction (left portion) was a non-stressful conversation, where stress likelihood remains below the daily average. The second interaction (right portion) presents a stressful conversation, where we observe several stress cycles that rise above the daily average of stress likelihood. This particular stressful conversation consists of five stress cycles. We next describe how we identify stress cycles automatically from the stress time-series data.

### 5.3 Stress Cycle Identification Algorithm

We propose a moving average-based method to identify each stress cycle with all four interesting points — stress rising, saturation, decay, and recovery point. We build upon the cycle identification model used to detect physiological phenomena such as breathing cycles [6]. Breathing signal follows some specific structure with inspiration and expiration phases driven by the physiological phenomenon. However, the stress cycle is guided mostly by the stressful situation and may not have any specific rules. Hence, the method developed for breathing cycle identification is not directly applicable to stress cycle identification. Therefore, we modify the algorithm to identify stress saturation and decay point.

First, we smooth the stress likelihood time-series using a 15 seconds moving average to remove spikes. Then another moving average centerline (MAC) curve is computed using a moving average of 2 minutes. The MAC appears as a center line (shown as a red dotted line in Figure 6a) that intercepts each stress cycle twice, once in the rising trend and then in the falling trend. Next, we identify the up and down intercepts where the MAC curve intercepts the rising and falling branch of smoothed stress time-series, respectively. The ‘rising point’ is the rightmost local minimum that lies below daily average found between consecutive down and up intercept pair. From this point, the signal rises monotonically towards saturation point.

The ‘saturation point’ lies between the up intercept and the following peak of that cycle where the rising trend reaches the peak. This point is the leftmost local maximum and must be above up intercept and MAC curve line.

‘Decay point’ lies between the saturation point and the following down intercept when the signal starts monotonically decreasing. This point is detected as the rightmost local maximum and must lie above the following down intercept and the MAC line. The falling trend reaches a recovery point when it decreases to the first local minimum value below the daily average of stress likelihood.

We annotated 160 stress cycles from several stress events, including stressful conversation, work, and commute related stress. We use the following metrics to evaluate the performance of the algorithm — the percentage of actual cycle detected, percentage of extra or spurious cycle found, and error in cycle duration due to mis-located rising and/or recovery point. Two coders independently labeled all the interesting points of a cycle. Inter-rater reliability was around 0.9 between the coders. The algorithm identified 96% cycles accurately and detected 3% cycles as extra or spurious. The mean absolute error in identifying cycle start or rising point is 8.86 seconds. The mean absolute error in identifying recovery point is 6.9 seconds. Therefore, mean error in cycle duration is 8.16 seconds. The rationale for calculating error in

cycle duration is that even if a rising or recovery point is identified correctly, their respective temporal position in the signal may introduce error in the resultant duration.

After applying this algorithm on all the annotated stress events, we find the average number of stress cycles per stress event are 4.42, 3.6, and 2.9 for stressful conversation, work, and commute, respectively, as depicted in Figure 6b. Number of cycles per stress event for stressful conversation is significantly higher compared to both work and commute related stress at 5% significance level (using  $t$ -test). But, no significance difference is found between work and commute related number of stress cycles per stress event. Average stress cycle duration is 3.7, 4.8, and 4.02 minutes for stressful conversation, work and commute, respectively depicted in Figure 6b (right portion). The cycle duration for stressful conversation is significantly lower compared to work related cycle duration with a  $p$ -value of 0.002 (using  $t$ -test).

#### 5.4 Distinguishing Patterns in Wrist Motion Sensors

Researchers have studied the role of gestures during conversational interaction in assessing stress. The more stressful the situation, the higher the proportion of speech that is accompanied by hand gestures [29]. We observe similar distinct patterns in the wrist-worn motion sensor signals (accelerometer and gyroscope) during stressful conversations compared to other stressful events such as work and commute. We observe that the frequency of wrist movement is higher during stressful conversations. While someone is working at a computer, the motion will be more guided towards typing or mouse movement. Similarly, hand motion during driving is expected to be dominated by the steering wheel movement. On the other hand, wrist motion is more random during an interaction, possibly due to communicative gesturing. Based on these insights, we extracted motion sensor features under each stress cycle to compare those differences to detect stressful conversations in daily life.

#### 5.5 Feature Computation

To capture differences in stress cycle characteristics during stressful conversations compared to other daily stressors, we identify new features from each stress cycle. From each cycle, we compute features from stress likelihood time-series and those from wrist-worn inertial sensors. In addition to computing features from individual cycles, we also compute features from two or three consecutive cycles, and all cycles in a stress event.

**5.5.1 Features from Individual Stress Cycle.**—We compute the following features from each stress cycle of a stress event: fractional rising and fractional falling time, rising and falling normalized area, ratio of rising and falling normalized area, elevation above daily average, rising and falling slopes and intercepts, skewness, kurtosis, and entropy. We now describe these features and how they are computed.

To compute these features, we first extract the following duration measurements from each stress cycle — stress cycle duration, saturation duration, and successive cycle distance as depicted in Figure 7a, where we show two stress cycles. Let  $S_j$  be the stress likelihood at time  $t_j$  with new values produced every  $\Delta t = t_j - t_{j-1} = 5$  seconds. A stress cycle is defined

by four 2D points, i.e.,  $C_i = (r_i, s_i, d_i, r_i^c)$  (as shown in 7a). Here,  
 $r_i = \langle t_{r_i}, S_{r_i} \rangle$ ,  $s_i = \langle t_{s_i}, S_{s_i} \rangle$ ,  $d_i = \langle t_{d_i}, S_{d_i} \rangle$ , and  $r_i^c = \langle t_{r_i^c}, S_{r_i^c} \rangle$ .

**Stress cycle duration ( $CD_i$ ):** Stress cycle duration is defined as the temporal distance between stress rising and recovery point, i.e.,  $CD_i = t_{r_i^c} - t_{r_i}$ .

**Saturation duration ( $SD_i$ ):** Saturation duration is the duration when the stress likelihood time-series stays in the upper region after reaching the saturation point before starting to decay, i.e.,  $SD_i = t_{d_i} - t_{s_i}$ .

**Successive cycle distance ( $SCD_i$ ):** Successive cycle distance is the distance between ending of one cycle and starting of next cycle, i.e.,  $SCD_i = t_{r_{i+1}} - t_{r_i^c}$ .

With these duration measurements, we compute the following features from each stress cycle.

1. *Fractional rising and falling time:* Fractional rising time is defined as the ratio of rising duration to stress cycle duration where rising duration is defined as the temporal distance between stress cycle start and saturation point. Similarly, fractional falling time is defined as the ratio of falling duration to stress cycle duration. Falling duration is the temporal distance between decay and recovery points. More specifically,  $t_{rise_i} = (t_{s_i} - t_{r_i})/CD_i$  and  $t_{fall_i} = (t_{r_i^c} - t_{d_i})/CD_i$ .

