Check for updates

Sensing Noticeability in Ambient Information Environments

Yi Fei Cheng
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
yifeic2@andrew.cmu.edu

David Lindlbauer
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
davidlindlbauer@cmu.edu

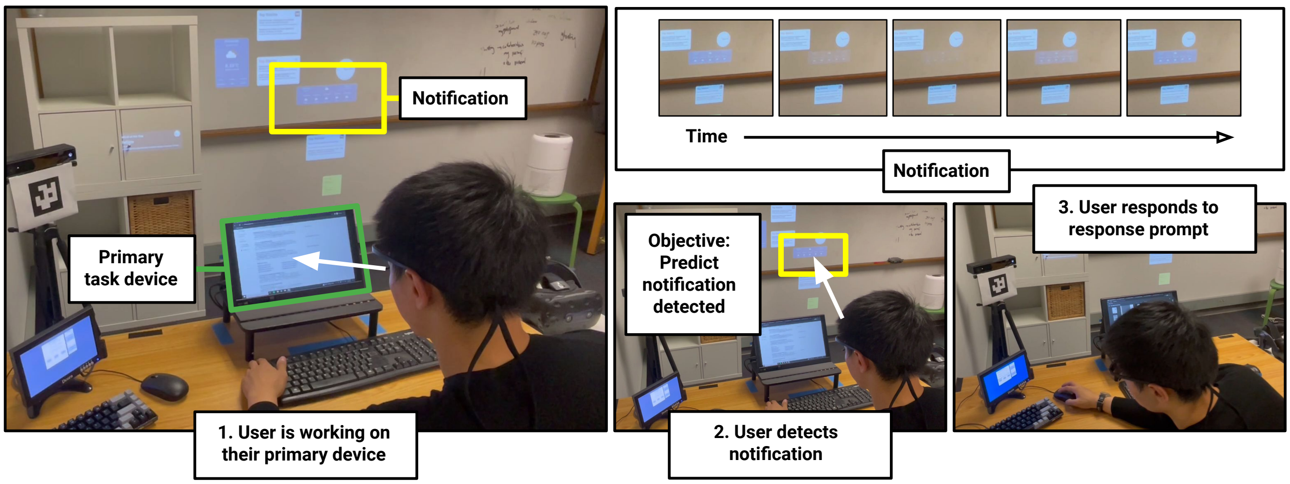

Figure 1: We study which features determine whether users detect peripherally presented notifications while working. To achieve this, we leverage data collected across 12 users, working an average of 8 hours in our setup (98 hours in total). (1) Participants engage in productive tasks using a primary device positioned in front of a projection-based ambient display. (2) Notifications, represented by 10-second opacity changes to a target information widget (top right), are presented at 5 to 10-minute intervals. (3) Users respond to a noticeability sampling prompt to indicate whether they detected a notification, thereby serving as our classification ground truth.

Abstract

Designing notifications in Augmented Reality (AR) that are noticeable yet unobtrusive is challenging since achieving this balance heavily depends on the user's context. However, current AR systems tend to be context-agnostic and require explicit feedback to determine whether a user has noticed a notification. This limitation restricts AR systems from providing timely notifications that are integrated with users' activities. To address this challenge, we studied how sensors can infer users' detection of notifications while they work in an office setting. We collected 98 hours of data from 12 users, including their gaze, head position, computer interactions, and engagement levels. Our findings showed that combining gaze and engagement data most accurately classified noticeability (AUC = 0.81). Even without engagement data, the accuracy was still high (AUC = 0.76). Our study also examines time windowing methods and compares general and personalized models.

This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713511

CCS Concepts

• **Human-centered computing** → **Field studies**; **Interactive systems and tools**; **Mixed / augmented reality**.

Keywords

Ambient displays, noticeability, computational interaction

ACM Reference Format:
Yi Fei Cheng and David Lindlbauer. 2025. Sensing Noticeability in Ambient Information Environments. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3706598.3713511

1 Introduction

New ubiquitous computing paradigms, such as Augmented Reality (AR), can present virtual elements at any time, anywhere, and with arbitrary appearance and behavior. By situating digital information, including notifications, within the users' ambient environments, AR systems facilitate efficient access to information while users engage in their daily activities within the physical world. For instance, notifications in AR can keep users informed about scheduled calendar entries, incoming messages, and other relevant events. This ultimately helps users maintain a productive awareness of their activities and communications within their operating contexts.

However, designing effective notifications in such ubiquitous information environments is challenging. Generally, notifications should be noticeable but not disruptive to the user's current tasks. Since AR notifications can be displayed anywhere within the user's environment, their visibility and disruptiveness depend on the user's context and overall viewing behavior. For example, a notification displayed in the user's periphery that subtly animates the color of a virtual element may be noticed when users are leisurely sketching, as their gaze occasionally wanders. The same notification might go unnoticed when they are engaged in a cognitively demanding task, such as programming, where their focus is on a display or AR window. Current AR systems have limited capabilities to customize notification designs to match contextual needs, partly because they lack the ability to automatically detect whether users notice a presented notification (i.e., its *noticeability*). This means systems either have to explicitly probe for user feedback to confirm noticeability (e.g., wait for users to dismiss a message by pressing a button) or optimize notification designs for maximum noticeability, both of which can be disruptive. Alternatively, systems may opt for a more subtle notification design, but that runs the risk of the notifications going unnoticed entirely. While there is substantial prior research on optimizing notification designs, including their placement, presentation, and timing, work on modeling whether a presented notification has been detected remains limited. We believe this functionality is crucial for providing timely notifications that are well-integrated with users' activities.

As a step towards addressing this challenge, we investigate the feasibility of using sensors to infer whether users detect peripherally presented notifications as they engage in everyday knowledge work. The ability to automatically determine the noticeability of notifications allows them to be acknowledged based on users' implicit behaviors, rather than requiring an explicit response that would otherwise interrupt the user's workflow. Additionally, this approach serves as a foundation for supporting more intelligent and adaptive notification displays. For instance, if the system classifies an important notification as missed, it can re-display the content with a more visually prominent design. Our approach also offers the potential for future AR systems to deliver notifications less obtrusively by gradually increasing the salience of notification while monitoring for noticeability. Finally, the ability to detect noticeability serves as proof-of-concept and a foundation for future approaches that predict whether virtual content will be perceived before it is presented.

To develop a sensor-based approach to infer noticeability, we first collected data from 12 participants, each working an average of 8 hours (98 hours of data in total) in front of a projection-based ambient display situated within a real office setting (Figure 1). The ambient display consisted of information widgets that were animated to simulate incoming notifications at randomized 5 to 10-minute intervals. Shortly after each peripherally presented notification, participants were prompted to report whether they detected the additional stimuli. These self-reports serve as our ground truth for performing noticeability classification. Besides this explicit stimulus-detection record, we collected data on the participants' head pose and gaze, their interactions with their primary device for knowledge work, and self-reports of their perceived level of engagement with their primary task.

Based on the data we collected, we demonstrated that it is feasible to build a classifier that can accurately infer the noticeability of notifications, achieving an Area Under the Curve (AUC) score of 0.81 by combining gaze and user-reported engagement metrics. With the exclusion of engagement measures, a high level of accuracy (AUC = 0.76) can still be achieved. Our analysis also revealed that using features extracted from time windows centered on the notification enabled the highest classification accuracy, but time windows preceding and subsequent to the notification also notably held predictive power. Furthermore, our study showed that a general classifier is a good starting point for inferring noticeability. Lastly, we discovered that using just one to two participants' data for training still produced general models of reasonable accuracy.

To the best of our knowledge, this is the first work sensing the noticeability of notifications based on longitudinally collected data in a close-to-in-the-wild setting. We conclude our work with a discussion of the implications of our findings for the design of AR systems and interfaces, the generalizability of our results, and directions for future studies, including the expansion of our method to fully uncontrolled settings. In summary, we contribute:

- A study involving 12 participants who each worked for an average of 8 hours, interacting with notifications on a projection-based ambient display. Throughout this study, we captured data on their gaze, head pose, computer interactions, and engagement levels.
- A machine learning-based analysis demonstrating the feasibility of sensor-based noticeability prediction. Insights include the effectiveness of various time windowing approaches, the predictive power of sensor and feature combinations, and the performance of general versus personalized models in predicting the noticeability of notifications.

2 Related Work

2.1 Mobile and Desktop Notifications

Visual notifications are used pervasively in mobile and desktop computing environments to proactively provide users with efficient access to information outside their current focus of attention [38, 47]. They can support users' situational awareness [68], but can conversely become a disruptive source of productivity loss [18], stress [66], and inattention [56] if delivered excessively. A significant body of literature has therefore studied notifications in a variety of usage contexts [48, 72] and devices [64, 86].

Prior research on notifications can generally be divided into two lines of work. There has been persistent interest in identifying *when notifications should be delivered* (e.g., [16, 17, 37, 39, 44]). A core consideration relating to this is the interruptibility of the recipient at a given moment. Early work (e.g., Czerwinski et al. [16, 17]) demonstrated that interruptibility depends on factors like relevance, timing, and user engagement. Based on these empirical intuitions, subsequent works have attempted to estimate human interruptibility by using sensors and learning-based models (e.g., [28, 69, 78, 103]). In addition to understanding when notifications should be delivered, significant prior research has investigated *how notifications should be presented* to redirect the user's attention to a target location [64].

This has involved exploring a variety of notification delivery modalities within various computing environments, including mobile phones [26], desktops [71], and multi-screen settings [34, 64].

Our work complements prior research by investigating *whether a notification was detected* (i.e., its noticeability). Sensing noticeability enables more implicit interactions, intelligent notification delivery, and opens the door to less obtrusive notification approaches.

2.2 Notifications beyond 2D screens

In contrast to conventional devices that rely on 2D screens, other paradigms like AR enable the integration of digital content into the user's physical environment in a ubiquitous manner, i.e., anytime, anywhere, with arbitrary appearance. This entails that notifications may appear in the user's visual periphery and vary in visibility depending on their physical backdrop. Several works have begun to build toward an understanding of how notifications may be perceived in these new computing environments. Gutwin et al. [34] and Mairena et al. [64], for instance, explored how factors like effect intensity, position, and primary task influenced people's perception of different pop-out effects in the periphery. Petford et al. [77] compared five attention-guiding techniques for directing users' attention to an out-of-view target in a full-coverage display.

Within augmented and virtual reality (AR/VR), research has also extensively investigated user preferences for its placement [42, 53, 58, 73, 80, 84, 85], representation [30, 32, 53, 54], and modality [14, 52, 57]. Building on AR's ubiquitous nature, these explorations have been undertaken in a variety of contexts, ranging from information consumption on-the-go (e.g., [58]) to usage during social interactions (e.g., [84]). While this body of literature has empirically mapped out the factors that determine the appropriate notification display, such as how its placement influences its noticeability and intrusiveness [80], effective computational approaches to implement these findings are generally lacking. The work of Chen et al. [11] on predicting opportune moments for notification delivery and Lindlbauer et al. [62] on adjusting the level of detail of virtual elements are steps in this direction. However, even if notifications can be delivered with optimal appearance and timing, current systems lack the means to automatically verify if they have been processed, which limits how fluidly they can integrate digital information with users' activities. To address this challenge, Li et al. [61] introduced an LSTM-based method to predict the noticeability of dynamic interface elements in a controlled VR setting.

