Remotion: A Motion-Based Capture and Replay Platform of Mobile Device Interaction for Remote Usability Testing

JING QIAN, Brown University, USA
ARIELLE CHAPIN, Brown University, USA
ALEXANDRA PAPOUTSAKI, Pomona College, USA
FUMENG YANG, Brown University, USA
KLAAS NELISSEN*, KU Leuven, Belgium
JEFF HUANG, Brown University, USA

Remotion is an end-to-end system for capturing and replaying rich mobile device interactions, comprising both on-screen video and physical device motions. The blueprints and software provided here allow an interface to be instrumented with Remotion's capture and visualization system. Remotion is able to mimic mobile device motion through a software 3D graphical visualization and a robotic mount that replicates the movements of a mobile device from afar. Deployed together, experimenters can emulate the mobile device postures of a remote user as if they were in the room. This is important since many usability studies are carried remotely and the contribution and scale of those studies are irreplaceable. We compared how HCI experts ("analysts") observed remote users behavioral data across three replay platforms: a traditional live time-series of motion, Remotion's software visualization, and Remotion's hardware visualization. We found that Remotion can assist analysts to infer the user's attention, emotional state, habits, and active hand posture; Remotion also has a reduced effect on mental demand for analysts when analyzing the remote user's contextual information.

CCS Concepts: • **Human-centered computing → Usability testing**; **Visualization systems and tools**; *Smartphones*;

Additional Key Words and Phrases: smartphone behavior; motion visualization; fabrication; remote studies; mobile sensing

ACM Reference Format:
Jing Qian, Arielle Chapin, Alexandra Papoutsaki, Fumeng Yang, Klaas Nelissen, and Jeff Huang. 2018. Remotion: A Motion-Based Capture and Replay Platform of Mobile Device Interaction for Remote Usability Testing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 77 (June 2018), 18 pages. https://doi.org/10.1145/3214280

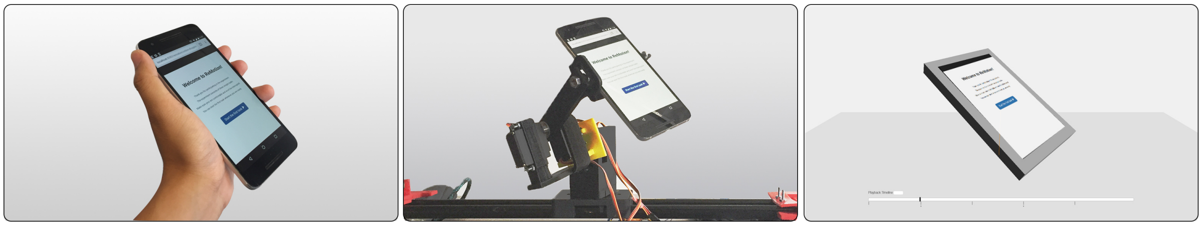
1 INTRODUCTION

Current approaches to user studies such as usability testing and participant observation require experimenters and participants to schedule time together in the same physical space. As an alternative and more convenient solution, participants can remotely perform the user study in their own space with the experimenter collecting data from afar (e.g., [2, 28]), or through crowdsourcing (e.g., [20]). Although remote usability testing has become easier to execute, it is still difficult to replicate the context of the user's behavior and the details of the interaction with their devices. When the target is a mobile device application, this lack of context can lead to the experimenter misinterpreting the user's intent. Remote video recording captures some of this context, but requires camera setup, high bandwidth and storage, and entails privacy concerns.

---

[*] Work completed while visiting Brown University

2474-9567/2018/6-ART77 $15.00
https://doi.org/10.1145/3214280



(a) A user holding a phone (b) Remotion hardware replays the holding gesture (c) Remotion software replays the holding gesture

Fig. 1. The concept of Remotion: the system can capture and replay the mobile device interactions in remote usability testing. For example, if (a) a user is holding a phone, the Remotion system can record data for this moment and replay using (c) software and (b) hardware visualizations.

We investigate whether motion sensors in mobile devices can provide cues for context data that could later be regenerated through replaying the captured hand motion with a physical replica. During a remote usability testing, when the user in a distant location moves, rotates, or shakes their device, an actual mobile device at the experimenter's location is moved, rotated, or shaken in the same manner by a robotic mount. The motion can be captured with an application we developed or purely through a JavaScript library if the target is a web application. The physical replay is complemented by a software component that projects the contents of the participant's screen on the replica's screen. This arrangement allows experimenters to run user studies of rich interactions remotely, and even simultaneously if desired, making them more affordable, scalable, and efficient.

In some cases, a remote user's hand motion reflects contextual information about their surroundings, points of attention, and even signals particular emotions. Imagine a scenario when a user is shopping online on a mobile device proceeds to the checkout page which asks for their credit card information. Some users will put down the phone and start to look for their wallet, while others may have the cards readily available without the need to place the phone away. None of this contextual information is captured on the screen, yet the different movements resulting from how a user handles their device reveal more about the situation than what is displayed on the screen. Having this concept in mind, we created the Remotion platform to further improve contextual information retrieval from various remote usability testing settings (see Fig. 1). At this stage, we focus on situations when a user is in a sitting or standing position and in a relatively comfortable place such as their home or office.

Remotion is a complete end-to-end platform comprising five components.

(1) The client software library for motion sensing on mobile devices (e.g., Android phones and Surface tablets) enables developers to collect sensing data by simply deploying an application on the participants' phone.
(2) The screen projection component[1] captures what the user sees on their screen and projects it to the screen in a software or hardware replay.
(3) The lightweight server receives and organizes the data collected with Remotion's client and Remotion's screen provides an interface to control the Remotion's software and hardware visualization replay.
(4) The software visualization displays on the experimenter's screen a replay of what the user sees on their screen and the phone's movements.
(5) Remotion's hardware (physical) replay replicates the remote user's movement and screen on a one-to-one scale. The robotic mount (with 4-axis degrees of freedom) is offered as blueprints for at-home construction.

---

[1] For simplicity in the user study, we used a third-party recording software which recorded the users' screens rather than a livecast.

Together, Remotion offers a system which enables richer context sharing in remote user studies, especially during in-home sessions. Remotion's website at https://remotion.cs.brown.edu/ provides instructions for setting up and deploying a remote user study, as well as hosts Remotion's open-source software and blueprints.

As part of the research, we conduct a study to assess effectiveness and usability of Remotion to reveal contextual information such as attention focus, emotions, and habits about the remote users through a smartphone's motion and orientation data. Researchers in the past primarily used smartphone motion data for behavioral [12] and input estimation [6] through quantitative methods such as supervised learning and pattern recognition. Our study explores the qualitative values from these motion data regarding estimating the users' habits, attention, and emotion through the visualization methods. We further discuss the implications of these findings, what insights can be learned, and how they can be used for improving future remote usability studies. As a result, we find that 1) replaying the smartphone's motion and orientation with Remotion provides diverse and rich contextual insights for researchers to infer the remote users' movement, orientation, behavior, emotion, and attention from the motion data logged from their phones, 2) Remotion helps researchers understand the remote user's active hold posture when using the phone; this could be used to improve interface design in applications, infer attention, and reveal habitual behavior, and 3) physical mirroring in Remotion's hardware visualization lowers researchers' mental load compared to a time-series visualization method and is an efficient medium for interpreting data from a remote user. Finally, we discuss broader implications on current remote usability testing methods.

