Touch-Move-Release: Studies of Surface and Motion Gestures for Mobile Augmented Reality

Ze Dong [1,2], **Jingjing Zhang** [1,2], **Xiaoliang Bai** [3], **Adrian Clark** [1], **Rob Lindeman** [2], **Weiping He** [3], **Thammathip Piumsomboon** [1]

[1] *School of Product Design, University of Canterbury, New Zealand*
[2] *HIT Lab NZ, University of Canterbury, New Zealand*
[3] *Cyber-Physical Interaction Lab, Northwestern Polytechnical University, Xi'an, China*

Correspondence*:
Ze Dong
ze.dong@pg.canterbury.ac.nz

ABSTRACT

Recent advancements in both hardware and software for mobile devices have allowed developers to create better mobile Augmented Reality (AR) experiences, which has led to an increase in the number of mobile AR applications and users engaging in these experiences. However, despite a broad range of mobile AR applications available to date, the majority of these applications that we surveyed still primarily use surface gestures, i.e., gesturing on the touch screen surface of the device, as the default interaction method and do not utilise the affordance of three-dimensional user interaction that AR interfaces support.

In this research, we have investigated and compared two methods of gesture interaction for mobile AR applications: Surface Gestures, which are commonly used in mainstream applications, and Motion Gestures, which take advantage of the spatial information of the mobile device. Our goal is to determine if motion gestures are comparable or even superior to surface gestures for mobile AR applications. To achieve this, we have conducted two user studies: an elicitation study and a validation study. The first study recruited twenty-one participants and elicited two sets of gestures, surface and mobile gestures, for twelve everyday mobile AR tasks. This yielded a total of five hundred and four gestures. The two sets of gestures were classified and compared in terms of goodness, ease of use, and engagement. As expected, the participants' elicited surface gestures are familiar and easy to use, while motion gestures were found more engaging.

Using design patterns derived from the elicited motion gestures, we proposed a novel interaction technique called ”TMR” (Touch-Move-Release). We developed a mobile AR game similar to Pokémon GO to validate this new technique and implemented a selected gesture chosen from the two gesture sets. A validation study was conducted with ten participants, and we found that the motion gesture enhanced engagement and provided a better game experience. In contrast, the surface gesture provided higher precision resulting in higher accuracy and was easier to use. Finally, we discuss the implications of our findings and give our design recommendations for using the elicited gestures.

Keywords: augmented reality, mobile device, gestures, elicitation study

1 INTRODUCTION

According to Azuma (Azuma, 1997), for an experience to be considered Augmented Reality (AR), it must be capable of demonstrating three characteristics. Firstly, it must be able to combine real and virtual imagery. Secondly, it has to support real-time interaction. Lastly, it must be able to register virtual content in 3D space. At the time of writing, arguably the most advanced commercial AR system available is the Microsoft HoloLens 2 (Microsoft Corporation, 2020). This head-worn AR device is capable of achieving all three characteristics, nevertheless, it is aimed toward professional uses. On the contrary, handheld devices such as mobile phones are ubiquitous, and they are currently the primary way people can experience AR in several domain applications (Billinghurst et al., 2015; Van Krevelen and Poelman, 2010).

Constant improvements in mobile hardware and software to support AR have led to an increasingly wide range of mobile applications which use AR as their core mechanic for visualisation and interaction. Mobile AR frameworks, such as Apple's ARKit (Apple Inc., 2020) and Google's ARCore (Google Inc., 2020), have made the development of mobile AR applications accessible to more developers than ever. This has led to a variety of AR applications in several domains, for example, IKEA Place (inter IKEA Systems B.V., 2020) allows customers to visualise virtual furniture in their home, while QuiverVision (QuiverVision, 2020) is the first application to introduce AR colouring books, and SketchAR (Sketchar.tech, 2020) teaches users how to draw by overlaying virtual drawings over a real canvas. By surveying popular mobile AR applications available on the Google Play and Apple App Stores, we found that those applications use interaction metaphors based on surface gestures designed for devices with touch-sensitive screens. While touch input is the dominant and familiar method of interaction for regular mobile users, past research has demonstrated that methods beyond those currently used in mobile AR applications can enrich mobile AR experiences.

Surface gestures based on touch input are the conventional interaction technique used in handheld mobile devices and have been adopted by consumer mobile AR applications. Previous studies have explored various design principles of surface gestures using different methodologies ranging from expert design (Wilson et al., 2008; Wu et al., 2006), participatory design by non-experts (Wobbrock et al., 2009), and comparative studies of both groups (Morris et al., 2010). Nevertheless, surface gestures have several drawbacks, e.g., a limited surface area for interaction of the handheld devices (Bergstrom-Lehtovirta and Oulasvirta, 2014), the interaction is limited to two dimensions (Bai, 2016; Hürst and Van Wezel, 2013), and only a limited number of fingers can fit in the interaction area (Goel et al., 2012). Furthermore, gesturing on the screen tends to cause occlusion (Forman and Zahorjan, 1994) and focusing on on-screen interaction may lead to dual perspectives (Čopič Pucihar et al., 2014).

Another type of gesture, the mid-air gestures, is widely used in a head-mounted display (HMD) based AR as they can offer 3D interaction (Piumsomboon et al., 2014). However, mid-air gestures are not ideal to use in public (Rico and Brewster, 2010), prolonged usage can lead to fatigue (Hincapié-Ramos et al., 2014), and bimanual gestures are not suitable for handheld mobile devices. The third type of gesture, known as motion gestures, utilises the mobile device's built-in sensors to detect the device's movements. Past research has proposed and demonstrated motion gestures as an interaction technique for handheld devices either alone (Ashbrook and Starner, 2010; Henrysson et al., 2005; Hinckley et al., 2000; Jones et al., 2010; Ruiz et al., 2011) or in conjunction with a secondary device (Chen et al., 2014; Stanimirovic and Kurz, 2014). In addition, previous work has explored motion gestures used in the context of AR, for example, direct camera manipulation (Henrysson et al., 2005), or virtual object manipulation (Ha and Woo, 2011; Mossel et al., 2013) however, to our knowledge, there has not been any research that explores the participatory design of motion gestures for a broad range of mobile AR applications nor compares them against the conventional interaction techniques to validate their usability.

Due to limited existing work on motion gestures for mobile AR interaction, we decided to conduct a study to explore various gestures suitable for different tasks in the mobile AR context. We chose to pursue a participatory design methodology, specifically, an elicitation study (Cooke, 1994), following the method of Wobbrock et al. (Wobbrock et al., 2009). While there have been elicitation studies conducted for motion gestures for handheld devices (Ruiz et al., 2011) and gestures for AR context (Piumsomboon et al., 2013b) in the past, the former study focused on eye-free interaction in a non-AR context and the latter emphasises on gestures for head-worn AR systems. Our goal is to explore a gesture design space for handheld devices, exploring gesture-based interaction sets that utilise 3D spatial information of mobile devices to create compelling AR experiences.

We surveyed common tasks users commonly perform in popular mobile AR applications. We selected twelve tasks for elicitation of the two gesture sets (surface and motion gestures) from the participants, and compared them using subjective ratings. We hypothesised that elicited surface gestures would be rated higher in terms of suitability and ease of use, while the elicited motion gestures would be more engaging. We developed a mobile AR game based on Pokémon GO for the follow-up study to validate the selected gestures in the task of throwing a ball. We hypothesised that there would be differences in terms of accuracy, subjective ratings, in-game experience, and system usability between the two interaction techniques. This research provides the following contributions to the field:

1. A literature review of surface, mid-air, and motion gestures, mobile AR interaction, and previous research with elicitation studies.

2. An elicitation study that yielded two sets of user-defined surface and motion gestures, anecdotal feedback, and a comparison between them in terms of suitability, ease of use, and engagement.

3. An overview of an example mobile AR application, a Pokémon GO-like game, and the implementation of the selected surface and motion gesture elicited in this game.

4. A validation study comparing the two gestures in terms of accuracy based on three levels of target sizes, subjective ratings from the previous study, an in-game experience questionnaire, system usability scale, and user preferences.

5. From the results of both studies, a discussion and summary of our findings have been provided, including implications of our work and guidelines for using this work in the future mobile AR applications. From the design pattern, we proposed the TMR (Touch-Move-Release) interaction technique for mobile AR applications.

In the sections to follow, we cover our literature review in Section 2. Section 3 reports the results of the elicitation study. Section 4 provides an overview of the mobile AR application development, gesture implementation and the results of the validation study. Our discussion of the outcomes of both studies and their implication is covered in Section 5. Finally, the conclusion and future work are presented in Section 6.

2 RELATED WORK

In this section, we cover background research on topics related to gesture-based interaction techniques and participatory design methodology, focusing on elicitation studies. The previous work has been categorised into four subsections. We provide a brief overview of research on surface and mid-air gestures in Section 2.1 and introduce motion gesture interaction technique in Section 2.2. Next, we cover past mobile AR interactions, including the current state of the art in Section 2.3. Previous elicitation studies are discussed in Section 2.4. Finally, Section 2.5 covers our research questions and goals.

2.1 Surface and Mid-Air Gestures

Surface gestures are fundamental methods of interaction for surface computing as they utilise the touch-sensitive screen as the primary input. Past research has provided various guidelines for designing and implementing surface gestures. Wobbrock et al. (Wobbrock et al., 2009) proposed taxonomy and a user-defined surface gesture set from an elicitation study of twenty participants who had no training in the area of interaction design. In their follow-up study (Morris et al., 2010), they compared the user-defined gestures set to the gesture set created by three interaction design experts. They found that the user-defined gestures were rated higher than the expert set. Moreover, although some of the expert's gestures were more appealing, the participants ultimately preferred simpler gestures that took less effort to perform.

Wu et al. (Wu et al., 2006) proposed three aspects of surface gesture design, including gesture registration, gesture relaxation, and gesture and tool reuse, which considered the interaction context, comfort level, and applicability of each gesture to different tasks, respectively. They developed a prototype application for a tabletop surface computing system and implemented four types of gestures, including annotate, wipe, cut, copy and paste, and pile and browse. In another approach, Wilson et al. (Wilson et al., 2008) focused on improving the realism of surface interaction through physics simulation by creating proxy particles to exert force on virtual objects. They experimented with six participants to complete three physics-based positioning, sorting, and steering tasks. They found that interaction through the proxy particles could shorten task completion time and received positive feedback for the proposed technique. Despite the progress, challenges remain when applying these surface gestures directly on the handheld devices, e.g., limited interaction space unable to support certain gestures (Bai, 2016), occlusion of the display (Forman and Zahorjan, 1994), and limited reach when operating single-handedly (Bergstrom-Lehtovirta and Oulasvirta, 2014). Further discussion on surface gestures in mobile AR is covered in Section 2.3.

Mid-air gestures, or gesturing in the air, refers to gestures that are performed while holding an arm or arms in front of one's body to perform actions such as pointing, pushing, waving. Through the physical nature of arm and hand movement, these gestures can provide a more engaging experience while interacting. For instance, Cui et al. (Cui et al., 2016) investigated the user's mental models while performing mid-air gestures for shape modelling and virtual assembly with sixteen participants. They found that users had different preferences for interaction techniques and felt more natural and comfortable with their preferred method, and bimanual gestures were more natural than unimanual. Kyriazakos et al. (Kyriazakos et al., 2016) designed a novel fingertip detection algorithm to extend this interaction method. The interaction between the user and the virtual object was achieved by tracking mid-air gestures using the rear camera of the mobile device. For example, the user could use a V-hand pose to move the virtual object using two fingers. Previous studies have also investigated mid-air gestures in various use cases and settings, such as on public displays (Walter et al., 2014), or using pairing between an armband sensor and a handheld device (Di Geronimo et al., 2017).

Although mid-air gestures can take advantage of the 3D interaction space, using them for an extended period can lead to fatigue and discomfort. Rico and Brewster (Rico and Brewster, 2010) also raised the issue of the social acceptability of performing mid-air gestures in public. Their study had the participants watch videos of different gestures and were asked to imagine performing those gestures in a social environment. They found that the thought of performing gestures in social environments does affect gesture preference, and device-based gestures were found to be more socially acceptable. To validate their findings, they had eleven participants perform chosen gestures in public and found that gestures that attracted less attention were preferable. They recommended that the designer avoid emblematic gestures, which might lead to confusion in social contexts, and support more familiar and socially acceptable gestures for use in public. Compared to surface gestures, which are strictly 2D in nature, mid-air gestures are performed in 3D space, bringing increased flexibility and more possibilities to the gesture design space. Previous research has found that two-handed mid-air gestures are commonly elicited, especially in AR tasks. However, these bimanual gestures are not suitable for mobile devices because users need to hold the device in their hands.

2.2 Motion Gestures on Handheld Devices

Motion gestures are interaction methods that utilise the hardware device's motion using measurements from sensors such as inertial measurement units (IMU). The orientation and linear acceleration obtained can determine the device's movement in 3D space. Compared to surface gestures, motion gestures support a broader range of hand and arm movement, allowing for various interactive experiences. Past research has utilised integrated inputs for richer user experiences. Hinckley et al. (Hinckley et al., 2000) integrated multiple sensors, including proximity range, touch sensitivity, and tilt sensors, to introduce novel functionalities on a handheld device. For example, the device would wake up when picked up, and the user can scroll by tilting the device. They found that sensors opened up new possibilities allowing a vast interaction design space for handheld devices. Later, Wigdor and Balakrishnan (Wigdor and Balakrishnan, 2003) proposed TiltText, an interaction technique that could reduce ambiguities of the text input process using a combination of a touch screen keypad and four tilting directions: left, right, forward, and back. TiltText was found to be faster to perform, but it had a higher error rate. GesText (Jones et al., 2010) was another system that made use of an accelerometer to detect motion gestures for text input. Again, they found that the area-based layout supported by simple tilt motion gestures was more efficient and preferred over the alphabetical layout.