Our work extends this research by exploring sensor-based methods to detect user notifications in more natural work environments. Unlike Li et al., who used noticeability data from artificial tasks such as transcription and arithmetic, our approach is grounded in longitudinal data collected from a productive office setting.

2.3 Models of Visual Attention

Modeling noticeability, or "perception" according to the model for situational awareness by Endsley et al. [23, 24], is in many ways equivalent to modeling visual attention, a topic with a rich tradition (e.g., comprising of theories like Feature Integration [89] and Guided Search [95] Theory) within psychology [96] and computer vision [8]. Generally speaking, there is agreement that visual attention is affected by both external *bottom-up* factors [50] (e.g., the saliency of a particular feature relative to its background) and internal *top-down* factors [98] (e.g., an individual's memory or objectives). Computational approaches for visual attention aimed to formalize *bottom-up* features through low-level image elements (e.g., color [63]) and *top-down* features using contextual information [81, 98]. More recent learning-based approaches have further tried to model both simultaneously [43].

Within HCI research, there has been persistent interest in leveraging such existing models as well as extending our understanding of attention when engaging with various interactive systems [1], such as web-pages [102] and public displays [19]. Research on understanding and directing user attention in ubiquitous computing and immersive environments is most relevant to our work. For example, Petford et al. [77] compared several visualization techniques to direct user attention to out-of-view targets on a projection-based display. Sitzmann et al. [87] analyzed viewing behavior in virtual reality. Veas et al. [91] and Grogorick et al. [33] explored various attention guidance mechanisms. Vortmann and Putze [92] further explored the use of EEG signals to classify internally and externally directed attention in a controlled ball and tube alignment task.

To our knowledge, there is no prior work on modeling the noticeability of changes to virtual elements in an ambient information environment, especially in-the-wild. This knowledge is essential for creating effective notifications in ubiquitous computing environments. Our work addresses this knowledge gap.

2.4 Ambient / Peripheral / Ubiquitous Displays

In recent decades, considerable research has focused on integrating computing into daily environments [93]. Within this literature, several terms have emerged to describe technologies that enable these blended virtual-physical environments, including *ambient* [49, 51, 65], *peripheral* [67] and *ubiquitous* [5, 51] displays. Research in this direction has explored many technical approaches, such as computationally-enabled surface materials (e.g., [2, 100]) and cognitively-responsive immersive work environments (e.g., [101]).

One popular approach involves enabling interactive applications by instrumenting environments with camera/projector systems [9, 55, 82, 97, 101], such as in Brooks's Intelligent Room [9] and Raskar et al.'s office of the future [82]. More recent work, like Roomalive [55] and Worldkit [97], sought to lower the barrier of entry for developing applications in these computing environments through open-sourcing system components.

In addition to projection-based augmented reality (AR), see-through head-mounted displays (HMDs) can also provide digitally embedded content and computing functions within the environment [7]. Previous research has demonstrated its potential benefits in a variety of contexts, including navigation [60], urban planning [94], and maintenance [36].

Our work introduces an approach to determine the noticeability of peripheral notifications presented in the users' workspace as part of a projection-based ambient display. We envision our work informing the design of the aforementioned computing environments, agnostic of the specific implementation approach (e.g., projection-based or HMD). With knowledge of the sensor-based features that influence noticeability in a workspace environment, interactive systems can support more fluid interactions with notifications.

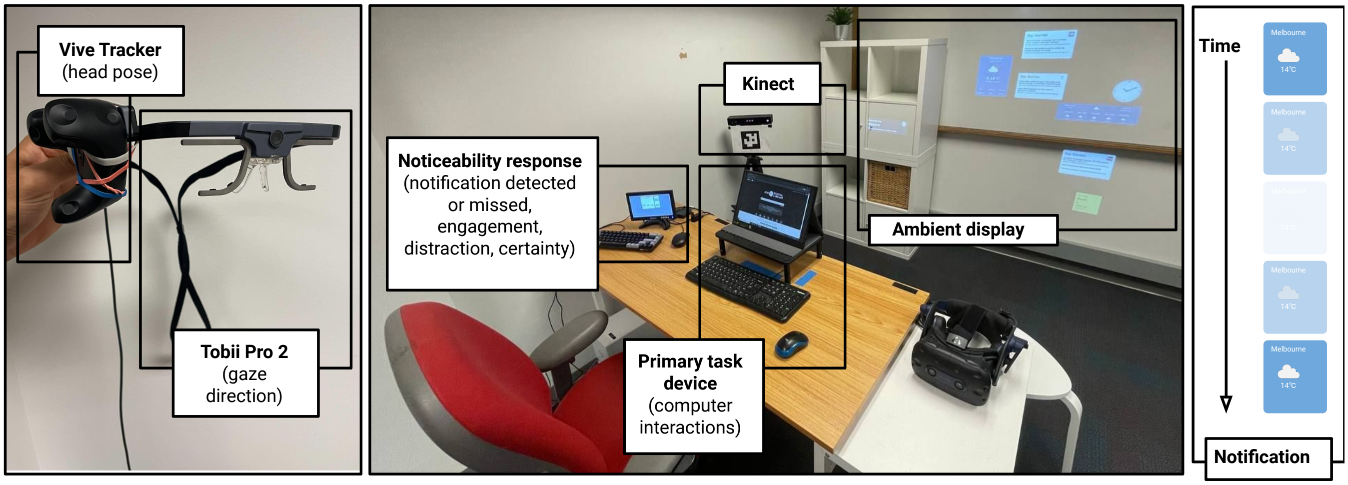

Figure 2: Experimental apparatus. Participants work on a primary task device situated in front of a projection-based ambient display in an office environment (center). They self-report detecting or missing notifications (an example illustrated in the right) using an auxiliary data collection device. For head pose and gaze tracking, participants wear a pair of Tobii Pro 2 glasses an attached HTC Vive Tracker (left).

3 Data collection

We collect data to create an approach that predicts users' detection of peripherally-presented notifications while primarily engaged in everyday knowledge work. We recruited 12 participants to work for around 8 hours each in an office environment equipped with an additional projection-based ambient information display. This yielded a total of 98 hours of data across all participants. The apparatus, shown in Figure 2, was designed to be minimally invasive, resembling Raskar et al.'s *Office of the Future* [82], where information widgets are embedded into the physical environment. The display presented 10-second notifications in the user's periphery every 5 to 10 minutes. After each notification, participants were prompted to report whether they had detected the notification via an auxiliary display on their desk. We compiled a comprehensive dataset, including the stimuli-detection record, data on head pose, gaze, computer interactions, and self-reported ratings of engagement levels.

3.1 Apparatus

We allocated a desk in one of our research offices for the study and equipped it with the following devices for participants to perform their self-selected primary tasks: an external monitor (ForHelp 15.6 inch), a keyboard, and a mouse. Additionally, we provided a projection-based ambient display for presenting notifications (ViewSonic PX701-4K), a small additional display (Dcorn 8 inch) for prompting participants to report whether they detected a notification, and a mouse for participants to input their response. We also installed a Kinect V2 camera for capturing the study session for later analysis. Lastly, participants wore a pair of Tobii Pro 2 Glasses with an attached HTC Vive Tracker for eye and head pose tracking. For the primary study task, participants could use their own laptop or one we provided. To maintain a consistent primary display for all participants, we required them to mirror their devices onto the ForHelp display for the study.

3.2 Ambient Display

In designing our projection-based ambient display, we drew inspiration from prior visions of ubiquitous displays (e.g., [9, 82]) and current widget-based interfaces on mobile phones (e.g., [45]). We were guided by the following specifications: (1) the display should effectively "disappear" [93] into the physical environment, and (2) it should enable quick access to information at a glance.

Our final display consisted of widgets for checking the weather, news headlines, word of the day, air quality, and time, representing common task-independent ambient applications (e.g., Han et al. [35], Cho et al. [13]). To increase the relevance of the presented information to our participants, all widgets are updated in real-time, with the weather and air quality applications tailored to the study location. Widgets are mapped to physical surfaces and scaled for readability. They are placed randomly, while avoiding occlusion and respecting surface boundaries, with placement limited to regions within the participant's peripheral vision when seated at the experimental desk, facing the monitor.

It is important to note that the widgets were intentionally designed to not be relevant to the participants' primary task. This decision was driven by privacy concerns, as the apparatus was deployed in a semi-public space. Additionally, as our focus was on predicting whether participants can detect additional peripheral stimuli, we reasoned that presenting generic auxiliary information was sufficient for our purposes.

Finally, we chose a projection-based ambient display over a head-mounted display due to the hardware constraints of current headsets, such as low resolution and high weight. The projection-based setup enables participants to perform their task with a familiar setup, which increases the ecological validity of the collected data. We hope to replicate our data collection with future headsets, once they allow participants to perform their primary task comfortably and efficiently over extended periods of time.

3.3 Office Environment

The office environment in which our apparatus was situated was simultaneously occupied by 3 to 4 other people who typically work in this space. Their daily activities usually involved some form of knowledge work, with some collaborating and engaging in discussions. The office also hosted a coffee machine frequented by members of the affiliated research team, serving as a daily gathering space. Occupants and visitors of the space were instructed to avoid physically interfering with the ambient display, but they were not told to avoid interaction with the participant entirely. We intentionally situated our apparatus in this comparatively less controlled environment to enable more ecologically valid working conditions, which included ambient sounds typical of office settings and unexpected bystander interruptions.

3.4 Notifications

To collect data for predicting noticeability, we introduced notifications by animating individual widgets on our projection-based display. Every 5 to 10 minutes, the system randomly chooses one widget as the notification target (Figure 4). We selected this time interval to balance collecting enough samples for machine learning and delivering notifications at an ecologically valid frequency (an average of 63.5 notifications per day [79]), i.e., infrequently enough so that participants cannot anticipate when the next notification will occur and focus on their main task.

We animated widgets by adjusting their opacity to smoothly create a subtle, oscillating fade. Each notification involved animating a selected widget for 10 seconds, roughly equivalent to 2-3 repetitions of a phone ringtone sequence. For our study, we included both a slow and a fast opacity change-based notification, drawing inspiration from Gluck et al.'s work [31]. The slow notification faded in and out with a 2.5-second period, while the fast notification had a 1-second period as a starting point. We adjusted these periods during the study to accommodate individual differences in attention, aiming to achieve an approximate 50% detection rate. These adjustments were intended to strike a balance between making notifications salient enough to be reliably detected and avoiding an overwhelming or intrusive level of stimuli.