Our paper is presented with the following structure. In section 2, we describe the related work that supports the Remotion's design and implementation basis. Section 3 outlines Remotion's five primary components and describes their functionality as well as integration. Section 4 and Section 5 introduces our experimental design and evaluation for showing the effectiveness and usability of Remotion in revealing the qualitative values of smartphone motion and orientation. Then, in Sections 6 and 7 we discuss the implications of Remotion for remote usability studies and its current limitations. We conclude our paper in Section 8.

2 RELATED WORK

2.1 Remote User Studies

Traditionally, user studies have been conducted in labs where researchers can observe and record the behavior of participants [5, 28]. Yet, the logistics of planning a study in the lab, the setup of necessary equipment, the time spent by the researcher during the test, and the analysis of the resulting video, can all create obstacles for the experimenters including privacy and bandwidth concerns. But as the use of mobile devices has grown and evolved, so has the possibility of performing remote user studies with these devices. Research attempts to measure the efficacy of various approaches for remote studies have found that a reasonable number of usability problems can be discovered at a fraction of the cost and effort of an in-lab study [3, 5, 20, 28]. Although these same studies have found that in-lab studies result in more discoveries of usability problems, they are not always a practical option for companies and researchers.

Remote user studies use a variety of methods to collect data, such as user-reported usability issues in diaries, forums, or forms [5, 7], screen recording [3], or clickstream logging [28]. Once the data has been amassed, analyzing the results takes at least two steps: (1) identifying the exact usability issues found and (2) rating the severity of each issue [3, 5, 28]. At each step, researchers ideally would have a clear understanding of the user's behavior while performing a particular task.

2.2 Getting Data from Remote User Studies

In a lab environment, experimenters collect qualitative results based on observations of participants. Bruun et al. [5] compared lab user studies with several remote methods that depend on user-reported problems and responses.

They found that the lab study identified a significantly greater number of usability issues than any of the user-reported methods, including a much higher percentage of “critical-rated” problems. A similar study compared self-reported issues from users to recorded videos, and although it found that the users identified more than half of all the problems found by the experts, the forms were designed to require detailed responses from the user [7]. This effectively transfers much of the unwanted effort onto the user. Such an approach can limit the amount of valid feedback or the number of participants, considering that a weakness of user-reported feedback is the integrity of the responses; participants, when self-directed, may not have a good enough incentive to give honest and thorough feedback [11]. Multiple responses from participants in Bruun et al. and Kittur et al. [5, 20] could not be used due to being missing, unintelligible, or obviously unhelpful. While this problem can be mitigated by better task design [20], it still presents a pitfall for this type of study.

Meanwhile, other remote studies depend on more “objective” data, such as screen recording or click-stream logging [3, 28]. While these studies did not rely on user-reported information, and so did not suffer from the associated missed observations, they still do not present the context of the user to the experimenter. Their behavior must be inferred through screen interactions alone. As reported in these studies, this lack of context can hinder observations, and therefore reduce the number of identified problems. Filming the user’s behavior in a remote study would theoretically be ideal for observations. This is because film from lab studies allows researchers to better analyze the user’s actions and reactions during the study [5, 28]. However, filming would require specific hardware, more setup, and more effort from the user.

Remote studies can benefit from an affordable and scalable way of enhancing how experimenters understand the behavioral context of remote participants. Remotion is designed to improve the qualitative understanding of a remote usability testing, without participants needing to set up equipment. It achieves this through visualizations of user movement and screen content, with no substantial cost to the user and the experimenter.

2.3 Physical Visualization

Multiple studies found that being able to directly interact with a physical representation of the data is a helpful cognitive aid to understand and perform tasks [17, 26]. For areas of design such as urban planning and architecture, physical models are used to represent spatial information, including building heights or the movement of people in a particular space [9, 16]. The interpretation of these spatial characteristics is made easier through visualizations, and often a 3D visualization is simpler to understand [16]. Kuzuoka et al. suggest visual realism as another aspect of what makes them accessible, beyond the ability to move around and interact with physical visualizations [21]. Urban planners often deal with complex sets of data; one study found that layering 2D and 3D visualizations simultaneously can help designers interpret the context of their project holistically. They hypothesized this layering lowers the cognitive load for urban planners using visualized data to inform their designs [16].

While often a physical visualization is a static model, another study used physical visualizations that changed over time, discovering an improvement in participants’ readings as the visualizations changed [25]. We can also observe precedence of dynamic physical visualizations in the field of robotics, especially in human-robot interaction. Users need the ability to interpret a robot’s actions for proper interactions to occur. One could conclude that these actions are merely dynamic physical visualizations of human communication. Multiple examples of robots performing recognizable human gestures can be found [18, 21, 23], whether they are executed programmatically or by a human controlling the robot directly. In the GestureMan study, those who interacted with the “robot” controlled remotely, could understand when the remote user was changing their attention from one area to another in the space, simply because the robot would seem to “look” at something else [21]. This interpretation of the robot’s movement signifies a human ability to interpret human-originated or human-simulated actions in technology that can replicate those actions, even if a human cannot be seen performing them.

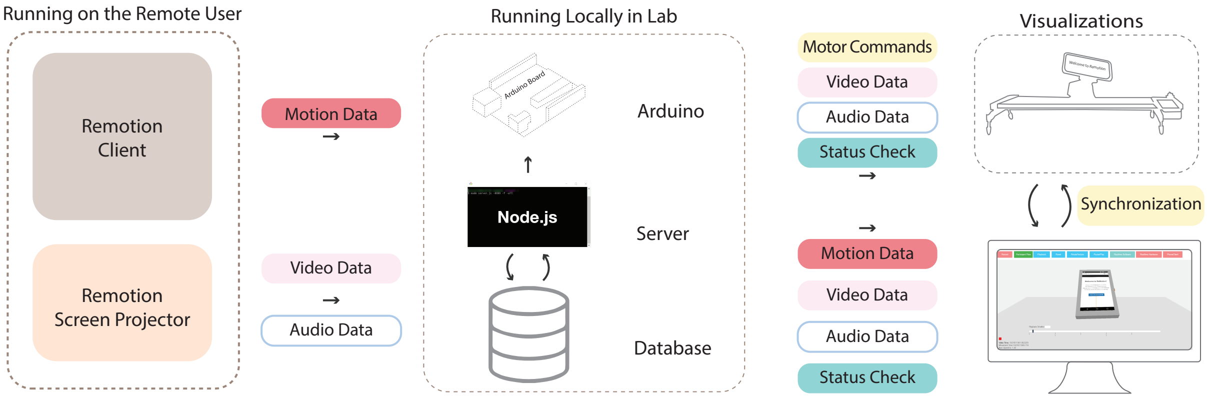


Fig. 2. Remotion's architecture showing how the remote user's data are captured (left) and saved to Remotion's server (middle), which can be visualized in Remotion's hardware visualization (upper right) and software visualization (lower right).

A dynamic physical visualization may be helpful for interpreting human movement to understand behavior. Remote user studies that gather user interactions and the context of the user's behavior like Remotion may contain data that could benefit from such a visualization. Spatial data that may represent user motion or gestures while completing the task may be physically visualized to aid interpretation. Non-spatial data (e.g., screen recordings) could be layered on the visualization as in Ishii et al. [16] to lower the cognitive load of analyzing these opposing types of data together.

2.4 Sensing Mobile Device Motions

Mobile devices include numerous motion sensors, enabling several useful applications. Kwapisz et al. classified different fitness activities that users performed, by using accelerometer data [22]. More recently, wrist-worn mobile devices, like the Microsoft Band, estimate certain fitness activities, such as the daily number of steps, using the GPS. Additionally, prior work investigated the combination of touch and motion to infer new information about users. For example, Tsandilas et al. detected gestures made by users through sensing touch and motion to perform different commands [27]. Goel et al.'s inferred the user's posture based on touch and motion sensors already in a smartphone [12]. Their approaches are for short-term detection, classifying behavior soon after the user interactions, while Remotion aims to provide a holistic view of users' mobile behaviors in user studies.

