Intelligent Interruption Management using Electro Dermal Activity based Physiological Sensor for Collaborative Sensemaking

NITESH GOYAL, Cornell University
SUSAN R. FUSSELL, Cornell University

Sensemaking tasks are difficult to accomplish with limited time and attentional resources because analysts are faced with a constant stream of new information. While this information is often important, the timing of the interruptions may detract from analyst's work. In an ideal world, there would be no interruptions. But that is not the case in real world sensemaking tasks. So, in this study, we explore the value of timing interruptions based on an analyst's state of arousal as detected by Electrodermal activity derived form galvanic skin response (EDA). In a laboratory study, we compared performance when interruptions were timed to occur during increasing arousal, decreasing arousal, at random intervals or not at all. Analysts performed significantly better when interruptions occurred during periods of increasing arousal than when they were random. Further, analysts rated process component of team experience significantly higher also during periods of increasing arousal than when they were random. Self-reported workload was not impacted by interruptions timing. We discuss how system designs could leverage inexpensive off-the-shelf wrist sensors to improve interruption timing.

CCS Concepts: • **Human-centered computing** → Human computer interaction (HCI); Collaborative and social computing; • **Computing methodologies** → Cognitive science; • **Hardware** → Sensor applications and deployments;

Additional Key Words and Phrases: Analytics, Sensemaking, Collaborative Sensemaking, EDA, Galvanic Skin Response, interruption, notification, interface design

ACM Reference format:
Nitesh Goyal and Susan R. Fussell. 2017. Intelligent Interruption Management using Electro Dermal Activity based Physiological Sensor for Collaborative Sensemaking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 52 (September 2017), 21 pages.
DOI: http://doi.org/10.1145/3130917

1 INTRODUCTION

In 1991 dentist Colin Hoswell murdered his wife, and his love interest's partner (co-conspirator Hazel Stewart who was also a police officer). It was not until 2010, 19 years later, that he was caught due to investigation mistakes. First, Howell (the murderer himself) made multiple attempts to give geolocation clues by directing the police to the garage where his victims were eventually found. The police officials ignored his clues as they continued on their analysis. Second, the police ignored information shared by multiple eyewitnesses when they challenged his testimonies. Finally, the investigation team also ignored multiple clues shared by other police officers about holes in their analysis and assumptions. In complex sensemaking tasks like crime-solving, it is hard to identify if a clue

This work is supported by the National Science Foundation, under grant IIS-0968450.
Authors' addresses: Dept. of Information Science, Gates Hall, Cornell University, Ithaca, NY 14850, US..

2474-9567/2017/9-ART52 $15.00
DOI: http://doi.org/10.1145/3130917

should be included in the analysis, owing to multiple factors like content, source, and timing etc. However, these factors randomly emerge and cannot be controlled. Ignoring information may act as a mechanism to pursue the task efficiently, without increasing the workload or negatively impacting the task performance. In retrospect, an intelligent interruption management mechanism that did not affect cognitive workload might have enabled analysts to appropriately include them in their analysis. Further, this tension between interruption and pertinence of information is challenging in dynamic sensemaking tasks where pertinence evolves with time. While pertinence to existing mental models has been explored using computational techniques like Natural Language Processing, Information Retrieval, and Machine Learning (when data is available), managing interruption to aid sensemaking remains unexplored.

As is evident, sensemaking is a complex task, especially when performed under strenuous conditions of limited time. In time-critical and dynamic situations, like medicine, crime solving, or crisis-response, sensemaking has been shown to be even harder [16, 41] because now analysts must not only make sense of their own data but also be receptive (or even perceptive) to emerging information from the automated systems or incoming data and insights from their partners; receive and parse the incoming information to aid knowledge development; update their own mental models by reconciling incoming and existing information; and pursue the foraging and sensemaking loop iteratively till the best possible solution is found.

In the current work, we focus on managing interruptions as notifications during a sensemaking task, similar to other challenging contexts like control rooms and cockpits [6, 14, 15, 33, 37, 39, 40, 50]. Previous work has shown that interrupting with notifications randomly can decrease performance[6, 13], yet it has remained unclear when to show the notifications to not affect performance negatively? Some researchers have actively pursued this question, and discovered that interruptions provided at the intersection of task boundaries, effectively manipulating the temporal dimension, can benefit by reducing mental workload [6, 14]. However, authors point that this approach requires developing, modeling and validating the task execution structures manually, prior to manipulating interruptions at task breakpoints [5].

In this work, we present empirical research on using physiological data to identify windows for interruption while participants performed an analytical task, shown to be significantly complex in previous studies [8, 21–25, 54]. In a carefully controlled experiment, users were interrupted using a mobile notification with information that would be useful, but not sufficient, to solve a complex crime-solving task. Prior research has shown that using physiological data from EDA (Electrodermal Activity derived form Galvanic Skin Response) can help identify arousal states [3, 42, 51]. Across four conditions, the arousal states were monitored using a wrist-worn EDA sensor, and interruption was introduced either randomly, when EDA arousal values increased significantly, decreased significantly, or no interruption was given. Results point to significant increase in task-performance and significant increase in reported collaboration experience when information was presented during arousal acceleration, compared to random interruption. In this work we show that designers of sensemaking tools may benefit from actively managing interruption notifications using off-the-shelf wrist worn EDA sensors.

In the remainder of the paper, we first discuss related work on interruptions in sensemaking, tools available to support interruptions in sensemaking, use of sensors to manage interruptions, and use of EDA specifically. We next outline our study research questions, and the design of the experiment used to answer those questions. We support our results with both quantitative and qualitative findings. Finally, we conclude with a discussion of the contributions of this work to the design of intelligent interruption management systems that can use cheap wrist-worn off-the shelf EDA sensors for higher task performance, and team experience.

Fig. 1. Off-the-shelf Affectiva Inc. Q Sensor, with electrodes to measure EDA at the back

2 RELATED WORK

2.1 Interruptions in Sensemaking

Process of sensemaking as described by Pirolli and Card [48] is an iterative process where analysts would iteratively forage for clues, identify pertinent pieces of information, and generate mental models that best depict the present state of knowledge. Domains like crime-solving are highly dynamic since collaborators try to share emerging information from analysis and field-research, in real-time [19]. Further, leveraging partner's cognition and insights is necessary to solve such complex problems [20, 21, 28, 53]. However, this iterative process is susceptible to an analyst's ability to accept new or emerging information, as it needs to be reconciled with the existing mental model and hypothesis.

This tension between sharing insights as early as possible for inclusion into mental model, and analyst's cognitive availability to parse such information when under data deluge is a legitimate challenge [1]. As pointed out previously [34], such difficult data analytic contexts lead to teammate inaccuracy blindness. This means that, owing to data deluge, analysts trust shared information and are too busy to identify data quality. Alternatively, they may choose to ignore this information until they can process it [43]. Consequently, interrupting when analysts are unable to attend and test the veracity of the notifications sets them up for failures, as in the case of Hoswell and Stewart introduced earlier.

In general, prior work has shown that random interruptions can have significantly negative impact on task completion time [13, 14], task performance [36], decision-making [49], and affective state [2, 6, 55]. Consequently, researchers have pursued multiple ways to overcome these negative impacts. First, one may schedule interruptions at breakpoints between tasks [2, 4, 31] by modeling the tasks. Alternatively, one may create statistical models that detect such breakpoints based on user activity [5, 17, 30, 31]. Finally, researchers have also attempted to identify the scheduling using correlation between interruptible moments and physiological data, such as pupil-dilation

[5], or studying how multiple sensor inputs can be used together to interrupt non-relevant distractions in a software engineering tasks [56].

As is evident, this interruption notification has been an active area of research. However, the past approaches have several limitations. First, modeling tasks requires extensive human effort and is only limited to very controlled non complex tasks [2, 4, 31]. Second, creating statistical models that detect breakpoints have shown to decrease frustration and reaction time [32] but have no impact on task performance.