Hartmann et al. (Hartmann et al., 2007) explored the role of sensors in handheld interaction, dividing the development process into three steps: connecting the appropriate hardware, defining the interaction logic, and establishing the relationship between the sensors and the logic. They proposed a tool to help interaction designers map the connections between the sensors and logic to support direct manipulation and pattern recognition. They showed that sensors had become a crucial tool for interaction designers to enhance interaction and overall user experience. Ashbrook et al. (Ashbrook and Starner, 2010) raised concerns regarding motion gesture design, with particular emphasis on two points: the practicality of the proposed gestures for the actual recognition technology and their robustness for avoiding false registration or activation. To address these issues, they proposed MAGIC, a motion gesture framework that defines the design process in three stages: requirement gathering, determining the function activation, and user testing. This process enabled non-experts to leverage sensors (e.g. accelerometers) in the design process of motion gestures.

The GripSense system (Goel et al., 2012) explored a possible solution to address difficulties with single-handed surface interaction, such as allowing the ”pinch-to-zoom” gesture using one hand. They used touch input in conjunction with an inertial sensor and vibration motor to detect the level of pressure exerted by the users on the screen. The technique allowed complex operations to be performed with a single hand. In another application, motion gestures could be performed on a handheld device to provide inputs and interaction with virtual objects on a large display system (Boring et al., 2009). In terms of human-centred design, Ruiz et al. (Ruiz et al., 2011) conducted an elicitation study with twenty participants to collect motion gestures for nineteen tasks on a handheld device. They categorised the gesture dataset based on gesture mapping and physical characteristics. They found that the mapping of commands influenced the motion gesture consensus. Past research has demonstrated that handheld device sensors can be leveraged to recognise motion gestures through the device's movement. The benefits of motion gestures inspired us to further explore the design space in the context of mobile AR applications. We propose a comparison between surface and motion gestures on a handheld device for mobile AR, which we believe has not been investigated prior to this research.

2.3 Mobile AR Interaction

There have been a number of research publications proposing various methods of interaction in mobile AR that offer different experiences for user interaction. Early research demonstrated that mobile AR could provide precise 6-DOF (degree-of-freedom) camera/viewpoint control through the movement of the handheld device through tracking a known target in the physical environment. In one of the first face-to-face collaborative mobile AR applications, Henrysson et al. (Henrysson et al., 2005) developed AR Tennis with two users sitting across a table who could play a game of virtual tennis on the table. The mobile device's rear camera registered an image marker placed on the table to determine the device's 6DOF pose relative to the marker. To interact with the virtual tennis ball, the user could move the camera in front of the ball's incoming path and nudge the device forward to exert force onto the ball in order to hit it back. Through feedback gained from a user study, they provided guidelines for designing games for mobile AR, including the recommendation to provide multi-sensory feedback, focus on the interaction, and support physical manipulation. They found that the combination of visual, tactile, and auditory outputs during the experience and the viewpoint control offered by AR experiences further increased the level of immersion and could improve collaboration and entertainment value.

The AR-Tennis prototype inspired Ha and Woo (Ha and Woo, 2011) to develop the ARWand, a 3-DOF device for mobile AR interaction based on the device's sensors. Users could use surface gestures on the device's screen to manipulate virtual objects - for example, when the device was held perpendicular to the ground, swiping up or down would move the object higher or lower. When it was held parallel to the ground, the swipes would manipulate the object forward and backward. This technique supported individual axis control; however, the built-in sensors lacked precision. Later, Mossel et al. (Mossel et al., 2013) proposed HOMER-S, a 6-DOF interaction technique to support object manipulation. The user could perform translation or rotation via the touch interface using surface gestures to change the virtual object's position and orientation. This technique supported single-handed operation and could complete tasks faster at the cost of lower accuracy.

To improve the accuracy of the manipulation and reduce the effect of shaky hands, Lee et al. (Lee et al., 2009) proposed a technique called ”Freeze-Set-Go”, which allowed the user to pause the current viewpoint of the AR system so that the user could manipulate the object in the current view. Once the task was completed, the user could resume the regular tracking of the viewpoint and update the virtual object's location accordingly. They found that this technique helped improve accuracy and reduce fatigue. Tanikawa et al.(Tanikawa et al., 2015) created a mobile AR system to support multimodal inputs of viewpoint, gesture, and device movement. They demonstrated the interaction in an AR Jenga game where the user could touch the screen to select the virtual wooden block to move. While holding the finger on the screen, the block was kept at a fixed distance from the screen. The user could move the block by moving the device, and when the finger was removed from the screen, the block would then be released. This technique provided reasonable accuracy and a better object manipulation experience in mobile AR.

In addition to the accuracy of manipulation techniques in mobile AR, the dual perspectives problem (Čopič Pucihar et al., 2014) is another issue that impacted mobile AR interaction. This problem occurs when the viewpoint captured by the device's rear camera and displayed on the screen does not match the scale of the real world as viewed from the user's actual perspective. Furthermore, ergonomics was also identified as another limitation of handheld-based mobile AR. Colley et al. (Colley et al., 2016) evaluated the ergonomics of the camera placement of mobile AR devices by comparing the level of tilt of the camera to the screen. They found that the screen size and a proper tilt level had a significant impact on comfort level while interacting.

Other approaches to enhancing mobile AR interaction included combining multiple devices for multi-device input and using mobile AR systems for visualising embedded sensors. Goldsmith et al. (Goldsmith et al., 2008) demonstrated SensAR, combining a mobile device with environmental detection sensors. When a marker has been scanned using the handheld device, the sensor shares the environment data displayed in AR. They found this method of interaction and visualisation to be seamless and immersive. Stanimirovic and Kurz (Stanimirovic and Kurz, 2014) took this idea further and used a smartwatch's camera to scan markers for hidden AR content scattered in the environment and notify the user to use their handheld device to access them. In another context, Chen et al. (Chen et al., 2014) used a smartwatch to edit text and send the updated text to the handheld device. Nevertheless, multi-device interaction requires additional devices to operate and might not be ideal for mobile AR applications in general.

Despite significant advancements in mobile AR technology, challenges still exist in the development of better mobile AR interactions(Kurkovsky et al., 2012). Hürst and Van Wezel (Hürst and Van Wezel, 2013) found that existing mobile AR interaction was limited to 2D screen interaction, such as touching and swiping, which could cause issues with occlusion of the screen hindering user performance and experience. To alleviate this, they proposed a technique to use the device's rear camera to track the user's thumb and index finger for direct manipulation of virtual objects. They found that this approach could offer more natural interaction that does not occlude the screen. Nevertheless, a constraint exists that the fingers must remain in the camera's view, and both hands are required to manipulate the object. Similarly, Bai (Bai, 2016) also explored mobile AR interaction behind the handheld device. When evaluating these mobile AR interaction methods, he found that 3D gesture-based manipulation was more intuitive and engaging than surface gestures and could be less fatiguing than motion gestures.

In a literature review of existing mobile AR interaction research, we found that interaction based on the device's movement could offer the users true 3D interaction(Piumsomboon et al., 2013a), but accuracy was an issue. We observed that interaction techniques that combined touch input and device movement could improve the precision of the interaction by utilising the touchscreen while preserving the 3D interaction experience. For this reason, we decided to investigate motion gestures, which combine touch input and device movement in this research.

2.4 Elicitation Study

Elicitation study is a method of collecting knowledge by analysing behaviour patterns and feedback of participants (Cooke, 1994). One such example is the work of Voida et al. (Voida et al., 2005), who explored interaction with projection displays in an office environment. They mapped user mental models by observing the user's manipulation of 2D objects in AR and asked the participants to propose gestures while interacting with multiple projection displays. Pointing gestures were commonly used to interact from afar, but the touch-based user interface was preferable when the virtual objects were situated closer to the participants. Epps et al. (Epps et al., 2006) studied user preferences for tabletop interaction through an elicitation study where they displayed images depicting different tasks on the desktop and asked twenty participants to propose gestures for the tasks. The study presented the guidelines for the hand poses for gestures and corresponding tasks for tabletop systems. They found that the index finger was frequently used in multiple tasks, such as tapping, drawing, or swiping, over seventy per cent of the time.

Several elicitation studies have yielded gesture taxonomies and collections of user-defined gestures. In surface computing, Wobbrock et al. (Wobbrock et al., 2009) conducted a study where non-expert participants were asked to design surface gestures, and the quality of those gestures was evaluated in terms of suitability and ease of use on a tabletop system. By eliciting one thousand and eighty gestures from twenty participants, they proposed a taxonomy and user-defined surface gesture set based on the gestures with a high consensus score. Using the think-aloud protocol, they could record the users' design process and provide guidelines for designers. Ruiz et al. (Ruiz et al., 2011) applied the same elicitation procedures to explore motion gestures for handheld devices. The elicited motion gestures exhibited characteristics of two dimensions of movement and command mappings. They provided anecdotal findings to help designers design better gestures that mimic everyday tasks and discussed how sensors could be used to better recognise those motion gestures. Later, Piumsomboon et al. (Piumsomboon et al., 2013b) adopted the methodology and elicited gestures for an AR head-mounted display system. They extended Wobbrock's taxonomy and identified forty-four user-defined gestures for AR. They found that the majority of the gestures were performed mid-air, and most of the gestures were physical gestures that mimicked direct manipulation of the objects in the real world. They also found that similar gestures often shared the same directionality with only variants of hand poses. The anecdotal findings and implications were provided to guide designers in deciding which gestures to support for AR experiences.

Past research has also included comparative elicitation studies. Hayati et al. (Havlucu et al., 2017) investigated two interaction techniques, on-skin and freehand gestures, which did not require an intermediate device for input. They compared two user-defined gesture sets on four aspects: social acceptability, learnability, memorability, and suitability. With a total of twenty participants, they found that on-skin gestures with small movements were better for social acceptance. In contrast, participants found freehand gestures to be better for immersion. Chen et al. (Chen et al., 2018) extended the research to explore inputs on the other body parts. They also collected user-defined gesture sets and validated them with another group of participants. Their method differed from the previous elicitation studies that combined the subjective and physiological risk scores. They found that gestures combined with the body parts helped enhance the naturalness of the interaction. Elicitation studies have also compared gestures under different use-cases and scenarios. May et al. (May et al., 2017) elicited mid-air gestures to be explicitly used within an automobile. They found that a participatory design process yielded easier gestures to understand and use than gestures designed by experts. Tran et al. (Pham et al., 2018) also elicited mid-air gestures for three different spaces: mid-air, surface, and room, for a virtual object of varying sizes. It was found that the scale of the target objects and the scenes influenced the proposed gestures.

Elicitation studies have been used to obtain user-defined gestures that were simple and easy to use, allowing designers to reuse the design patterns and apply them in various settings and constraints, which reflect the user's mental model under the given circumstance. Previous studies have shared design guidelines for various gestures, whether surface, motion, or mid-air, for different tasks, systems, and scenarios. Nevertheless, we have not encountered any research that has elicited gestures for mobile AR experiences. As a result, there is limited knowledge of design practices and guidelines for motion gestures in mobile AR interaction. Moreover, there has not been any comparison of the performance differences between surface and motion gestures for mobile AR interaction. Therefore, we have conducted an elicitation study to elicit surface and motion gestures for mobile AR to explore possibilities in the proposed design space to address this shortcoming.

2.5 Research Questions and Goals

In this research, we are interested in enhancing the user experience of mobile AR applications through novel interaction techniques. From the literature review, we discovered three common mobile AR interaction methods using surface, mid-air, or motion gestures. The touchscreen-based surface gestures were the most widely used method on handheld devices and were highly familiarised to regular users. However, this type of interaction is limited to the 2D screen, lacking the utilisation of the 3D spatial environment enabled in mobile AR. Furthermore, surface gestures were also limited by the screen space for interaction (Bergstrom-Lehtovirta and Oulasvirta, 2014) and suffered from screen occlusion (Forman and Zahorjan, 1994).

On the other hand, mid-air gestures could not be used to their full potential on mobile AR as one hand is required to hold the handheld device, unlike head-mounted display systems where the users could operate using both hands. For this reason, we decided not to further explore mid-air gestures. Instead, we decided to explore motion gestures that combine touch input and device movement to provide 3D interaction for mobile AR interaction.

To our knowledge, such a combination of interaction techniques has not been well-explored. Furthermore, there have been few comparisons between the performance of surface gestures and motion gestures in mobile AR settings. Through investigation into existing mobile AR applications, we have identified common tasks suitable for the elicitation process and compared the two categories of gestures. This research aims to answer these research questions:

RQ1 – Are there gesture-based interaction methods utilising the three-dimensional space that mobile AR applications support on handheld devices?

RQ2 – What are the perceived suitability, ease of use, and engagement by the users between the surface and motion gestures?

RQ3 – What would be the performance differences between the surface and motion gestures in an actual mobile AR application?