3 THE REMOTION PLATFORM

Remotion is designed to replicate a mobile device's screen content, movement, and rotation in a remote usability test, especially when the user is stationary (in a sitting or standing position). We introduce five different components that when deployed together can replicate hand-motion related contextual information that is otherwise not captured in a conventional remote usability testing session (see Fig. 2). Detailed instructions and open source files needed to deploy Remotion fully are available online.

3.1 Remotion's Client Software Library

The Remotion client is the first component of the Remotion platform. It enables collecting mobile sensing data (e.g., gyroscope and accelerometer) and user interaction data (e.g., taps and scrolls) from remote users. The Remotion client records these data in real time and uploads them to the Remotion server upon the completion of the task. The remote user can choose to start, stop, and upload data at will. The client provides the motion-encoded contextual data that can later be revealed to the experimenters.

3.2 Remotion's Screen Projection Component

The second component provides additional information about the remote user's screen content to the experimenters. The Remotion screen projection component captures the phone's screen changes and sends it to the Remotion server as a compressed image stream. For our user study, we used a simplified scheme which captured the screen with a third-party software [15] and later uploaded the video file to the Remotion's server.

3.3 Remotion's Lightweight Server

The server functions as a hub to organize, configure, and control the Remotion's visualization ends. The Remotion server supports two modes: 1) offline replay and 2) real-time replay. In the offline replay mode, the server loads data from the Remotion client and the Remotion screen projection components described above, relaying them to Remotion's visualization endpoints within a fixed interval of time. The real-time component decodes the incoming data from Remotion's client and screen projection components and forwards them directly to the visualization endpoints.

The Remotion server has a back-end (i.e., Node.js) and a front-end graphical user interface (GUI) for the experimenter to load, re-save, stream the data, and control the visualization playback. The control panel on the Remotion server is HTML-based; it does not require additional configuration or installation once the back-end is running. Remotion's interface has been successfully tested on mainstream browsers without functionality or layout issues.

3.4 Remotion's Software Visualization

Recreating the remote user's motion is the primary functionality the Remotion platform provides. It allows distant users' hand movement data to be presented to the experimenters. Remotion's software visualization projects the mobile device motions and renders live video and audio content on the experimenter's computer screen. The Remotion software visualization contains a 3D model of a mobile phone and a virtual screen showing the video stream, as shown in Fig. 1. During a replay, the 3D phone model's and the virtual screen's orientations are updated from on incoming data.

The accuracy of motion replay is ensured with a sensor fusion algorithm package[2] that calculates quaternion rotation. The use of quaternion rotations avoids "dead zones", where the visualized results may be incorrect. Without using sensor fusion, repeated mobile device movements and fast changes in its orientation will render the raw data visualization drastically drifted and may not reflect its actual orientation.

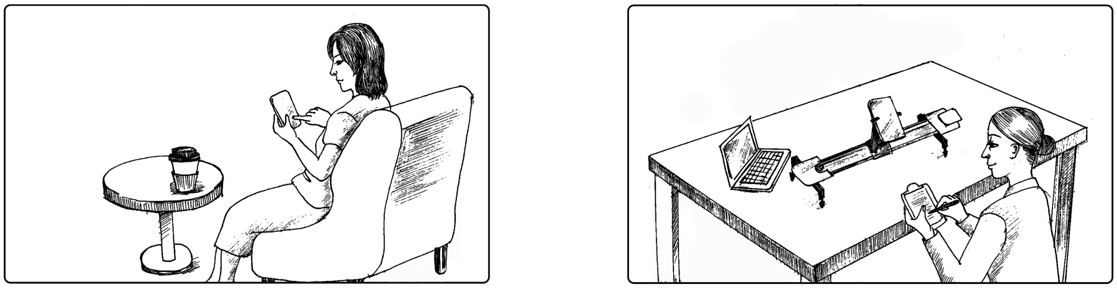
3.5 Remotion's Hardware (Physical) Visualization

Remotion Hardware visualization is a fully customizable and low-cost Arduino-controlled robotic mount for motion visualization. Its main component consists of a circuit-board, three motors, a 3D-printed mounting system, and a track. Remotion hardware can render the motion data and screen content in the experimenter's physical space after the experimenter mounts a mobile device onto the robotic arm. Unlike its software counterpart, the physical visualization offers a one-to-one scale motion playback (i.e., replicating the degree of movement, the form of a real phone, and the contents of the screen mirror the user's screen). This physicality further enables an understanding of the spatial depth and a more naturalistic way of observation.

Remotion's hardware visualization uses the same sensor fusion algorithm with the software counterpart to ensure consistency. On occasion, Remotion needs to swap the x and y-axis in the visualization to make the motion

---

[2] https://www.npmjs.com/package/cordova-plugin-device-sensor-fusion



(a) In the Pre-study data collection, an actor uses a phone.

(b) In the user study, an analyst uses Remotion.

Fig. 3. Illustration of our pre-study data collection (left) and the evaluation procedure (right). Four “actors” used a smartphone to perform four tasks. Then, nine HCI experts which we call “analysts” used Remotion to analyze how the actors used the smartphones through replaying the data on Remotion system.

seamless. This is not the case with the Remotion’s software visualization since its rotation is quaternion-based. The Remotion’s hardware visualization uses Euler angles to rotate the device. Once the user rotates the phone from portrait to landscape, Remotion automatically swaps the x and y-axis to match the visualization.

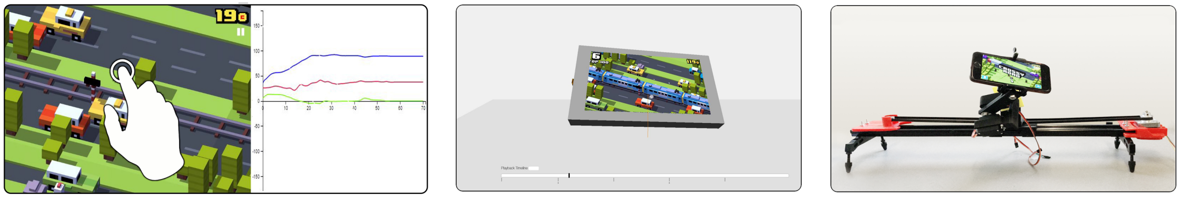
4 EXPERIMENT

To understand what insights can be gleaned from the smartphone motion and orientation data when using Remotion, we designed a user study to 1) explore the unknown qualitative values of smartphone’s motion and orientation data, 2) evaluate the overall effectiveness and usability of Remotion compared to an existing visualization method. We conducted our study in a similar fashion to how Remotion would be used in real life (see Fig. 3). First, we collected motion, orientation and screen recording data with Remotion’s client software and screen projection; then, we recruited HCI experts to analyze the data using Remotion. For evaluation purposes, we added two Canon HD cameras to record the remote users and asked them to think-out-aloud during the process. The think-out-aloud protocol provides important emotional and behavior cues of the remote users that can later be used for comparison. These data were not available to the participants and used to evaluate HCI experts’ ability to recreate qualitative insights based on visual evidence only. For clarity, we refer to the remote users as “actors”, and the HCI experts as “analysts”.