Finally, using extensive equipment like pupil dilation measurement, or a multiple customized set of sensors is detrimental to ease of use, requires high maintenance and significant technical knowledge. This paper differentiates its contribution in multiple ways. First, we study the impact of scheduling interruptions that are relevant to the task at hand and hence require attention. Second, unlike previous research task contexts like programming, diagramming, document editing etc., our work focuses on interruptions in a sensemaking task that has been shown to be robust and complex. Third, our goal has been to leverage minimally distracting off the shelf wrist worn sensors (Affectiva inc. Q sensor as shown in fig. 1[1]) requiring relatively little technical knowledge to construct new hardware sensors.

2.2 Sensemaking Tool Support for Interruptions

Prior research in sensemaking tools has shown that explicit notifications through chats or comment threads[29] or annotations [35] are necessary to build a shared mental model, when collaborating. The tools designed to enable real-time notification vary from notification of each activity by the partner [10] to shared workspaces [18, 26] where other's activities in the task are listed out [12, 47]. Instead of explicitly notifying activities, tools may also remind analysts to learn about their partner's analysis [43]. Alternatively, tools have been designed to enable implicit notifications that are shared automatically [21]. Such tools have improved user experience but limited gains in task performance. Further, these works have focused on the design of tool, and have not studied the impact of notification timing as an interrupt on sensemaking, as we propose.

Perhaps one of the more recent tools aimed at interruption scheduling is OASIS, a system that defers notifications until breakpoints are reached to reduce interruption costs [32]. The tool identifies the level of breakpoint needed for a notification, and then accordingly interrupts when a notification arrives. While this tool has been shown to work with six users on non-sensemaking tasks, it requires continual user-activity monitoring, via software installation at workplace, which has privacy implications. Further, the tool also required extensive training thought automated and manual coding. In our work, we present use of EDA sensors, without significant training as a mechanism to support interruption scheduling for sensemaking tasks.

2.3 Interruption Management using Sensor

Sensors are increasingly being explored as avenues to identify interruptibility. Prior work has focused on multiple sensors like electroencephalographs (EEG), eye tracking systems, electrocardiogram (ECG) blood volume pulse (BVP), body temperature sensors, and electrodermal activity (EDA) or Galvanic Skin Response (GSR). We present a brief overview of how these sensors have been used for interruption scheduling and consequently focus on EDA/GSR, the sensor we used to inform our results.

For example, heart variability and muscular activity have been used as indicator for interruptibility and the message notification levels were adjusted accordingly [11]. Past work has also shown that EEG data can be used to retrospectively classify interruptibility successfully during US military training [38]. As discussed previously, eye-tracking can be used to identify context-switching and consequently identify opportune moments of interruptions [5].

[1]As received on August 20th, 2017 from http://www.prweb.com/releases/2010/11/prweb4745334.htm

GSR or EDA sensor is used to measure the electrical skin conductivity that varies due to sweat production in skin. It is calculated by applying a small current on the skin surface using two electrodes and is a relative measure. Previous work has shown correlation between EDA and arousal, attention, emotional states, stress and anxiety [9]. In problem solving tasks involving mathematics, a strong link has been discovered between EDA and cognitive load [44].

Recently, using a Naìve Bayes classifier, researchers have discovered that in a software programming context, using multi-sensor setup: EEG, EDA, and Temperature sensor together, task-irrelevant interruptions to a programing tasks could be successfully classified at 91.5% accuracy [56]. However, as researchers pointed out that EEG bands might be too obtrusive for long-term, and a better solution is needed. Further researchers have shown that analyzing a smartphone data trace to identify context, including their activity, location, time of day, emotions and engagement, can help time interruptions that increase user satisfaction and shorter response times [46]. However, this has significant privacy implications owing to access to multi-dimensional personal information.

2.4 EDA and Affective State

As discussed, GSR or EDA sensor when used in a multi-sensor setup enables successful classification [56]. However, the researchers used the raw values of Mean Phasic Peak Amplitude (MPA), Sum Phasic Peak Amplitude (SPA), and Phasic Peak Frequency (PPF) in a Naìve classifier in this work to identify two levels of affective state: interruptible or not-interruptible. Since EDA measures a composite value for anxiety, arousal, anxiousness, and emotional state, previous work has ignored the impact of significant increase (acceleration) or decrease (deceleration) of the MPA, and focused on raw values. In this work, we focus on the former and conjecture that a significant change in direction of the values is as important as the raw value itself and when used together, might offer previously ignored benefits. For example, higher MPA values than baseline might suggest a change in affective state, but a significant acceleration in a short time-window suggests the intensity of the change. Hence, in this work we study the impact of this intensity of change, as opposed to the change itself. In the next sections, we discuss how our works compares to the most recent relevant research.

We find that our work differentiates from Iqbal and Bailey [32] because the authors used models on relatively simpler non-sensemaking tasks to identify the task-breakpoints when cognitive workload would be potentially lower and introduced interruptions then. On the contrary, our task is far more complex and involves multiple cognitive parts that occur concurrently including but not limited to perusing, foraging, shoeboxing, connecting, and validating. Second, the authors could model task-breakpoints for their tasks, but it is very difficult to identify breakpoints in our task. Third, the authors used breakpoints to identify low cognitive workload and study their impact as markers for interruption. Conversely, we use EDA values to mark interruption windows and study impact of interruption on cognitive workload.

We find that our work differentiates from Okoshi et. al.[Okashi 2015] because the authors created Attelia II, a middleware that identified breakpoints in mobile and wearable devices without any psycho-physiological sensors as opposed to using physiological sensors like EDA in our case. Further, their work is limited to situations where one is allowed to install custom software on android devices only. Finally there is significant privacy concern, as one must allow the software to access the physical and UI activity as well. On the other hand, our approach enables users to use an off-the shelf sensor synchronized with a local server to control interruptions without any custom software installation on personal devices.

We find that our work differentiates from Zueger and Fritz [56] because the authors were interested in non-task-related interruption where 10 software developers were interrupted by mathematical quizzes while coding. We find this setup unlikely to exist in real world, and instead, we focus on task-related interruptions that would help analysts perform data analytical tasks, yet would distract them from their sensemaking loop. Second, the authors used EDA raw score as a measure as opposed to acceleration value in the scores as indicator of interruption.

Finally, while the authors measured the perceived utility of interruption through disruptiveness, time-lag, and cognitive load measures, we further analyze the impact of interruption on task-performance and collaborative experience.

We tested direction of intensity of change in value of EDA Mean Phasic Amplitude as indicator of two states: interruptibility or not. The direction could be positive (suggesting a sharp increase in MPA) or negative (suggesting a sharp decrease in MPA). MPA was calculated over a sliding window of 5 seconds. Based on heuristics outlined in previous work [52, 56], every 5 seconds a MPA value was calculated. Adjusting after pilot testing, a sharp change was defined as a change in at least 10% of the value of MPA in the current window as compared to the previous 5-second window, sustained for at least 30 seconds. When this trend was identified to be sustainable, the state was defined as interruptible under acceleration or deceleration of MPA values. We next describe why we chose 5 seconds, and 10% change over 30 second windows, as defining characteristics for this task.