We conducted an elicitation study for both surface and motion gestures to answer *RQ1* and *RQ2*, which we discuss in Section 3. Next, we validated our gesture sets by comparing the selected gestures in a chosen mobile AR game to answer *RQ3* in Section 4. Next, we summarise our findings and discuss our results in Section 5. Finally, we conclude our research outcomes and our plan for future work in Section 6. We believe that this research's outcomes will show the benefits and drawbacks of surface gestures and motion gestures for mobile AR interaction so that designers can make better design decisions in choosing the most suitable interaction methods for their applications.

3 ELICITATION STUDY

To answer *RQ1*, we conducted an elicitation study for two sets of gestures, surface and motion, for various common tasks for mobile AR applications running on a handheld device. The term ”motion gestures” are those defined by Ruiz et al. Ruiz et al. (2011). However, we extend the definition to be broader and do not limit these gestures to just the movement of the handheld device in 3D space but also consider interaction which combines device movement and touch inputs from the device's touchscreen.

Ruiz et al. proposed that motion gestures could be recognised using built-in sensors of the handheld devices. Current AR technology combines software and hardware techniques, computer vision and sensor fusion to localise the device's 6 DOF (degree-of-freedom) position and orientation in the physical environment. We propose that by incorporating changes in the device's 6 DOF pose and touch inputs with mobile AR capability, novel motion gestures that have not been previously explored may be possible. As surface gestures have been the dominant form of interaction for handheld devices with touchscreen input, it is essential to compare these two classes of gestures in the context of mobile AR to answer *RQ2*.

The remainder of the section describes our elicitation study's methodology and results. First, we discuss our methodology and task selection in Section 3.1. Next, we provide details of participants, experimental setup, and procedure in Sections 3.2, 3.3, and 3.4. Next, we propose our hypotheses in Section 3.5, report the study results in Section 3.5, and finally give a summary in Section 3.6.

3.1 Methodology and Task Selection

We adopted the elicitation technique proposed by Wobbrock et al. (Wobbrock et al., 2009), which requires an initial selection of standard tasks that the targeted system should support. Researchers must first develop a set of descriptions of the tasks or short animations or videos that depict the manipulation's effects, which removes the need to develop a gesture recogniser and thus removes any limitation of the underlying technology and related constraints in the design process. During the elicitation study, each task is explained to the participants using either the animations or videos of the selected task displayed to the participant using the system's display (e.g. a large surface computing touchscreen (Wobbrock et al., 2009), AR headset (Piumsomboon et al., 2013b)), or simply through descriptions of the task (Ruiz et al., 2011). Once the participant understands the purpose of the task, they have to perform the gesture that they think would yield the outcome. After eliciting the gesture, the participant rates their chosen gesture for the goodness-of-fit (*Goodness*) and ease-of-use (*Ease of Use*) on a 7-point Likert scale.

Our study asked participants to watch two videos and design one surface gesture and another motion gesture for each task. Furthermore, apart from rating each gesture based on the goodness-of-fit and ease-of-use, we introduced a third measure of engagement. To decide on appropriate tasks for the elicitation study, we surveyed sixteen mobile AR applications on Google Play and the Apple App Stores. After surveying these applications, we selected twelve tasks that were commonly used in the applications and had appeared in the relevant past research (Wobbrock et al., 2009; Ruiz et al., 2011; Piumsomboon et al., 2013b; Goh et al., 2019). For each task, we also selected an application where the task is commonly performed, resulting in twelve tasks and six applications as shown in Table 1. Finally, we prepared a set of videos for the twelve tasks by recording the screen during the interaction from the six chosen mobile AR applications to elicit the gestures. Two videos were recorded for each task, the first for surface gestures showing object manipulation with minimal movement of the mobile device and the second for motion gestures showing object manipulation through the movement of the mobile device. From the two videos, we collected the two sets of user-defined gestures, surface gestures and motion gestures, for the different mobile AR tasks.

3.2 Participants

Twenty-one participants (ten female, eleven male) were recruited, aged 18 to 59 years old, with an average age of 29 (SD=10.7) years. All participants were right-handed. While all participants owned a touch-screen mobile device, eight had no prior experience with mobile AR applications. The remainder had some experience, but none were frequent mobile AR users. Participants signed a consent form containing experiment details to participate in the study. The participants were told that they could discontinue the experiments without penalty and that there were no serious health and safety risks. In addition, participants were given a gift voucher for their participation in the study.

3.3 Experimental Setup

The setup for this study was kept simple; the participants were seated in front of a television screen, while the experimenter was seated to the right of the participant, as shown in Figure 1. The participants were given a mobile phone to hold, a Samsung Galaxy S9, as a prop during the design process. Pre-recorded videos were displayed on a 32” television screen placed before the participants. This way, the participants could watch the video and perform the gesture on the mobile phone simultaneously to overcome the limited screen size of the mobile phone and any finger occlusion issues during the gesture design process. The participants were asked to follow the think-aloud protocol, and their gestures were recorded with a camera rig set up behind and to the right-hand side as they were all right-handed.

3.4 Procedure

The procedure of the study is as follows:

a. The experimenter introduces themselves and makes safety recommendations to the participants. The participants are informed that they can stop the study at any time.

b. Participants are asked to read the information sheet and sign the consent form. Next, the experimenter answers any questions or concerns the participants might have—finally, the participants consent to the video recording of their interaction during the experiment.

c. Participants are informed of the procedure of the elicitation study. A pre-experiment questionnaire is then presented to the participants to collect demographic information and any previous experience with mobile AR applications.

d. Before starting the elicitation process, the participants are given two minutes to familiarise themselves with the setup and are permitted to ask any questions. Then, as the process begins, the participants are asked to watch the video and design their gestures for the task shown. Each participant has to watch twenty-four videos from twelve selected tasks to propose the two types of gestures, one surface and one motion.

e. After a gesture is elicited for each task, the participants rate their gesture on a 7-point Likert scale in terms of *Goodness* (how suitable is the gesture for the task?), *Ease of Use* (how easy is the gesture to perform?), and *Engagement* (how engaging is the gesture to use?). Each task takes approximately 4 minutes to elicit both gestures.

f. Finally, after completing the elicitation process, the experimenter presents a post-experiment questionnaire to the participants to collect their general feedback for the experiment. The study takes approximately an hour to complete.

3.5 Result

With twenty-one participants, we elicited a total of 504 gestures. The number of common surface and motion gestures elicited for each task are shown in Figure 2. The level of agreement and characteristics for the elicited gestures is discussed in Sections 3.6.1 and 3.6.2, respectively. The two sets of gestures, were classified and their subjective ratings are compared in terms of *Goodness*, *Ease of Use*, and *Engagement* in Section 3.6.3. Section 3.6.4 discusses our feedback and observations, including our proposed interaction technique based on motion gestures called TMR (Touch-Move-Release), utilising the design pattern observed.

3.5.1 Level of Agreement

As previously described by Wobbrock et al. (Wobbrock et al., 2009), user-defined gesture sets are based on the largest set of identical gestures performed by participants for a given task. In gesture analysis, we found both similar gestures proposed for the same task and similar gestures used across multiple tasks. Similar gestures for each task were combined, and a record was kept of the number of gestures combined for each task. For example, for the *Slingshot* task, twenty participants proposed the same *Swipe-Down* surface gesture while one participant proposed a *Tap* gesture. As a result, we classified two groups for the surface gesture in the *Slingshot* task, the former group with 20 points and the latter with 1 point.

We compared the level of consensus for elicited gestures in each task. The agreement score was calculated using Equation 1 for both sets of gestures based on (Wobbrock et al., 2009). P_t represents the total number of gestures elicited in the selected task, and P_s is the number of similar gestures categorised into the same group for that task.

$$A = \sum_{p_s} \left(\frac{|p_s|}{|p_t|} \right)^2 \tag{1}$$

The results of agreement scores for each task have been plotted and illustrated in Figure 3. The difference between the agreement scores of the two sets of gestures is notable for Task 1 - Slingshot (A_s=.91, A_m=.48) and Task 2 – Throw (A_s=.83, A_m=.22). While Task 9 - Open Drawer, Task 10 - Close Drawer, Task 11 - Open Door, and Task 12 - Close Door had low agreement scores for both sets of gestures. Based on the highest scored consensus group of gestures in each task, we constructed two sets of user-defined gestures: a surface gesture set with 13 gestures and a motion gesture set with 12 gestures, as shown in Figure 4.

3.5.2 User-defined Gesture Characteristics

Figure 4 shows that some gestures can be used to perform multiple tasks. For example, in the surface gesture set, Task 9 - *Open Drawer*, 10 - *Close Drawer*, 11 - *Open Door*, and 12 - *Close Door* shared the *Double-Tap* gesture. Additionally, *Swiping* and *Holding* gestures were also common occurrences across multiple tasks for the surface gesture set. We found that the agreement scores for some of the tasks are lower than the surface gesture set for the motion gesture set (see Figure 3). This matches our expectation as motion gestures allow for handheld device movement in 3D space and supports greater possibilities in a larger design space.

Nevertheless, we observed some common characteristics and interaction patterns in the elicited motion gestures. Firstly, the trajectory of the gestures, i.e. the device's movement direction, varied but was generally aligned with the desired movement direction of the manipulated virtual object. Secondly, participants utilised the touch-sensitive screen to initiate and terminate their actions. These observations led us to propose the Touch-Move-Release (TMR) technique, which involves three steps of action corresponding to the functions of initiating, performing and terminating an interaction.

3.5.3 Comparisons of Subjective Ratings

We analysed the three subjective rating scores between surface and motion gestures: *Goodness*, *Ease of Use*, and *Engagement*. We applied the Friedman test followed by a post-hoc pairwise comparison using Wilcoxon signed-rank tests with Bonferroni correction (with p-value adjusted) to compare the two sets of ratings. Figure 5 shows the plots for all the twelve tasks in terms of *Goodness*, *Ease of Use*, *Engagement*, and *Overall*. The *Overall* score was an average of all the three ratings. We indicate the mean rating using $\bar{x_s}$ for the surface gestures and $\bar{x_m}$ for motion gestures and show standard deviation values in brackets.

Goodness Ratings: We found significant differences for Task 5 – *Scale Up* (V=123.5, p=.03, $\bar{x_s}$=6.1(1.2), $\bar{x_m}$=5.1(1.4)), Task 7 – Move (V=6, p=.03, $\bar{x_s}$=5.8(1.2), $\bar{x_m}$=6.4(0.6)), Task 9 – Open Drawer (V=91, p=.02, $\bar{x_s}$=6.1(1.0), $\bar{x_m}$=4.7(1.8)) and Task 10 – Close Drawer (V=75.5, p=.04, $\bar{x_s}$=6.1(0.9), $\bar{x_m}$=5.1(1.5)).

Ease of Use Ratings: Significant differences were found for Task 9 – Open Drawer (V=91, p=.001, $\bar{x_s}$=6.7(0.5), $\bar{x_m}$=5.3(1.6)), Task 10 – Close Drawer (V=101.5, p=.002, $\bar{x_s}$=6.6(0.5), $\bar{x_m}$=5.2(1.6)), Task

11 – Open Door (V=85, p=.04, $\bar{x_s}$=6.3(0.8), $\bar{x_m}$=5.7(1.1)), and Task 12 – Close Door (V=78, p=.02, $\bar{x_s}$=6.3(0.8), $\bar{x_m}$=5.4(1.3)).

Engagement Ratings: We found significant differences for Task 1 – Slingshot (V=9, p=.006, $\bar{x_s}$=5.1(1.3), $\bar{x_m}$=6.1(1.0)), Task 2 – Throw (V=39.5, p=.04, $\bar{x_s}$=5.0(1.2), $\bar{x_m}$=5.7(1.1)), Task 7 – Move (V=13.5, p=.005, $\bar{x_s}$=5.3(1.1), $\bar{x_m}$=6.4(0.7)), and Task 11 – Open Door (V=33.5, p=.04, $\bar{x_s}$=4.7(1.7), $\bar{x_m}$=5.8(1.2)).

Overall Score: The average yielded significant differences for Task 7 – Move (V=21.5, p=.02, $\bar{x_s}$=5.7(1.1), $\bar{x_m}$=6.4(0.7)), and Task 9 – Open Drawer (V=164.5, p=.03, $\bar{x_s}$=6.1(1.0), $\bar{x_m}$=5.0(1.7)).

3.5.4 Feedback and Observation

We asked participants to think aloud and explain their design decisions and ratings of their gestures during the elicitation process. At the end of the study, we also conducted a short semi-structured interview for additional feedback. When asked which gesture sets the participants would like to use in mobile AR applications, thirteen chose motion gestures, while the remainder picked surface gestures. From the information collected, we identified design patterns. We summarised our results into six themes: (1) versatility of gestures elicited, (2) multiple fingers or trajectories, (3) trade-offs between ease of use and engagement, (4) functionality-focused, (5) context-focused, and (6) Touch, Move, and Release.