4.1 Pre-Study Data Collection

We first collected data from the remote users (i.e., actors) needed for the actual study with Remotion’s client software and two extra cameras. We recruited four actors (24–60 years old) from a university-wide mailing list, who were asked to use a mobile device we provided and performed four tasks: 1) online shopping, 2) a game “crossy-roads”, 3) a game “fruit-ninja”, and 4) a photo capture task. Remote usability testing focused on real life tasks and the usefulness of applications [8], and the tasks we chose covered a wide range of applications (e.g., web activity, entertainment, photography) and behaviors (e.g., tapping, swiping and scrolling) that smartphone users frequently do for both indoor home and work environments [10]. The tasks were also chosen to cover different device orientations. Thus, we could collect and observe the effect of orientation on user behaviors. Furthermore, many tasks such as reading text on a screen and browsing websites are intrinsically less “active” than other tasks such as playing games or taking pictures.

Once the actors signed a consent form, we explained the tasks to them. Each task was limited to five minutes and we stopped the trial if the time was exceeded. The first task, *online shopping*, consisted of a typical-looking online shopping application. We asked them to shake the phone three times to confirm synchronization among



(a) The time-series visualization (b) Remotion software visualization (c) Remotion hardware visualization

Fig. 4. Our evaluation experiment compared three methods of replay: (a) the screen replay with a time-series line chart visualization; (b) Remotion software visualization replaying the screen and the phone's movement as a 3D model; (c) Remotion hardware visualization replaying the screen and the phone's movement with a physical replica.

motion and video data. The actors were given $20 in virtual store credit and were asked to order any combination of items totalling less than the provided store credit. The actors needed to choose from a list of products, read the descriptions, add them to the shopping cart, and eventually "checkout" when they were done. The checkout button was deliberately fickle (working 15% of the time) to inject confusing moments to elicit a range of reactions. Both the second (*Crossy-roads*) and third task (*Fruit Ninja*) were games, selected because they used different interaction techniques (i.e., touch and swipe) and generated different motion patterns. The goal of *Crossy-roads* is to control a virtual character and cross as many streets as possible while not being hit by oncoming traffic. Actors were instructed to obtain their best scores within the five minute limit using interactions of tapping or swiping. *Fruit Ninja*, is another game that allows the user to swipe the screen to "cut" virtual fruits and increase a score counter. Again, actors needed to play the game for five minutes and score as high as possible. The last task used the smartphone's built-in camera to take photos within the office. The actors were asked to locate and take a photograph of two different items that were placed around the office. While taking the photos, the actors were instructed not to walk around but they were able to rotate their torso and move their arms.

The actors were encouraged to use the smartphone in a naturalistic manner; we rearranged the data collection space, added furniture, and placed the cameras in the corner of the room so that they were inconspicuous to the actors. In addition, we stayed outside the actors' field of view so that their behavior would not be altered based on the presence of another person. Actors often did not follow the think-aloud protocol during the gaming tasks, presumably because these tasks required fast responses and more concentration.

Remotion's client software recorded actors performing the four tasks, while their actions and think-out-aloud responses were recorded with two external video cameras placed at two different angles to capture the full scene. During the task performance, we asked the actors to describe their actions, intentions and thought process (e.g., what they were seeing on the screen, what action they were about to take, and what they believed would be the results of the actions). The actors' final recorded data contained the smartphone screen capture, gyroscope, accelerometer, and compass values along with voice and video recordings. The resulting data was separated into two parts. The first contained the smartphone sensor readings of the gyroscope, accelerometer, and compass, along with screen recordings (no audio), and was used for evaluating Remotion. The second consisted of the video, audio, and actors' think-aloud data used for generating the ground truth for validation.

4.2 User Study Design

Remotion aims to help HCI researchers understand contextual information from remote usability testing. For this reason, we limited our participants (i.e., the analysts) to people who had past experience with usability studies and were knowledgeable about Human-Computer Interaction research. The user study was estimated to take about 120 minutes and the participants were compensated $30 at the end of the study. The analysts were asked to take a break after the first half of the study, and they could pause the study at anytime to write notes or take short breaks to relieve their mental load.

To test if the analysts could obtain qualitative information from the recorded actors' data, we used three visualization methods to replay the data and asked the analysts to write down observations and insights. The three visualization consists of: 1) an existing time-series visualization of motion, 2) Remotion's software visualization, 3) Remotion's hardware visualization.

The first visualization was a time-series of motion next to the actor's phone screen recording on a computer screen. This type of visualization is an existing method that researchers used to study smartphone movements [1, 24]. To our best knowledge, this is still the popular form of visualization to show effect of motion and orientation from smartphones, and a time-series of motion can clearly show slight movements happening on the smartphone. Therefore, we compare the effectiveness of observing contextual information and the usability of this visualization to Remotion.

The second and third methods are Remotion's software and hardware visualizations. Remotion's software visualization used a virtual smartphone model to represent the actor's device and rendered on the model's screen the data capture from the actor's screen. The virtual smartphone's orientation and movement was synchronized with the recorded data. The Remotion's hardware visualization used a robotic mount and a physical mobile device to replicate the actor's hand movement while the physical device's screen replayed the actor's screen.

Evaluating the visualization methods required analysts to watch the entire set of actors' recordings. With four actors performing four tasks, we have 16 trials for each analyst. We randomized the pair between the actor's task and the method used to minimize learning effect. The four tasks' appearance order to the analysts were also randomized. We generated tables for every analyst with unique sequences of trials throughout the study.

During each trial, we asked analysts to annotate their observations on a timeline with our predefined coding scheme (Fig. 5), similarly to multimedia annotation, e.g., in [19]. The coding scheme suggested that the analysts should respond to the following criteria for the given visualization in the trial: a) if the actor was making progress or not, b) label events from the smartphone's movement and orientation (e.g., smartphone pickup from a surface), c) impression or habitual information of the actor, d) emotion or attention learned from the data. The interpretation of these criteria is dependent on the analysts and we do not limit the number of their responses.

We prepared a modified NASA-TLX questionnaire to record analysts' impression of working with the baseline and Remotion's visualization method. NASA-TLX [14] is a survey instrument widely used to assess workload and measure performance. Questions from the original NASA-TLX that were not relevant to our tasks were removed, and we added one question regarding preference.

4.3 Ground Truth

In order to understand if the analysts' responses were reasonably derived from the actors' data, two authors performed similar tasks to the analysts. We refer them as *judges*. In addition, they were given video recordings and the actors' think-aloud audio. These data provide accurate information on gestures, attention, emotions, and habits, as well as visual and audio references to our authors. One author labelled all 16 trials with the same predefined coding scheme before the user study, and a second author performed the same series of tasks separately. Both authors' results were combined to form the ground truth which is used as a reference to evaluate the analysts' responses. The responses not captured by our ground truth were checked again by the authors to determine their validity.

We noticed some disagreement when comparing the judges' annotations using a computed Cohen's Kappa for measuring inter-rater reliability. Between the judges, the kappa index was $\kappa = 0.33$ for behavior labels, $\kappa = 0.60$ for emotion labels, $\kappa = 0.36$ for attention labels, and $\kappa = 0.50$ for motion events (e.g., picking up and putting down the phone). Normally, these results are considered fair or moderate agreement.

However, many of such disagreements are different interpretations of the same event. For example, one author marked one motion event for actor 4's task 2 with continuous movement for almost 10 seconds. The other author marked few motion events within the same time period. When discussing later, they both agreed that they were referring to the same event (i.e., moving and shaking for 10 seconds). Eventually, differences in authors' opinions were resolved through discussion and different labels that indicated the similar insights were merged.