The sliding window is based on some of the most recent work [56] that examined which time window of psycho-physiological data per interruption works best for the classification? Authors applied machine learning to several time windows, ranging from ten seconds to three minutes. Based on lab and field study results, a time window of ten seconds was found to work generally better than longer ones. Authors also show that there were no significant differences in accuracy across various time windows. We suggest readers to refer to [56] for further details. Having been informed by [56], during pilot testing with 7 participants, we discovered that owing to the nature of our 45-minute task, as opposed to the 60-minute task for [56], 5-second window works better to determine the interruptible state as a 10 second window misses opportunities for interruptions in such a tight time window. Further, research [52] has shown that EDA values are related to Pulse Oxymetry based Inter-Beat Interval (IBI) for information processing tasks signifying emotional arousal under unpleasant/stressful conditions. Authors found that over a 4 second window detecting acceleration and deceleration of IBI slope at 12.5% change has been shown to be effective at identifying a significant difference between increasing and maintaining, and maintaining and decreasing arousal level. During retrospective analysis of the pilot-testing of the same concept, but for EDA, we found that 10% change over a 5 second window in EDA is effective for our task. We next outline our research questions

3 STUDY RESEARCH QUESTIONS

As described in the previous section, we measured the direction: positive or negative for MPA change. Subsequently, an interruption can be given when the direction is positive or an interruption can be given when the direction is negative or randomly or we can block interruptions completely. A positive change in MPA would suggest higher arousal, or potentially higher attention, rendering availability to accept interruption. This shows that as a compact score, impact on workload is unclear. So, we pose the following research question:

RQ1: How will the Interruption at different directions (positive, negative, random, none) of EDA MPA change affect participant's workload?

Similarly, one may argue that increase in MPA value reflects increasing anxiety or increasing attention. This shows that as a composite score, interruption might cause higher anxiety or would help catch the attention of the user. Consequently, the impact on task performance would be unclear. So we present our second Research Question:

RQ2: How will the interruption at different directions (positive, negative, random, none) of EDA MPA change affect participant's task performance?

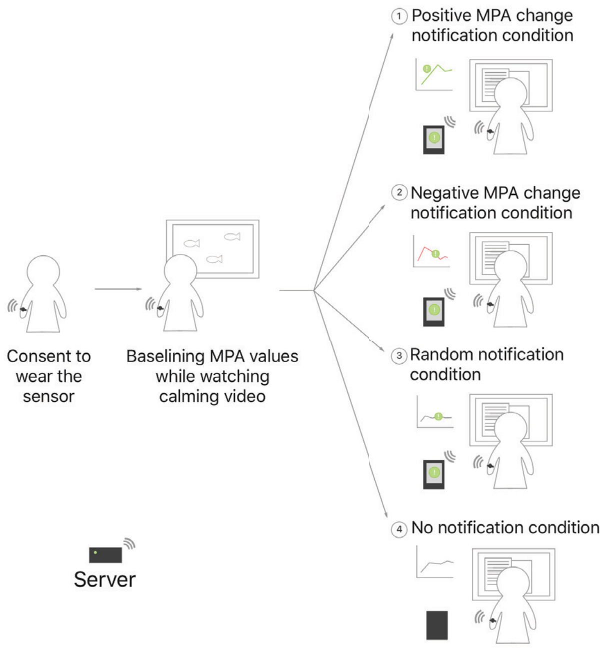

Fig. 2. The Experimental Setup. Participant signed a consent form prior to wearing Q Sensors (by Affectiva Inc.). Next, the Q Sensor was synched with a server and MPA values were baselined at the end of watching a calming swimming fish video. Based, on one of the four conditions, user then was explained the task, give the task-set, and was notified appropriately (or not at all in the fourth condition). The sensor was removed at the end of task. The user then filled out a report about serial killer cases and serial killer identity. Next, user filled out a survey about workload, and clue utility and timing. Finally, the experimenter performed the retrospective video analysis of the sensemaking process associated with MPA acceleration, deceleration, and clue timing.

Third, we wanted to reason if users perceived interruption as wasteful, when we manipulated the timing of the interruption. As in the previous cases, we had no significant indicator to hypothesize the effect of direction change on the perceived utility of the interruption. So we present our third research question:

RQ3: How will the interruption at different directions (positive, negative, random, none) of EDA MPA change affect participant's perceived utility of notification?

Finally, as this was a non-solo task, user's perception of their partner, who interrupted them with the notification, was measured as collaborative experience and presents our final research question:

RQ4: How will the interruption at different directions (positive, negative, random, none) of EDA MPA change affect participant's perceived collaboration experience?

4 METHOD

In this work, we present data from a controlled lab experiment where an analyst was presented with a crime-solving task and their job was to identify a hidden pattern in the crime dataset. As shown in fig. 2, Analysts were randomly assigned to one of the four conditions: random interruption, no interruption, interruption with a significant positive MPA change, and interruption with a significant negative MPA change. In each condition, they received a set of PDF crime documents, and PDF worksheets to fill out. They worked on solving these crimes using a simulated geographically distributed environment. The analyst was instructed that they had a partner in the field who is actively collecting insights and clues. If any new clues or insights were found, they would receive a notification on an android smartphone that was provided to them during the task, based on the condition. The smartphone would vibrate, ring and show the notification for 10 seconds before it disappeared but still could be accessed by clicking on the mailbox icon. At the end of the task, we measured task performance, cognitive workload, and perceived utility of the interruption.

4.1 Participants

Fifty participants participated in the experiment described as a "Solve Crimes Together Study" and played the role of a crime analyst. These participants were recruited through advertisement on the internal mailing list of the Psychology Department where volunteers sign up to participate in a study. Participants were assigned a condition randomly, at sign up. Data for one of the participants was corrupted due to a break to go the restroom and one participant had to leave early. Finally, forty-eight participants participated in the experiment such that twelve participants were assigned to each of the four conditions: No interruption, Random interruption, Interruption during a Positive Change, and Interruption during a Negative change. All students (16 male, 32 female; 85% U.S. born; age range 18-28, median age approximately 22; 85% spoke English as first language) were undergraduate or graduate students at a large U.S. university. Participants were paid $15 for their participation in the 1.5-hour experiment. All conditions received 3 to 5 male participants and no other significant demographic differences were found between the participants across the four conditions.

4.2 Materials

4.2.1 Serial Killer Task. The task was based on a paradigm used in a number of previous studies of collaborative sensemaking [7, 8, 22–25, 34]. In this task, each participant receives a set of crime case documents. They receive 6 cold murder cases, and 1 current murder case, bus route information, and maps of the areas of the crimes. In total, there were seven murders, with about 40 potential suspects, hidden across 20 documents.

They were asked to solve as many cases as they can. However, there was a pattern hidden in the crime cases, suggesting that 5 cases were related to each other and there was a serial killer involved in 4 of them. Connecting this information was necessary to identify the name of the serial killer. In the previous studies, this task has been shown to be extremely difficult and correctly identifying the Serial Killer has remained elusive to most of the participants [7, 8, 22–25, 34]. In this study, participants had to find multiple clues and one such necessary but not fully solving clue was shared with them as an interruption notification during the task.

4.2.2 Post-task report. At the end, participants filled out a paper report identifying the name of the serial killer and associated cases.

4.2.3 Post-task survey. After filling out the report, an online post-task survey asked participants about their perceived utility of the interruption, cognitive load (TLX), analytic ability, and demographic information.

4.3 Equipment

A Mac Mini (Intel Core i7 processor, 16 GB RAM) was connected to the Internet and powered a 25 inch monitor, where participants could explore, and read crime cases, and make associated notes. A GSR/EDA sensor was placed next to the workstation, along with sanitization alcohol swipes. The EDA sensor was connected via Bluetooth to a Macbook Air running as server, where a continuously running python script received EDA values, computed MPA, and sent a notification according the experimental condition. The participants also received a Moto G Play Droid Android smartphone with a 5 inch 720p screen running Gmail application in the background, where interruptions were received as email notifications by the previously mentioned python script.

4.4 Procedure

Upon arrival, the participant was given a consent form seeking written consent for using sensors and the study participation. Next participant was seated at the workstation. They were reminded that any restroom break or consumption of beverage or food over the next one hour would alter body composition and subsequently disqualify them from participation. So, they were encouraged to perform such activities now, instead of the middle of the experiment.

The sanitized Q sensor was then tied to their dominant wrist and connected to the local server. Participants were asked to ignore the sensor and "*think of it like a future apple watch*". Then they were shown a 5 minute HD YouTube video of fish swimming in water to baseline their arousal levels while the experiment was setup. The EDA MPA values from the sensor were sent wirelessly via Bluetooth to the server that measured MPA Skin conductance at 0.1μS values continuously at 4 Hz. At the end of the video (5 minutes), the baseline MPA was saved by the script.