We chose our opacity change-based notification approach after exploring alternative notification mechanisms, such as motion and flag-based animations, in pilot studies. During these initial explorations, we found that single-state change-based notifications, like introducing a flag, were too subtle and users missed them most of the time. Compared to other continuous state-change notifications, we ultimately selected the opacity-based notification for its aesthetic appeal and subtlety. Finally, we settled on the 2.5-second and 1-second periods for the slow and fast opacity change-based notifications, respectively, via experimenting with different oscillation speeds in additional pilots.

3.5 Noticeability Sampling

To collect a ground truth for noticeability, we employed a cued recall paradigm, retrospectively prompting participants to self-report whether they noticed a previously presented notification. This approach is notably different from instructing participants to respond immediately. We chose this paradigm after observing in our pilots

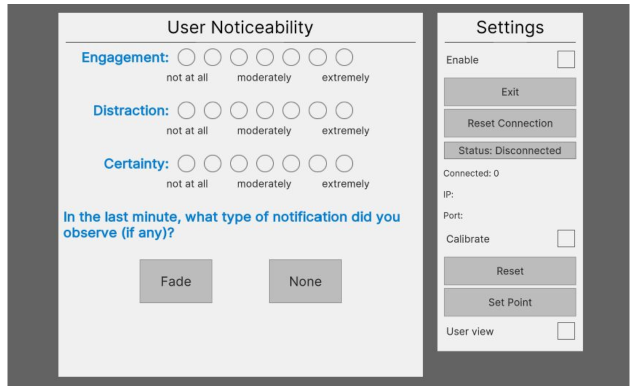

Figure 3: Noticeability sampling prompt and reporting interface shown on the additional display. Participants retrospectively reported whether they detected the notification, their level of engagement, the degree to which the notification distracted them, and their certainty in their detection.

that requiring immediate responses introduced a significant amount of noise to the data. Specifically, it was challenging to distinguish signals arising from participants detecting the notification from those associated with confirming that they noticed them, i.e., reaching over to our data collection device to indicate their response. Disentangling the moment of detection from when participants needed to react to a notification resolved this challenge.

Prompts asking participants to recall detecting a notification were shown after a 30- to 60-second delay on a small screen to the left of their monitor (Figure 3), which flashed blue. Positioned near their mid-peripheral vision, the prompt aimed to be more noticeable than the notifications without causing annoyance. We used an additional device to ensure that participants could easily access the recall prompt. In addition to asking whether they detected the notification, we used a 7-point Likert scale to assess their current level of engagement with their primary task. We also assessed their level of certainty in their detection and the extent to which the presented notification was distracting.

The flow of the notification and sampling procedure is presented in Figure 4. To summarize, 10-second notifications were presented at randomized 5 to 10-minute intervals. Notifications were followed by a 30-second to 1-minute buffer, after which a prompt was presented to determine whether the participant had detected the notification. The system advances to the next notification (i.e., starts the timer for the next interval) either upon receiving a user response or if left unanswered for over a minute. We henceforth refer to each notification-prompt pair as one trial. In each trial, we denote the start of a notification as $t = 0s$. Notifications are continuously displayed between $t = 0$ and $t = 10s$.

3.6 Implementation

We implemented the ambient display and noticeability sampling prompt in Unity3D. All widgets were updated with real-time information from the web using JSON APIs, including the News API for headlines, the Open Weather API for weather-related information,

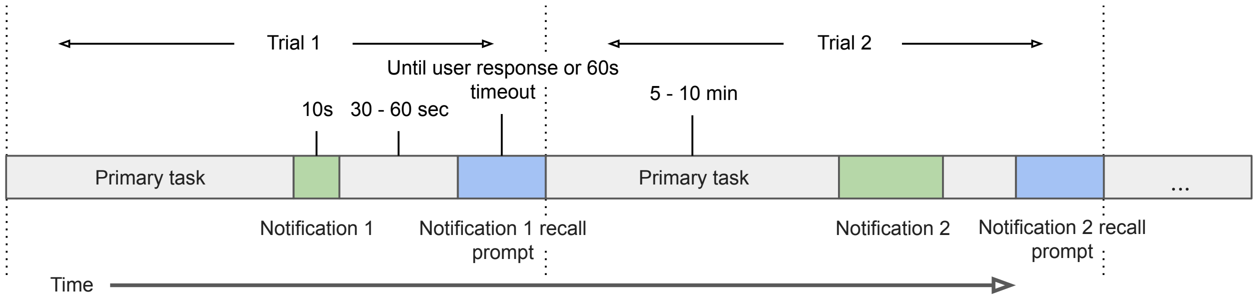

Figure 4: Procedure. 10-second notifications are presented randomly at 5 to 10-minute intervals. After each notification, there is a 30 to 60-second delay, after which participants are retrospectively prompted to report whether they detected notification. The system proceeds to the next notification upon a user response or if it is left unanswered for over a minute. One trial refers to a notification-prompt pair.

and the Words API for words and definitions. To map the widget display to the surrounding physical environment, we pre-calibrated the projector to the environment using an additional Kinect V2 sensor and the RoomAlive Toolkit [55]. The ambient display and notification prompt application was deployed on an Alienware Aurora R12 desktop computer (Windows 10, Intel Core i9 119000KF, NVIDIA GeForce RTX 3090 24GB, 32GB RAM).

3.7 Head Pose and Gaze Tracking

To perform head pose and gaze tracking, we attached an HTC Vive Tracker to a pair of Tobii Pro 2 Glasses. We implemented a separate Unity3D application for logging the HTC Vive Tracker position and orientation (sampled at 30Hz). We adapted De Tommaso and Wykowska's Python controller [21] for logging the relevant gaze data (sampled at 50Hz). We manually calibrated the coordinate systems of the Vive Tracker and Tobii Glasses. To calibrate the head pose and gaze with the ambient display, we performed a three-point calibration before the start of each study session. Both applications were deployed on the same desktop computer as the ambient display and notification prompt. We connected and synchronized all three applications via a web socket (e.g., starting and stopping logging).

3.8 Monitoring Tool

To collect computer interaction data, we adapted the monitoring tool developed by Meyer et al. [70]. We tracked participant's mouse and keyboard interactions, as well as the active window. For mouse interactions, we recorded clicks (button) and movement (distance moved in pixels), along with corresponding timestamps. Regarding the keyboard, we logged keystroke types (normal, navigating, or delete key) with timestamps. For the active window, we captured the name of the active process and its activity category (e.g., coding, emailing, writing documents), along with the timestamp of user window switches. To categorize activities, we followed the method described by Züger et al. [103]. For privacy purposes, we did not record participants' specific key strokes and the active window title. The monitoring tool was deployed on the device participants used for their primary task during the study, and deleted after the study. We connected the monitoring tool with the ambient display and notification sampling application through a web socket.

3.9 Procedure

At the beginning of the study, we explained the purpose and process of the study, with detailed information about the signals we would collect. The participants then signed a consent form and completed a demographic questionnaire inquiring about their age, gender, experience with spatially immersive interfaces, daily time spent in front of a display, and notification engagement practices.

Subsequently, participants were guided to their assigned work desk, where the researcher showed them the apparatus, introduced the sensors and the data we collected in more detail, and demonstrated the study procedure with a sped-up version of the notification response task. Following, the researcher assisted participants in connecting their personal devices or our provided computers to the external monitor on the desk. They also helped install and set up the monitoring tool, had the participants put on the eye-tracking glasses, started the ambient display and notification sampling application, and performed calibration between all the sensors. Finally, after confirming that all system components were correctly set up and logging data as required, we instructed participants to begin their work. We told participants to mainly do productive work on their primary device; however, they are also free to take breaks at any time. We emphasized that they should concentrate on their primary device and refrain from actively monitoring the ambient display, as well as avoid interacting with auxiliary devices like smartphones and tablets. Beyond these considerations, we intentionally did not specify any constraints around what task the participant should engage in, to encourage natural, everyday working behaviors. Upon receiving a notification, participants are instructed to simply continue with their work while remaining cognizant of the notification to report when prompted.

A full session of the study lasted approximately 8 hours, with start and finish times tailored to each participant's preferences and work routine. Most participants opted for a break every one to two hours, returning for multiple sessions to complete the study. Participants were compensated with a $150 gift card. The study was approved by the local IRB.

3.10 Participants

We recruited 12 participants through personal contacts and university communication channels. All were graduate or undergraduate

	Total	**per Participant**
Notifications	740 occurrences	62 occurrences (7)
Gaze & head pose	90.6 hours	7.6 hours (1.2)
Computer input	94.1 hours	7.8 hours (1.1)
Computer activity	51.6 hours	4.3 hours (3.8)

Table 1: Dataset after pre-processing. Per participant metrics are reported as M (SD).

students with an average age of 25 years (SD = 3). 7 were female and 5 were male. 10 participants were researchers from a range of disciplines, including education, design, mechanical engineering, computational fabrication, and human-computer interaction. We focused on researchers and students as one community of knowledge workers because of the diversity of their tasks, extensive computer usage in their work, and availability [4]. Our participants reported spending an average of 8 hours (SD = 2) in front of a computing display on a daily basis. On a 1 (none) to 7 (expert) scale, participants reported a relatively low amount of experience with spatially immersive interfaces (M = 2.5, SD = 1.4). This was expected as such interfaces are currently still far from ubiquitous. Participants also reported often checking for notifications while working, M = 4.8 (SD = 0.9), on a scale from 1 (never) to 7 (always). 2 out of 12 participants used their own device connected to the primary display, 5 participants opted to use a laptop we provided, and 5 set up a Remote Desktop connection to their own device via our laptop.

3.11 Data Summary

We collected a total of 98 hours of data from 12 participants (M = 8 hours per participant, SD = 1). Participants spent the majority of their time browsing the internet for work-related purposes (52%), followed by reading and editing documents and other artifacts (26%), miscellaneous tasks (5%), and planning-related tasks (3.6%; e.g., editing calendar entries). Within this time, participants were collectively presented with 793 notifications (M = 66 per participant, SD = 7). They responded to 87% (SD = 11%) of the noticeability sampling prompts they received. Based on our observations, the remaining prompts were missed entirely. We speculate this may have occurred because participants were either highly focused on their primary task, leading to "tunnel vision" [20], or leaning closer to their display, making the notifications and prompt appear further in their peripheral vision. On average, participants detected 49% (SD = 16%) of the notifications.

4 Features and Data Processing

Prior to the main analysis, we performed several pre-processing steps that are summarized as follows.

4.1 Basic Pre-processing

We initiated our pre-processing by manually inspecting the data. Figure 4 illustrates the notion of a trial within the context of our data collection. We examined the Kinect-captured video of the participants' activities in the 70-second interval centered around each notification (i.e., including the 10-second notification display and the 30 seconds before and after), the computer interaction (i.e., input and activity), gaze, and head pose data logs.