4.4 Procedure for the User Study

Upon the analysts' arrival, we handed out consent forms and they were asked to review the details of the trials and understand their responsibilities. Then, we explained the predefined coding process and provided highlighters and pens. Once the study began, the analysts were informed about the current trial number which was recorded in a spreadsheet and passed out on a blank piece of paper with a printed timeline for the coding scheme. The analyst could start the visualization playback and annotate insights based on their understandings and observations. When annotating, they may pause or resume the visualization. But rewind and restart functionality were not used in the procedure. At the end of the 16 trials, the analysts were provided the modified NASA TLX questionnaire and asked to complete it. The study was completed when the form was filled out and they were compensated.

We recruited 9 analysts and collected a total of 144 trials, with each analyst labeling 16 trials, of which 4 were excluded from the analysis because they duplicated conditions. Two trials in total were filtered out due to synchronization errors between the video and motion playback. We proceeded by comparing the analysts' logs with our ground truth to compare with their judgment for timeline marks and insights individually. In the following evaluation, we describe our findings generated from analysts' marks, insights, and their reported experience with Remotion. The included usability analysis is based on the NASA TLX score from the analysts and anecdotal feedback about the Remotion system.

5 ANALYSIS

We used thematic analysis [4, 13] to analyze the responses collected from the 9 analysts and evaluate them. We used this approach for two reasons: 1) analysts' annotations contributed rich observations that do not necessarily align with our ground truth due to the open-ended responses. Using quantitative methods such as counting the number of insights overlapped with our ground truth would reduce the diversity of the responses. In contrast, using a flexible qualitative method, we can additionally check for annotations not in the ground truth, and therefore capture essential details in the data. 2) Most of the responses are qualitative (e.g., reporting insights).

Using thematic analysis, judges independently annotated timelines generated by the analysts, as shown in Fig. 5. The data was digitized and organized via our custom online analysis tool to filter by different tasks, actors, and visualization condition. Then the authors generated initial codes from the timeline labels and annotations, and grouped them into themes that addressed similar observations. Each formed the themes from the codes individually and later merged them through discussion; the authors reduced them into five distinct themes and noted the relevance evidence in each theme. The analysis of these five themes below suggest that Remotion reproduced richer contextual information, such as the actor's current holding posture, personality, habits, and emotion. Finally, the analysts' NASA TLX scores indicated that Remotion was more effective and reduced mental load for analysts during the labelling task.

5.1 Theme 1: Active Holding Posture

Remotion helped analysts learn the actors' current holding posture in a direct and efficient way. Analysts inferred whether the actor was holding the smartphone with their left, right or both hands. Throughout the study, analysts monitored different cues to determine which hand the actor was actively using. Analysts 1, 2, 3, and 5 all used the "the tilt of the phone" as a primary motion cue for deducing the holding posture. Analyst 3 justified their observation based on the assumption that the phone orientation changes during a touch interaction. The tilt-based strategy worked well for the reading and gaming task, but was less effective for the camera search task, where hand movement is used as a primary input, and motion here is no longer merely contextual information.

Analysts 2 and 5 used a touch-point based strategy as a cue for the current holding posture. They relied on the tap location, with the idea that some parts of the screen were easier to touch based on the grip. This strategy worked well for Tasks 1 and 2 where the actor's touch points are concentrated. However, it is affected by different types of interactions, such as swipes or random taps across the whole screen. These interactions impeded analysts 2 and 5 from inferring the active holding posture and eventually moved to a tilt-based strategy.

In some situations, actors used both hands to grip the phone. When creating the ground truth data, the judges noticed that actor 1 switched from using their right hand to holding the phone with both hands when they were particularly focused on a task. This switching was noticed by analysts 2 and 3, and analyst 2 used the tilt-based strategy to determine that the user was initially holding the device with their right hand, and used a combination of the tilt-based and the touch-point strategies to infer when the actor switched to using both hands.

In a real-world remote user study session, a person may put down and pick up their phone at different times due to interruptions. We asked our actors to pick up the phone at the beginning of a task and put it down once they were done. All analysts (1–9) were able to indicate these events correctly and mark them on the timeline. From our observations, analysts required more focus and exerted more cognitive load when deducing that a participant picked up the phone in the baseline visualization (the time-series of motion sensor data); doing this required interpreting data in 2D coordinates and mapping them to 3D motion. With Remotion, all analysts noticed the initial move when the actors picked up and later put down the phone the first time. The analysts almost instantaneously reported the pick up event when they saw the Remotion Hardware Visualization moved the physical phone from a lay-flat position to a vertical one.

Once analysts become aware of this motion, it can be used to determine when the phone is left on a surface rather than actively held by the user; as an additional cue, if the screen is static at the time, then the user is probably distracted at that moment. In a way, the active holding posture is a proxy for an actor's attention. The information about the way users hold their phones can also be used for interface improvements. Many analysts reported that some actors' tapping points are all concentrated in certain areas of the screen when holding the phone with one hand. This is useful to adjust the size, layout, and hierarchy of an application's interface, potentially providing enough contextual information for an interface that adapts to each persons' habits.

5.2 Theme 2: Inferring Actors' Habits

Habits often go unnoticed by the actors themselves. The judges observed that each actor had a distinct way of handling their phone. Some adjusted their body position at the end of a task, and some tapped on the screen harder while others are more gentle with their interactions. When an actor changed to a sitting or standing position, the mobile device moved drastically and ended up in a different orientation. During this movement, the interaction on the screen is often motionless. Both analysts 1 and 6 noticed that actor 1 changed their sitting position during the task, as analyst 1 noted, "The device is tilted to a very large degree. It seems that the user is playing while lying on the side." Analyst 6 wrote after marking on the timeline that the phone "rotate[s] to right; maybe they are sitting in new position." Although Remotion at its current stage is unable to reproduce position changes, it is capable of demonstrating if an actor's position has changed and to which direction. This can be important when a remote researcher requires their participant to remain still within a certain range or needs to determine if they have changed their body posture.

Tapping on the interface could vibrate the phone depending on its intensity. Some actors tapped harder consistently, while others tapped harder only when they were emotionally engaged. Analysts noticed that the smartphone moved whenever actors 3 and 4 scrolled or tapped on the screen during the online shopping task

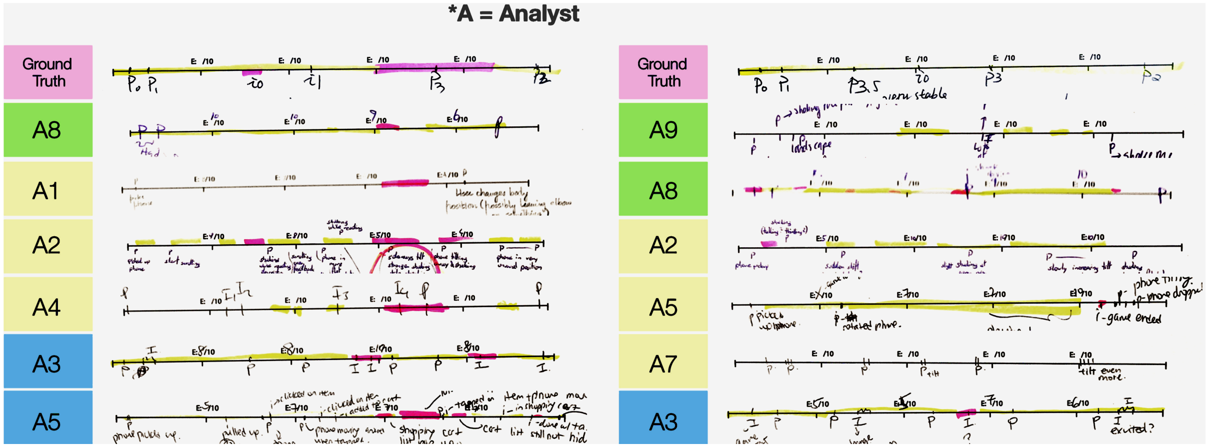


Fig. 5. Pink: The ground truth as annotated by judges. Green: Time-series of Motion. Yellow: Remotion software visualization. Blue: Remotion hardware visualization. Annotated timelines from analysts representing what they inferred as actors were performing tasks. More detailed notes written by the analysts are not shown in this figure.