Next, the experimenter explained that participant would role-play as a detective. Participant received training for finding clues including but not limited to motive, opportunity, and lack of alibi. Participant was also explained how to differentiate between a serial crime and non-serial crimes. Next participant used a 10-minute practice task to identify motive, opportunity and (lack of) alibi in a laptop theft crime case. After another 5-minute break of the fish-video, participants were then given the experimental task to begin solving as many cases as possible over 45 minutes. They were also instructed that they are collaborating with another analyst remotely on the crime task who might send them a clue via email notifications on the provided smartphone. This clue was necessary but not sufficient to solving the task, hence rendering the content of the clue lesser important to the interruption itself and associated effects.

As explained previously, MPA was calculated continuously in real-time over a sliding window of 5 seconds. When a sharp change of at least 10% of MPA for 30 seconds between consecutive windows was identified, the state was defined as interruptible. Based on the condition, the script then sent out an email with the clue as the subject and body to the provided smartphone. The smartphone would then ring, vibrate and show the notification for 5 seconds. If it was not interacted with, the users could open the email app. with a single tap and access the notification. Due to the nature and complexity of the task, the first and last 10 minutes of the tasks were classified as non-interruptible and no notification was sent then. This gave the users sufficient time to act on the interruption, if any.

At the end of the task, users filled out the report of task-performance, and a survey for perceived cognitive workload, perceived quality of collaborative experience and perceived utility of the interruption, details of which are provided next.

5 MEASURES

5.1 Cognitive Workload

NASA TLX [27] was used to measure self-reported mental demand, temporal demand, effort, subjective performance, and frustration with the task using five questions in the post-task survey. Principal Component analysis showed that inverted subjective performance loaded onto a separate component, while the other four items loaded into a single component, referred to as TLX subsequently (chronbach alpha = 0.74)

5.2 Task Performance

Task Performance was measured using the post task report form using two measures: First, a binary variable: Serial Killer Identification that was set as 1 when correctly identified and 0 otherwise, based on previous work [20–25]. Second, Serial Killer Case Identification refers to the number of correctly identified serial killer connected cases out of the maximum 4 cases. To reduce the effect of the timing of the clue, the clue timing was used as covariate. No effect of clue-timing was found, and it was subsequently removed from the model.

5.3 Team Experience

Team Experience was measured using post-task survey consisting of 10 items about how participants felt about their collaboration experience, as used previously in [21–25]. The 10 items loaded onto 3 components: *communication, collaboration process* and *goal-oriented collaboration*. Bartlett's test confirmed the non-correlation of the variables ($\tilde{\chi}^2$ = 235.03, df = 45, p = 0.00).

Kaiser's Sampling adequacy was 0.701, and none of the variables showed a communality value below 0.54. Kaiser's criterion and Scree test suggested 3 components, explaining 72.86% of the total variance.

Communication referred to questions like "*My partner kept me up to date about what they were doing*"and explained 44.99% variance. *Collaboration Process* referred to questions like "*My partner and I often agreed about who should be doing what task*" and explained 16.06% variance. *Goal-Oriented Collaboration* referred to questions like "*My partner had a definite sense of direction and purpose*" and explained 11.80% of the variance.

5.4 Perceived Clue Utility

Perceived Clue Utility was measured, as a 5 Point Likert scale to rate how useful the content of the information presented during the interruption was perceived while performing the task.

5.5 Perceived Clue Timing

Perceived Clue Timing was measured, as a 5 Point Likert scale to rate the timing of the content of the information presented during the interruption while performing the task.

6 RESULTS

This section is divided into five parts. First we present the effects of manipulating interruption based on MPA on subjectively measured cognitive workload via TLX scale Second we share the impact on the Task Performance objectively measured by analyst's ability to identify the serial killer cases, and the serial killer itself. Next, we discuss the team experience subjectively measured across thee components: communication, collaboration process, and goal-oriented collaboration. Finally, we share the effect on perceived clue utility and timing.

6.1 Cognitive Workload

RQ1 questioned how MPA informed interruption would affect the analyst's perceived cognitive workload, measured via TLX. To test this, we conducted a two-way ANOVA that examined the effect of MPA's direction, as

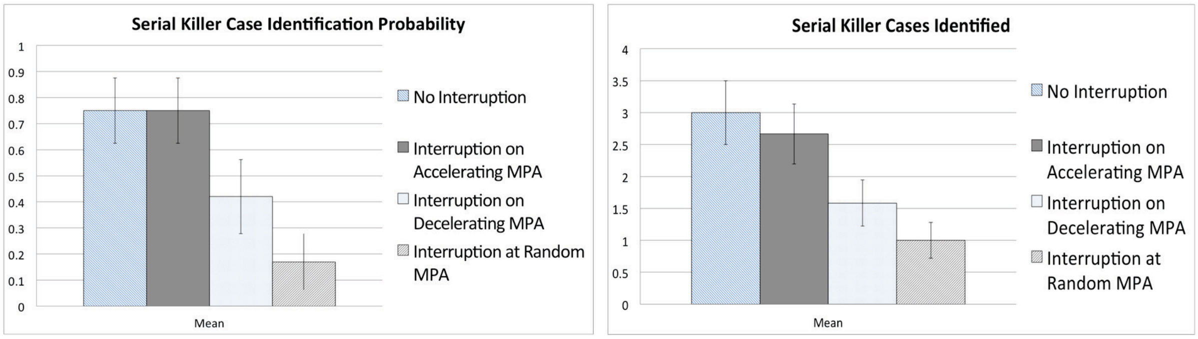

Fig. 3. Results. Task Performance. The first graph on the left shows that participants interrupted during accelerating MPA were significantly more likely to identify the serial killer than those receiving random interruptions, in a 0-1 probability across the Y-axis. The second graph on the right shows that participants were also significantly more likely to identify all the 4 serial killer connected cases (plotted 0 - 4 on Y axis) when interrupted during accelerating MPA values as opposed to random interruption. In both the graphs, there is a weak trend that points to accelerating MPA also performing better than decelerating MPA interruptions.

independent variable, on TLX as dependent measure. We found no statistically significant effect (F [3, 44] = 1.50, p = .22) between them, suggesting that interruption in either condition was no worse than the other.

6.2 Task Performance

Our second research question, RQ2 questioned the impact of MPA informed interruption on objective performance of identifying the serial killer or not, and identifying the correct number of associated serial killer cases. As shown in fig. 3, we found significant effects.

6.2.1 Serial Killer Case Identification. To test the former, we conducted a Poisson regression and ascertained the effects of MPA's direction on the likelihood of participants identifying number of serial killer cases correctly. Significant effects of MPA's direction were found (F[3,44]=15.99, p=.00).Post Hoc Tests for pairwise comparisons revealed that participants who were interrupted when MPA was increasing (M=2.66, SE=.47, p=.01) and who received no interruption (M=3.00, SE=.50, p=.00), performed significantly better than randomly interrupted participants (M=1.00, SE=.28; F[3, 44]=16.94, p=.00).

6.2.2 Serial Killer Identification. To test latter, we conducted a binomial logistic regression and ascertained the effects of MPA's direction on the binary likelihood that participants identify the serial killer correctly. Significant effects of MPA's direction were found (F[3,44]=12.35, p=.00). Post Hoc Tests for pairwise comparisons revealed that participants who were interrupted when MPA was increasing (M=.75, SE=.12, p=.00) and who received no interruption (M=.75, SE=.12, p=.00), performed significantly better than randomly interrupted participants (M=.17, SE=.10; F[3, 44]=17.94, p=.00).

6.3 Team Experience

RQ3 questioned the impact of MPA informed interruption on subjectively reported experience of working collaboratively with an interrupting one-way communicating remotely located partner on three measures: *Communication, Process*, and *Goal*.