Versatility – Many gestures elicited were used for multiple tasks. These gestures are more common for the surface gestures set, for example, the *Tap* gesture was used for Task 1 – *Slingshot* (1 vote), Task 2 – *Throwing* (2 votes), Task 4 – *Erase* (2 votes), Task 5 – *Scale-Up* (1 vote), Task 9 – *Open Drawer* (5 votes), Task 10 – *Close Drawer* (10 votes), Task 11 – *Open Door* (12 votes), and Task 12 – *Close Door* (11 votes). Other gestures that were considered versatile included *Double-Tap*, *Long Press*, *Swipe Up*, *Swipe Down*, *Swipe Left*, and *Swipe Right*. In comparison, versatile motion gestures were not as common; with the only versatile gestures being *Tap-Phone*, *Swing-Release* and *Tap-Pull Back-Release*. When designing gesture sets for applications, designers should be aware that supporting versatility of the same gesture for multiple tasks might help reduce user's mental effort, as the user will have to learn and recall fewer gestures.

Multiple fingers or trajectories – We observed that some participants consciously differentiated how many fingers they were using for different surface gestures, and some participants tried to minimise the number of fingers used. For instance, in task 6 – *Scale Down*, seven participants initially designed the gesture with five fingers, *Pinch Together*, on the screen but then changed to the conventional two-finger pinch. A possible reason might be due to the finger's occlusion of the screen, as stated by *Participant 4* as follows:

> *"I feel that in surface interaction, the extra fingers will obstruct my screen view."* – P_4

Two participants also asked if a traditional graphical user interface (GUI) could be provided, so they could directly tap an on-screen button to scale instead. Below was a comment made by Participant 14 on providing a GUI to assist with the interaction.

> *"Can I imagine a slide bar button in the scene? When I slide the button, the virtual object will scale down automatically."* – P_{14}

For motion gestures, we observed that all the participants performed the gestures using a single hand, their right hand, and only used their thumb for touch input. The participants mainly focused on working out the appropriate movement trajectories of the handheld device for different motion gestures.

Trade-Offs Between Ease of Use and Engagement – When participants were asked to rate their gestures in terms of *Ease of Use* and *Engagement*, they often based their decision on the duration of the interaction. Some participants felt that surface gestures might be less fatiguing and more efficient than motion gestures for certain tasks for prolonged usage. *Participant 15* mentioned that moving the device around could be quite tiring:

"Holding the phone for a long time makes my palms sweat, and more physical movements will exacerbate the situation." – P_{15}

Nevertheless, some participants felt that surface gestures could be quite boring to use for a long duration, and motion gestures could deliver a better experience for some tasks in 3D space, as *Participant 17* described:

"I have played Pokémon Go before… I like its story more than swiping up the screen to capture the Pokémon." – P_{17}

Functionality-focused – Some participants gave different subjective ratings for dichotomous tasks with similar gestures. For example, for Task 9 – *Open Drawer*, *Participant 5* rated their *Double Tap* surface gesture 7 / 7 / 7 (Goodness/Ease of Use/Engagement), while only rating their *Tap-Pull-Release* motion gesture 1 / 2 / 3. On the contrary, for Task 10 – *Close Drawer*, they then rated their *Swipe Up* surface gesture 4 / 6 / 5, but their *Tap-Down-Forward-Release* motion gesture 6 / 6 / 6. *Participant 5* explained that, when opening the drawer using the surface gesture, the content inside the drawer could be seen immediately, however with the motion gesture the user might lose the view of the drawer during the manipulation. Nevertheless, when the drawer needed to be closed, the content inside the drawer was not important, and they found motion gestures more engaging to use.

Context-focused – Some participants felt that their ratings would depend on the context of the application. For example, for gaming applications, motion gestures might be rated highly for *Goodness*. However, they might be rated much lower for non-entertainment applications. Therefore, the application context should be considered when choosing between surface and motion gesture sets.

"While playing a mobile AR game, I feel motion gestures have an irreplaceable charm, like the Joy-Con of Nintendo Switch, it will bring a more realistic user experience. But for some applications scene that requires accuracy, surface gestures are more appropriate." – P_1

Touch, Move, and Release – For the elicited motion gestures, a consistent design pattern emerged. We found that when the participants wanted to initiate their gestures, they first used their thumb to *touch* the device's screen to initiate the sequence of actions. Next, they would *move* the device in the desired trajectory to manipulate the virtual object. Finally, once they completed their action, the participant would then *release* their thumb from the screen, indicating the completion of the motion gesture cycle. From this pattern, we propose ***TMR (Touch, Move, and Release)*** interaction technique for mobile AR to guide the designers in developing more engaging interactions for their mobile AR applications.

3.6 Summary

In this section, we have answered our first two research questions, *RQ1* and *RQ2*. For *RQ1*, we elicited both surface and motion gestures for twelve tasks. We found an interaction pattern in the elicited motion gesture set and proposed an interaction technique called TMR (Touch-Move-Release), which utilises the device's touchscreen to initiate and terminate interaction and track the device's movement in 3D space for a better AR interaction experience. For *RQ2*, we analysed how the participants rated the two gesture sets in terms of *Goodness*, *Ease of Use*, and *Engagement* and found significant differences for different tasks. We further discuss these results in Section 5.1.

4 VALIDATION STUDY

To answer our final research question, *RQ3*, ”What would be the performance differences between the surface and motion gestures in an actual mobile AR application?” we conducted a validation study comparing the two types of gestures. For the mobile AR application, we decided to implement a game modelled after Pokémon Go, arguably the most popular mobile AR game at the time of this research. We chose the popular game mechanics of throwing a ”Poké Ball” action to capture the creature for the task. We implemented the surface gesture most comparable to the original experience in Pokémon Go, and we implemented our own motion gesture based on our proposed TMR (Touch, Move, and Release) interaction technique.

Released by Niantic in 2016 (Niantic, Inc., 2020), the main goal of Pokémon Go is for players to collect various Pokémon virtual creatures with different abilities, which are scattered using geospatial information throughout the real world. In the game, once the player has found a Pokémon, they can catch it by throwing a virtual Pokéball at the creature, with this interaction actioned by performing an onscreen *Swipe Up* gesture. The ball’s velocity (speed and direction) is controlled by the speed and direction of the player’s swipe from the bottom toward the top of the screen.

For our validation study, we implemented the two highest-scoring gestures elicited for Task 2 – *Throw* from the elicitation study, the *Swipe Up* surface gesture, which we chose at the baseline condition in this evaluation, and the *Push Forward* & *Change Axis* TMR motion gesture. We give an overview of the development of our system in Section 4.1, describe the study design in Section 4.2, participants in Section 4.3, and experimental procedure in Section 4.4, discuss our hypotheses in Section 4.5 and the results of the study in Section 4.6, and end with a summary of the experiment in Section 4.7.

4.1 System Prototyping

The Pokémon GO clone that we build can be played indoors or outdoors and utilises geospatial information of the player’s physical location. Similar to the original, the goal of our game was for the player to capture a collection of Pokémon-like creatures. However, as our focus was not on gameplay but on comparing the two gesture interaction techniques, we chose not to spawn creatures in random locations in the physical world. Instead, we created a menu that allowed the player to spawn creatures around their current location. The creatures become visible when the player is within a proximity threshold, approximately 2 meters. The player must not move any closer and must throw the ball at this distance to capture the creature. Once captured, the creatures would be indexed in a Pokédex-like encyclopedia so the player can keep track of the creatures they have captured. The game has been developed using the Unity Game Engine (Unity Technologies., 2020) and Vuforia SDK (PTC, Inc., 2020b). Screenshots of our game and further descriptions are given in Figure 6.

4.1.1 The Unity Game Engine

The Unity Game Engine (Unity Technologies., 2020) is a popular game development engine developed by Unity Technologies. It supports cross-platform game development, enabling developers to create applications on a personal computer for various devices and platforms, including Apple and Android mobile devices. Developers build 2D and 3D applications by authoring virtual scenes using a tree-based hierarchical layout of ”GameObjects”. ”Components” can be attached to these GameObjects to define and change their behaviour, with a range of predefined components available in Unity and the option to create new components written using the C# programming language. For our prototype development, we used Unity Game Engine version 2017.3.1f1.

4.1.2 The Vuforia Software Development Kit

The Vuforia Software Development Kit (SDK) (PTC, Inc., 2020b) is a toolkit for the creation of mobile AR experiences. Initially developed by Qualcomm (Qualcomm Technologies, Inc. and/or its affiliated companies, 2020), it was acquired by PTC Software (PTC, Inc., 2020a) in 2015. Vuforia can determine the position and orientation of a mobile device in 3D space and identify the real-world ground plane using the device's camera and other embedded sensors. We used Vuforia version 8.0 for Unity to develop our game prototype, which allowed us to place virtual creatures to appear localised in the real world even when the device moved and rotated in real space.

4.1.3 Location-Based Services

In addition to the relative localisation of the device to its immediate surroundings provided by Vuforia, we also integrated location-based services (LBS) (Schiller and Voisard, 2004) in our game to better replicate the experience of Pokémon Go. During the game, LBS obtains the real-time absolute location information of the player through the network provider and Global Positioning System (GPS) subsystems. By combining LBS and Vuforia, we can accurately determine the device's location in 3D space outdoors and indoors.

4.1.4 Gesture Implementation

For both gestures, the motion of the virtual ball being thrown is based on the built-in physics engine in Unity. The ball was implemented as a spherical rigid body influenced by gravity and following a projectile motion. The implementation details for the two gestures are described below.

Baseline Surface Gesture (*Swipe Up*) – Our system registers and records the position where the player first touches the ball on the screen. Their touch's final position and duration are recorded as the player slides their finger upward and eventually lifts it off the screen. The distance vector between the final and the initial touch position is divided by the touch duration yielding the velocity vector of the swipe. A force vector is created using this velocity vector multiplied by a constant and applied to the ball, which improves the realism of throwing. The force is applied upon release of the swipe action. This causes an unnoticeable delay between the player's action and the ball being thrown as the finger also occluded the screen.

TMR Motion Gesture (*Push Forward & Changing Axis*) – The velocity vector for the TMR motion gesture is calculated based on two factors: the time when the thumb is lifted off the screen (*Release*), and the angular velocity of the device. To measure the device's angular velocity, we used the device's inertial measurement unit (IMU), which combines an accelerometer and a gyroscope to calculate the orientation and linear acceleration of the device. The TMR motion gesture begins with the device being held almost vertically to the ground. The player would then touch and hold their thumb to the screen to simulate holding onto the ball. Then the player swings the phone forward and releases their thumb to throw the ball. When the player first touches the screen, we set the update rate of the IMU to 0.1 seconds and started monitoring the angular velocity of the device. If the angular velocity (tilt forward with speed) exceeds a threshold, a linear force is then calculated based on the linear velocity of how far and how fast the device is moving forward. This linear velocity is then applied to the ball. The resulting force vector and experience are similar to the surface gesture but use the device movement instead of the finger movement.

4.1.5 Device Hardware

We used the Samsung Galaxy S9, an Android-based smartphone, in this study. With a powerful processor, built-in motion sensors, camera, and running Android 9.0 Pie, it was more than capable of running our prototype game at a suitable frame rate.

4.2 Study Design

This study used a 2×3 within-subjects factorial design, where the two independent variables were the interaction techniques (surface gesture and TMR motion gesture) and the target creature sizes (*Small*, *Medium*, and *Large*). This combination yielded six conditions, as shown in Table 2. The two interaction techniques were counterbalanced, and the three sizes were then counterbalanced for each method. The dependent variables were the accuracy of the throw, measured by counting the number of balls thrown before successfully hitting the target, the system usability scale (SUS) (Bangor et al., 2009), game experience questionnaire (GEQ) (IJsselsteijn et al., 2013), and subjective ratings on *Goodness*, *Ease of Use*, and *Engagement*.

Participants were tasked with capturing two target creatures of each size (*Small*, *Medium*, and *Large*), as shown in Figure 7. Each targeted creature was automatically spawned when the player was within a two-meter radius of the indicated location of the creature, with the target creature placed 2 meters in front of the player's location. We recorded how many balls participants used in each condition and asked our participants to answer three sets of questionnaires (SUS, GEQ, and subjective ratings) after completing each interaction technique. A short unstructured interview was conducted at the end of the study.

4.3 Participants

We recruited 10 participants (six females) with an age range of 18 to 40 years old ($\bar{x}$ = 26.4, SD = 5.95). All participants owned touchscreen smartphones, used them every day, and were highly familiar with surface gestures. Five participants had no previous experience with mobile AR applications, and the remaining participants had only limited experience with the original Pokémon GO game. Participants signed a consent form that contained information about the experiment. Participants were told that they could discontinue the experiment and that no serious health and safety concerns had been identified.

4.4 Procedures

The procedure of the study was as follows:

a. The participants are informed of the study through an information sheet and sign the consent form if they are willing to participate.

b. Participants answer a pre-experiment questionnaire to collect demographic information and their previous experience with mobile AR applications.

c. Participants are given two minutes to familiarise themselves with the Pokémon GO clone game and its user interface. They are encouraged to ask any questions about the game.

d. Participants are tasked with capturing six creatures, two of each size (*Small*, *Medium*, and *Large*), by throwing a ball at the creature using the interaction method provided and the number of balls thrown in each condition is recorded.

e. After successfully capturing all six creatures with one of the interaction methods, participants have to complete the questionnaires on subjective rating, system usability scale (SUS), and the game experience questionnaire (GEQ).

f. Steps d and e are then repeated for the second interaction technique.

g. Once the participants have completed all tasks, they are asked to complete the post-experiment questionnaire, choose their preferred interaction technique, and comment on the game's overall experience.