Since we were primarily interested in predicting whether users can detect peripherally-presented notifications while engaged in everyday work, we excluded trials where participants were not focused on their primary device. This included scenarios like the participant engaging in conversation with others in the office (21 occurrences) or using their personal mobile devices (32 occurrences). From our initial analysis, we also encountered issues with recording computer input and activity data. First, when participants established a Remote Desktop connection, the recorded active process was the Remote Desktop client, rather than their device's actual active process. Consequently, we excluded the computer activity data collected from these five participants. Secondly, for a specific set of trials, an oversight by the experimenter resulted in the monitoring tool not being initiated. Errors in gaze and pose data were primarily attributed to manual miscalibrations.

Out of 793 notifications (i.e., trials), we excluded 53 trials entirely, along with computer input data from 8 trials, computer activity data from 362 trials (5 participants), and gaze and head pose data from 42 trials. One participant misunderstood the instructions and used the certainty Likert scale to indicate whether they detected a notification. For them, a high score indicated that they noticed the notification and a low score indicated otherwise. We, therefore, determined their noticeability labels by dividing their Likert scale into two states (1234 for missed and 567 for noticed).

We provide a summary of the dataset after pre-processing in Table 1. Our final dataset comprised 740 notifications. In this pre-processed dataset, participants responded to 88% (SD = 10%) of the noticeability sampling prompts they received, detecting 46% (SD = 15%) of the notifications.

4.2 Feature Extraction

After pre-processing the data, we extracted features from the raw sensor logs to build the noticeability classifier. We identified features previously linked to cognitive and attentional states, as well as interruptibility. This includes metrics characterizing participants' gaze, head pose, computer input and activity, and perceived level of engagement. Table 2 provides a summary of the extracted features.

4.2.1 Gaze and head pose. The Tobii Pro 2 glasses and Vive Tracker continuously capture participants' head position, orientation, and gaze direction. By combining these data with the known positions of information widgets in the environment, we estimated the visibility of notifications in relation to the participants' visual attention using two metrics: a *head-to-widget* angle and a *gaze-to-widget* angle. The head-to-widget and gaze-to-widget angles refer to the angles between the normalized direction vector from the participant's head position to the notification and the directions of the participant's head and gaze, respectively. We also characterize the general movements of the head and gaze of the participants with their *head angular velocity*, *head positional velocity*, *gaze angle* [11], and *gaze-shift speed* [11].

4.2.2 Computer input and activity. Using the computer interaction and activity data, we compute a feature set similar to Züger et al. [103]. We extracted *keystrokes* and mouse events (i.e., *mouse clicks*, *mouse movement distance*) to approximate the user's level of interaction with their device. We also extracted application window

Feature type	Feature	Reference
Gaze (widget)	gaze-to-widget angle (min, max, mean, std)	[6, 11, 22, 25]
Gaze (general)	gaze angle (min, max, mean, std), gaze shift speed (min, max, mean, std)	
Head pose (widget)	head-to-widget angle (min, max, mean, std)	
Head pose (general)	head positional velocity (min, max, mean, std), head angular velocity (min, max, mean, std)	
Computer input	keystrokes (sum), mouse clicks (sum), mouse movement distance (sum)	[3, 40, 41, 46, 103]
Computer activity	window focus duration (max), window switches (sum), activity focus duration (max), activity switches (sum)	
Engagement	engagement	[11, 29, 59, 76, 103]

Table 2: Extracted features are grouped by type, accompanied by references to prior work where they have been defined or used. *Gaze (widget)* and *head pose (widget)* metrics define their respective relative spatial relationships to the notification, while *gaze (general)* and *head pose (general)* metrics characterize their general behavior. The bracketed and colored values additionally indicate the statistics used to characterize each feature in the time windows we analyzed.

features to measure engagement in windows and specific *activity* categories, focus duration (i.e., *in window* or *in activity*), as well as window and activity switching events (i.e., *window switches*, *activity switches*). We follow Züger et al.'s semi-automatic methodology [103] for obtaining categories, which maps windows and process names to categories like coding and email.

4.2.3 Engagement. Our noticeability sampling prompt included Likert ratings for participants to report their level of engagement with their primary task, their level of certainty in their detection, and the extent to which they perceived the notification as distracting. Since prior research has shown that user engagement influences their receptiveness to notifications (e.g., [11]), we include it as an additional feature in our analysis. We did not include participants' reports of certainty and distraction as features because these reflect their perception of the notification instead of their state at the time the notification is presented.

4.3 Time Windows

To further prepare our data for noticeability classification, we transformed our continuous time-series data streams into discrete input variables. This involved computing summary metrics for each of our extracted features within different time windows (e.g., sum, mean, max, min, standard deviation). We focused on analyzing time windows closer to the notification display (i.e., within 30 seconds) because we expected significant fluctuations in participants' attentional states and activities throughout the study [4]. The time windows we examined include: $t = 0s$ to $10s$ (i.e., during the notification display), $t = -10s$ to $20s$ (i.e., a 30-second time window centered on the notification display), $t = -30s$ to $40s$ (i.e., a 70-second time window centered on the notification display), $t = -10s$ to $0s$ (i.e., the 10-second time window prior the notification display), $t = 10s$ to $20s$ (i.e., the 10-second time window following the notification display), $t = -30s$ to $0s$ (i.e., the 30-second time window prior the notification display), and $t = 10s$ to $40s$ (i.e., the 30-second time window following the notification display).

5 Sensing Noticeability

To assess the predictive capabilities of our data on participants' detection of peripherally-presented notifications, we applied machine learning on pre-processed features, with participants' self-reports serving as the ground truth. In the following, we compare the performance of different classification algorithms, explore feature extraction from continuous data streams with varying time windows, and investigate different feature combinations. Additionally, we assess the scalability of our approach and compare the performance of a personalized model to a general classifier.

5.1 Evaluation Method & Metrics

In all our experiments, unless otherwise specified (e.g., Section 5.6), we adopted a leave-one-participant-out cross-validation method (LOOCV) to evaluate different approaches. One participant's data was used as test data, while the data from the remaining 11 participants were used as training data. The results were averaged over 10 runs. We quantify performance using standard machine learning metrics, including accuracy, recall, precision, F1-score, and the area under the receiver operating characteristic curve (AUC). For the sake of conciseness, we focus on reporting the AUC. Additional metrics and experiments are detailed in Appendix A.

5.2 Overview of Results

Overall, our results suggest that inferring whether participants detected a peripheral notification with sensor data is feasible. Specifically, a gradient boosting pipeline achieved an AUC of 0.78 using all available sensor features. By refining our time windowing approach and feature selection, we achieved a higher AUC of 0.81.

Through our analysis, we observed that users' gaze relative to the notification display and their engagement at the time a peripheral notification is shown were the most informative for predicting noticeability. Additionally, while a general model may not fully capture all individual behavioral differences, it still demonstrates reasonable performance, even when trained on limited user data.

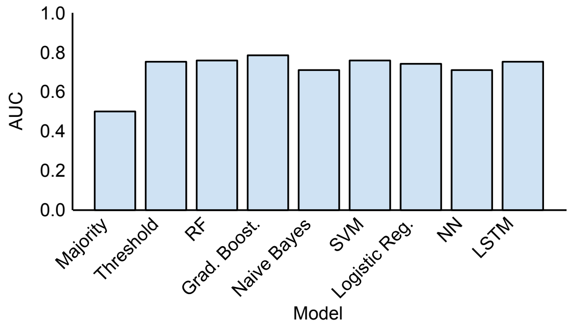

Figure 5: Comparison of classification approaches.

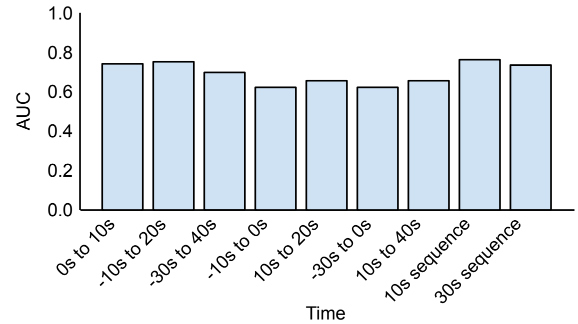

Figure 6: Comparison of models implemented using features extracted from different time windows.

5.3 Classification Approaches

We initially evaluated several traditional machine learning classifiers, as cited in the interruptibility and notification literature [11, 90]. These classifiers include a random forest classifier, gradient boosting classifier, AdaBoost classifier, Naive Bayes classifier, Support Vector Machine (SVM), neural network (NN), and logistic regression. All classifiers were implemented using scikit-learn [75].

Next, given that approaches based on Long Short-term Memory (LSTM) neural networks have demonstrated potential in time-series classification tasks such as interruptibility prediction [11] and gaze pattern recognition [88], we decided to investigate their classification performance as well. We implemented a LSTM model using PyTorch [74]. As input, we experimented with features extracted from consecutive $10s$ and $30s$ time windows.

Lastly, we evaluated two baseline models: a majority classifier and an optimal threshold classifier. The majority classifier always predicts the more prevalent class, which in this case is that the participant missed the notification. Our optimal threshold classifier predicts whether a notification has been detected or missed based on a single variable. Through experiments, we found that thresholding using the minimum gaze angle extracted from the time window $t = 0s$ to $t = 10s$ yielded the best performance.

Details on the design and tuning of the aforementioned models are provided in Appendix B.

5.3.1 Results. Figure 5 compares the classification approaches we evaluated. Overall, the gradient boosting classifier achieved the highest performance, with an AUC of 0.78. Excluding the majority classifier, all other approaches, including our naive unimodal thresholding baseline, also demonstrated reasonable performance, with AUCs of at least 0.71. These results not only suggest that **it may be feasible to predict whether participants will detect a peripheral notification in AR while engaged in knowledge work**, but also indicate that the collected **sensor data contains features with high predictive power that do not necessarily require a complex model to learn**.

For the remainder of the paper, we present results using the gradient boosting model, as it achieved comparatively better performance without requiring any feature pre-selection.

5.4 Time Windows

To investigate the predictive power of our features extracted from different time windows, we trained separate gradient boosting classifiers using each window feature set (i.e., seven models, one for each time window listed in Section 4.3). Additionally, we examined the performance of the model when using a combination of features extracted from $10s$ and $30s$ time window sequences.

5.4.1 Results. The prediction performance using different time windows is presented in Figure 6. Overall, models using features from a time window that overlapped with the notification generally performed best, with the most accurate achieving a 0.78 AUC (i.e., using features from $t = -10s$ to $20s$). This suggests that **users' activities near the time a peripheral notification is presented are the most informative for predicting noticeability**. However, features from time windows capturing activities before or after a notification also demonstrated reasonable performance. For example, signals extracted from 10-second windows before or after a notification achieved AUCs of 0.64 and 0.68, respectively. This suggests that **behaviors preceding the notification window may serve as predictors of participants' impending notification detection**. Similarly, the predictive signal conveyed by features from the time window following the notification indicates that **receiving the notification may have influenced the participants' subsequent behaviors**.