2. *Rising and falling normalized area, Ratio of rising and falling normalized area:* Rising normalized area is computed as the area under rising region divided by the rising duration. Similarly, falling normalized area is computed as the area under falling region divided by the falling duration. We also use the ratio of these two values as a feature. The variation of this feature values are depicted in Figure

$$7b \text{ for different stressors, } A_{rise_i} = \frac{\sum_{k=r_i}^{s_i} S_k}{t_{rise_i}} \text{ and } A_{fall_i} = \frac{\sum_{k=d_i}^{r_i^c} S_k}{t_{fall_i}}.$$

3. *Elevation above daily average ( $E_i$ ):* The amplitude difference between maximum value or the peak of a cycle and the daily average is defined as elevation above daily average. Peak amplitude of a cycle  $C_i$  is  $peakAmp_i = \max(S_{r_i}, S_{r_{i+1}}, \dots, S_{r_i^c})$ . Then, elevation above daily average is  $E_i = (peakAmp_i - Avg(S_j, \forall j))$ .
4. *Rising and Falling slopes and intercepts:* We fit a least square regression line in the rising phase. That is, we find slope and intercept  $c$  of equation  $y = mx + c$  using the sequence of points  $t_k, S_k$  between  $t_{r_i}$  and  $t_{s_i}$ . Similarly, falling slope and intercept are computed in the decay region. The variation of intercept values are depicted in Figure 7c for different stressors.
5. We also compute skewness, kurtosis, entropy for each stress cycle. Since, a stress cycle is defined by four points, i.e.,  $C_i = (r_i, s_i, d_i, r_i^c)$  therefore all the stress

likelihood within the cycle  $S_{r_i}, S_{r_i+1}, \dots, S_{r_i^c}$ . More specifically, skewness is

$$\frac{\sum_{i=r_i}^{r_i^c} (S_i - \bar{S})^3}{\|r_i^c - r_i\| * std^3} \text{ and kurtosis is } \frac{\sum_{i=r_i}^{r_i^c} (S_i - \bar{S})^4}{\|r_i^c - r_i\| * std^4}.$$

**5.5.2 Wrist Motion Features in Each Stress Cycle.**—From inertial sensor data coinciding with each stress cycle, we compute several time domain features from both accelerometer and gyroscope signals — mean, median, standard deviation, quartile deviation, skewness, and kurtosis of three axes of accelerometer and gyroscope. For wrist orientation features, we compute roll, pitch, and yaw that provide information about the orientation of the wrist with respect to gravity on the window of data. We also computed energy as the magnitude of the accelerometer and magnitude of the gyroscope, i.e.,

$$a_{mag} = \sqrt{a_x^2 + a_y^2 + a_z^2} \text{ and } g_{mag} = \sqrt{g_x^2 + g_y^2 + g_z^2}.$$

**5.5.3 Whole Stress Event Features.**—To compute features from the entire stress event, we compute — number of stress cycles per event, duration of stress cycles per minute, and average stress likelihood across the entire event.

**5.5.4 Features from Multiple Stress Cycles.**—We compute features from consecutive stress cycles (i.e., two cycles or three cycles) to determine the degree of performance improvement with more information. We note that using features from the entire event may delay the detection of stressor until after the stress event is over. The combination features include differential features from successive individual stress cycle features and statistical features such as mean and standard deviation across selected number of cycles. For wrist motion features, we compute only statistical features across selected number of cycles.

## 5.6 Model Selection and Training

We group the stress events into two categories — stressful conversation and non-stressful conversation. The non-stressful conversation group includes all the stress events due to other common daily stressors, i.e., work and commute. Our aim is to identify whether a stress cycle is induced due to stressful conversation related activity. To do so, we identify each stress cycle automatically from the continuous stress time-series using the previously mentioned stress cycle algorithm. Next, we compute the features from each stress cycle and train a machine learning model to identify whether the current stress event is due to interpersonal interactions.

In total, we obtain 13 features from each stress cycle and 42 features from the motion sensor data. To avoid over-fitting, we use selected features for modeling. The idea behind feature selection is to remove highly correlated and non-informative features. We use the Correlation-based Feature Selection (CFS) [45] to select a subset of 15 features. CFS selects features that are mutually uncorrelated but highly indicative of the interaction and non-interaction classes.

As the trained model to detect stressful conversations should be suitable for real-time implementation on smartwatches that already include sensors for both stress inference and wrist motion, we narrow down our choice of models to simple and efficient models. Therefore, we evaluate Logistic Regression (LR) for a linear model and Random Forest (RF) for a non-linear model to determine how much improvement in performance we can obtain by using a non-linear model. We assess the performance of the model using standard classification metrics such as precision, recall, and F1 score. We use labeled stress cycles to train the models.

## 6 PERFORMANCE EVALUATION

In this section, we compare the model performance when using stress cycle features only in each cycle, performance improvement when adding features from wrist-worn inertial sensors, and further performance gain when using features from the entire stress event. We then compare our model's performance with a baseline model that is based on [27] and analyze the trade-off between accuracy and how many cycles to use for detection.

### 6.1 Performance with Individual Stress Cycle Features Only

We consider a stress cycle as the smallest unit for detecting an on-going stressful event. The Logistic Regression (LR) classifier can distinguish whether a stress cycle belongs to stressful conversation class with an F1 score of 0.77 using stratified 10-fold cross-validation method using stress cycle features from one cycle (see Figure 8a). Precision and recall values are 0.65 and 0.95, respectively. On the other hand, Random Forest (RF) classifier achieves precision, recall, and F1-score of 0.66, 0.85, and 0.74, respectively.

### 6.2 Performance Improvement from Using Wrist Motion Features

After fusing wrist motion features with individual stress cycle features, Logistic regression (LR) can classify with 0.75 precision, 0.89 recall, and 0.82 F1-score. The Random Forest (RF) model can classify the two classes with precision, recall, and F1 score of 0.78, 0.92, 0.85, respectively. This shows that adding computationally inexpensive and less power-hungry motion sensors, which are already part of wrist devices, can significantly improve the accuracy of detecting stressful conversation, especially with a random forest model.

### 6.3 Performance Improvement Using Stress Event Features

When we augment whole stress event features with the individual stress cycle features, the precision improves to 0.83, recall improves to 0.97, and F1-score becomes 0.89 using Random Forest (RF) model as shown in Figure 8b. The stress event features provide a significant improvement in classification performance. But, to achieve this performance gain, we need to observe the whole duration of the stressful event. We get fewer false alarms but at the cost of longer waiting times for interventions.

### 6.4 Comparison with a Baseline Model

To put the performance of our model in perspective, we construct a baseline model. We consider a natural model that compares the percentage of stress event duration that is detected to be spent in conversation by user from another data source motivated by the work

presented in [27]. To detect the start and end times of a conversation, we use an audio-based model available from LENA [1] and separately a respiration-based model from [6]. We find overlap duration between these two events that result in optimal performance (see Figures 9b and 9c).