4.5 Hypotheses

The two interaction techniques, a baseline surface gesture and the TMR motion gesture are compared based on throwing accuracy with varying target sizes, subjective user rating of goodness, ease of use and engagement, user rating of the overall game experience, and the System Usability Scale (SUS). Our hypotheses are ***there would be differences in accuracy (H1)***, ***subjective ratings (H2)***, ***game experience (H3)***, and ***system usability (H4) between the two interaction techniques***.

4.6 Results

In the following sections, we present the comparison results from various dependent variables measured between the two interaction methods. First, the throwing accuracy for various target sizes are discussed in Section 4.6.1, the subjective ratings of goodness, ease of use, and engagement in Section 4.6.2, the outcomes of the game experience questionnaire in Section 4.6.3, the scores of the system usability scale in Section 4.6.4, and the preferences of the participants between the two interaction methods for the chosen task in Section 4.6.5.

4.6.1 Accuracy with Various Target Sizes

For the *Small* targets, participants used a total of 51 balls ($\bar{x}$ =5.1, SD =2.3), and 85 balls ($\bar{x}$ = 8.5, SD = 5.4) for the *Surface* and *TMR* interaction technique respectively. Similar trends were found for the *Medium* and *Large* target sizes, with 23 balls ($\bar{x}$ = 2.3, SD =0.61) for surface gesture and 57 balls ($\bar{X}$ = 5.7, sd = 2.3) for TMR motion gesture for the *Medium* targets and 21 balls ($\bar{x}$ = 2.1, SD = 0.3) for surface gesture and 37 balls ($\bar{x}$=3.7, SD=2.1) for TMR motion gesture for the *Large* targets. Figure 8 shows the number of balls used for each target size by all the participants for the two interaction methods.

We performed two-way ANOVA to examine the effect of interaction methods and target size on the accuracy of the throwing task. Figure 9 shows the number of balls used for each condition. No significant interaction effect was found between the combination of the two variables, *interaction methods* × *target sizes*, and the accuracy of the throw. We found that the *Surface* gesture was significantly more accurate than the *TMR* interaction technique (F_1,59=14.7, p=.0003). We also found that there were significant differences between target sizes (F_2,59=10.1, p=.0002).

Wilcoxon signed-rank (WSR) tests with Bonferroni correction (p-value adjusted) were used to perform post-hoc pairwise comparisons. We found that there were significant differences for the surface gesture between *Small-Medium* targets (W=85, p=.006), an *Small-Large* targets (W=78, p=.002), and for the TMR motion gesture between *Small-Large* targets (W=116.5, p=.01) and *Medium-Large* (W=95.5, p=.04).

4.6.2 Subjective Ratings: Goodness, Ease of Use, and Engagement

As in the elicitation study, we asked participants to rate the interactions based on *Goodness*, *Ease of Use*, and *Engagement*. For *Goodness*, the *Surface* gesture was rated $\bar{x}$ = 4.8 (SD = 1.3) and the elicited *Motion* gesture was rated $\bar{x}$ = 5.4 (SD=1.4). For *Ease of Use*, the *Surface* gesture was rated $\bar{x}$ = 6.8 (SD=0.4) and the elicited *Motion* gesture was rated $\bar{x}$ = 4.7 (SD=1.3). Finally for *Engagement*, the *Surface* gesture was rated $\bar{x}$ = 3.9 (SD=1.2) and the elicited *Motion* gesture was rated $\bar{x}$ = 5.3 (SD= 1.5). Figure 10 shows the average ratings for the two interaction methods. We applied WSR tests and found significant differences for *Ease of Use* (W=36, p=.01) in favour of the *Surface* gesture, and for *Engagement* (W=0, p=.03) in favour of the *TMR* interaction technique.

4.6.3 Game Experience

We collected participant feedback of the game using the Game Experience Questionnaire. Seven aspects are measured using 14 questions: *Competence*, *Sensory* and *Imaginative Immersion*, *Flow*, *Tension*, *Challenge*, *Negative Affect*, *Positive Affect*.

Figure 11 shows the results of the game experience questionnaire. With WSR tests, we found that the *TMR* interaction technique was rated significantly higher for ***Competence*** (*TMR* – $\bar{x}$ = 3.1 (SD=0.4),*Surface* – $\bar{x}$ = 2.3 (SD=1.2), W=5, p=.04), ***Sensory and Imaginative Immersion*** (*TMR* – $\bar{x}$=2.6 (SD=0.6), *Surface* – $\bar{x}$ = 2.4 (SD=1.2), W=2, p=.04), ***Flow*** (*TMR* – $\bar{x}$ = 2.1 (SD=0.9), *Surface* – $\bar{x}$ = 1.6 (SD = 1.1), W=5, p=.04), and ***Challenge*** (*TMR* – $\bar{x}$ = 2.9, SD=0.9, *Surface* – $\bar{x}$ = 0.9, SD=0.7, W=0, p=.006). There was no significant difference found for *Tension*, *Negative Affect*, and *Positive Affect*.

One of the participants gave their opinion on the reason for the difference in the ***Challenge*** rating as follows:

"TMR is quite challenging for me, especially when the size of the Pokémon is extremely small." – P_5

"TMR can bring a sense of victory for me. I am proud of myself when I completed the task of capturing a small size Pokémon." – P_5

4.6.4 System Usability Scale

For the System Usability Scale (SUS), the *surface* gesture received scores between 62.5 to 95 with the average of $\bar{x}$ = 77 (SD=9.9), while the *TMR* motion gesture scores ranged from 37.5 to 87.5 with $\bar{x}$ = 68 (SD=15.9). Figure 12 shows the comparison of the average and SD of the SUS scores. A WSR test yielded a significant difference (W = 28, p-value = 0.02)

"Both interaction methods can complete the task. I feel good for the whole cycle of the demo. I think I need more levels to feel challenged." – P_8

"TMR lets my vision off the phone screen. If I try to capture the screen with my eyes, it makes me dizzy. Honestly, I prefer simple surface gesture." – P_3

4.6.5 Preference

In the post-experiment interview, we asked our participants which interaction technique they preferred. Half of the participants preferred the baseline *Surface* gesture, while the other half chose the *TMR* motion gesture, as shown in Figure 13. Those who preferred the *Surface* gesture found it easier to use, while those who liked *TMR* motion gesture found it more challenging and felt more engaged in the game.

"I have been using surface interaction for a long time, and it is a habit for me. TMR has unnecessary physical consumption for me." – P_7

"TMR makes me feel a challenge when using every Pokeball, which greatly increases the fun of the game. I hope that more complex and diverse TMR can be used in the future." – P_2

4.7 Summary

In this section, we presented a validation study to compare two types of interaction methods, a baseline *Surface* gesture and the *TMR* motion gesture, based on our elicitation study. We described the mobile AR game development, Pokémon GO clone, including implementation to support both interaction techniques. We conducted a 2×3 within-subjects factorial design study with ten participants to compare the two gestures on a throwing task for three different target sizes. We found significant differences in accuracy, subjective ratings, system usability, and game experience. Participants were equally split in their gesture preferences. We will further discuss these results in Section 5.2.

5 DISCUSSION

In this section, we summarise our key findings from the two user studies and discuss their implications. Section 5.1 focuses on the elicitation study and Section 5.2 on the validation study. Next, we discuss the implications of this research in Section 5.3. Finally, we describe the limitations of this research in Section 5.4.

5.1 Discussion of Elicitation Study

For our elicitation study, we found that the participants perceived and rated their gestures based on the nature of the tasks, which involved factors beyond the consideration of the study. Despite this limitation, further examination of the results yielded some insights into the tasks that were found to have significant results.

In terms of ***Goodness***, surface gestures were rated significantly higher than motion gestures for Tasks 5 – *Scale-Up*, 9 – *Open Drawer*, and 10 – *Close Drawer*, while motion gestures were significant higher for Task 7 – *Move*. Most elicited surface gestures were those commonly found in existing applications. Thus participants found them very familiar and suitable for many tasks, reflected in the high average scores for all tasks. In addition, as discussed in Section 3.6.4, surface gestures were described as highly versatile and so easier to recall. Despite this, the elicited motion gestures were also highly rated, particularly for the moving task done in 3D space, which indicates that 3D tasks which require precise 3D inputs benefit from the elicited motion gestures. Thus, as discussed in Section 3.6.4 – *Functionality-Focused* and *Context-Focused*, it is crucial to consider what are the key interactions for a specific experience and the type of gestures that would be best to support them.

For the ***Ease of Use*** comparison, surface gestures were found easier to perform than motion gestures for Tasks 9 – *Open Drawer*, 10 – *Close Drawer*, 11 – *Open Door*, and 12 – *Close Door*. We suspect that the lower physical demand of surface gestures, coupled with our experiment design which treated task completion as a binary option, made movement tasks easier to perform than motion gestures. For example, whether it was opened slowly or only halfway in the Open Drawer task, it did not matter how the drawer was opened. Therefore, any gesture that executed the action would satisfy the goal. We speculate that the results may differ in different contexts, for example, in a stealth game where players have to quietly open the doors or drawers to avoid detection.

For the ***Engagement*** comparison, motion gestures were rated significantly higher than surface gestures for Tasks 1 – *Slingshot*, 2 – *Throw*, 7 – *Move*, and 11 – *Open Door*. As discussed in Section 3.7.1 - *Trade-Offs Between Ease of Use and Engagement*, participants felt that the motion gestures made the interaction more engaging because the elicited motion gestures reflected the physical movement involved in performing similar physical tasks in the real world. In the post-experiment questionnaire, most participants indicated that they would prefer motion gestures in a gaming context. However, they preferred surface gestures for non-gaming applications, requiring less physical effort. Given these findings, we recommend using the TMR motion gesture to improve the level of engagement in mobile AR gaming applications.

5.2 Discussion of Gesture Validation

The interaction pattern found in the elicitation study led us to propose an interaction technique for motion gestures called TMR (Touch-Move-Release). We conducted a validation study to compare a TMR motion gesture with a surface gesture, our baseline, in a chosen mobile AR gaming application based on Pokémon GO. We have chosen the ball throwing task (*Task 2 – Throw with three target sizes*), which is a popular activity in Pokémon GO for capturing virtual creatures. For the surface gesture, we used the *Swipe Up* gesture as a baseline, and the *Push Forward & Changing Axis* to represent a TMR motion gesture. We hypothesised that there would be differences in *accuracy (**H1**) subjective ratings (**H2**)*, *game experience (**H3**)*, and *system usability (**H4**)*. The results of the study supported our hypotheses.

In terms of *accuracy*, we found significant differences between the two interaction methods across various target sizes. Generally, the *Surface* gesture was more accurate than *TMR* motion gesture, and the smaller targets had lower accuracy scores. We found that for the *Surface* gesture, participants had better accuracy in holding the device in their non-dominant hand and swiping with their dominant hand. One participant attempted holding the device and swiping with a single hand but then changed to operate with both hands. Furthermore, with *Surface* gestures, the participants could see the screen at all times while interacting, although their fingers may occlude the screen at times. In comparison, when using the *TMR* motion gesture, participants could not see the screen clearly as the device was in motion. However, the TMR motion gesture could be performed easily using a single hand.

For the subjective ratings, we found that *Surface* was significantly easier to perform than the TMR motion gesture. However, *TMR* motion gesture was more engaging than the surface gesture, which coincided with our findings in the elicitation study. For the in-game experience rating, *TMR* motion gesture was rated significantly higher than *Surface* for *Competence*, *Sensory and Imaginative Immersion*, *Flow*, and *Challenge*. This matched with our findings in Section 3.6.4 that *TMR* motion gesture could improve interaction in a gaming context. The *TMR* motion gesture made participants feel challenged to improve the accuracy of their physical gestures, and, at the same time, they reported increased sensory immersion due to the physical movement of the device matching the virtual object. Participants also reported increased engagement, that they felt skilful and competent, and that they had an increased sense of accomplishment when they successfully achieved their goal. *TMR* motion gesture required participants to pay more attention and increase their focus, and combining this with the sense of challenge not only increased feelings of immersion but also led to an increased sense of flow during the experience.

The results of the system usability scale (SUS) showed that the *Surface* gesture was more usable than the *TMR* motion gesture in the game. We believe this is due to familiarity with the surface gesture and the ease of use compared to the physical TMR motion gesture. We observed that some participants learned to use *TMR* quickly and found it novel and interesting to learn, while the others took longer to learn and felt irritated using it. This observation reflected the equal split in the number of votes for gesture preference for the two interaction techniques.

5.3 Implications

This research elicited two user-defined gestures, surface and motion gestures, for mobile AR applications. Mobile AR application designers can use these gesture sets to determine the most suitable gestures for their application based on user ratings and preferences. Furthermore, from identifying interaction patterns in this initial study, we propose a novel motion gesture technique called Touch-Move-Release (TMR) motion gesture. The elicited motion gestures in this research only represent a subset of all possible TMR motion gestures, and any motion gesture can be considered TMR if it follows the three-step process of Touch – to register or initiate the beginning of the action; Move – to perform six-dimensional (three positions and three orientation) motion input through device movement; and Release – to conclude and execute the action. Participants also found that TMR might be best suited for entertainment applications such as games, while surface gestures are more suitable for the other applications.