6.3.1 Communication. To test experience from communication perspective, we conducted a two-way ANOVA that examined the effect of MPA's direction, as independent variable, on Communication as dependent measure.

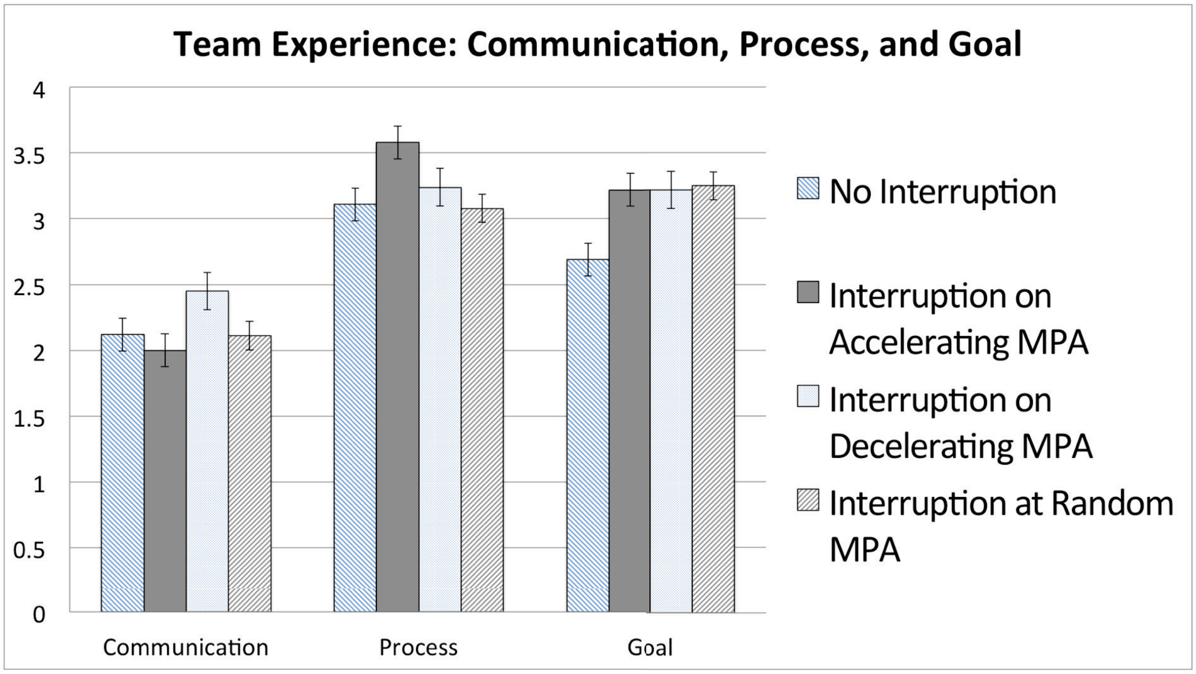

Fig. 4. Results. Team Experience. The three graphs show the effect of interruption timed based on MPA acceleration on the perceived team experience, on a 5-point Likert scale on Y axis. The first graph on the left shows that participants reported no significant effect of MPA induced timing on communication. The second graph in the middle shows that participants rated collaboration process to be significantly better when interrupted during accelerating MPA values as opposed to random interruption. The third graph, on the right, shows that participants reported no significant effect of MPA induced timing on goal-oriented collaboration.

We found no statistically significant effect (F [3, 43] = .90, p = .44) between them, suggesting that interruption in either condition impacted perceived communication no worse than the other. As shown in fig. 4, no significant effects were reported.

6.3.2 Process. To test experience from collaboration-process perspective, we conducted a two-way ANOVA that examined the effect of MPA's direction, as independent variable, on process as dependent measure. We found statistically significant effect (F [3, 43] = 3.63, p = .02) between the two, suggesting that interruption impacted perceived experience with the collaboration process. As shown in fig. 4, Post Hoc Tests for pairwise comparisons revealed that participants who were interrupted when MPA was increasing (M=3.58, SE=.12) rated process significantly better than those who received no interruption (M=3.11, SE=.12, p=.03), and randomly interrupted participants (M=3.08, SE=.12, p=.02).

6.3.3 Goal. To test experience from Goal-oriented collaboration perspective, we conducted a two-way ANOVA that examined the effect of MPA's direction, as independent variable, on goal-oriented collaboration as dependent measure. We found no statistically significant effect (F [3, 44] = .86, p = .16) between them, suggesting that interruption in either condition impacted perceived goal-oriented collaboration no worse than the other.

6.4 Perceived Clue Utility

RQ4 questioned how MPA informed interruption would affect the analyst's perceived utility of the clue. To test this, we conducted a two-way ANOVA that examined the effect of MPA's direction, as independent variable, on Clue Utility rating as dependent measure. Removing the control condition (when no clue was shared) from the analysis, we found no statistically significant effect ($F [2, 30] = 1.23$, $p = .30$), confirming that content of the clue itself was not perceived significantly different across the conditions.

6.5 Perceived Clue Timing

RQ5 questioned how MPA informed interruption would affect the analyst's perceived timing of the clue. To test this, we conducted a two-way ANOVA that examined the effect of MPA's direction, as independent variable, on Clue Timing rating as dependent measure. Removing the control condition (when no clue was shared) from the analysis, we found no statistically significant effect ($F [2, 30] = .66$, $p = .52$), confirming that timing of the clue itself was not perceived significantly different across the conditions.

7 QUALITATIVE FEEDBACK

This section is divided into three parts and describes the results of qualitative analysis based on post-task video analysis with the participants. The digital workspace of the participants was video recorded using a screen recording software on the monitor for retrospective analysis that has been used previously by researchers to identify the effects of participant's cognitive activity on the behavior. Post-task completion, and survey closure, the participants were shown the recorded video of their task for retrospective analysis at 5x speed to query what was the state of analysis, identify the "aha moments", and impact of interruption. Participant feedback during the retrospective analysis helped us understand the effects of interruption on perceived clue utility and timing, on the sensemaking activity when MPA values decelerated significantly and accelerated significantly.

7.1 Understanding Perceived Clue Utility and Timing

While the quantitative result for RQ4 showed that MPA informed interruption had no affect on the analyst's perceived utility of the clue, and it's timing. We wanted to understand through qualitative feedback the reasoning behind this. Inline with the quantitative results, some of our participants did indeed found no value in the clue itself, and ignored the interruption owing to lack of relevance to the current object of sensemaking or due to the timing itself:

> *"…I got it during the current cases. I kind of just ignored it because it wasn't relevant. Yeah, I wasn't really thinking about it at first but then I, like, remembered it and then I was like okay, this could make sense…"* by P36, Random condition

> *"…I had already kind of finished everything and there was not any information that was relevant…"* by P21, Ascending condition

> *"…It was after. I already knew that. I already figured out that there was a serial killer…and I felt very confident in the fact that there was a serial killer involved so I didn't…I kind of was just like okay, that's dumb, like good for you for figuring that out…"* by P35, Random condition

Alternatively, despite the lack of relevance, users reported saving the information for retrospective access:

> *"…Yeah, I hadn't gotten to the cold case so I think I was probably in the beginning of the second current case. So I think I probably just made a mental note…"* by P25, Descending condition

> *"…That was during the cold cases when I was trying to figure out why there were two gun kills and then there were the four blunt instrument kills so it was actually very helpful. Because it said it seemed*

that some of the cold cases had a serial killer, and it kind of seemed that it all fit together, it kind of seemed like the right thing…" by P19, Ascending condition

"… Yeah [notification make you change the strategy]. It made me rush. It increased [possibility of identifying the serial killer connection], cause it made me bring more attention to it." by P15, Ascending condition

Finally, multiple participants reported that they did not base their analysis solely on the shared clue but used the shared clue to confirm their findings:

"… I think I probably would have figured it out anyways that there was a connection between four of the cases. But getting an outside confirmation of someone else who already thinks that, it probably saved me some time of having to piece it together myself, and it also confirmed for me the conclusion I might have come to anyways…" by P25, Descending condition

"… The message came at a time when I was suspecting so it just reaffirmed my beliefs. It definitely made me pay less attention to the details and go with my assumption that there was a serial killer, and not trying to distinguish between killers…" by P30, Descending condition

"… Yeah so it came when I had almost identified the pattern, that means I had read through all five cases and I knew there was a serial killer. So it just came like a [validation] there…" by P34, Random condition

Our participant's feedback shows that despite lack of significant differences in self-reported Likert scale measure about clue utility and timing across the conditions, some of our participants did indeed not just ignore the interruption. Instead, they either bookmarked the information shared during interruption and used it to inform their strategy when relevant data was accessed. Others used this information as a confirmation to their hunch and reported as way to save time.