Lastly, our results show that using a $10s$ window sequence (AUC = 0.79) instead of a single 10- (AUC = 0.77) or 30-second window (AUC = 0.78) provides a slight boost in performance. Though the difference is comparatively minor, this nonetheless suggests that explicitly providing the model with inputs representing participants' interactions and behaviors before and after the notification may beneficially enhance contextualization.

5.5 Feature Importance

To explore which combinations of features best predicted noticeability, we implemented gradient boosting classifiers using data from each sensor individually and combinations thereof. We utilized features extracted from $10s$ window sequences, guided by the results in Section 5.4. In addition, we quantify feature importance using Gini impurity from scikit-learn [75]. This measure captures each

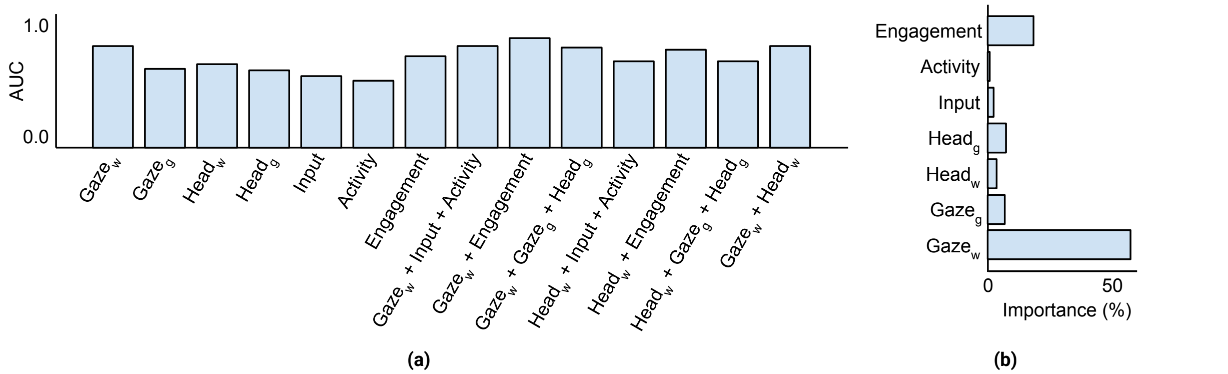

Figure 7: Feature Importance. (a) Comparison of models implemented using different feature combinations. (b) Gini impurity. Subscripts w and g denote widget and general.

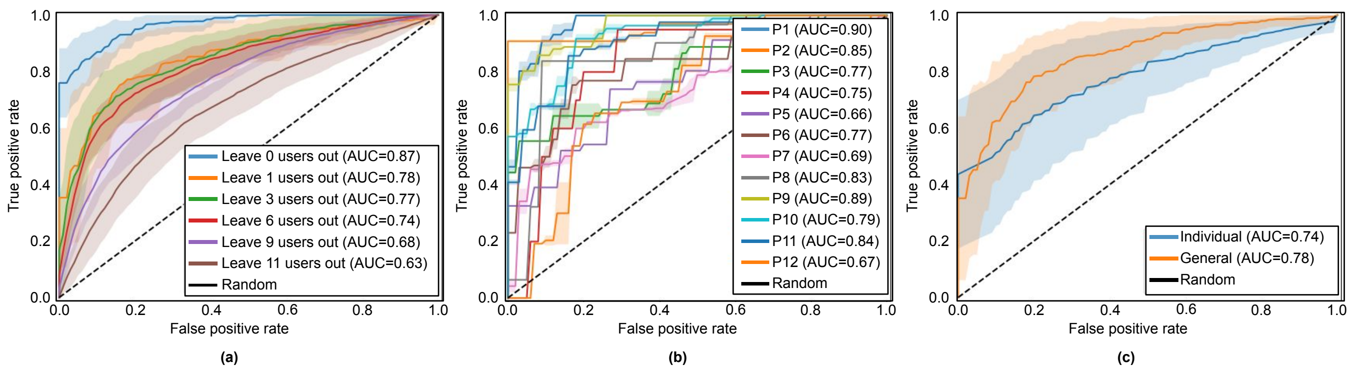

Figure 8: ROC curves. (a) Cross-validation results obtained by leaving different numbers of users out. (b) Results from leave-one-participant-out cross-validation using a general model. (c) Comparison of a general versus an individual model.

feature's contribution in a gradient boosting classifier to influence the classification outcome. We calculated Gini impurity using a gradient boosting classifier constructed with all features extracted from the 10s window sequence.

5.5.1 Results. Figure 7a presents the prediction results using different feature combinations. Figure 7b presents the computed feature importances. Overall, the results suggest that **gaze relative to the notification display is the strongest predictor of noticeability**, achieving the highest AUC (0.75) and emerging as the most important feature (58%) by a significant margin. **Engagement emerged as the second most predictive feature**, achieving an AUC of 0.69 and Gini impurity of 18.8%. **Combining the aforementioned gaze and engagement features yielded our best-performing model**, achieving an AUC of 0.81. However, combining additional features did not improve performance; in some cases, it even resulted in decreased prediction accuracy. This suggests that **not all sensors were complementary, and introducing certain sensors may instead add noise, adversely impacting performance**.

Excluding gaze-related metrics, features characterizing participants' head pose relative to the notification display still achieved reasonable above-chance performance (AUC = 0.63), particularly when combined with engagement (AUC = 0.73). On the other hand, computer input and activity were the least important features, with Gini impurity scores of 2.9% and 1.0%, respectively. Using a classifier based solely on computer input and activity similarly yielded the lowest AUC scores, with 0.53 and 0.51, respectively.

5.6 Scalability

We assessed the scalability of our gradient boosting classifier to new users by conducting a leave-1, 2, 4, 6, 8, 10, 11-participants-out cross-validation. For exploration, we also performed a leave-0-participants-out cross-validation.

5.6.1 Results. Figure 8a presents the results of our experiments on scalability. Leaving out data from 10 participants resulted in an AUC of 0.68, compared to 0.78 when excluding one user. This indicates that **while the model benefits from training on data from more participants, it may still generalize well enough to provide reasonable performance even with limited user data**. Our leave-0-participants-out cross validation resulted in an AUC of 0.89, suggesting some degree of ambiguity in the data, as the model could not perfectly match the data of users even when they are included in the training set.

5.7 Personalization

To examine the extent to which a general model may account for individual differences, we analyzed individual ROC plots generated during leave-one-participant-out cross-validation. We also investigated whether access to individual data could enhance model performance by training personalized models for each participant. For this, we used the same gradient boosting pipeline to generate learning curves for each individual through shuffle split cross-validation (100 splits, with a test size of 20% of the available samples).

5.7.1 Results. As shown in Figure 8b, while our classifier achieves reasonable accuracy on average, its performance varies significantly across participants. This suggests that **a general model may not effectively capture all individual behavioral differences**.

Figure 8c shows ROC curves of a general versus individual model. On average, individual classifiers achieved a lower AUC (0.74) compared to the general models. However, on a participant-by-participant basis, individual models sometimes outperformed the general model, as evidenced by overlaps in the learning curves. This suggests that while personalized models did not significantly improve overall performance, accounting for individual differences may still be beneficial. An important caveat to consider, however, is that the general models were trained on over 10 times the amount of data available to the individualized models (i.e., data from 11 participants compared to only 80% of one participant's data), which may have biased these results.

6 Discussion

In this work, we investigated the feasibility of predicting people's detection of peripherally presented notifications using gaze, head pose, computer activity, and self-reported levels of engagement. To this end, we collected and performed a machine learning-based analysis of data on interactions with notifications across 12 participants. Based on our results, we discuss implications for ambient interfaces and opportunities for future work.

6.1 Usage

In general, a classifier for the noticeability of interface elements supports several novel interactive functionalities. First, it enables acknowledgment of notifications based on implicit behaviors rather than explicit feedback, thus reducing unnecessary interruptions to the user's workflow. Second, it provides the system with knowledge of whether the user should be prompted again (i.e., in cases where the notification has been classified as missed). Finally, it opens the door to less obtrusive notification delivery by gradually increasing the visual saliency of a notification while monitoring whether it is noticed. Specifically, with our current approach, we may gradually increase the frequency at which a virtual element fades while checking the sensor signals at 10-second intervals to determine if it has been detected. This allows notifications to be presented in a just-noticeable manner.

A classifier that can predict a user's susceptibility to a notification (i.e., whether they will notice an upcoming notification) offers further support for the intelligent delivery of notifications. Based on predictions of whether a user will notice a notification, the system may adjust the saliency of its presentation to strike a balance between being noticeable and not overly obtrusive. If the classifier determines that a user is unlikely to notice an upcoming notification, it can be presented in a more salient manner (e.g., increased oscillation speed for our transparency adjustment effect). In the converse case, the presentation can be reduced in saliency.

6.2 Implications

Our current results demonstrate that constructing a classification-based approach for noticeability using sensors is feasible. While our sensors provide some signals for predicting future noticeability, this task proves to be more challenging in comparison. Our learning-based analysis further provides the following insights for enabling the use of sensor-based approaches to classify noticeability:

First, since features extracted from the 10s time window surrounding the notification were most predictive, future systems should **prioritize capturing user behavior signals overlapping with the notification for optimal noticeability detection**.

Second, based on our ablation studies and analyses of feature importance, future systems should **focus on capturing the user's gaze or head pose relative to the notification**. In contrast, the user's interactions with their devices may be less valuable as a signal to capture, as they offer less predictive power.

Our results further suggest that **understanding the user's level of engagement with their primary task will benefit noticeability prediction**; however, capturing this information may pose additional challenges. While the metrics like gaze, head pose, and input can be collected directly from off-the-shelf sensors if the users' workspace and devices are appropriately instrumented, engagement is a property of the user's cognitive state and can currently only be self-reported or imperfectly inferred from proxy metrics like EEG [15] and video data [83]. That said, while incorporating engagement achieved the best performance ($AUC = 0.81$), excluding the engagement feature still enables the development of reasonably performing models (i.e., achieving an AUC of 0.76 using gaze alone). Notably, an AUC of 0.76 is comparable to performance achieved by prior work on modeling noticeability [61] and interruptibility [103], some of which were conducted in more controlled contexts.

The combination of sensors can ultimately be tailored to specific usage requirements. For instance, our gaze-based model may be sufficient for lower-stakes scenarios where classification errors have minimal consequences for the user experience, such as displaying a non-critical weather alert. In this example, if the notification is set to be dismissed implicitly based on the model classification result, a false positive prediction (e.g., incorrectly identified as noticed) may be acceptable, as the information does not require urgent attention. We believe that it is, in fact, in these scenarios where subtlety is the objective that our approach may be most applicable. In contrast, in higher-stakes scenarios (e.g., a reminder for an interview), notifications should arguably be optimized for maximum noticeability, which is how they are effectively designed now. That said, awareness of noticeability can still be valuable in these settings, highlighting future opportunities to improve our models. Our approach lays the groundwork for more complex and, hopefully, more accurate models that include measures such as engagement.