5.4 Limitations

While conducting these studies, we encountered several issues that could use some improvements. Firstly, although we surveyed the most popular current mobile AR applications for suitable tasks for this study, we could only pick twelve tasks for the elicitation study, limiting the number of gestures that could be elicited, especially for the motion gestures, which offer a larger design space. Future studies could focus on specific contexts to elicit a set of gestures for better-suited gestures. Secondly, we chose to implement and test a single task of the original twelve in the validation study. Implementing other experiences to validate additional gestures from the gesture sets would be necessary to further validate the generalisability of our results. Finally, although our focus has been on the user experience, we could have also measured the task load in performing the tasks using questionnaires such as the NASA TLX, which would have indicated the cognitive and physical effort to perform these gestures. The SUS results predict that the TMR interaction technique would likely be more demanding than surface gestures, but this must be confirmed.

In addition to the limitations of the studies, we have also identified some limitations in the proposed TMR motion gestures. First, there is limited visual feedback due to the device's movement. Second, prolonged interaction might lead to fatigue as the TMR motion gesture requires more physical movement, making the TMR motion gesture unsuitable for long-duration experiences. Finally, there is a risk of damaging the device if the user loses their grip when performing actions.

6 CONCLUSION AND FUTURE WORK

This research explored gesture-based interaction for mobile Augmented Reality (AR) applications on handheld devices. Our survey of current mobile AR applications found that most applications adopted conventional surface gestures as the primary input. However, this limits the input to two dimensions instead of utilising the three-dimensional space made possible by AR technology. This gave rise to our first research question, which asked if there might be a gesture-based interaction method that could utilise the 3D interaction space of mobile AR. This, in turn, led to our second research question, which asked how such a method would compare to the surface gestures.

We conducted a study to elicit two sets of gestures, surface and motion, for twelve tasks found in six popular AR applications to answer these two questions. The study yielded 504 gestures, from which we categorised a total of 25 gestures for the final user-defined gesture set, including 13 surface and 12 motion gestures. We compared the two sets of gestures regarding Goodness, Ease of Use, and Engagement. We found that the surface gestures were commonly used in existing applications and participants found them significantly easier to perform for some tasks. In contrast, the motion gestures were significantly more engaging for some other tasks. We observed an interaction pattern and proposed the Touch-Move-Release (TMR) motion gestures to improve engagement for mobile AR applications, especially games.

Our third research question asked how our proposed TMR motion gestures would compare to the surface gestures in an actual mobile AR application. We developed a Pokémon GO clone and conducted a validation study to compare the two gestures in terms of accuracy with three different target sizes, subjective ratings from the participants, an in-game experience questionnaire, system usability, and user preference. We found that surface gestures were more accurate and easier to use. However, the TMR interaction technique was more engaging and offered a better in-game experience and increased feelings of competence, immersion, flow, and challenge for the player. We discussed our results, shared the implications of this research, and highlighted the limitations of our studies and the TMR motion gestures.

In future work, the limitations highlighted the need for further validation studies to verify the validity of the entire user-defined gesture sets in various domain applications. In addition, more tasks, including context-specific tasks, should be included in the elicitation study to cover a broader scope of possible AR tasks. There is also an opportunity to explore the TMR interaction technique for exergames, i.e. games that encourage players to do more physical activities. Finally, there should be investigations into improving accuracy and visual feedback, reducing user fatigue, and lowering the risk of damaging the device for TMR motion gestures.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Ze Dong, Tham Piumsomboon, Adrian Clark, Rob Lindeman provided concepts and directions for this research. Ze Dong developed and improved the system with suggestions from Tham piumsomboon, Adrian Clark, Xiaoliang Bai, Weiping He. Ze Dong and Tham Piumsomboon conducted the research study and analyzed the data. This manuscript was written by Ze Dong and revised by all authors.

FUNDING

This work is supported by National Natural Science Foundation of China (NSFC), Grant No: 6185041053.

ACKNOWLEDGMENTS

We thank the members of HIT Lab NZ, School of Product Design, and Cyber-Physical Interaction Lab, as well as the participants who participated in our user study for their invaluable feedback. The results of Experiment 1 had been presented in the extended abstract format in ACM CHI2020 Late-Breaking Work (Dong et al., 2020a) and the results of Experiment 2 part A was published as a poster at ACM SUI2020 (Dong et al., 2020b).

REFERENCES

[Dataset] Apple Inc. (2020). Augmented reality - apple developer

Ashbrook, D. and Starner, T. (2010). Magic: a motion gesture design tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2159–2168

Azuma, R. T. (1997). A survey of augmented reality. *Presence: teleoperators & virtual environments* 6, 355–385

Bai, H. (2016). Mobile augmented reality: Free-hand gesture-based interaction

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 114–123

Bergstrom-Lehtovirta, J. and Oulasvirta, A. (2014). Modeling the functional area of the thumb on mobile touchscreen surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1991–2000

Billinghurst, M., Clark, A., and Lee, G. (2015). A survey of augmented reality

Boring, S., Jurmu, M., and Butz, A. (2009). Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*. 161–168

Chen, X., Grossman, T., Wigdor, D. J., and Fitzmaurice, G. (2014). Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 159–168

Chen, Z., Ma, X., Peng, Z., Zhou, Y., Yao, M., Ma, Z., et al. (2018). User-defined gestures for gestural interaction: extending from hands to other body parts. *International Journal of Human–Computer Interaction* 34, 238–250

Colley, A., Van Vlaenderen, W., Schöning, J., and Häkkilä, J. (2016). Changing the camera-to-screen angle to improve ar browser usage. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 442–452

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International journal of human-computer studies* 41, 801–849

Čopič Pucihar, K., Coulton, P., and Alexander, J. (2014). The use of surrounding visual context in handheld ar: device vs. user perspective rendering. In *Proceedings of the SIGCHI Conference on Human Factors* *in Computing Systems*. 197–206

Cui, J., Kuijper, A., Fellner, D. W., and Sourin, A. (2016). Understanding people's mental models of mid-air interaction for virtual assembly and shape modeling. In *Proceedings of the 29th International Conference on Computer Animation and Social Agents*. 139–146

Di Geronimo, L., Bertarini, M., Badertscher, J., Husmann, M., and Norrie, M. C. (2017). Exploiting mid-air gestures to share data among devices. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–11

Dong, Z., Piumsomboon, T., Zhang, J., Clark, A., Bai, H., and Lindeman, R. (2020a). A comparison of surface and motion user-defined gestures for mobile augmented reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8

Dong, Z., Zhang, J., Lindeman, R., and Piumsomboon, T. (2020b). Surface vs motion gestures for mobile augmented reality. In *Symposium on Spatial User Interaction*. 1–2

Epps, J., Lichman, S., and Wu, M. (2006). A study of hand shape use in tabletop gesture interaction. In *CHI'06 extended abstracts on human factors in computing systems*. 748–753

Forman, G. H. and Zahorjan, J. (1994). The challenges of mobile computing. *Computer* 27, 38–47

Goel, M., Wobbrock, J., and Patel, S. (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*. 545–554

Goh, E. S., Sunar, M. S., and Ismail, A. W. (2019). 3d object manipulation techniques in handheld mobile augmented reality interface: A review. *IEEE Access* 7, 40581–40601

Goldsmith, D., Liarokapis, F., Malone, G., and Kemp, J. (2008). Augmented reality environmental monitoring using wireless sensor networks. In *2008 12th International Conference Information Visualisation* (IEEE), 539–544

[Dataset] Google Inc. (2020). Arcore overview —google developers

Ha, T. and Woo, W. (2011). Arwand: Phone-based 3d object manipulation in augmented reality environment. In *2011 International Symposium on Ubiquitous Virtual Reality* (IEEE), 44–47

Hartmann, B., Abdulla, L., Mittal, M., and Klemmer, S. R. (2007). Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 145–154

Havlucu, H., Ergin, M. Y., Bostan, I., Buruk, O. T., Göksun, T., and Özcan, O. (2017). It made more sense: Comparison of user-elicited on-skin touch and freehand gesture sets. In *International Conference on Distributed, Ambient, and Pervasive Interactions* (Springer), 159–171

Henrysson, A., Billinghurst, M., and Ollila, M. (2005). Face to face collaborative ar on mobile phones. In *Fourth ieee and acm international symposium on mixed and augmented reality (ismar'05)* (IEEE), 80–89

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., and Irani, P. (2014). Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1063–1072

Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E. (2000). Sensing techniques for mobile interaction. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*. 91–100

Hürst, W. and Van Wezel, C. (2013). Gesture-based interaction via finger tracking for mobile augmented reality. *Multimedia Tools and Applications* 62, 233–258

IJsselsteijn, W. A., de Kort, Y. A., and Poels, K. (2013). The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* 46

[Dataset] inter IKEA Systems B.V. (2020). Ikea apps - ikea

Jones, E., Alexander, J., Andreou, A., Irani, P., and Subramanian, S. (2010). Gestext: accelerometer-based gestural text-entry systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2173–2182

Kurkovsky, S., Koshy, R., Novak, V., and Szul, P. (2012). Current issues in handheld augmented reality. In *2012 international conference on communications and information technology (ICCIT)* (IEEE), 68–72

Kyriazakos, V., Nikolakis, G., and Moustakas, K. (2016). Natural interaction with 3d content on mobile ar systems using gesture recognition. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (Springer), 348–357

Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J. H., et al. (2009). Freeze-set-go interaction method for handheld mobile augmented reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. 143–146

May, K. R., Gable, T. M., and Walker, B. N. (2017). Designing an in-vehicle air gesture set using elicitation methods. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 74–83

[Dataset] Microsoft Corporation (2020). Hololens2

Morris, M. R., Wobbrock, J. O., and Wilson, A. D. (2010). Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*. 261–268

Mossel, A., Venditti, B., and Kaufmann, H. (2013). 3dtouch and homer-s: intuitive manipulation techniques for one-handed handheld augmented reality. In *Proceedings of the virtual reality international conference: laval virtual*. 1–10

[Dataset] Niantic, Inc. (2020). pokemongolive

Pham, T., Vermeulen, J., Tang, A., and MacDonald Vermeulen, L. (2018). Scale impacts elicited gestures for manipulating holograms: Implications for ar gesture design. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 227–240

Piumsomboon, T., Altimira, D., Kim, H., Clark, A., Lee, G., and Billinghurst, M. (2014). Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (IEEE), 73–82

Piumsomboon, T., Clark, A., and Billinghurst, M. (2013a). Kite: Platform for mobile augmented reality gaming and interaction using magnetic tracking and depth sensing. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (IEEE), 289–290

Piumsomboon, T., Clark, A., Billinghurst, M., and Cockburn, A. (2013b). User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction* (Springer), 282–299

[Dataset] PTC, Inc. (2020a). Ptc software

[Dataset] PTC, Inc. (2020b). Vuforia - engine

[Dataset] Qualcomm Technologies, Inc. and/or its affiliated companies (2020). Qualcomm semiconductor company

[Dataset] QuiverVision (2020). Quivervision 3d augmented reality coloring apps

Rico, J. and Brewster, S. (2010). Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 887–896

Ruiz, J., Li, Y., and Lank, E. (2011). User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 197–206

Schiller, J. and Voisard, A. (2004). *Location-based services* (Elsevier)

[Dataset] Sketchar.tech (2020). Start drawing easily using augmented reality - sketchar

Stanimirovic, D. and Kurz, D. (2014). Smartwatch-aided handheld augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (IEEE), 373–374

Tanikawa, T., Uzuka, H., Narumi, T., and Hirose, M. (2015). Integrated view-input ar interaction for virtual object manipulation using tablets and smartphones. In *Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology*. 1–8

[Dataset] Unity Technologies. (2020). Unity engine

Van Krevelen, D. and Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality* 9, 1–20

Voida, S., Podlaseck, M., Kjeldsen, R., and Pinhanez, C. (2005). A study on the manipulation of 2d objects in a projector/camera-based augmented reality environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 611–620

Walter, R., Bailly, G., Valkanova, N., and Müller, J. (2014). Cuenesics: using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 299–308

Wigdor, D. and Balakrishnan, R. (2003). Tilttext: using tilt for text input to mobile phones. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*. 81–90

Wilson, A. D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., and Kirk, D. (2008). Bringing physics to the surface. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 67–76

Wobbrock, J. O., Morris, M. R., and Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1083–1092

Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. (2006). Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'06)* (IEEE), 8–pp

FIGURE CAPTIONS

Table 1. A list of selected tasks from six common mobile AR applications.