7.2 Understanding the effect of MPA decrease on Sensemaking Activity

We now describe the sensemaking activity as reported by the participants using retrospective analysis when the MPA decreased significantly as shown by the sensor logs. We particularly scrubbed the video over the time when MPA decreased significantly to identify the impact of MPA deceleration on sensemaking. Some of our participants reported cognitive activity while they perused, foraged, sandboxed, and reasoned with the information as MPA values decelerated significantly. It is perhaps best captured in this quote when the MPA values decreased significantly by P16:

"… Here I was just kind of highlighting little things, I knew it wouldn't really highlight it, I just kept highlighting things because I knew it would come back to my mind. That she was poisoned with the drink and that it would take 20 minutes to take into effect, so it wouldn't happen, that same, during the game, and that it would happen when no one was there. So to give the boss sort of his alibi. But at the same time he was the one that made the drinks when he normally has people bring it to him, which was another major flag because he wouldn't want anyone to be able to rat him out… I don't think so (knew the killer), I think, well I knew that she was poisoned by a drink, so I was looking out for who made the drink and who gave her the drink and if anyone else had contact with the drink and he was the only one…" by P16, Ascending condition

Further, when the interruption was received as the MPA values decreased significantly, our participants reported no significant change in their strategy due to the interruption, as suggested by this recollection:

"… Not that it was too easy but I guess I was doubting myself and wondering if I was doing it correctly. If this was actually just the level that it was at. Yeah so I was filling out the details, just to keep track in my head since I was about to start the cold cases and with the first two I kind of narrowed it down to what I

thought it was. I thought this (the shared clue) was a little harder because there was less information to go off of. I'm not necessarily having any strong inclination as to who did it but I'm just kind of writing down my best guess and then I'll analyze it further later on. So right now I'm not copying per se, but just putting down who I think is my first guess and then just going through more data and seeing if it sticks or if there's anything else…" by P26, Descending condition

According to our participants, significantly decreasing MPA values reflect sensemaking activity related to perusal and foraging. Our participants did not report finding a groundbreaking "aha" moment, with or without interruption during the MPA decrease, while performing the analytic task. This shows that perhaps a significantly decelerating MPA reflects a situation when analysts are busy parsing information and interruptions laden with useful information may not have a positively consequential impact on the analytic task performance.

7.3 Understanding the effect of MPA increase on Sensemaking Activity

In this section we draw attention to the sensemaking activity as reported by the participants using retrospective analysis when the MPA increased significantly as shown by the sensor logs. We particularly scrubbed the video over the time when MPA increased significantly and an interruption was sent to identify the impact of MPA acceleration on sensemaking. Some of our participants reported cognitive activity about searching and identifying connections or patterns across the dataset when MPA values accelerated significantly. It is perhaps best captured in this quote when the MPA values increased significantly by P16:

"…I think I was connecting all the blunt instruments. The cases, and kind of putting them aside and seeing if they were all female, I was looking at their ages because I saw that two were young and two were older, so I thought maybe at first like oh, maybe this is some coincidences, but then they were all, uh, done, either outside the house or while they were entering or had just gotten in. But there was a case where the door was still locked, so I thought that maybe that wasn't connected…" by P16, Ascending condition

Other participants reported an "aha" moment when they had identified the clue (toolbox) or the serial killer (Wayne Millican) as the MPA values were accelerating significantly:

"…I think it was at the very end like when I was filling out these. It was the last case. Actually it was this one, the Mueller. When it said access to like a toolbox, I said okay, so they have the weapon basically. Right there. Okay, so at approximately the 20th minute …" by P21, Ascending condition

"…I was going through potential suspects for the Raffield case, and this guy who's her husband knew Wayne Millican, and I thought it was weird that he said he worked at the hospital that one of the victims worked at, and police had already said that he was unemployed. So then I clearly kinda found out that he was lying, and then it also said that he was on the bus, so I kinda confirmed that he rides the bus often." by P19, Ascending condition

"…Uh, I'm forgetting, there's so many cases, so many names…What I remember clearly was like a gunshot in the car, that was like the first one Alspach case. And I think here the key moment happened when I read through this guy Isherwood's thing, at that moment, because he said that I saw two people a bit before that I was reading, so at that moment I thought I had found the person …" by P35, Random condition

As evident from the illustrative quotes, our participants described how the sensemaking activity was significantly different during the ascending condition compared to the descending condition described in the previous section. These quotes refer to how interruptions sent at the right moment lead to an "aha" moment, and subsequently increased the task performance. In the next section, we unpack our quantitative and qualitative findings and discuss design implications.

8 DISCUSSION

In this work, we presented an empirical study about impact of interrupting a sensemaking process based on the accelerating/decelerating Mean Peak Amplitude (MPA) EDA values using a useful clue to help aid analysis. We found that an interruption sent to the analyst when MPA is accelerating significantly, causes least distraction and the task-performance is significantly higher than a random interruption (RQ2). Further, the analysts reported the communication process team experience about the shared clue from their partner significantly higher when MPA accelerated significantly than a random interruption (RQ4). Finally, we found no significant quantitative variance in perceived cognitive workload (RQ1), clue utility and clue timing (RQ3) across the conditions.

Qualitative feedback highlighted how the sensemaking process differed significantly with MPA changes. Participants reported that MPA acceleration was timed with identifying connections, patterns and “aha” moments. Alternatively, Deceleration was timed with mechanical sensemaking processes of highlighting and copying information. These quantitative and qualitative results point to significant effect of MPA acceleration on interruption timing for collaborative analytics task performance and team experience. In short, interruptions at significantly increasing arousal are significantly better at improving task performance and process, than a random interruption. In the rest of this section we will discuss how our findings relate to similar work in the past that uses EDA data.

In our work, we found that interruption-timing informed by measuring the acceleration/deceleration of MPA values, had no significant effect on the self-reported NASA TLX scores when interruption was given during significant acceleration, deceleration, randomly, or no interruption at all. Past work has shown that interruptions given during task-breaks cause the least disruption since workload is the least [32]. Further, recent work [44] investigated different time and frequency domain features of EDA in simple arithmetic and reading experiments. Authors found that binary cognitive load levels could be differentiated from mean and accumulative EDA and the spectral features.

Our qualitative feedback does show that interruptions made during acceleration does lead to participants perceiving that they are working “faster ” and “saving time”, and do not have to contribute extensive resources to redundant fact-checking as they see the information as “confirmation” to their analysis. However, our sensemaking task was not simple enough to identify the task-breaks or to utilize the components referred here. Further, we only have access to EDA values. Even more recently, researchers have found that in mobile multi-device environment, use of tools like Attelia II can indeed reduce workload [45] but requires access to physical activity, UI activity, works only on Android devices and requires installation. These results point to the complexity of using only EDA values for interruption management in complex tasks such as sensemaking and further research is required.

Further, based on qualitative data, it seems that the information sent in the clue was not the deciding factor to solving the task for all of them. In fact, people could still do the task well without getting the clue (no interruption performed just as well). So, it was not the clue's information, but the timing of the interruption caused by the clue that had the impact on how well the participants performed the task. This was an important factor in our study because we did not want the clue information to bias and interact with the timing.