We observe an F1-score of 0.51 for the audio-based model with around 32% overlap with the stress event and an F1 score of 0.60 for the respiration based model with 58% overlap with the stress event. Lower F1-scores for these two baseline models can be explained by the fact that models detecting conversations are not perfect (F1-score 0.7). Secondly, people usually multitask and therefore, even when a user may be in conversation during a stress event, (s)he might be stressed for other reasons. For example, a driver may be in conversation with a co-passenger during driving, but stressor can be traffic-related.

In addition to 15–30% performance improvement over baseline models, our model also has the advantage of detecting stressful conversation from the stress time series itself, without needing concurrent detection of potential stressors (e.g., conversation from audio, work status from computer logs, etc.).

## 6.5 Trade-offs for Delivery of Just-In-Time Stress Intervention

A just-in-time intervention needs information on most opportune moments for delivering the intervention. In this section, we investigate the trade-off between the accuracy of detecting the stressful conversation and how quickly since the start of the stress event, the stressful conversation can be detected.

The model performance is expected to improve when features are computed over longer intervals, but it also comes at the cost of delayed detection. We also observe that when we use features spanning multiple cycles (i.e., two cycles, three cycles), the number of instances for classification reduces and the model may tend to overfit. The total number of instances while taking features from one cycle, two cycles and three cycles are 346, 258 and 183, respectively. After computing features from three cycles, the dataset becomes too small to test any further combination of featureset. This issue does not arise when using whole event features as the unit of analysis is still each cycle within the stress event.

The F1-score using features from one cycle, two cycles, three cycles and whole stress event based featureset is 0.74, 0.78, 0.84, and 0.89, respectively. Fusing the wrist motion features with stress cycle features increases the F1-score to 0.83, 0.86, 0.88 and 0.89 for one cycle, two cycles, three cycles, and whole stress event based feature set, respectively. Figure 10b shows these results.

Stress intervention designers can consider the trade-offs between the timing of intervention and the accuracy of detecting stressful conversations based on one or multiple stress cycles. For example, if a quicker intervention is called for, then they can consider intervening after one cycle, which will allow them to intervene within 3.9 minutes (on average) from the stress rising point with an F1-score of 0.83, when the stress cycle features are used with wrist motion features (see Figure 10b). To achieve higher accuracy, one can use the model that fuses two cycles together. In that case, the F1 score improves to 0.86, but the timing of



the delivery will be further delayed (on average, 9.3 minutes from the stress rising point shown in Figure 10a). We note that the best accuracy is achieved when using whole event features, but this will further delay the detection of stressor, to an average of 19 minutes. These analyses can help intervention designers find a suitable operating point.

## 7 LIMITATIONS AND FUTURE WORKS

There are several limitations to this work that can inspire future research.

First, this work used stress event detection from chest-worn ECG and respiration sensors. These sensors provide a firmer attachment than pulse plethysmography or electrodermal sensing from conveniently worn wrist devices. Wearing electrodes or a chest belt in the field for long term to monitor stress is burdensome and sometimes interferes with daily activities. Therefore, it is a challenge to deploy such systems in the field. However, smartwatches are becoming increasingly popular, and recent research work shows the feasibility of detecting stress from wrist-worn sensors. Future work can assess how well the presented model can be adapted to work with potentially noisier stress time series obtained from wrist-worn sensor data and test its real-time usage.

Second, this study used data from 38 participants, but the data was collected only for one full day (due to privacy concerns with audio capture in the natural environment). Future work can investigate the generalizability of the presented model on data collected in longer-term studies and those involving a more general population.

Third, the limited size of dataset (in terms of number and diversity of detected stress events) in this work was insufficient to develop and test a three class classifier to distinguish interaction, work, and commute. In future, a larger dataset can enable identification of other stressors as well as support the construction of data-driven features in a deep learning model.

Fourth, in addition to stressful conversations, work, and commute, there are numerous other sources of stress such as financial difficulties, health issues, news about friends, family, colleagues, region, country, and the world, among others. Future work can investigate the possibilities of detecting these and other stressors, by potentially exploring novel methods to combine the data collected from other sources with the stress dynamics data.

Fifth, the labeling of stress events was done based on participants interviews. As we asked the participants to recall what was causing the stress after showing the detected stress events using the visualization, this may have introduced some bias. To better assess recall and detect false positives in stress event detection, a future study can present the surrounding contexts and time segments (both when stress is detected and not detected) without disclosing whether stress event was detected at those times. Another way to reduce bias may be to first let the participants recall major periods of stress and then show them the visualizations to verify the stress events.

Sixth, this work assigned each detected stress event to a single stressor. Some real-life situations involve multitasking, where a stress event may be due to the confluence of

multiple concurrent factors. Future work can investigate methods for detection of multiple concurrent stressors.

## 8 CONCLUSION

This work introduced the concept of stress cycles and presented an algorithm to identify them in a stress likelihood timeseries and characterize points of interest in them. It further showed that features derived from stress cycles have sufficiently distinct patterns to distinguish stressful conversations from other stressors (with improved accuracy when combined with features derived from hand gestures). This work opens the doors to future research that can collect larger datasets consisting of a large number of other daily stressors and develop models to identify each of them. Such models can be used to determine various sources of stress for each stress event detected by wearable sensors. This information can not only inform the timing of intervention delivery, but also the right content, the adaptation mechanisms for personalizing it to the individual and the user's context, and selecting the right modality for delivery (e.g., smartphone or smartwatch).

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for significantly improving the organization and presentation of this manuscript. We sincerely thank Dr. Moushumi Sharmin from Western Washington University for assisting with the study protocol and the design of visualization for study participants. Research reported here was supported by the National Institutes of Health (NIH) under awards P41EB28242, R01CA224537, R01MD010362, R01CA190329, R24EB025845, U01CA229437, and U54EB020404 (by NIBIB) through funds provided by the trans-NIH Big Data-to-Knowledge (BD2K) initiative, and by the National Science Foundation (NSF) under awards IIS-1722646, ACI-1640813, CNS-1823221, and OAC-2019085.

## REFERENCES

- [1]. 2019. LENA Research Foundation. <http://www.lenafoundation.org/>. (Accessed: February 2019).
- [2]. al' Absi Mustafa, Bongard Stephan, Buchanan Tony, Pincomb Gwendolyn A, Licinio Julio, and Lovallo William R. 1997. Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors. *Psychophysiology* 34, 3 (1997), 266–275. [PubMed: 9175441]
- [3]. Almeida David Mand Kessler Ronald C. 1998. Everyday stressors and gender differences in daily distress. *Journal of personality and social psychology* 75, 3 (1998), 670. [PubMed: 9781406]
- [4]. Almeida David M, Wethington Elaine, and Kessler Ronald C. 2002. The daily inventory of stressful events: An interview-based approach for measuring daily stressors. *Assessment* 9, 1 (2002), 41–55. [PubMed: 11911234]
- [5]. Anderson Anne H, Bader Miles, Bard Ellen Gurman, Boyle Elizabeth, Doherty Gwyneth, Garrod Simon, Isard Stephen, Kowtko Jacqueline, McAllister Jan, Miller Jim, et al. 1991. The HCRC map task corpus. *Journal of Language and speech* (1991).
- [6]. Bari Rummana, Adams Roy J, Rahman Md Mahbubur, Parsons Megan Battles, Buder Eugene H, and Kumar Santosh. 2018. rConverse: Moment by Moment Conversation Detection Using a Mobile Respiration Sensor. *ACM IMWUT* 2, 1 (2018), 2. [PubMed: 30417165]
- [7]. Birjandtalab Javad, Cogan Diana, Pouyan Maziyar Baran, and Nourani Mehrdad. 2016. A non-EEG biosignals dataset for assessment and visualization of neurological status. In *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*. 110–114.
- [8]. Bohannon Richard W. 1997. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age and Aging* (1997), 15–19.
- [9]. Bone Daniel, Mertens Julia, Zane Emily, Lee Sungbok, Narayanan Shrikanth S, and Grossman Ruth B. 2017. Acoustic-Prosodic and Physiological Response to Stressful Interactions in Children with Autism Spectrum Disorder.. In *INTERSPEECH*. 147–151.