(task 1). From the video recording, our judges noticed that actors 3 and 4 tapped on the screen harder than others which caused the vibration, while actors 2 and 1 were rather calm when tapping. In later tasks, however, the tap pattern changed, as tasks 2 and 3 required more concentration, often accompanied by intense activity. Actors 1 and 2 who did not display any noticeable tapping effect on task 1, exhibited, as analysts 2, 6, and 7 noticed, increased shakes when their character was lost or they were about to gain a point. These changes in motion are visible when using Remotion, but are not apparent in the time-series visualization condition.

5.3 Theme 3: Attention

While emotional cues discussed above may indicate that the actors are engaged and focused on the active task, other motion-related cues can provide insights to an actor's engagement level. Remotion helped to identify two types of distractions: 1) when an actor is not as engaged as before and 2) when an actor has temporarily paused the study. Analyst 2, when working with Remotion indicated that actor 1 lost interest towards the end of the task based on decreased phone movement. Events such as the phone lying flat on the surface without any actor inputs indicated that the actor was away. Conventional methods such as screen recordings or click logging are more limited than Remotion in inferring this information. Without knowing that the phone lied flat on a surface, analysts' interpretations tended to be mistaken. Analysts 1, 3, and 6 all noted that actor 3 was stuck at the beginning of task 2 and marked this as a distraction. In fact, when comparing with video recordings, actor 3 was merely asking for details and observing the interface.

When it comes to capturing the level of engagement, Remotion can assist analysts identify when an actor is focused or distracted. Analysts 2, 6, and 7 noticed that actor 2 gradually tilted their phone towards the end of task 2. Analyst 2 believed it was due to heightened concentration, and analyst 7 wrote it was associated with the task becoming harder. We mentioned above that actor 2 had a habit of leaning sideways. But here, we found that analysts were correct in inferring that the actor was more focused, based on video recordings and the ground truth observations made by the judges. It is important to note that not all events of phone tilting are signals for heightened focus. Rather, a combination of the screen content together with the degree of tilting and nuanced movement allowed analysts to identify the level of engagement using Remotion visualizations.

5.4 Theme 4: Emotions

Emotional or shaking moments were reported by analysts 2, 3, 5, 6, and 9. Most of these moments were associated with actors losing the game, receiving a bonus, or opening a mystery box. Analyst 9 reported that actor 2 "seems very calm until they lose, and then they shake their phone a lot." Analyst 6 mentioned a few times that actor 4 was "swirling" after the character was lost, while analyst 2 wrote "shakes on game over" for actor 2. When comparing these observations to the judges' annotations, judges noted that actors 1, 2, and 4 became more intense and moved more intensely after their character "crushed" or "died" during the game. Actor 1 moved back-and-forth and actor 4 laughed loudly.

Sometimes, the analysts reported emotions in a greater detail: two noticed frustration, three found evidence of confusion, four noted calmness, and one indicated that the actor was excited. Whether these emotions matched the judges' observations varied based on the type of emotion. Many analysts reported that the actors were confused by the "check-out" interface in Task 1 (online shopping). As explained, we deliberately created a confusing moment in the shopping task to gauge the analysts' reactions. Analyst 2 wrote "side way tilt, stronger shaking while checkout out, losing interest." and marked that the actor was stuck. Although we are not sure if the motion indicated the confusion or vice versa, it is possible that the sudden increase in movement captured the analyst's attention in noticing this event from all other interactions. We noticed that for calmness, the signal is the degree of device motion. Many analysts indicated that actor 3 was "very calm and steady" throughout all tasks. This was confirmed through the judges' report. Analyst 5 identified by noticing actor 3's phone "slightly tilts while reading, otherwise pretty steady" and analyst 9 also acknowledged that actor 3 remained calm.

There are cases when the device motion can actually cause confusion for emotion categorization. The analysts did not correctly differentiate emotions of frustration and excitement, since both displayed similar motion patterns. From analyzing the reports and labels, we realized that device motion cues need to be coupled with screen recordings and other features to be meaningful, and the context for further refining the observed emotion into frustration and excitement require additional experiments. Both analysts labeled the frustration and excitement emotions without a clear reference from the screen and their observations were incorrect.

5.5 Theme 5: Orientation and Movement

Perhaps the most straightforward benefit of the Remotion software and hardware visualization components are their ability to show the motion of the device. Analysts could identify the device orientation in a naturalistic way. Our baseline method, the time-series visualization, used a 2D graph to represent motion on an x-y canvas. But this indirect representation of orientation requires extra mental processing to associate the line movements with the phone's orientation, as seen in Fig. 4. Analyst 2 repeatedly mentioned that "I cannot guess how the phone is held" with motion data logs visualized in the baseline method. Analysts working with this visualization method are also prone to misinterpretations of motion data. Analysts 1, 5, and 9 all had mistakes in their interpretation of the phone orientation when working with time-series visualizations of motion. This is rarely the case when they used the Remotion software and hardware visualization and they were able to correctly identify the orientation of the phone with similar tasks. The direct representation of orientation visualized by Remotion allows analysts to understand the phone's movement in a direct way.

The one-to-one scale movement replay of Remotion's hardware visualization enables analysts to further distinguish motion patterns. Recall from the pre-study data collection section, we requested that actors shake the phone three times at the beginning of Task 1. When working with the baseline method, analyst 3 noticed this movement and labeled it, "lots of movement, but unable to tell what it is." Analyst 7 marked it as "tilted," analyst 4 as "tilt the phone to horizontal" and analyst 1 as "orientation change to right." These analysts all noticed the movement but were not able to identify their source. The other four analysts who were assigned Remotion visualizations were able to tell that the phone was "moving/tilting up and down." Although no analyst reported

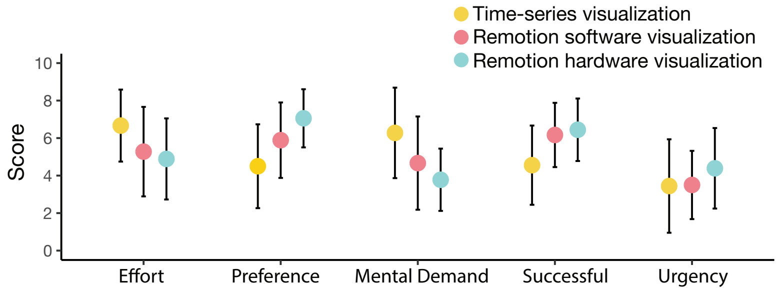


Fig. 6. The results of the modified NASA TLX questionnaire. Error bars represent the standard deviation.

the exact number of upward and downward movements, their observations were closer to the actual events. Since the movement was swift and unanticipated by the analysts, in a real-life deployment situation, a second replay (not provided in our study) may help analysts capture fine details such as exact number of shakes.