Further, as opposed to using traditional raw scores or aggregated scores [5, 44, 56], we focused on the significant acceleration/deceleration of the MPA. This novel metric-approach offers researchers an opportunity to utilize cheap wrist-worn EDA devices in contexts it has been found to produce limited results. Researchers can further create far more sophisticated complex models including acceleration as a metric too. Finally, our work contributes a carefully designed lab-experiment to tease apart effects on EDA values on interruption that could occur in similar workspaces.

To summarize, we found that timing an interruption with accelerating MPA performs significantly better than a random interruption by significantly reducing distraction. This is understandable because while no-interruption means no distraction, and hence lesser efforts involved in refocusing on the task. However, accelerating MPA performs just as well because it overcomes the distraction caused by the interruption. Further, the process component in team experience is significantly improved when interruption occurs at accelerating MPA over no-interruption or random interruption. While, this makes sense since no collaboration really ended up happening in the case of "no-interruption", it is more interesting to find that when interrupted appropriately, perception of collaboration may improve over a random interruption. Finally, participants referred to "aha" moments during accelerating MPA as significant indicators of conclusive sensemaking activity in qualitative feedback. Further research should be pursued to better understand the advantages and limitations of accelerating MPA.

9 DESIGN IMPLICATIONS

The results of this study have implication for collaborative analytics research in CSCW, HCI and UbiComp communities. Our study shows that monitoring physiological activity and timing interruptions accordingly can help analysts perform their task better, and improves collaborative work experience without significantly altering the cognitive workload. We could use these findings in multiple ways, some of which are outlined below:

9.1 Managing Implicit and Explicit Interruption

While we presented a time-critical, content-relevant interruption in this work, we envision ability to monitor and schedule information sharing requests opens up a broad design space. At one end, we could design to pause and batch process interruptions implicitly by the system based on user's EDA values, and at the other end enable users to explicitly vet each interruption irrespective of their body signals. Past work has also shown that sharing information, can improve task performance and team experience, over sharing information explicitly (with significant interruption) alone in sensemaking tasks [23]. Designers can find the right mix in between the two extremities based on the domain (critical vs. non-critical), interruption content (relevant or not) and EDA physiological data (accelerating vs. random). For example, medical doctors can be sent relevant interruptions implicitly when they are off-duty but sent relevant interruptions explicitly when they are on-duty.

9.2 Managing Collaboration Availability/Interruptibility State with Collaborators

Using better-informed interruption notification mechanisms, as shown in this work, systems can now implicitly manage the interruption state of a collaborator as "available" or "not available" instead of explicit self-reported status updates, which can be tiring owing to significant continuous manual effort. When "available", the collaborators may choose to implicitly or explicitly share information to reduce impact on collaborator's sensemaking performance and team experience. For example, as crime-solving intelligence analysts sit at their desk in Head Quarters (HQ) and solve crimes, their partners who are out in the field collecting clues and gathering data/evidences could be better informed about when to share the information for maximum impact back to their analysts in the HQ.

9.3 Managing for Attention Competition with connected devices

For open-ended sensemaking tasks, systems can now leverage physiological signals to seed the intermediate results of sensemaking process with ideas as MPA accelerates significantly instead of seeding randomly and indiscriminately. As more devices connect to IoT environments, competition for our attention is only expected to increase. Developing intelligent notification systems that can enable "aha" moments, without the excessive data streams that disrupt the user's attention is needed. In our case, the disruption was designed to occur through a peripheral device (smartphone) to the workspace (desktop). These are not just different devices, but afford different modalities too (auditory + visual vs. visual). So cognitive resources needed for one modality did not hinder use of cognitive resources for the other modality. Future system designs could benefit from exploiting multiple modalities as multiple devices become a source for interruption.

9.4 Enabling Passive Consumption in Transit or Driving

Users in relatively stationary transit venues like self-driving cars where hands do not require significant movement may now be interrupted when it is the right time. Automobile driving is a sensemaking task requiring significant effort in understanding the dynamic context and the response by the vehicle. Self-driving car drivers can now employ EDA to identify the most optimal time to be interrupted by any useful information that requires displaying in the car dashboard. Similar ubiquitous situations may exist in other (relative to velocity) stationary positions like commuting in public transit etc.

9.5 Managing visibility of devices

Future devices are expected to work “behind the curtains” and emerge only when needed. Such devices may now be controlled to emerge and interrupt a user only when the user is ready to be interrupted. The devices connected as IoT may access the interruptible state and identify if the present time is the right time. If so, they may present themselves and offer new pieces of information. For example, public displays and advertisements may just disappear and not distract unless one is ready to interact with them.

10 LIMITATIONS AND FUTURE WORK

To summarize, this is one of the first few attempts at utilizing EDA acceleration/deceleration values to identify interruption timing in a sensemaking task. As one of the first attempts, we chose to utilize slope of MPA calculated over a sliding window of 5 seconds. When, at least 10% change in the value of MPA in the current window compared to the previous 5-second window was sustained for at least 30 seconds, interruption state was defined as interruptible or not. While, these numbers are heuristics based on previous work [52, 56] and pilot-testing during our task, we do not claim that these numbers are the most optimal. Other sliding window lengths, or relative change percentages might work better. This is a limitation of our work, and this also opens up a new direction for future work, in identifying more optimal window sizes.

In this work, we focused on a time-limited collaborative converging sensemaking task. Our results might not hold true for other kinds of sensemaking tasks that might be exploratory or creative. Future work is needed to understand the interruption timings in such cases. Our work also focused on relevant interruption from a remote collaborator. However, there might be cases when users might receive non-relevant interruptions or a mix of relevant and non-relevant interruptions. Further, we focus on a single interruption in 25 minutes. In our work, the number of interruptions and number and quality of clues was set such that the clue's content did not become the deciding factor but it was an aid essentially. The goal was to understand the effect of clue timing at varying arousal states. Anymore clues, and task would have become too easy, offering no way to study the impact on timing. This is limitation. Future work should focus on different kinds of tasks.

Longer sensemaking or analytic tasks might suffer from multiple interruptions. More research will be needed to understand implications for using EDA to time dynamically over longer sensemaking tasks. Further, instead of focusing on a single analyst, future research is needed to understand the impact of team size, and scaling across multiple collaborators that might send different types or timed interruptions and can receive interruptions too. Finally, using the lab-setting enabled us to vary interruption timings in a controlled setting to understand the effects of our experimental design across multiple measures, field research with analysts would help understand the impact of EDA informed interruptions in real life crime solving teams.

11 CONCLUSION

In this paper, we presented findings from an experiment in which participants played the role of crime analysts collaborating with a remote analyst to identify a serial killer in a distributed synchronous setting. We found that interruptions from partner informed by acceleration in EDA's MPA values significantly reduced distraction by improving both identification of the hidden pattern across serial killer cases and success at identifying the hidden serial killer, without increasing cognitive workload. Further, we found that when interruptions were sent during accelerating MPA, the users rated their team-experience process significantly higher. Encouraged by these results, future research is needed to further explore the utility of increasing arousal measured by EDA in longer and more interrupting tasks.

REFERENCES

[1] Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 33.

[2] Piotr D Adamczyk and Brian P Bailey. 2004. If not now, when?: the effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 271–278.

[3] Albert F Ax. 1953. The physiological differentiation between fear and anger in humans. *Psychosomatic medicine* 15, 5 (1953), 433–442.

[4] Brian P Bailey, Piotr D Adamczyk, Tony Y Chang, and Neil A Chilson. 2006. A framework for specifying and monitoring user tasks. *Computers in Human Behavior* 22, 4 (2006), 709–732.

[5] Brian P Bailey and Shamsi T Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 4 (2008), 21.

[6] Brian P Bailey and Joseph A Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* 22, 4 (2006), 685–708.