- [10]. Said Can Yekta, Chalabianloo Niaz, Ekiz Deniz, and Ersoy Cem. 2019. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors* 19, 8 (2019), 1849.
- [11]. Choi Jongyoon, Ahmed Beena, and Gutierrez-Osuna Ricardo. 2011. Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE transactions on information technology in biomedicine* 16, 2 (2011), 279–286. [PubMed: 21965215]
- [12]. Coleman Lester, Mitcheson Jan, and Lloyd Gareth. 2013. Couple relationships: Why are they important for health and wellbeing? *Journal of Health Visiting* 1, 3 (2013), 168–172.
- [13]. de Geus Eco J, Van Doornen Lorenz J, and Orlebeke Jacob F. 1993. Regular exercise and aerobic fitness in relation to psychological make-up and physiological stress reactivity. *Psychosomatic medicine* 55, 4 (1993), 347–363. [PubMed: 8416085]
- [14]. Dissing Agnete Skovlund, Jørgensen Tobias Bornakke, Gerds Thomas Alexander, Rod Naja Hulvej, and Lund Rikke. 2019. High perceived stress and social interaction behaviour among young adults. A study based on objective measures of face-to-face and smartphone interactions. *PloS one* 14, 7 (2019), e0218429. [PubMed: 31348787]
- [15]. Ejupi Andreas and Menon Carlo. 2018. Detection of talking in respiratory signals: A feasibility study using machine learning and wearable textile-based sensors. *Sensors* 18, 8 (2018), 2474.
- [16]. E Ertin N Stohs S Kumar, Raj A, al' Absi M, Kwon T, Mitra S, Shah S, and Jeong J. 2011. AutoSense: Unobtrusively Wearable Sensor Suite for Inferencing of Onset, Causality, and Consequences of Stress in the Field. In *ACM SenSys*.
- [17]. Fuligni Andrew J, Telzer Eva H, Bower Julienne, Cole Steve W, Kiang Lisa, and Irwin Michael R. 2009. A preliminary study of daily interpersonal stress and C-reactive protein levels among adolescents from Latin American and European backgrounds. *Psychosomatic Medicine* (2009).
- [18]. Gujral Aditya, Chaspari Theodora, Timmons Adela C, Kim Yehsong, Barrett Sarah, and Margolin Gayla. 2018. Population-specific Detection of Couples' Interpersonal Conflict using Multi-task Learning. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM.
- [19]. Healey Jennifer A and Picard Rosalind W. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* (2005), 156–166.
- [20]. Hovsepian Karen, al' Absi Mustafa, Ertin Emre, Kamarck Thomas, Nakajima Motohiro, and Kumar Santosh. 2015. cStress: towards a gold standard for continuous stress assessment in the mobile environment. In *ACM UbiComp*.
- [21]. Ishii Ryo, Ren Xutong, Muszynski Michal, and Morency Louis-Philippe. 2020. Can Prediction of Turn-management Willingness Improve Turn-changing Modeling?. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.
- [22]. Jungbluth Chelsy, MacFarlane Ian M, Veach Patricia McCarthy, and LeRoy Bonnie S. 2011. Why is everyone so anxious?: an exploration of stress and anxiety in genetic counseling graduate students. *Journal of Genetic Counseling* 20, 3 (2011), 270–286. [PubMed: 21264500]
- [23]. Juster Robert-Paul, McEwen Bruce S, and Lupien Sonia J. 2010. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews* 35, 1 (2010), 2–16. [PubMed: 19822172]
- [24]. Kahneman Daniel, Krueger Alan B, Schkade David A, Schwarz Norbert, and Stone Arthur A. 2004. A survey method for characterizing daily life experience: The day reconstruction method. *Science* (2004), 1776–1780.
- [25]. Kiecolt-Glaser Janice K, Gouin Jean-Philippe, and Hantsoo Liisa. 2010. Close relationships, inflammation, and health. *Neuroscience & Biobehavioral Reviews* 35, 1 (2010), 33–38. [PubMed: 19751761]
- [26]. Lamichhane Bishal, Großekathöfer Ulf, Schiavone Giuseppina, and Casale Pierluigi. 2017. Towards stress detection in real-life scenarios using wearable sensors: normalization factor to reduce variability in stress physiology. In *eHealth 360°*. 259–270.
- [27]. Lee Munhee, Moon Junhyung, Cheon Dongmi, Lee Juneil, and Lee Kyoungwoo. 2020. Respiration signal based two layer stress recognition across non-verbal and verbal situations. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. 638–645.