5.6 Usability of Remotion

We further compared Remotion's usability with the baseline method, and gathered responses from all 9 analysts with a modified NASA TLX survey at the end of the study, as shown in Fig. 6. We used one-way repeated measures ANOVA tests on each of the five dimensions and found a significant difference in "Mental Demand" ($F(2, 16) = 4.71$, $p = 0.025$, $\eta^2 = 0.20$) and marginal differences in "Effort" ($p = 0.059$, $\eta^2 = 0.12$), "Preference" ($p = 0.089$, $\eta^2 = 0.24$), and "Successful" ($p[GG] = 0.095$, Greenhouse-Geisser adjustment because of violating the sphericity assumption, and $\eta^2 = 0.19$). We then employed simultaneous tests and used a Bonferroni correction ($\alpha = 0.017$), finding a significant difference among "Mental Demand" when comparing Remotion hardware visualization to the baseline method ($p = 0.007$, $\mu = 3.38$ vs. 6.31, $\sigma = 0.52$ vs. 0.76); we also found a marginally significant difference among "Preference" when comparing these two conditions ($p = 0.015$, $\mu = 7.06$ vs. 4.50, $\sigma = 0.52$ vs. 0.75); there is no significant difference found for other three dimensions.

We conclude that the Remotion hardware visualization can reduce analysts' mental load over the baseline method, but the effect is small; it is also preferred by the analysts, but the effect size is unknown given the small sample size. The Remotion hardware visualization may reduce effort and increase successful judgements, but we are not able to draw a conclusion.

6 DISCUSSION

6.1 Differences Between Visualization Conditions

There were some qualitative differences between the baseline time-series visualization, the Remotion hardware visualization, and the Remotion software visualization. Analysts reported differences between the two Remotion's visualization methods. Remotion made these strong emotional moments more visible. Many traditional remote usability methods (e.g., screen recordings, click-through logs, surveys) cannot recreate these moments, and this important contextual information is usually not captured. Being able to tell which events trigger strong emotional responses can be useful for improving the user experience, as well as creating emotionally engaging interfaces.

Analysts 3 and 9 reported that the Remotion hardware visualization enabled a realistic 3D observation. The physical depth and one-to-one scale movement replay enabled direct observation of the actors' hand movement; analysts needed to just move their head to see around the device. The Remotion software visualization required them to use a mouse to adjust the viewing angle. This is an indirect method and could cause disorientation in practice (e.g., analysts could not find the original direction). The Remotion hardware visualization did not require constant attention and analysts could quickly recover from distractions since the sound of the motor movement signals ongoing mobile device movements. While the hardware made noises from the motors as they moved the device, this was not distracting. The Remotion software visualization on the other hand, lacked the mechanical sound, so rapid or small movements could pass unnoticed. Similarly to the times-series visualization, the Remotion software visualization required constant attention to the screen. This may explain the lower NASA TLX score for mental demand for the Remotion hardware visualization.

Some analysts reported earlier that they could not focus both on the time-series visualization and the screen replay at the same time, but they could do a better job when watching both Remotion's visualization methods. This was not the case for every analyst, perhaps due to various degrees of familiarity with this type of visualization.

6.2 Implications on the Deployment of Remotion

Mobile devices are often involved in user studies where the user needs to perform in-lab style studies at home (e.g., on crowdsourcing platforms like Amazon Mechanical Turk). If well-designed, these remote usability testings are as effective as the in-lab counterparts, as Andreason et al. argue that the "remote usability testing in the RS condition is feasible compared to the conventional laboratory-based method" [2]. A remote user study can be completely anonymous, with the user visiting a website or using a mobile application at home, while only having their behavior tracked. This allows studies to be conducted on a more geographically and demographically diverse set of users, rather than primarily local users with the time flexibility to come into a lab for experiments. In many cases, it is inconvenient for the user to have to return to the experimenter's lab frequently, creating extra burdens for both the participants and the experimenters.

Remotion's visualization, as already presented, has benefits for observing remote users' active hand posture, level of engagement, emotion, and habits. In fact, we expect that the deployment of Remotion in the real-world can produce more effective and a wider range of insights compared to our study, since we truncated the audio cues to evaluate how purely visual and motion signals benefit remote usability testing. With audio recordings, researchers are one step closer to eliminating ambiguities in interpreting detailed emotions. A person who speaks a lot at the beginning of the task but gradually becomes silent with increased movement and task progress can be assumed to be concentrating more, a behavior observed in actor 2. Enabling audio does not require the remote user to do any extra work, except consent to providing these data. Consequently, researchers are able to use the audio feedback with Remotion to cross-reference the findings, and in turn gain a better understanding of the contextual information of a remote user.

Due to system constrains and security reasons, many scenarios require the Remotion client and Remotion screen projection component to be installed on the remote user's mobile device. But when the remote usability is tested through a web-browser, Remotion does not require any additional installation on the remote user's side. This is achieved by the website simply including the Remotion client JavaScript library.

6.3 Informational Value of Motion Data

Smartphone motion data, when coupled with screen recordings, can be valuable to understand the changing status of remote users. Active hand use, for example, is a dynamic indicator different from the user's dominant hand use. We observed the actors switched hands during a prolonged task, perhaps indicating physical fatigue; at times, they start to hold the phone with both hands when the tasks get harder or they become more engaged. As a result, the active hand use tells us more information about the actor's (i.e., remote user) dynamic status beyond knowing their general dominant hand use, and can help designers infer usability problems. Factors such as emotion, attention and orientation events (e.g., laying down the smartphone) from the motion data can provide further detail for usability decisions; detecting frustration and confusion periods can guide onboarding interfaces, while periods of inattention may lead to a reminder of what the user was doing when they return.

Traditional methods such as video and audio recording offer a viable solution in capturing contextual information from a remote in-home user study session. However, the requirement for extra hardware and breaching of anonymity makes it difficult to deploy this technology at scale. Due to privacy concerns, recording the home environment of the participants can be challenging. The use of the Remotion client, however, poses almost no extra effort on the participants' side and does not require the remote users to install any additional hardware.

Remotion could enrich traditional methods such as the think-aloud protocol. The underlying hand movement can be compared against what the remote participants say or do. This can also lead to insights into the personal habits of mobile devices' use. For example, in our study some actors did not speak when they were nervous, firmly gripping their phones instead. Others did the reverse, increasing verbal output and hand motion simultaneously. We believe this is an interesting future direction; in combination with traditional methods, Remotion can yield more insights from interpreting rich contextual information from remote in-home user studies.

6.4 Future Work

The current study had Remotion use actors' replays that did not take place live. It would be interesting to observe how an experimenter might use Remotion differently in a live remote user study scenario. The ability to watch the remote user interact with research tasks in real time and interact with the user would enable a better understanding of the user's performance. This could make providing instructions easier, and the experimenters could guide the participants through the tasks in real time, creating a virtual presence for both parties.

Furthermore, while Remotion currently operates in a one-analyst-one-user mode, multiple users can be run simultaneously. Our findings suggest that Remotion is less mentally demanding while being similarly productive in terms of insights compared to the typical time-series visualization, which creates opportunities for analysts to observe multiple devices each used by a different user at once. Hence, one person can give instructions to many users at home and visually see aggregated mobile activity.

6.5 Limitations

With its current design, Remotion is not effective for situations where a user is moving around with the phone while performing a task. Body movement is not captured by the phone and if it were, replicating that in a replay is challenging. Therefore, Remotion is effective for a mobile device, but not when it is used while mobile. In fact, by itself, the motion data is difficult to directly make inferences from without also seeing the screen, a cognitively challenging task; the sensor readings are sensitive to drift and the data provides only additional context but not a clear statement about what the user is doing.