[7] Aruna D Balakrishnan, Susan R Fussell, and Sara Kiesler. 2008. Do visualizations improve synchronous remote collaboration?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1227–1236.

[8] Aruna D Balakrishnan, Susan R Fussell, Sara Kiesler, and Aniket Kittur. 2010. Pitfalls of information access with visualizations in remote collaborative analysis. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 411–420.

[9] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.

[10] AJ Brush, David Bargeron, Jonathan Grudin, and Anoop Gupta. 2002. Notification for shared annotation of digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 89–96.

[11] Daniel Chen, Jamie Hart, and Roel Vertegaal. 2007. Towards a physiological model of user interruptability. In *IFIP Conference on Human-Computer Interaction*. Springer, 439–451.

[12] Gregorio Convertino, Helena M Mentis, Mary Beth Rosson, Aleksandra Slavkovic, and John M Carroll. 2009. Supporting content and process common ground in computer-supported teamwork. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2339–2348.

[13] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant messaging and interruption: Influence of task type on performance. In *OZCHI 2000 conference proceedings*, Vol. 356. 361–367.

[14] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant messaging: Effects of relevance and timing. In *People and computers XIV: Proceedings of HCI*, Vol. 2. 71–76.

[15] RK Dismukes, Grant E Young, Robert L Sumwalt III, and Cynthia H Null. 1998. Cockpit interruptions and distractions: Effective management requires a careful balancing act. (1998).

[16] Steven A Egger. 2002. *The killers among us: An examination of serial murder and its investigation*. Prentice Hall Upper Saddle River, NJ.

[17] James Fogarty, Andrew J Ko, Htet Htet Aung, Elspeth Golden, Karen P Tang, and Scott E Hudson. 2005. Examining task engagement in sensor-based statistical models of human interruptibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 331–340.

[18] Darren Gergle, Robert E Kraut, and Susan R Fussell. 2004. Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of language and social psychology* 23, 4 (2004), 491–517.

[19] Steven Gottlieb, Sheldon I Arenberg, Raj Singh, et al. 1994. *Crime analysis: From first report to final arrest*. Alpha Publishing Montclair, CA.

[20] Nitesh Goyal. 2015. Designing for Collaborative Sensemaking: Using Expert & Non-Expert Crowd. *arXiv preprint arXiv:1511.06053* (2015).

[21] Nitesh Goyal and Susan R Fussell. 2015. Designing for Collaborative Sensemaking: Leveraging Human Cognition For Complex Tasks. *arXiv preprint arXiv:1511.05737* (2015).

[22] Nitesh Goyal and Susan R Fussell. 2016. Effects of Sensemaking Translucence on Distributed Collaborative Analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 288–302.

[23] Nitesh Goyal, Gilly Leshed, Dan Cosley, and Susan R Fussell. 2014. Effects of implicit sharing in collaborative analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 129–138.

[24] Nitesh Goyal, Gilly Leshed, and Susan R Fussell. 2013. Effects of visualization and note-taking on sensemaking and analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2721–2724.

[25] Nitesh Goyal, Gilly Leshed, and Susan R Fussell. 2013. Leveraging partner's insights for distributed collaborative sensemaking. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, 15–18.

[26] Carl Gutwin and Saul Greenberg. 1998. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. ACM, 207–216.

[27] SG Hart and LE Staveland. 1988. Development of a Muli-dimensional Workload Rating Scale: Results of Empirical and Theoretical Research. Human Mental Workload. PA Hancock and N. Meshkati. (1988).

[28] Stephen C Hayne, Lucy J Troup, and Sara A Mccomb. 2011. âĂIJWhereâĂŹs Farah?âĂİ: Knowledge silos and information fusion by distributed collaborating teams. *Information Systems Frontiers* 13, 1 (2011), 89–100.

[29] Jeffrey Heer, Fernanda B Viégas, and Martin Wattenberg. 2009. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM* 52, 1 (2009), 87–97.

[30] Eric Horvitz, Andy Jacobs, and David Hovel. 1999. Attention-sensitive alerting. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 305–313.

[31] Shamsi T Iqbal and Brian P Bailey. 2006. Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 741–750.

[32] Shamsi T Iqbal and Brian P Bailey. 2008. Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 93–102.

[33] Thomas Jackson, Ray Dawson, and Darren Wilson. 2001. The cost of email interruption. *Journal of Systems and Information Technology* 5, 1 (2001), 81–92.

[34] Ruogu Kang, Aimee Kane, and Sara Kiesler. 2014. Teammate inaccuracy blindness: when information sharing tools hinder collaborative analysis. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 797–806.

[35] Ruogu Kang and Sara Kiesler. 2012. Do collaborators' annotations help or hurt asynchronous analysis. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*. ACM, 123–126.

[36] Kara A Latorella. 1998. Effects of modality on interrupted flight deck performance: Implications for data link. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 42. SAGE Publications Sage CA: Los Angeles, CA, 87–91.

[37] John D Lee, Joshua D Hoffman, and Elizabeth Hayes. 2004. Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. ACM, 65–72.

[38] Santosh Mathan, Stephen Whitlow, Michael Dorneich, Patricia Ververs, and Gene Davis. 2007. Neurophysiological estimation of interruptibility: Demonstrating feasibility in a field context. In *In Proceedings of the 4th International Conference of the Augmented Cognition Society*. 51–58.

[39] D Scott McCrickard, Richard Catrambone, Christa M Chewar, and John T Stasko. 2003. Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies* 58, 5 (2003), 547–582.

[40] C McFarlane Daniel and A Latorella Kara. 2002. The scope and importance of human interruption in HCI design. *Human-Computer Interaction* 17, 1 (2002), 1–61.

[41] Helena M Mentis, Paula M Bach, Blaine Hoffman, Mary Beth Rosson, and John M Carroll. 2009. Development of decision rationale in complex group decision making. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1341–1350.

[42] Yoko Nagai, Laura H Goldstein, Hugo D Critchley, and Peter BC Fenwick. 2004. Influence of sympathetic autonomic arousal on cortical arousal: implications for a therapeutic behavioural intervention in epilepsy. *Epilepsy research* 58, 2 (2004), 185–193.

[43] Syavash Nobarany, Mona Haraty, and Brian Fisher. 2012. Facilitating the reuse process in distributed collaboration: a distributed cognition approach. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1223–1232.

[44] Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A Calvo. 2012. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*. ACM, 420–423.

[45] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K Dey, and Hideyuki Tokuda. 2015. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 475–486.

[46] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 897–908.

[47] Nicholas J Pioch and John O Everett. 2006. POLESTAR: collaborative knowledge management and sensemaking tools for intelligence analysts. In *Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, 513–521.

[48] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. 2–4.

[49] Cheri Speier, Joseph S Valacich, and Iris Vessey. 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences* 30, 2 (1999), 337–360.

[50] Neville A Stanton. 1994. *Human factors in alarm design.* CRC Press.
[51] Jonathan Sykes and Simon Brown. 2003. Affective gaming: measuring emotion through the gamepad. In *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 732–733.
[52] Heather L Urry, Carien M van Reekum, Tom Johnstone, and Richard J Davidson. 2009. Individual differences in some (but not all) medial prefrontal regions reflect cognitive demand while regulating unpleasant emotion. *Neuroimage* 47, 3 (2009), 852–863.
[53] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. 2011. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 3131–3140.
[54] Paweł Wozniak, Nitesh Goyal, Przemysław Kucharski, Lars Lischke, Sven Mayer, and Morten Fjeld. 2016. RAMPARTS: Supporting Sensemaking with Spatially-Aware Mobile Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2447–2460.
[55] Fred RH Zijlstra, Robert A Roe, Anna B Leonora, and Irene Krediet. 1999. Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology* 72, 2 (1999), 163–185.
[56] Manuela Züger and Thomas Fritz. 2015. Interruptibility of software developers and its prediction using psycho-physiological sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2981–2990.

Received November 2016; revised May 2017; accepted July 2017