- [28]. Lee Youngki, Min Chulhong, Hwang Chanyou, Lee Jaeung, Hwang Inseok, Ju Younghyun, Yoo Chungkuk, Moon Miri, Lee Uichin, and Song Junehwa. 2013. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. 375–388.
- [29]. Lefter Iulia, Burghouts Gertjan J, and Rothkrantz Léon JM 2015. Recognizing stress using semantics and modulation of speech and gestures. *IEEE Transactions on Affective Computing* 7, 2 (2015), 162–175.
- [30]. Lu Hong, Fraudentorfer Denise, Rabbi Mashfiqui, Mast Marianne Schmid, Chittaranjan Gokul T, Campbell Andrew T, Gatica-Perez Daniel, and Choudhury Tanzeem. 2012. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 351–360.
- [31]. Montoliu Raul, Blom Jan, and Gatica-Perez Daniel. 2013. Discovering places of interest in everyday life from smartphone data. *Springer Multimedia Tools and Applications* (2013), 179–207.
- [32]. Muaremi Amir, Arnrich Bert, and Tröster Gerhard. 2013. Towards measuring stress with smartphones and wearable devices during workday and sleep. *BioNanoScience* (2013).
- [33]. Munzner Tamara. 2014. *Visualization analysis and design*. AK Peters/CRC Press.
- [34]. Myrtek Michael and Brünger Georg. 1996. Perception of emotions in everyday life: studies with patients and normals. *Biological psychology* 42, 1–2 (1996), 147–164. [PubMed: 8770376]
- [35]. Natarajan Annamalai, Ganesan Deepak, and Marlin Benjamin M. 2019. Hierarchical Active Learning for Model Personalization in the Presence of Label Scarcity. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 1–4.
- [36]. Nath Rajdeep Kumar, Thapliyal Himanshu, and Caban-Holt Allison. 2020. Validating Physiological Stress Detection Model Using Cortisol as Stress Bio Marker. In *2020 IEEE International Conference on Consumer Electronics (ICCE)*. 1–5.
- [37]. Plarre Kurt, Raij Andrew, Hossain Syed Monowar, Ali Amin Ahsan, Nakajima Motohiro, al’absi Mustafa, Ertin Emre, Kamarck Thomas, Kumar Santosh, Scott Marcia, et al. 2011. Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *ACM IPSN*.
- [38]. Powers Sally I, Pietromonaco Paula R, Gunlicks Meredith, and Sayer Aline. 2006. Dating couples’ attachment styles and patterns of cortisol reactivity and recovery in response to a relationship conflict. *Journal of personality and social psychology* 90, 4 (2006), 613. [PubMed: 16649858]
- [39]. Rahman Md Mahbubur, Ali Amin Ahsan, Plarre Kurt, Al’Absi Mustafa, Ertin Emre, and Kumar Santosh. 2011. mconverse: Inferring conversation episodes from respiratory measurements collected in the field. In *Proceedings of the 2nd Conference on Wireless Health*. 1–10.
- [40]. Rahman Md. Mahbubur, Ali Amin Ahsan, Plarre Kurt, al’Absi Mustafa, Ertin Emre, and Kumar Santosh. 2011. mConverse: Inferring Conversation Episodes from Respiratory Measurements Collected in the Field. In *ACM Wireless Health*.
- [41]. Rahman Md Mahbubur, Bari Rummana, Ali Amin Ahsan, Sharmin Moushumi, Raij Andrew, Hovsepian Karen, Hossain Syed Monowar, Ertin Emre, Kennedy Ashley, Epstein David H, and Others. 2014. Are we there yet?: feasibility of continuous stress assessment via wireless physiological sensors. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 479–488.
- [42]. Raja G Senthil and Dandapat S. 2010. Speaker recognition under stressed condition. *International Journal of Speech Technology* 13, 3 (2010), 141–161.
- [43]. Ramos-Garcia Raul I, Tiffany Stephen, and Sazonov Edward. 2016. Using respiratory signals for the recognition of human activities. In *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 173–176.
- [44]. Sarker Hillol, Hovsepian Karen, Chatterjee Soujanya, Nahum-Shani Inbal, Murphy Susan A, Spring Bonnie, Ertin Emre, Al’Absi Mustafa, Nakajima Motohiro, and Kumar Santosh. 2017. From markers to interventions: The case of just-in-time stress intervention. In *Mobile health*. 411–433.

- [45]. Sarker Hillol, Tyburski Matthew, Rahman Md Mahbubur, Hovsepian Karen, Sharmin Moushumi, Epstein David H, Preston Kenzie L, Furr-Holden C Debra, Milam Adam, Nahum-Shani Inbal, et al. 2016. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. In Proceedings of the 2016 CHI conference on human factors in computing systems. 4489–4501.
- [46]. Sas Corina, Challioner Scott, Clarke Christopher, Wilson Ross, Coman Alina, Clinch Sarah, Harding Mike, and Davies Nigel. 2015. Self-Defining Memory Cues: Creative Expression and Emotional Meaning. In ACM CHI Extended Abstract.
- [47]. Schmidt Philip, Reiss Attila, Duerichen Robert, and Van Laerhoven Kristof. 2018. Wearable affect and stress recognition: A review. arXiv preprint arXiv:1811.08854 (2018).
- [48]. Sharmin Moushumi, Raij Andrew, Epstien David, Nahum-Shani, Beck J Gayle, Vhaduri Sudip, Preston Kenzie, and Kumar Santosh. 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. 505–516.
- [49]. Snyder Douglas K and Halford W Kim. 2012. Evidence-based couple therapy: Current status and future directions. *Journal of Family Therapy* 34, 3 (2012), 229–249.
- [50]. Spanier Graham B. 1976. Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family* (1976).
- [51]. Stadler Gertraud, Snyder Kenzie A, Horn Andrea B, Shrouf Patrick E, and Bolger Niall P. 2012. Close relationships and health in daily life: A review and empirical data on intimacy and somatic symptoms. *Psychosomatic Medicine* (2012).
- [52]. Ståhl Anna, Höök Kristina, and Kosmack-Vaara Elsa. 2011. Reflecting on the Design Process of Affective Health. In IASDR.
- [53]. Taelman Joachim, Vandeput Steven, Vlemincx Elke, Spaepen Arthur, and Van Huffel Sabine. 2011. Instantaneous changes in heart rate regulation due to mental load in simulated office work. *European journal of applied physiology* 111, 7 (2011), 1497–1505. [PubMed: 21188414]
- [54]. Tang Karen P, Hong Jason I, and Siewiorek Daniel P. 2011. Understanding how visual representations of location feeds affect end-user privacy concerns. In ACM UbiComp.
- [55]. Timmons Adela C, Chaspari Theodora, Han Sohyun C, Perrone Laura, Narayanan Shrikanth S, and Margolin Gayla. 2017. Using multimodal wearable technology to detect conflict among couples. *Computer* 3 (2017), 50–59.
- [56]. Vhaduri Sudip, Ali Amin, Sharmin Moushumi, Hovsepian Karen, and Kumar Santosh. 2014. Estimating drivers' stress from GPS traces. In Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. 1–8.
- [57]. Wilson Stephanie J, Bailey Brittney E, Jaremka Lisa M, Fagundes Christopher P, Andridge Rebecca, Malarkey William B, Gates Kathleen M, and Kiecolt-Glaser Janice K. 2018. When couples' hearts beat together: Synchrony in heart rate variability during conflict predicts heightened inflammation throughout the day. *Psychoneuroendocrinology* 93 (2018), 107–116. [PubMed: 29709758]
- [58]. Wyatt Danny, Choudhury Tanzeem, and Kautz Henry. 2007. Capturing spontaneous conversation and social dynamics: A privacy-sensitive data collection effort. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Vol. 4. IEEE, IV–213.
- [59]. Xu Chenren, Li Sugang, Liu Gang, Zhang Yanyong, Miluzzo Emiliano, Chen Yih-Farn, Li Jun, and Firner Bernhard. 2013. Crowd++ unsupervised speaker count with smartphones. In Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. 43–52.