Our study design also has limitations, partly due to the small number of participants. We recruited analysts who had experience with user studies and interpreting behavior data. However, this small sample size has limited statistical power and external validity. We are unable to provide compelling quantitative analysis for making comparisons in the study due to a low number of trials. Our findings are mainly qualitative and show what are potential advantages of various visualizations, and further experimentation to understand issues of cognitive load in comparison to simpler baselines of screen-only replays or click log data.

7 CONCLUSION

This work describes Remotion, a system that captures and replays remote studies such as usability testing on mobile devices. Remotion extends existing technology that remotely casts the screen display by replaying the device's motions made by users while they were using the device. Through different application tasks performed by actors, the nine analysts were able to interpret users' changes in attention, emotion, habits, and how a remote user actively holds the device. These characteristics enable a richer understanding of a user's mental model, corroborating and supplementing their think-aloud feedback.

With the rise in usage of mobile devices, the ability to replay user activity captured in naturalistic environments can enable new and valuable remote user studies. In addition, Remotion extracts more information about a user's performance in a remote usability testing, without the privacy or data size issues of video recordings. We seek to make remote user studies feel as if the user is invisible but in the room manipulating the device.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. IIS-1552663. The authors would like to further thank Shaun Wallace and Sachin Pendse for helping with editing, Ren Fang Xu for creating the illustrations, and Ishaan Agarwal for contributing to the programming of Remotion's client software.

REFERENCES

[1] Sabir Akhadov, Marcel Lancelle, Jean-Charles Bazin, and Markus Gross. 2016. Motion based remote camera control with mobile devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 428–433.

[2] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What Happened to Remote Usability Testing?: An Empirical Study of Three Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1405–1414. https://doi.org/10.1145/1240624.1240838

[3] Adriana Holtz Betiol and Walter de Abreu Cybis. 2005. Usability Testing of Mobile Devices: A Comparison of Three Approaches. In *Human-Computer Interaction - INTERACT 2005: IFIP TC13 International Conference, Rome, Italy, September 12-16, 2005. Proceedings*, Maria Francesca Costabile and Fabio Paternò (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 470–481. https://doi.org/10.1007/11555261_39

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[5] Anders Bruun, Peter Gull, Lene Hofmeister, and Jan Stage. 2009. Let Your Users Do the Testing: A Comparison of Three Remote Asynchronous Usability Testing Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1619–1628. http://doi.acm.org/10.1145/1518701.1518948

[6] Liang Cai and Hao Chen. 2011. TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion. *HotSec* 11 (2011), 9–9.

[7] José C. Castillo, H. Rex Hartson, and Deborah Hix. 1998. Remote Usability Evaluation: Can Users Report Their Own Critical Incidents?. In *CHI 98 Conference Summary on Human Factors in Computing Systems (CHI '98)*. ACM, New York, NY, USA, 253–254. https://doi.org/10.1145/286498.286736

[8] Constantinos K Coursaris and Dan J Kim. 2011. A meta-analytical review of empirical mobile usability studies. *Journal of usability studies* 6, 3 (2011), 117–171.

[9] Jan Dijkstra and Harry Timmermans. 2002. Towards a multi-agent model for visualizing simulated user behavior to support the assessment of design performance. *Automation in Construction* 11, 2, 135 – 145. https://doi.org/10.1016/S0926-5805(00)00093-5 ACADIA '99.

[10] Trinh Minh Tri Do, Jan Blom, and Daniel Gatica-Perez. 2011. Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 353–360.

[11] Susan Dray and David Siegel. 2004. Remote Possibilities?: International Usability Testing at a Distance. *interactions* 11, 2, 10–17. https://doi.org/10.1145/971258.971264

[12] Mayank Goel, Jacob Wobbrock, and Shwetak Patel. 2012. GripSense: Using Built-in Sensors to Detect Hand Posture and Pressure on Commodity Mobile Phones. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 545–554. https://doi.org/10.1145/2380116.2380184

[13] Greg Guest, Kathleen M MacQueen, and Emily E Namey. 2011. *Applied thematic analysis*. sage.

[14] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.

[15] Hecorat. 2014. AZ Screen Recorder. https://play.google.com/store/apps/details?id=com.hecorat.screenrecorder.free. (2014). [Online; accessed 2017-09-18].

[16] Hiroshi Ishii, John Underkoffler, Dan Chak, Ben Piper, Eran Ben-Joseph, Luke Yeung, and Zahra Kanji. 2002. Augmented urban planning workbench: overlaying drawings, physical models and digital simulation. In *Proceedings. International Symposium on Mixed and Augmented Reality*. 203–211. https://doi.org/10.1109/ISMAR.2002.1115090

[17] Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2013. Evaluating the Efficiency of Physical Visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2593–2602. https://doi.org/10.1145/2470654.2481359

[18] Takayuki Kanda, Hiroshi Ishiguro, Takahiko Ono, Masayoshi Imai, and Ryohei Nakatsu. 2002. Development and evaluation of an interactive humanoid robot "Robovie". In *Proceedings 2002 IEEE International Conference on Robotics and Automation*, Vol. 2. 1848–1855

vol.2. https://doi.org/10.1109/ROBOT.2002.1014810

[19] Michael Kipp. 2010. Multimedia annotation, querying and analysis in ANVIL. *Multimedia information extraction* 19 (2010).

[20] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 453–456. https://doi.org/10.1145/1357054.1357127

[21] Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki, and Mamoru Mitsuishi. 2000. GestureMan: A Mobile Robot That Embodies a Remote Instructor's Actions. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, New York, NY, USA, 155–162. https://doi.org/10.1145/358916.358986

[22] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. 2011. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explorations Newsletter* 12, 2 (March 2011), 74–82. https://doi.org/10.1145/1964897.1964918

[23] Kun Qian, Jie Niu, and Hong Yang. 2013. Developing a Gesture Based Remote Human-Robot Interaction System Using Kinect. In *International Journal of Smart Home*. 203–208.

[24] David Sachs. 2010. Sensor Fusion on Android devices: A Revolution in Motion Processing [Video]. Google Tech Talk. http://www.youtube.com/watch?v=C7JQ7Rpwn2k. (2010).

[25] Simon Stusak, Aurelien Tabard, Franziska Sauka, Rohit Ashok Khot, and Andreas Butz. 2014. Activity Sculptures: Exploring the Impact of Physical Visualizations on Running Activity. *IEEE Transactions on Visualization and Computer Graphics* 20, 12, 2201–2210. https://doi.org/10.1109/TVCG.2014.2352953

[26] Faisal Taher, Yvonne Jansen, Jonathan Woodruff, John Hardy, Kasper Hornbaek, and Jason Alexander. 2017. Investigating the Use of a Dynamic Physical Bar Chart for Data Exploration and Presentation. *IEEE Transactions on Visualization and Computer Graphics* 23, 1, 451–460. https://doi.org/10.1109/TVCG.2016.2598498

[27] Theophanis Tsandilas, Caroline Appert, Anastasia Bezerianos, and David Bonnet. 2014. Coordination of Tilt and Touch in One- and Two-handed Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2001–2004. https://doi.org/10.1145/2556288.2557088

[28] Sarah Waterson, James A. Landay, and Tara Matthews. 2002. In the Lab and out in the Wild: Remote Web Usability Testing for Mobile Devices. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 796–797. https://doi.org/10.1145/506443.506602

Received February 2018; revised April 2018; accepted May 2018