Lastly, our results suggest that **constructing a general model offers reasonable performance**. Therefore, while a general model is not robust to individual differences, systems can likely begin with a model constructed from data captured from multiple users rather than relying solely on personalized data collection.

Our results offer practical insights for both projection- and HMD-based AR systems. For projection-based displays, our results suggest that additional instrumentation for gaze and head-pose tracking is needed to facilitate noticeability sensing. Our apparatus was also designed to mimic the characteristics of a lightweight HMD, free from the current device and ergonomic constraints (e.g., limited field of view) and extended with additional sensing modalities (e.g., gaze). Our notifications effectively represent world-fixed AR elements, a common placement pattern where the information is fixed the physical environment and displayed in situ [27, 62, 85]. Current commercial HMDs (e.g., Apple Vision Pro) can similarly characterize a user's gaze and head pose relative to world-fixed AR elements, and we therefore anticipate that our findings may translate directly. Additionally, though the sensors on current HMDs capture similar information (e.g., gaze, head pose), they are not identical to the ones used in our apparatus, as discussed below.

6.3 Limitations and Future Work

Our current work is subject to several limitations, which we describe in the following.

6.3.1 Towards "In-the-Wild" conditions. The first concern is whether the study conditions were sufficiently realistic. We conducted our study in an office where the usual occupants worked alongside the participants and engaged in their typical activities, provided they did not physically interfere with the ambient display. While this setup enabled conditions closer to real working environments, such as prompting cases of bystander interference, it may have nonetheless precluded a fully "in-the-wild" experience. For instance, although other office occupants were not explicitly instructed to avoid interacting with the study participants, they may have felt disinclined to do so due to concerns about interfering with the research. Additionally, the study required participants to work in a foreign environment under consistent monitoring, which may have also inhibited their natural work behaviors. We believe the extensive eight-hour study period we used would have allowed participants to acclimate; however, future iterations could always consider extending the deployment duration. To addressing these challenges, we are currently modularizing the setup to support deployment participants' own productivity environments.

6.3.2 Generalizability across environments, tasks, and displays. We conducted our study with participants working in a specific office. This precludes many other work settings. In addition to an office setting, for example, knowledge workers typically work in libraries, coffee shops, and other public spaces. Even if we were to constrain the target environment to an office, offices differ in furniture arrangement and size.

Besides differences in environments, while our participants performed a variety of canonical knowledge work tasks, such as browsing and reading documents, these tasks are by no means exhaustive. Furthermore, beyond knowledge work, AR has also been applied to a variety of usage scenarios, such as social settings (e.g., [10]) and on-the-go information consumption (e.g., [58]).

Lastly, in AR settings, virtual interface elements can also vary significantly in their presentation. In addition to the surface-embedded method of presentation we use, virtual elements can be presented in mid-air [12] or anchored to the user's field of view [62].

Ultimately, additional research is required to validate its relevance across a broader range of environments, tasks, and displays.

6.3.3 Notification timing. Our study randomly introduces notifications at 5 to 10-minute intervals. This decision was intended to strike a balance between collecting sufficient samples for performing learning-based analysis and presenting notifications at an ecologically valid frequency. While we somewhat perturbed the interval between notifications, the periodic delivery of notifications may not be representative of real life, where notifications can arrive in bursts or not at all for extended durations. This also raises concerns about participants' ability to anticipate notification arrivals in advance. From our observations and informal conversations with participants, we did not find that participants were actively expecting notifications; nonetheless, for greater ecological validity, future iterations can consider adopting more varied and naturalistic patterns of notification delivery.

6.3.4 Projection-based approach. In our current apparatus, we used a projection-based approach to circumvent the limitations of current head-mounted AR displays, such as their weight and limited field of view. We captured signals similar to those available on HMDs (e.g., gaze, head pose) under quasi-ideal tracking conditions. Our decisions were guided by the goal of ensuring that our findings would remain relevant not only to present technologies but also to future advancements, offering more seamless and ubiquitous digital information displays. That said, this raises questions about whether our findings can directly inform the design of current head-mounted displays, where the sensors used may differ and suffer from inaccuracies. Therefore, we see value in replicating the experiment with current and future headsets, particularly in more challenging environments that introduce sensing difficulties.

6.3.5 Single device vs multi device. In our study, we focused on contexts where participants were fully engaged in productive work on a single desktop device. However, current users engage in a significantly wider variety of tasks within multi-device ecologies that include diverse screen form factors and arrangements [99]. Extending the considered activities and screen configurations within our model, therefore, serves as a valuable direction for future work.

6.3.6 Recall vs immediate response. Our study might also be constrained by the selection of notifications and the procedures used for their presentation. First and foremost, in our study, we assessed the noticeability of 10-second animations using a cued recall approach. The choice of a 10-second duration was justified by its equivalence to receiving two to three phone ringtones. We specifically chose a cued recall paradigm after pilot testing revealed that asking for an immediate response made it challenging to differentiate between signals arising from detecting the notification and participants' responses. This approach, however, suffers from several limitations. First, while we could determine if participants detected a notification within the 10-second window, this method prevented us from pinpointing the exact moment of detection. Second, this approach required participants to keep the notification in mind until the recall prompt, which may have influenced their subsequent behavior and involved a memory bias. To address this, future studies may benefit from employing alternative methodologies that provide a more granular understanding of participants' notification detection. A direct comparison of our cued recall design, having participants respond immediately, and other methodologies may further offer interesting insights.

6.3.7 Notification type. Our study exclusively evaluated one type of notification: a periodic opacity change effect. We chose this specific design for its aesthetic appeal and avoided introducing diverse notification designs as additional variables. This decision was motivated by the data requirements of our learning-based approach, aiming for simplicity and clarity in the assessment of noticeability. Our evaluated notification is solely visual, and this characteristic may lead to different responses compared to alternative modalities such as audio and haptics. For instance, while our results suggest that gaze is a strong predictor for visual notifications, this predictive power may not extend to auditory notifications. Therefore, we suggest that future research explore a broader range of multimodal notification designs to enhance the generalizability of the findings to diverse user interfaces and scenarios.

6.3.8 Missing features. As discussed in Section 4.1, not every one of the collected samples contained all features, which might have influenced our analysis of their comparative predictive value. We imputed missing values by replacing them with the mean before classification, as this technique can lead to better results than discarding them, which would decrease the sample size.

6.3.9 Ethical Considerations. While the objective of our work was to establish the technical foundations for creating ambient information interfaces that adapt the noticeability of virtual elements based on user context, it is important to note that such an approach could inadvertently be misused, potentially contributing to further notification overload. In particular, with the ability to predict whether a user will detect a notification, systems could also be optimized to make notifications subtly inescapable. We, therefore, believe it is critical for users to retain agency over when such adaptations are applied, including more granular control over which applications or types of information are affected. Future work should consider these risks and integrate protection mechanisms to mitigate them.

7 Conclusion

In this paper, we present the results of a study involving 12 knowledge workers who each worked for an average of 8 hours, where we examined gaze, head pose, computer interaction, and self-reported features to predict interactions with an ambient notification display. Our analysis demonstrates that, through the implementation of a model relying on gaze and engagement metrics, we can infer noticeability with up to 0.81 AUC. Even when excluding engagement metrics, which might be challenging to acquire, we can still achieve an AUC of 0.76. We further show that models benefit from a characterization of the user's level of engagement, while features measuring their interactions and activities on their primary device contributed comparatively less. Finally, we show that a general model provides a good starting point for noticeability prediction and that it does not degrade to unusable even when only trained on a single participant's data. Overall, our results demonstrate the feasibility of identifying when users either detect or miss notifications using sensor data. We believe future approaches to predicting noticeability can serve as the foundations of ambient information interfaces that automatically adapt virtual elements to be noticeable when needed and unobtrusive otherwise.

Acknowledgments

We thank our project sponsors and collaborators from Honda Research Institute USA, particularly Erin Clepper, Joan Smith, and Duane Detwiler. We also thank all involved peers, participants, and anonymous reviewers, especially Nikolas Martelero, Howie Wang, Karan Ahuja, and Riku Arakawa for their input throughout the project. Yi Fei Cheng received partial support from the Croucher Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 58 (jul 2018), 27 pages. https://doi.org/10.1145/3214261

[2] Nivedita Arora, Ali Mirzazadeh, Injoo Moon, Charles Ramey, Yuhui Zhao, Daniela C. Rodriguez, Gregory D. Abowd, and Thad Starner. 2021. MARS: Nano-Power Battery-Free Wireless Interfaces for Touch, Swipe and Speech Input. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 1305–1325. https://doi.org/10.1145/3472749.3474823

[3] Ernesto Arroyo and Ted Selker. 2011. Attention and intention goals can mediate disruption in human-computer interaction. In *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part II 13*. Springer, 454–470.

[4] Ebrahim Babaei, Namrata Srivastava, Joshua Newn, Qiushi Zhou, Tilman Dingler, and Eduardo Velloso. 2020. Faces of Focus: A Study on the Facial Cues of Attentional States. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376566

[5] Patrick Bader, Alexandra Voit, Huy Viet Le, Paweł W. Woźniak, Niels Henze, and Albrecht Schmidt. 2019. WindowWall: Towards Adaptive Buildings with Interactive Windows as Ubiquitous Displays. *ACM Trans. Comput.-Hum. Interact.* 26, 2, Article 11 (mar 2019), 42 pages. https://doi.org/10.1145/3310275

[6] Marilou Beyeler, Yi Fei Cheng, and Christian Holz. 2023. Cross-Device Shortcuts: Seamless Attention-Guided Content Transfer via Opportunistic Deep Links between Apps and Devices. In *Proceedings of the 25th International Conference on Multimodal Interaction* (Paris, France) *(ICMI '23)*. Association for Computing Machinery, New York, NY, USA, 125–134. https://doi.org/10.1145/3577190.3614145

[7] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272. https://doi.org/10.1561/1100000049

[8] Ali Borji and Laurent Itti. 2013. State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1 (2013), 185–207. https://doi.org/10.1109/TPAMI.2012.89

[9] R.A. Brooks. 1997. The Intelligent Room project. In *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age*. 271–278. https://doi.org/10.1109/CT.1997.617707

[10] Runze Cai, Nuwan Nanayakkarawasam Peru Kandage Janaka, Shengdong Zhao, and Minghui Sun. 2023. ParaGlassMenu: Towards Social-Friendly Subtle Interactions in Conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.