**CCS Concepts:**

- **Human-centered Computing** → Ubiquitous and Mobile Computing;
- **Information Systems** → *Data Mining*,



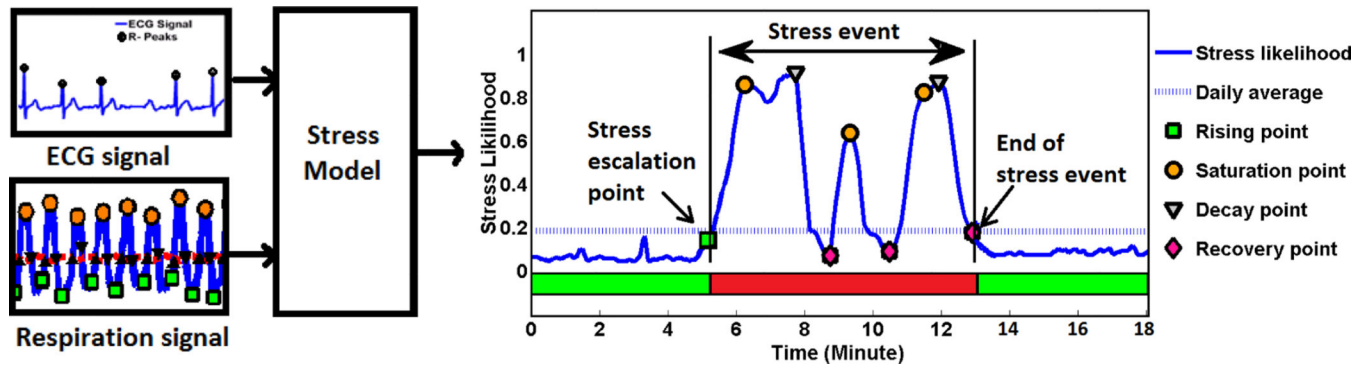


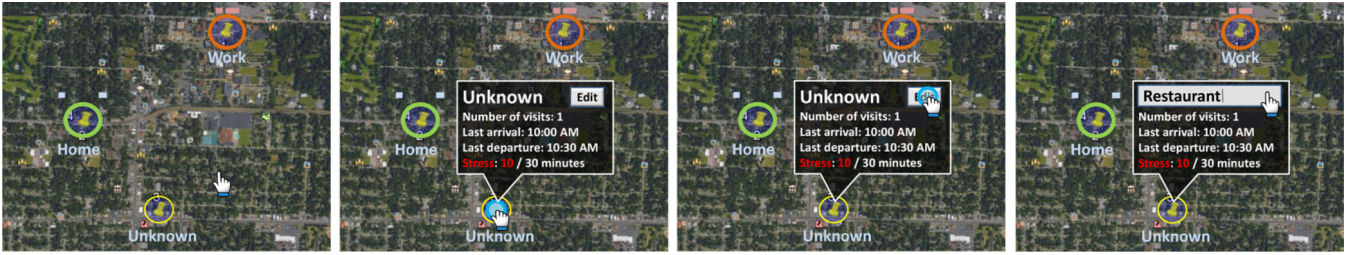
Fig. 1. Inferences of continuous stress likelihood using ECG and respiration signal.

Author Manuscript

Author Manuscript

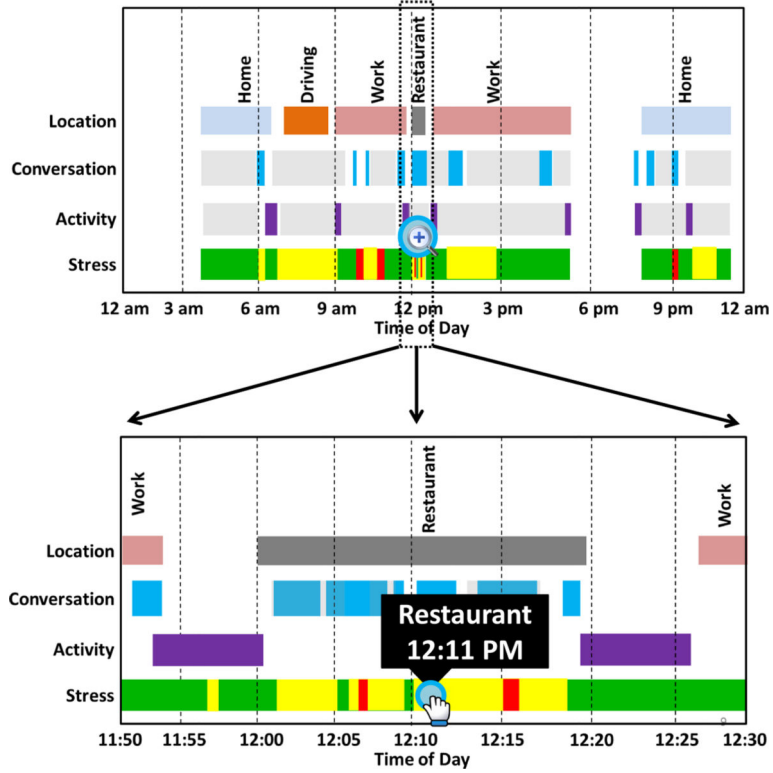
Author Manuscript

Author Manuscript



**Fig. 2.**

The circles represent significant locations visited by a user in a day. At a given location, the thickness of the circle corresponds to the duration of time spent and its color indicates the intensity of the average stress. Significant places can be labeled by the user. Clicking on a pushpin displays the frequency of visit to the location, start and end times of the last visit, and the duration of time that the user was stressed at that location. Users can edit and relabel the unknown locations, as shown in the picture.



**Fig. 3.** Stress timeline visualization consists of four channels of inferences. Significant locations are marked with corresponding semantic location labels (e.g., *Home*). Dark color represents the presence of conversation (blue) and activity (purple) and grey color implies its absence. The bar display for Stress has three colors (Green = No stress, Yellow = Medium stress, and Red = High stress). Gaps in any of the channel indicate unavailable data. The interface has zoom-in (e.g., restaurant is zoomed-in in the lower figure) and zoom-out plus info-tip features (shown in black box with exact time in the lower zoomed-in part) to precisely pinpoint each stress events and corresponding contexts.