#	**Apps**	**Tasks**	**#**	**Apps**	**Tasks**
1	Angry Birds	Slingshot	7	IKEA Place	Move
2	Pokémon Go	Throw	8	IKEA Place	Rotate
3	Just a Line	Draw	9	ARia	Open Drawer
4	Just a Line	Erase	10	ARia	Close Drawer
5	CooolAR	Scale Up	11	CooolAR	Open Door
6	CooolAR	Scale Down	12	CooolAR	Close Door

Table 2. Conditions of Surface Gesture and TMR Interaction Technique Validation

Target Size **Interaction Methods**	**Small (height=30 cm)**	**Medium (height=60 cm)**	**Large (height=120 cm)**
Surface Gesture (Baseline)	*Swipe Up* - 30 cm target	*Swipe Up*- 60 cm target	*Swipe Up*- 120 cm target
TMR Motion Gesture	*Push Forward*- 30 cm target	*Push Forward*- 60 cm target	*Push Forward*- 120 cm target

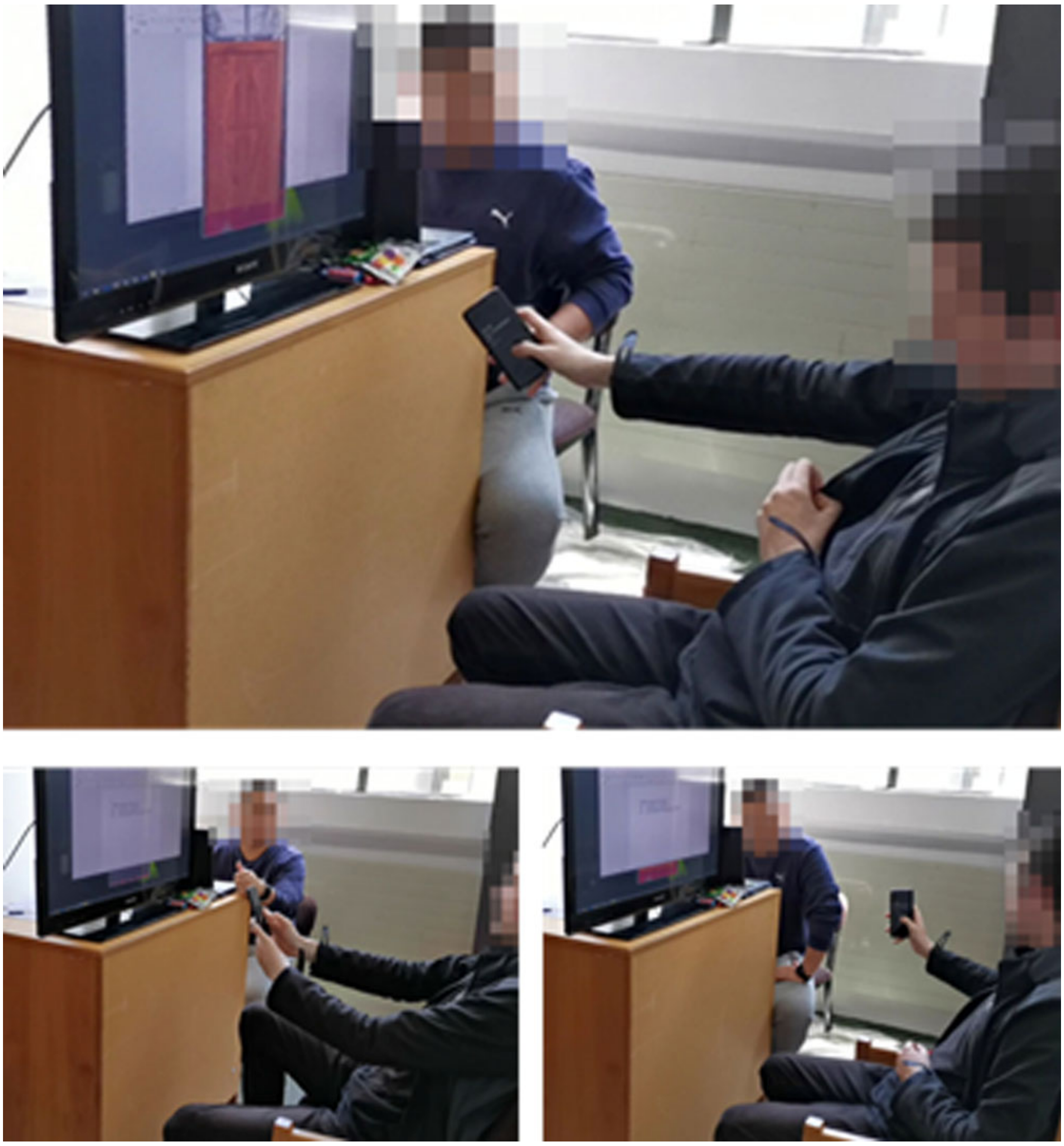


Figure 1. Experimental Setup - a participant is performing a gesture while watching a video displayed on a TV screen.

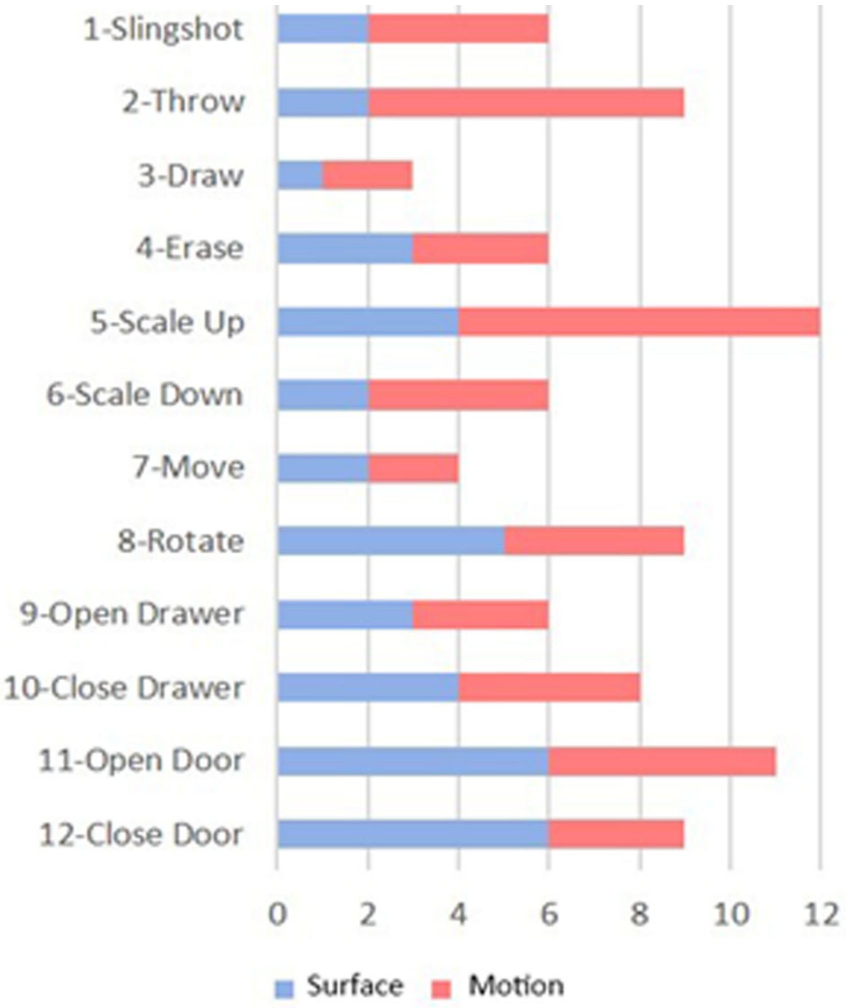


Figure 2. The number of types of gestures elicited for each task, surface gestures in blue and motion gestures in red

Figure 3. Agreement scores for each task in descending order, surface gestures (blue line) and motion gestures (red line).

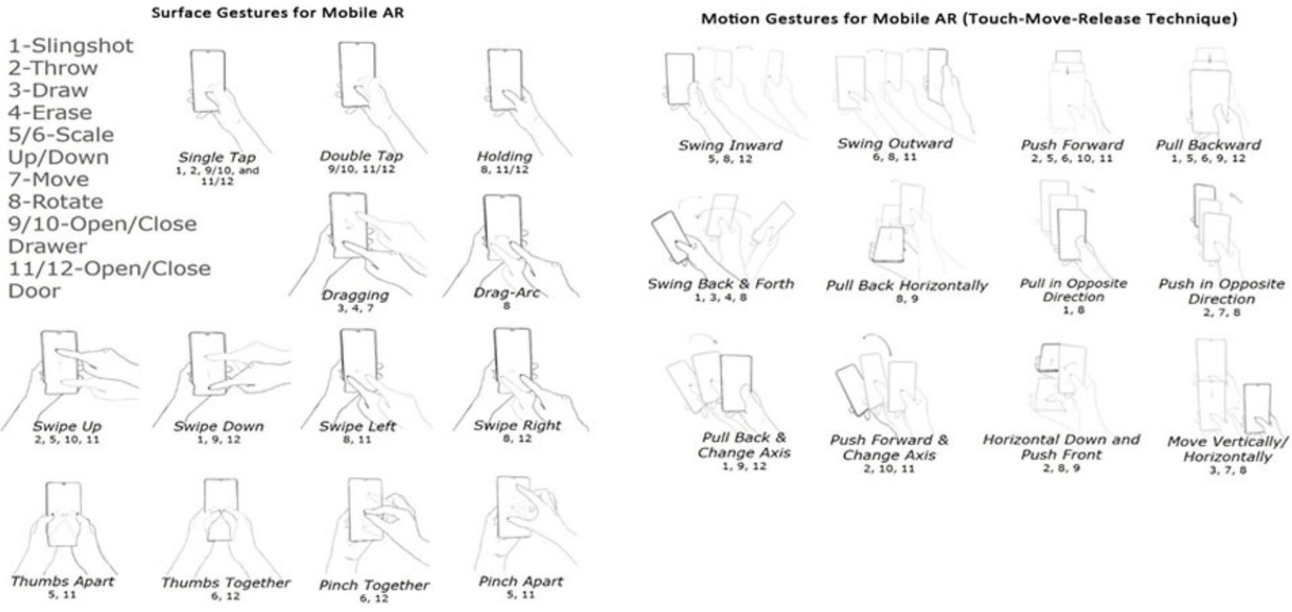


Figure 4. Two sets of user-defined gestures, surface gestures for mobile AR (top), motion gestures for mobile AR (bottom). Motion gestures demonstrate the concept of TMR (Touch-Move-Release) interaction technique.

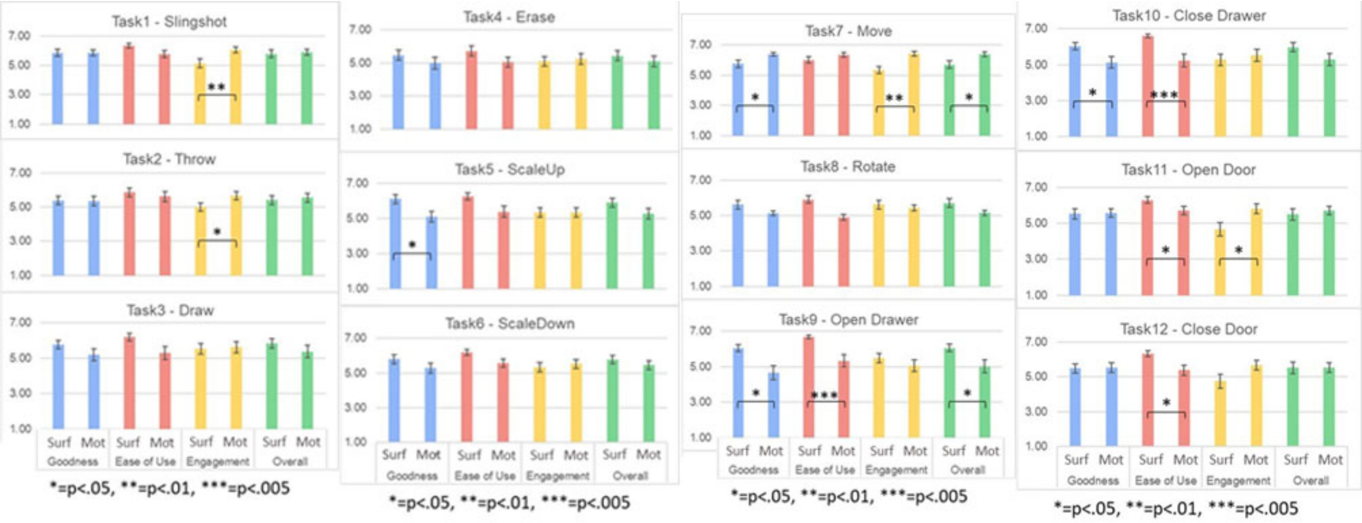
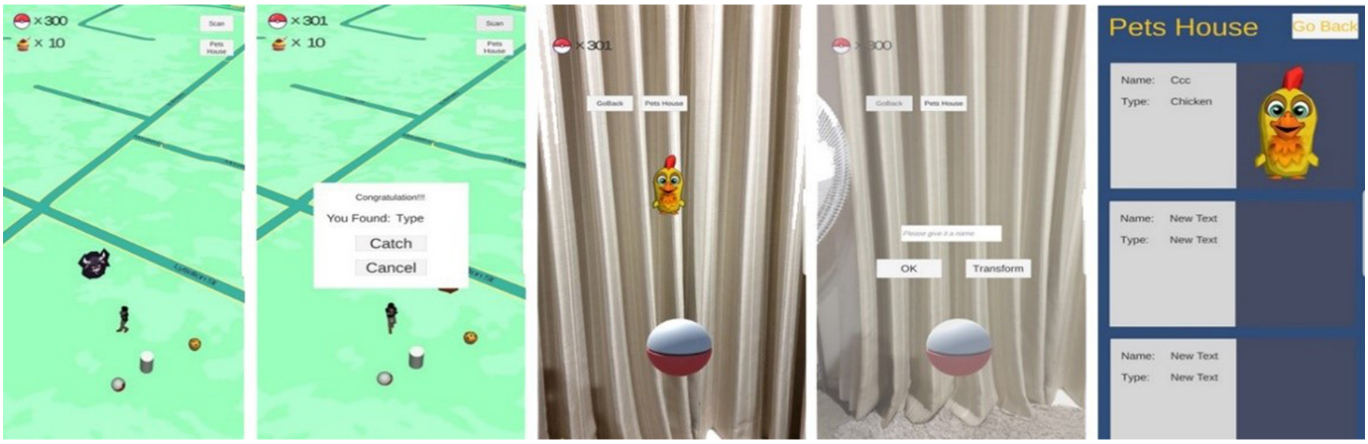


Figure 5. Subjective ratings in terms of Goodness, Ease of Use, Engagement, and Overall ratings for Task 1 to 12.