[11] Kuan-Wen Chen, Yung-Ju Chang, and Liwei Chan. 2022. Predicting Opportune Moments to Deliver Notifications in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 186, 18 pages. https://doi.org/10.1145/3491102.3517529

[12] Yi Fei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 282–297. https://doi.org/10.1145/3472749.3474750

[13] Hyunsung Cho, Yukang Yan, Kashyap Todi, Mark Parent, Missie Smith, Tanya R. Jonker, Hrvoje Benko, and David Lindlbauer. 2024. MineXR: Mining Personalized Extended Reality Interfaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 609, 17 pages. https://doi.org/10.1145/3613904.3642394

[14] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR Using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proceedings of the 7th Augmented Human International Conference 2016* (Geneva, Switzerland) *(AH '16)*. Association for Computing Machinery, New York, NY, USA, Article 3, 8 pages. https://doi.org/10.1145/2875194.2875204

[15] Federico Cirett Galán and Carole R Beal. 2012. EEG estimates of engagement and cognitive workload predict math problem solving outcomes. In *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings 20*. Springer, 51–62.

[16] 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. Citeseer, 361–367.

[17] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant messaging: Effects of relevance and time. In *Proceedings of HCI 2000*. British Computer Society, 71–76.

[18] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A Diary Study of Task Switching and Interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 175–182. https://doi.org/10.1145/985692.985715

[19] Nicholas S. Dalton, Emily Collins, and Paul Marshall. 2015. Display Blindness? Looking Again at the Visibility of Situated Displays Using Eye-Tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3889–3898. https://doi.org/10.1145/2702123.2702150

[20] Anwesha Das, Zekun Wu, Iza Skrjanec, and Anna Maria Feit. 2024. Shifting Focus with HCEye: Exploring the Dynamics of Visual Highlighting and Cognitive Load on User Attention and Saliency Prediction. *Proc. ACM Hum.-Comput. Interact.* 8, ETRA, Article 236 (May 2024), 18 pages. https://doi.org/10.1145/3655610

[21] Davide De Tommaso and Agnieszka Wykowska. 2019. TobiiGlassesPySuite: An Open-source Suite for Using the Tobii Pro Glasses 2 in Eye-tracking Studies. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. ACM, New York, NY, USA, Article 46, 5 pages. https://doi.org/10.1145/3314111.3319828

[22] Anup Doshi and Mohan M Trivedi. 2012. Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of vision* 12, 2 (2012), 9–9.

[23] Mica R Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human factors* 37, 1 (March 1995), 32–64. https://doi.org/10.1518/001872095779049543

[24] Mica R Endsley, Betty Bolté, and Debra G Jones. 2003. *Designing for situation awareness: An approach to user-centered design*. CRC press. https://www.taylorfrancis.com/books/mono/10.1201/9780203485088/designing-situation-awareness-mica-endsley-betty-bolte-debra-jones

[25] Julie Epelboim, Robert M Steinman, Eileen Kowler, Zygmunt Pizlo, Casper J Erkelens, and Han Collewijn. 1997. Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision research* 37, 18 (1997), 2597–2607.

[26] Anja Exler, Christian Dinse, Zeynep Günes, Nadim Hammoud, Steffen Mattes, and Michael Beigl. 2017. Investigating the Perceptibility Different Notification Types on Smartphones Depending on the Smartphone Position. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 970–976. https://doi.org/10.1145/3123024.3124560

[27] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology* (Atlanta, Georgia, USA) *(UIST '93)*. Association for Computing Machinery, New York, NY, USA, 145–155. https://doi.org/10.1145/168642.168657

[28] James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. 2005. Predicting Human Interruptibility with Sensors. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (mar 2005), 119–146. https://doi.org/10.1145/1057237.1057243

[29] James Fogarty, Amy 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* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 331–340. https://doi.org/10.1145/1054972.1055018

[30] Sarthak Ghosh, Lauren Winston, Nishant Panchal, Philippe Kimura-Thollander, Jeff Hotnog, Douglas Cheong, Gabriel Reyes, and Gregory D. Abowd. 2018. NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1447–1456. https://doi.org/10.1109/TVCG.2018.2793698

[31] Jennifer Gluck, Andrea Bunt, and Joanna McGrenere. 2007. Matching Attentional Draw with Utility in Interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/1240624.1240631

[32] Matt Gottsacker, Nahal Norouzi, Kangsoo Kim, Gerd Bruder, and Greg Welch. 2021. Diegetic Representations for Seamless Cross-Reality Interruptions. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 310–319. https://doi.org/10.1109/ISMAR52148.2021.00047

[33] Steve Grogorick, Michael Stengel, Elmar Eisemann, and Marcus Magnor. 2017. Subtle Gaze Guidance for Immersive Environments. In *Proceedings of the ACM Symposium on Applied Perception* (Cottbus, Germany) *(SAP '17)*. Association for Computing Machinery, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3119881.3119890

[34] Carl Gutwin, Andy Cockburn, and Ashley Coveney. 2017. Peripheral Popout: The Influence of Visual Angle and Stimulus Intensity on Popout Effects. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 208–219. https://doi.org/10.1145/3025453.3025984

[35] Violet Yinuo Han, Hyunsung Cho, Kiyosu Maeda, Alexandra Ion, and David Lindlbauer. 2023. BlendMR: A Computational Method to Create Ambient Mixed Reality Interfaces. *Proc. ACM Hum.-Comput. Interact.* 7, ISS, Article 436 (nov 2023), 25 pages. https://doi.org/10.1145/3626472

[36] Steven J. Henderson and Steven Feiner. 2008. Opportunistic Controls: Leveraging Natural Affordances as Tangible User Interfaces for Augmented Reality. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology* (Bordeaux, France) *(VRST '08)*. Association for Computing Machinery, New York, NY, USA, 211–218. https://doi.org/10.1145/1450579.1450625

[37] Joyce Ho and Stephen S. Intille. 2005. Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 909–918. https://doi.org/10.1145/1054972.1055100

[38] ECMCE Horvitz. 2001. Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In *Human-Computer Interaction: INTERACT*, Vol. 1. 263.

[39] Eric Horvitz and Johnson Apacible. 2003. Learning and Reasoning about Interruption. In *Proceedings of the 5th International Conference on Multimodal Interfaces* (Vancouver, British Columbia, Canada) *(ICMI '03)*. Association for Computing Machinery, New York, NY, USA, 20–27. https://doi.org/10.1145/958432.958440

[40] Eric Horvitz, Paul Koch, and Johnson Apacible. 2004. BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 507–510. https://doi.org/10.1145/1031607.1031690

[41] Eric J Horvitz, Paul Koch, Carl Kadie, and Andy Jacobs. 2012. Coordinates: probabilistic forecasting of presence and availability. *arXiv preprint arXiv:1301.0573* (2012).

[42] Ching-Yu Hsieh, Yi-Shyuan Chiang, Hung-Yu Chiu, and Yung-Ju Chang. 2020. Bridging the Virtual and Real Worlds: A Preliminary Study of Messaging Notifications in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376228

[43] Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao. 2015. SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

[44] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. ACM, New York, NY, USA, 257–264. https://doi.org/10.1145/642611.642657

[45] Apple Inc. 2023. Add, edit, and remove widgets on iPhone. https://support.apple.com/guide/iphone/add-edit-and-remove-widgets-iphb8f1bf206/ios. Accessed: 2022-10-20.

[46] Shamsi T. Iqbal and Brian P. Bailey. 2007. Understanding and Developing Models for Detecting and Differentiating Breakpoints during Interactive Tasks.

In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 697–706. https://doi.org/10.1145/1240624.1240732

[47] Shamsi T. Iqbal and Brian P. Bailey. 2011. Oasis: A Framework for Linking Notification Delivery to the Perceptual Structure of Goal-Directed Tasks. *ACM Trans. Comput.-Hum. Interact.* 17, 4, Article 15 (dec 2011), 28 pages. https://doi.org/10.1145/1879831.1879833

[48] Shamsi T. Iqbal and Eric Horvitz. 2010. Notifications and Awareness: A Field Study of Alert Usage and Preferences. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 27–30. https://doi.org/10.1145/1718918.1718926

[49] Hiroshi Ishii, Craig Wisneski, Scott Brave, Andrew Dahley, Matt Gorbet, Brygg Ullmer, and Paul Yarin. 1998. AmbientROOM: Integrating Ambient Media with Architectural Space. In *CHI 98 Conference Summary on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. Association for Computing Machinery, New York, NY, USA, 173–174. https://doi.org/10.1145/286498.286652

[50] L. Itti, C. Koch, and E. Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259. https://doi.org/10.1109/34.730558

[51] Nassim Jafarinaimi, Jodi Forlizzi, Amy Hurst, and John Zimmerman. 2005. Breakaway: An Ambient Display Designed to Change Human Behavior. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1945–1948. https://doi.org/10.1145/1056808.1057063

[52] Nuwan Janaka, Jie Gao, Lin Zhu, Shengdong Zhao, Lan Lyu, Peisen Xu, Maximilian Nabokow, Silang Wang, and Yanch Ong. 2023. GlassMessaging: Towards Ubiquitous Messaging Using OHMDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–32.

[53] Nuwan Janaka, Chloe Haigh, Hyeongcheol Kim, Shan Zhang, and Shengdong Zhao. 2022. Paracentral and Near-Peripheral Visualizations: Towards Attention-Maintaining Secondary Information Presentation on OHMDs during in-Person Social Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 551, 14 pages. https://doi.org/10.1145/3491102.3502127

[54] Nuwan Nanayakkarawasam Peru Kandage Janaka, Shengdong Zhao, and Shardul Sapkota. 2023. Can icons outperform text? understanding the role of pictograms in ohmd notifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–23.

[55] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-Camera Units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 637–644. https://doi.org/10.1145/2642918.2647383

[56] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1011–1020. https://doi.org/10.1145/2858036.2858359

[57] May Jorella Lazaro, Sungho Kim, Jaeyong Lee, Jaemin Chun, and Myung-Hwan Yun. 2021. Interaction Modalities for Notification Signals in Augmented Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 470–477. https://doi.org/10.1145/3462244.3479898

[58] Hyunjin Lee and Woontack Woo. 2023. Exploring the effects of augmented reality notification type and placement in AR HMD while walking. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 519–529.