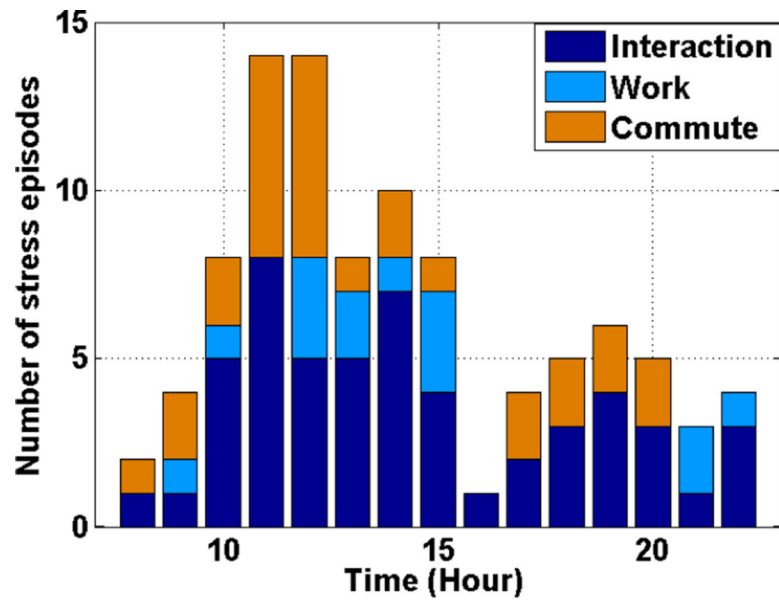
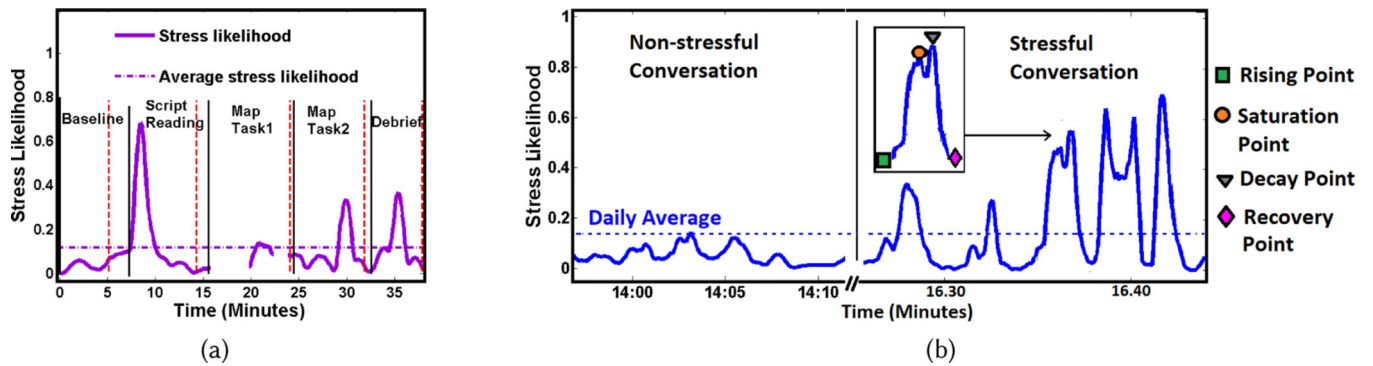
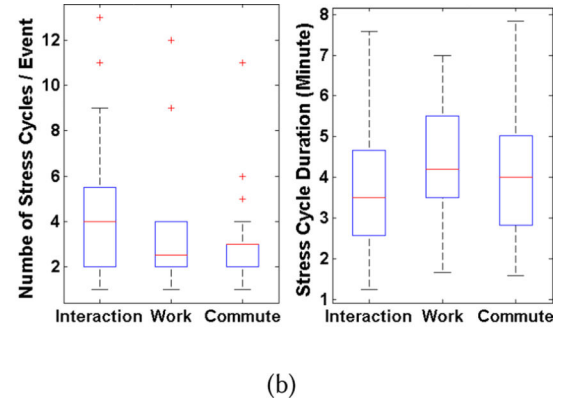
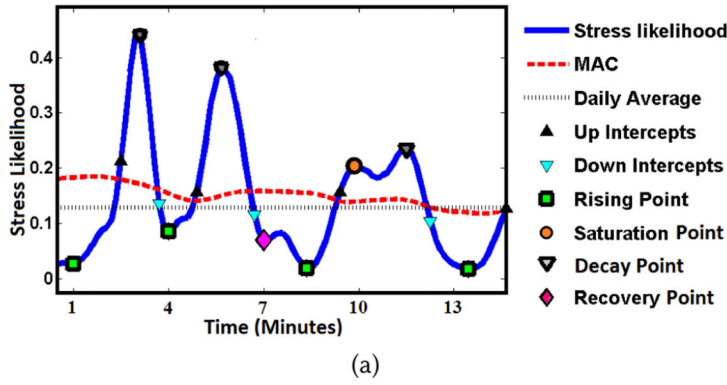


Fig. 4.  
Distribution of stress events throughout the whole day.



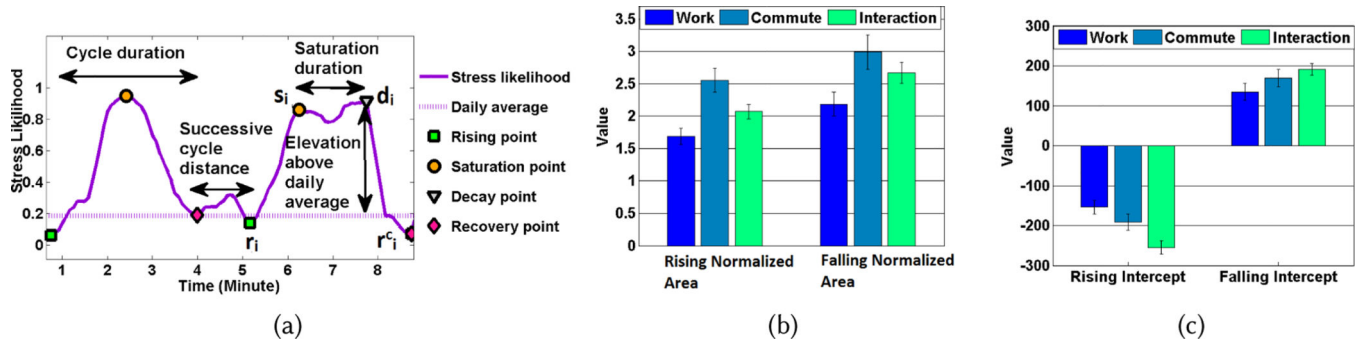
**Fig. 5.**

(a) Cyclical pattern of stress likelihood observed in lab data. The vertical solid black and dotted red lines depict start and end times of each tasks, respectively. Stress cycles are visible during scripted reading, Map Task2 and debrief session. In between 15 and 25 minutes, we observe a portion of missing data. (b) Variation in stress likelihood pattern during stressful and non-stressful conversations in field. The horizontal blue dotted line shows the daily average of stress likelihood. During stressful conversation, we observe several high arousal stress cycles.