Congratulation!!!
You Found: Type

Name: Ccc
Type: Chicken

Name: New Text
Type: New Text

Name: New Text
Type: New Text

Figure 6. Screenshots of the Pokémon GO clone game: **A**) World map showing the player's current geolocation in the real-world and the locations of the creatures, **B**) once within the vicinity of the Pokémon, the player is prompted to attempt to catch the creature, **C**) the creature appears in the AR view, registered in the real-world 2 meters in front of the device camera, **D**) the creature is captured, **E**) the encyclopedia shows information about the captured creatures.

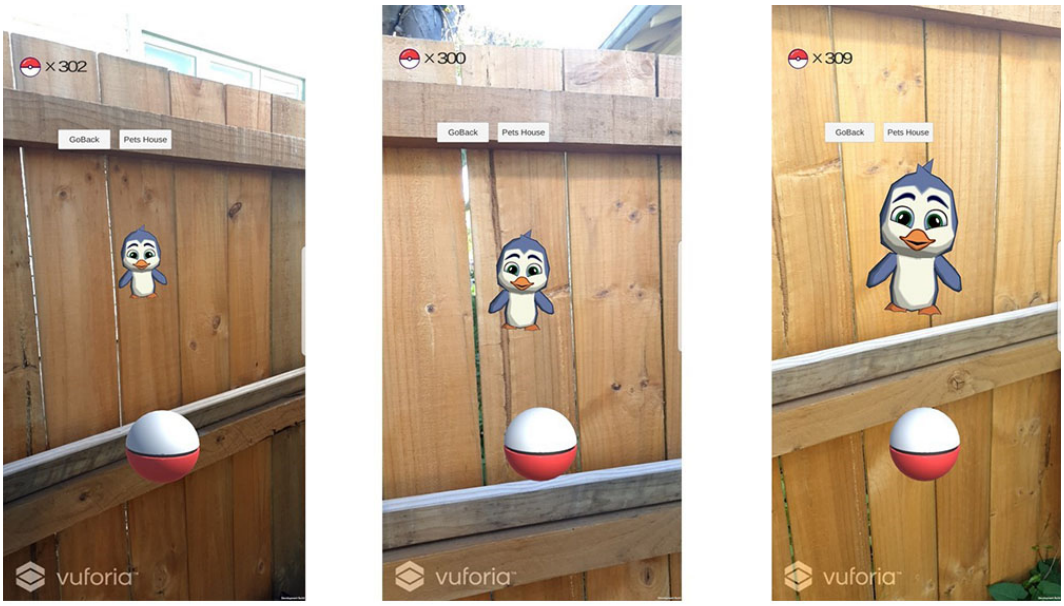


Figure 7. The three sizes of Pokémon-like creatures.

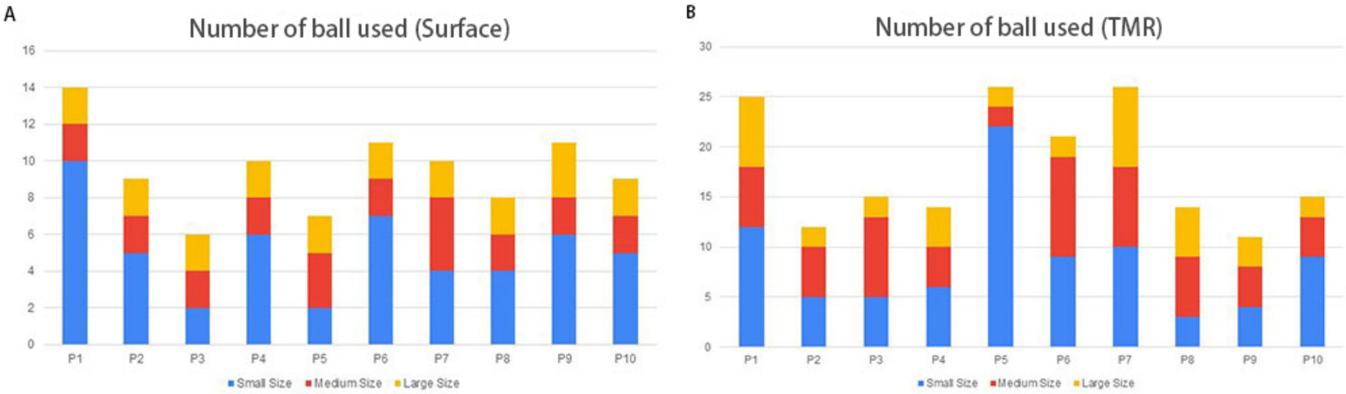


Figure 8. The number of balls used in the **(A)** Surface and **(B)** TMR motion gesture conditions.

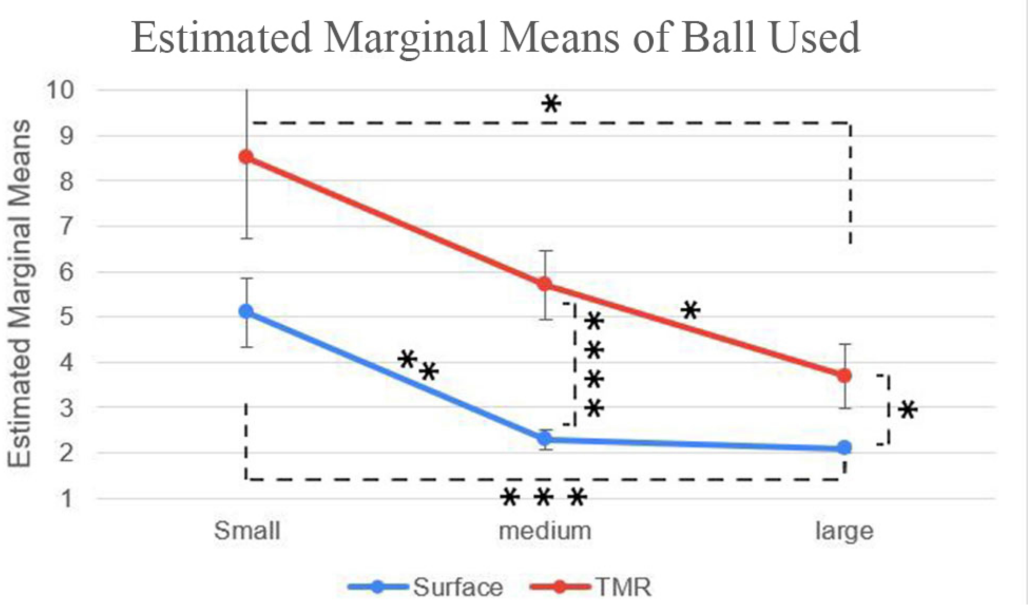


Figure 9. Number of balls used in each condition (*=p <.05, **=p <.01, ***=p <.005, ****=p <.001).

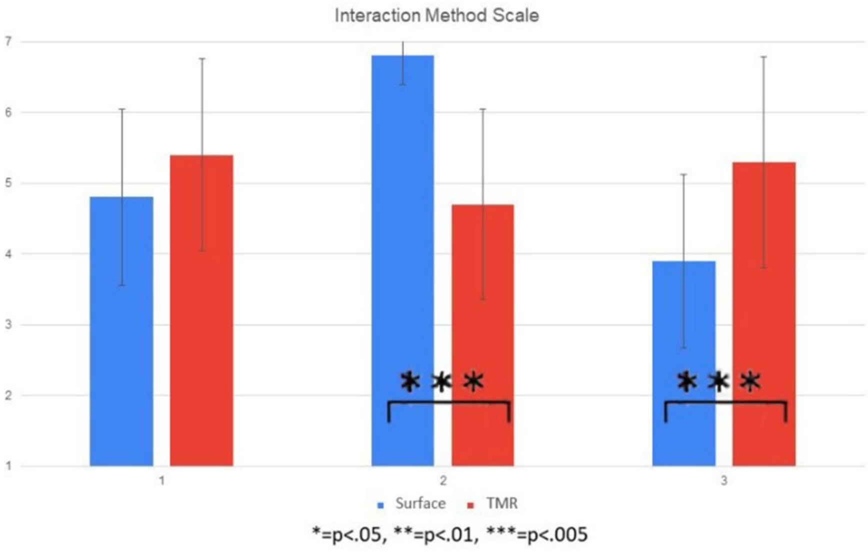


Figure 10. Subjective ratings between Surface gesture and TMR motion gesture.

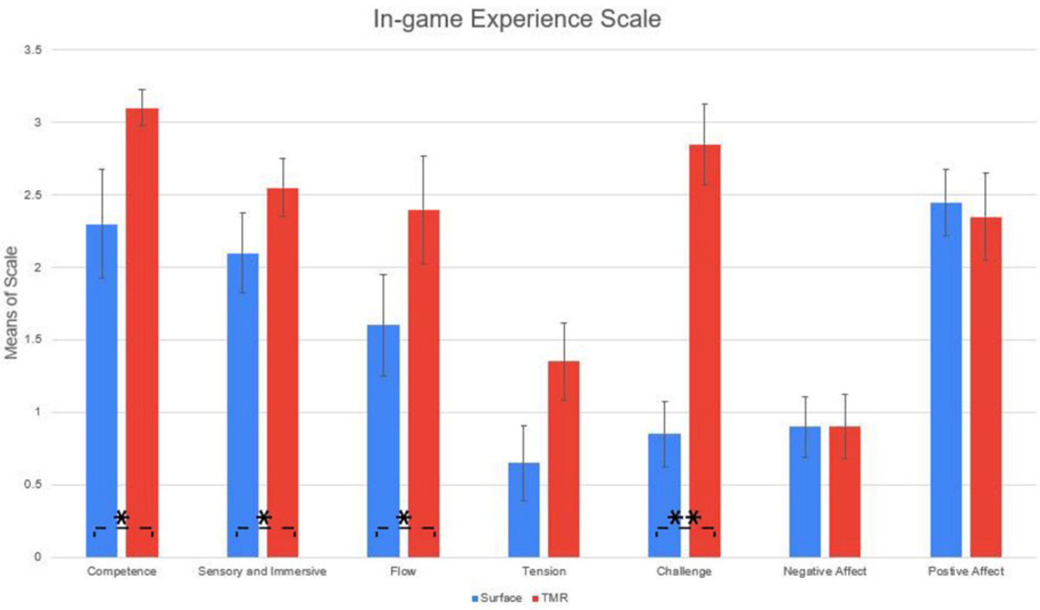


Figure 11. In-game experience scale of surface gesture and TMR interaction technique. (*=p <.05, **=p <.01)

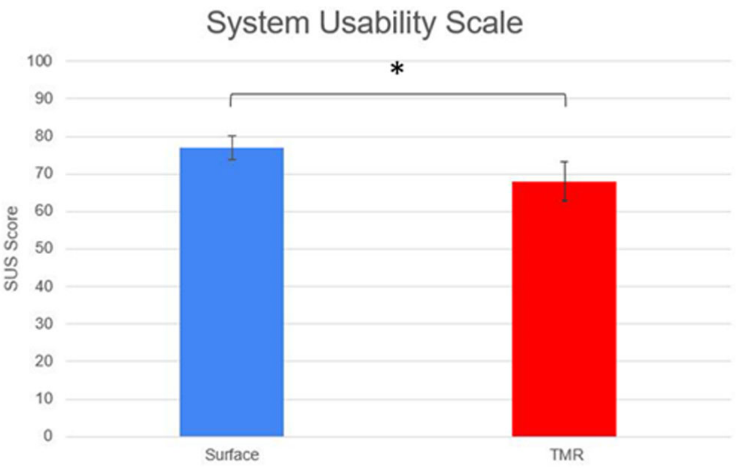


Figure 12. SUS scores for surface gesture and TMR motion gesture.

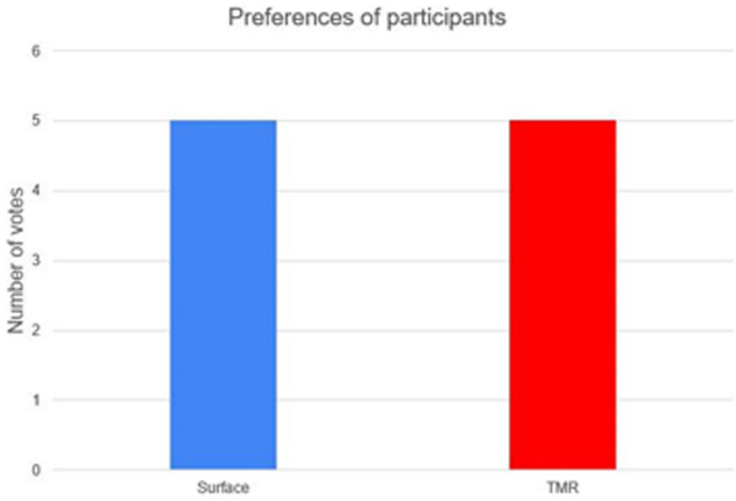


Figure 13. User preference of surface gesture and TMR motion gesture