[59] Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, Chia-Yu Chen, Yu-Lin Chung, Yung-Ju Chang, and Chien-Ru Sun. 2019. Does Who Matter? Studying the Impact of Relationship Characteristics on Receptivity to Mobile IM Messages. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300756

[60] Jaewook Lee, Fanjie Jin, Younsoo Kim, and David Lindlbauer. 2022. User Preference for Navigation Instructions in Mixed Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 802–811. https://doi.org/10.1109/VR51125.2022.00102

[61] Zhipeng Li, Yi Fei Cheng, Yukang Yan, and David Lindlbauer. 2024. Predicting the Noticeability of Dynamic Virtual Elements in Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 954, 17 pages. https://doi.org/10.1145/3613904.3642399

[62] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 147–160. https://doi.org/10.1145/3332165.3347945

[63] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum. 2011. Learning to Detect a Salient Object. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 2 (2011), 353–367. https://doi.org/10.1109/TPAMI.2010.70

[64] Aristides Mairena, Carl Gutwin, and Andy Cockburn. 2019. Peripheral Notifications in Large Displays: Effects of Feature Combination and Task Interference. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300870

[65] Jennifer Mankoff, Anind K. Dey, Gary Hsieh, Julie Kientz, Scott Lederer, and Morgan Ames. 2003. Heuristic Evaluation of Ambient Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 169–176. https://doi.org/10.1145/642611.642642

[66] Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The Cost of Interrupted Work: More Speed and Stress. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 107–110. https://doi.org/10.1145/1357054.1357072

[67] Tara Matthews, Anind K. Dey, Jennifer Mankoff, Scott Carter, and Tye Rattenbury. 2004. A Toolkit for Managing User Attention in Peripheral Displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (Santa Fe, NM, USA) *(UIST '04)*. Association for Computing Machinery, New York, NY, USA, 247–256. https://doi.org/10.1145/1029632.1029676

[68] D. Scott McCrickard and C. M. Chewar. 2003. Attuning Notification Design to User Goals and Attention Costs. *Commun. ACM* 46, 3 (mar 2003), 67–72. https://doi.org/10.1145/636772.636800

[69] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[70] Andre N. Meyer, Gail C. Murphy, Thomas Zimmermann, and Thomas Fritz. 2017. Design Recommendations for Self-Monitoring in the Workplace: Studies in Software Development. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 79 (dec 2017), 24 pages. https://doi.org/10.1145/3134714

[71] Philipp Müller, Sander Staal, Mihai Bâce, and Andreas Bulling. 2022. Designing for Noticeability: Understanding the Impact of Visual Importance on Desktop Notifications. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 472, 13 pages. https://doi.org/10.1145/3491102.3501954

[72] Kristine S. Nagel, James M. Hudson, and Gregory D. Abowd. 2004. Predictors of Availability in Home Life Context-Mediated Communication. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 497–506. https://doi.org/10.1145/1031607.1031689

[73] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review* 18, 2 (2014), 20–31.

[74] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[75] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[76] 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* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[77] Julian Petford, Iain Carson, Miguel A. Nacenta, and Carl Gutwin. 2019. A Comparison of Notification Techniques for Out-of-View Objects in Full-Coverage

Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300288

[78] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 91 (sep 2017), 25 pages. https://doi.org/10.1145/3130956

[79] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014. An In-Situ Study of Mobile Phone Notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 233–242. https://doi.org/10.1145/2628363.2628364

[80] Lucas Plabst, Sebastian Oberdörfer, Francisco Raul Ortega, and Florian Niebling. 2022. Push the red button: comparing notification placement with augmented and non-augmented tasks in AR. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction.* 1–11.

[81] Vasili Ramanishka, Abir Das, Jianming Zhang, and Kate Saenko. 2017. Top-Down Visual Saliency Guided by Captions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[82] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)*. Association for Computing Machinery, New York, NY, USA, 179–188. https://doi.org/10.1145/280814.280861

[83] Ognjen Rudovic, Hae Won Park, John Busche, Björn Schuller, Cynthia Breazeal, and Rosalind W Picard. 2019. Personalized estimation of engagement from videos using active learning with deep reinforcement learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 217–226.

[84] Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 30, 11 pages. https://doi.org/10.1145/3419249.3420095

[85] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 199–211. https://doi.org/10.1145/3311350.3347190

[86] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-Scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[87] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1633–1642. https://doi.org/10.1109/TVCG.2018.2793599

[88] Mikhail Startsev, Ioannis Agtzidis, and Michael Dorr. 2019. 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods* 51 (2019), 556–572.

[89] Anne M. Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97 – 136. https://doi.org/10.1016/0010-0285(80)90005-5

[90] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 801–812. https://doi.org/10.1145/2750858.2807514

[91] Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg. 2011. Directing Attention and Influencing Memory with Visual Saliency Modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1471–1480. https://doi.org/10.1145/1978942.1979158

[92] Lisa-Marie Vortmann and Felix Putze. 2021. Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 80 (jun 2021), 27 pages. https://doi.org/10.1145/3463507

[93] Mark Weiser. 1991. The Computer for the 21 st Century. *Scientific american* 265, 3 (1991), 94–105.

[94] Sean White and Steven Feiner. 2009. SiteLens: Situated Visualization Techniques for Urban Site Visits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1117–1120. https://doi.org/10.1145/1518701.1518871

[95] Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. 1989. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance* 15, 3 (1989), 419.

[96] Jeremy M Wolfe, Keith R Kluender, Dennis M Levi, Linda M Bartoshuk, Rachel S Herz, Roberta L Klatzky, Susan J Lederman, and Daniel M Merfeld. 2006. *Sensation & perception.* Sinauer Sunderland, MA.

[97] Robert Xiao, Chris Harrison, and Scott E. Hudson. 2013. WorldKit: Rapid and Easy Creation of Ad-Hoc Interactive Applications on Everyday Surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 879–888. https://doi.org/10.1145/2470654.2466113

[98] Jimei Yang and Ming-Hsuan Yang. 2017. Top-Down Visual Saliency via Joint CRF and Dictionary Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 3 (2017), 576–588. https://doi.org/10.1109/TPAMI.2016.2547384

[99] Ye Yuan, Nathalie Riche, Nicolai Marquardt, Molly Jane Nicholas, Teddy Seyed, Hugo Romat, Bongshin Lee, Michel Pahud, Jonathan Goldstein, Rojin Vishkaie, Christian Holz, and Ken Hinckley. 2022. Understanding Multi-Device Usage Patterns: Physical Device Configurations and Fragmented Workflows. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 64, 22 pages. https://doi.org/10.1145/3491102.3517702

[100] Yang Zhang, Chouchang (Jack) Yang, Scott E. Hudson, Chris Harrison, and Alanson Sample. 2018. Wall++: Room-Scale Interactive and Context-Aware Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3173574.3173847

[101] Nan Zhao, Asaph Azaria, and Joseph A. Paradiso. 2017. Mediated Atmospheres: A Multimodal Mediated Work Environment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 31 (jun 2017), 23 pages. https://doi.org/10.1145/3090096

[102] Quanlong Zheng, Jianbo Jiao, Ying Cao, and Rynson W.H. Lau. 2018. Task-driven Webpage Saliency. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

[103] Manuela Züger, Sebastian C. Müller, André N. Meyer, and Thomas Fritz. 2018. Sensing Interruptibility in the Office: A Field Study on the Use of Biometric and Computer Interaction Sensors. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174165

A Analysis

We present the detailed results of our analysis of classification approaches (Table 3), time windows (Table 4), and feature combinations (Table 5).

Model	**AUC**	**F1**	**Prec.**	**R.**	**Acc.**
Majority	0.5000	0.0000	0.0000	0.0000	0.5257
Threshold	0.7519	**0.7444**	0.7262	0.7635	0.7513
RF	0.7636	0.7251	0.7521	0.7472	0.7619
Grad. Boost.	**0.7835**	0.7289	**0.7659**	**0.7531**	**0.7721**
Naive Bayes	0.7083	0.6982	0.6392	0.8255	0.6955
SVM	0.7569	0.7233	0.7217	0.7653	0.7583
Logistic Reg.	0.7417	0.7134	0.7006	0.7584	0.7469
NN	0.7134	0.6739	0.6780	0.7531	0.6914
LSTM	0.7553	0.7149	0.7267	0.7412	0.7523

Table 3: Performance comparison of classification approaches.

B Machine Learning Tuning

We evaluated the following learning-based classifiers (subsection 5.3) by applying them to our feature set and testing various hyperparameter values through a grid search: gradient boosting (30 estimators, max. depth=3, no prior feature selection), random forest (300 estimators, selected 5 best features prior to classification

Time	AUC	F1	Prec.	R.	Acc.
0s to 10s	0.7662	0.7045	0.7526	0.7179	0.7573
-10s to 20s	0.7789	0.7279	0.7552	0.7526	0.7708
-30s to 40s	0.7197	0.6802	0.6865	0.7298	0.7098
-10s to 0s	0.6431	0.5673	0.5890	0.6682	0.6017
10s to 20s	0.6777	0.6242	0.6471	0.6778	0.6617
-30s to 0s	0.6405	0.5926	0.6087	0.6897	0.5997
10s to 40s	0.6768	0.6232	0.6331	0.7058	0.6463
10s sequence	**0.7889**	**0.7421**	**0.7816**	**0.7543**	**0.7820**
30s sequence	0.7600	0.7162	0.7432	0.7366	0.7582

Table 4: Comparison of models implemented using features extracted from different time windows.

Features	AUC	F1	Prec.	R.	Acc.
Gaze_w	0.7577	0.7088	0.7326	0.7357	0.7505
Gaze_g	0.5891	0.5329	0.5665	0.5687	0.5754
Head_w	0.6279	0.5601	0.5996	0.6221	0.5905
Head_g	0.5757	0.4933	0.5532	0.5156	0.5462
Input	0.5298	0.4303	0.4893	0.4334	0.5330
Activity	0.5033	0.3420	0.4297	0.4318	0.5130
Engagement	0.6854	0.5935	0.6095	0.7585	0.6158
Gaze_w + Input + Activity	0.7633	0.7119	0.7425	0.7375	0.7535
Gaze_w + Engagement	0.8116	0.7566	0.8097	0.7618	0.7984
Gaze_w + Gaze_g + Head_g	0.7480	0.6990	0.7236	0.7267	0.7373
Head_w + Input + Activity	0.6405	0.5660	0.6016	0.6311	0.6067
Head_w + Engagement	0.7324	0.6654	0.7078	0.7528	0.6818
Head_w + Gaze_g + Head_g	0.6430	0.5899	0.6200	0.6537	0.6106
Gaze_w + Head_w	0.7551	0.7339	0.7422	0.7861	0.7361

Table 5: Results using different feature combinations. Subscripts w and g denote widget and general.

based on the ANOVA F-value between the label and feature), SVM (kernel=RBF, C=1, gamma=0.03, using the 10 best features), Naive Bayes (using the 5 best features), logistic regression (using the 5 best features), NN (solver=adam, alpha=0.0001, hidden size=128, layers=1, 5 best features), and LSTM (constructed from a $N \times 3$ input matrix from N features extracted during the following time intervals: $t = -30s$ to $t = 0s$, $t = -20s$ to $t = 10s$, and $t = 10s$ to $t = 40s$; solver=adam, alpha=0.0001, hidden size=64, layers=1, pre-selected 1 best feature). In all our machine learning experiments, we first imputed missing values by replacing them with a mean and standardizing features to comparable scales [103].

For our optimal threshold classifier, we exhaustively searched for a threshold that yielded the highest prediction performance on normalized statistical feature values computed from each window, uniformly sampling at an increment of 0.001. Our experiments showed that thresholding with the minimum gaze angle extracted from the time window $t = 0s$ to $t = 10s$ achieved the best results.