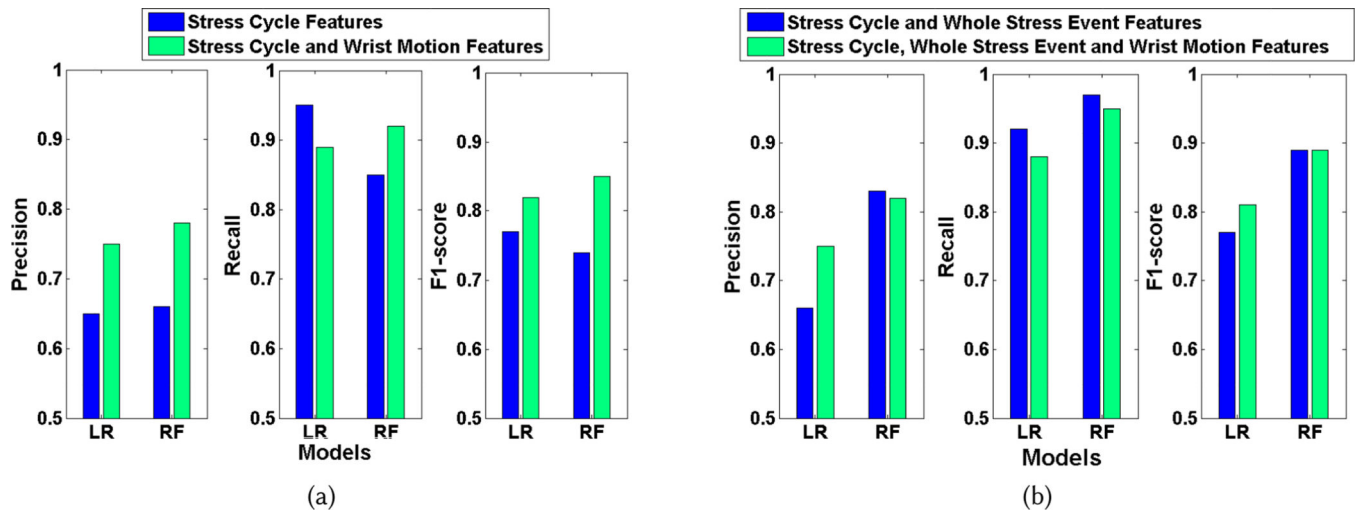


**Fig. 6.** (a) Moving average based cycle identification method to identify all four interesting points- (1) stress rising point; (2) stress saturation point; (3) stress decay point; and (4) stress recovery points are detected. (b) Summary of detected stress cycles for stressful conversation, work and commute related stress. Median numbers of stress cycles/event are 4, 2.5 and 3 for stressful conversation, work and commute, respectively, where number of cycles/event is significantly higher for stressful conversation related stress compared to other two stressors. Median stress cycle duration's are 3.5, 4.2 and 4.0 minutes for stressful conversation, work and commute related stress, respectively. Significant difference is found between stressful conversation and work related stress cycle duration.



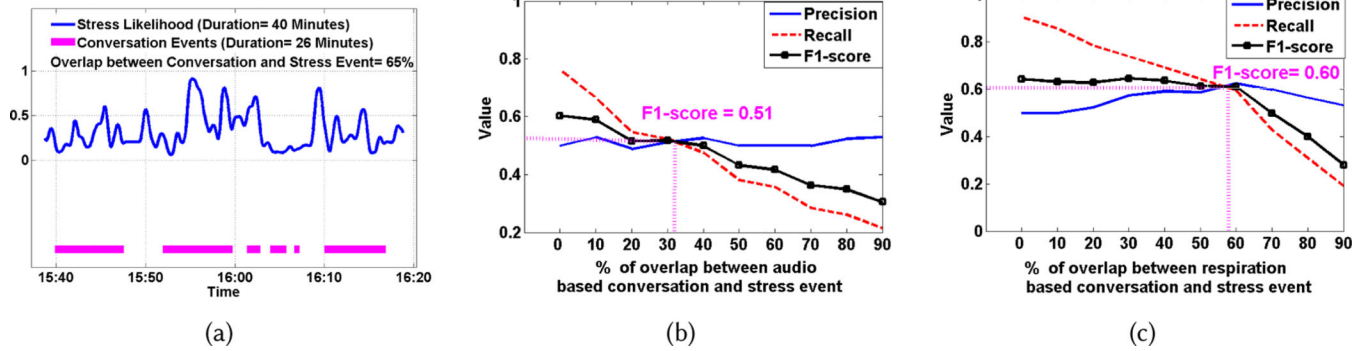


**Fig. 7.** (a) Several duration values from individual stress cycle. (b) Rising and falling normalized area. (c) Rising and falling intercepts of stress cycles.

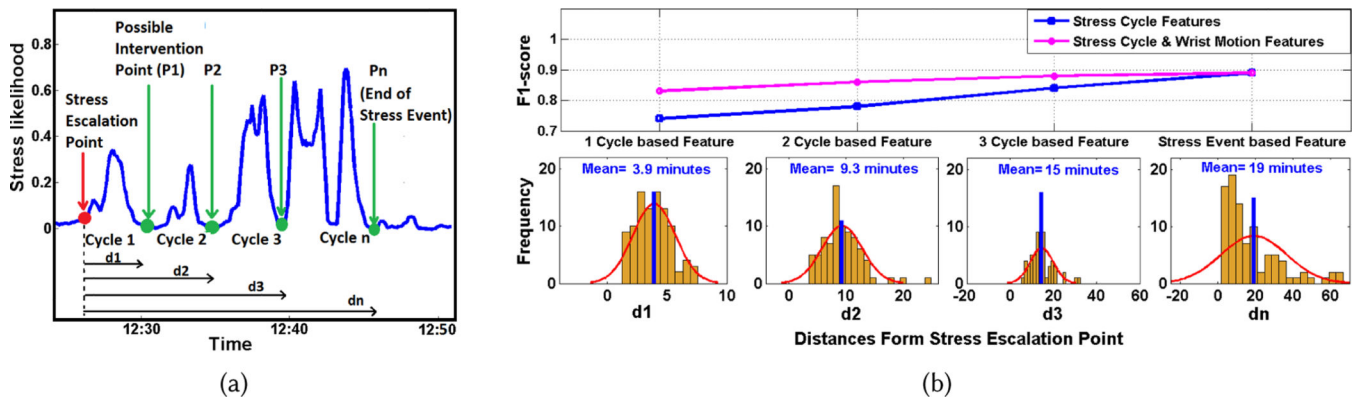


**Fig. 8.**

(a) Performance of LR (Logistic Regression) and RF (Random Forest) models using individual stress cycle features and then fusing wrist motion features. (b) Performance using whole stress event features fusing with individual stress cycle features and wrist motion features.

**Fig. 9.**

(a) Whether a stress event is due to stressful conversation depending on the overlap ratio between detected conversation and stress events. Here conversation is measured from the LENA audio device. (b) If conversation is measured using respiration based model, then the accuracy of finding whether a stress event is due to interaction.



**Fig. 10.**

(a) Possible timing/points ( $P_1$ ,  $P_2$ ,  $P_3$  and  $P_n$ ) of delivering intervention depending on features computed from one stress cycle or span of multiple stress cycles.  $d_1$ ,  $d_2$ ,  $d_3$  and  $d_n$  are representing the temporal distances from stress escalation point or beginning of a stress event. (b) Trade-off between stressful conversation detection accuracy and timing of stress intervention delivery based on features computed from one stress cycle or features from multiple cycles.

**Table 1.**

Stressors observed in participant's daily life for which sensor data is fully available and usable for modeling.

Stressors	Number of stressful events	Average event duration (Minute)	What was going on during stressful events
Stressful Conversations	53	22.68 (3.83)	Conversations with partner, friends, colleagues, supervisor
Commute	30	12.74 (2.28)	Time pressure, other driver's behavior, construction on road
Work	14	18.23 (3.54)	Deadline, answering work related email/text

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript