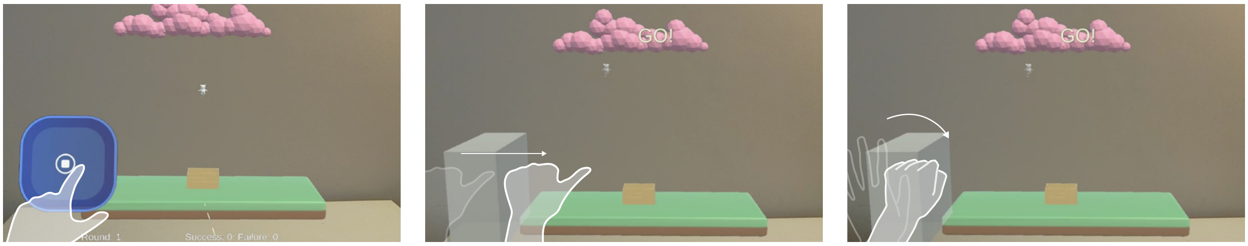
On the Benefits of Sensorimotor Regularities as Design Constraints for Superpower Interactions in Mixed Reality

Jingyi Li and Per Ola Kristensson



(a) Mid-Air Button (Tap) (b) SRs-Incongruent (Directional Swipe) (c) SRs-Congruent (Grab/Release)

Fig. 1: (a) MID-AIR BUTTON: a baseline method using a mid-air MRTK3 button to control the pace of time in a set of object-catching games in mixed reality. A single tap will slow down time for the falling object, and tapping it again will reset the time to its normal pace. (b) SRS-INCONGRUENT: a gestural control method incongruent with sensorimotor regularities in the context of temporal manipulation. Swiping the thumb right slows down time, and swiping the thumb left resets the speed of time. (c) SRS-CONGRUENT: a gestural control method congruent with sensorimotor regularities in the context of temporal manipulation. A grabbing gesture (from an open hand to a closed fist) slows down time, and a releasing gesture (from a closed fist to an open hand) resets the speed of time.

Abstract— Mixed Reality (MR) systems enable users to perform augmented superpowers that transcend real-world limitations. However, it remains unclear what types of action-outcome mappings can enable users to easily learn, control, and feel a sense of ownership of these augmented superpowers. Humans develop a set of sensorimotor regularities (i.e., image schemas and lawful relations between them) from recurring bodily experiences since early infancy, and use them to predict the outcome of our actions, or choose actions based on the desired outcome. We investigate whether sensorimotor regularities (SRs) can serve as effective design constraints for superpower interactions, by comparing three temporal manipulation methods in MR games: (1) mid-air button control; (2) gestures incongruent with SRs embedded in the concept of temporal manipulation; and (3) gestures congruent with these SRs. A within-subject study with 18 participants reveals that the SRs-congruent method enables significantly improved task performance, lower overall workload, and a greater sense of agency and presence compared to both an SRs-incongruent method and a mid-air button-based method. The SRs-congruent method also enabled faster mastery of the augmented superpower. No significant difference was observed in any of the above-mentioned metrics between the SRs-incongruent and mid-air button-based methods. These results empirically demonstrate multiple benefits of SRs as design constraints for superpower controls in MR, and encourage future research to explore their wider applicability in superpower interaction design.

Index Terms—Human augmentation, Mixed Reality, Image Schemas, Sensorimotor Contingencies

1 INTRODUCTION

Imagine having the ability to slow down or speed up time, make distracting objects vanish from your sight without physically removing them, or create a spatial-temporal clone of a moment and place you wish to remember and revisit. While these capabilities seem impossible in the real world, they become attainable within immersive environments like Mixed Reality (MR). Repeated calls have been made for utilizing these immersive spaces as playgrounds for designers and developers to create novel experiences that *feel real* yet transcend human capabilities and what the physical reality can offer [1,2,40,55,57], such as teleoperation [42] or working in an expansive mid-air office using MR headsets [43].

However, it is no trivial task for designers to create effective action-outcome mappings for augmented superpowers, which have no direct real-world equivalents for designers to draw on. For example, it is relatively straightforward for designers to mimic the action of using a spray can in real world when designing for the same function in Mixed Reality [51]. However, selecting appropriate actions for superpower outcomes like "creating a spatial-temporal clone" or "slowing down time" presents a much greater challenge. When designing such augmented superpowers, the absence of effective design constraints on the action-outcome mapping results in an unbounded design space with limitless possibilities, making it difficult to select mappings that align with general user expectations. Misalignment between designer-created mappings and users' expectations can make it difficult for users to learn, adapt to, and feel a sense of ownership over the augmented superpower. Moreover, the nature of MR can induce another type of misalignment between action-outcome mappings in the superimposed virtual space and those established in the real world, potentially weakening users' sense of "nonmediation" (i.e., the feeling of *presence* [35]).

To address the action-outcome mapping challenges, researchers have suggested drawing inspiration from science fiction and a wide range of superhero narratives [40,57]. However, mappings based on these superhero themes may not be universally recognized and can be difficult to generalize across users with varied backgrounds and cultural contexts. Other researchers suggested utilizing skills users have already developed in the real world to make beyond-real interactions easier to learn [2,24]. The challenge, however, is that users have a wide range of motor skills developed in the real world, and it remains unclear how to determine which specific well-practiced motor action should be mapped to a particular superpower outcome. Abtahi et al. [2] argue

• *Jingyi Li and Per Ola Kristensson are with University of Cambridge. E-mail: {jl2231 | pok21}@cam.cac.uk.*

Received 18 September 2024; revised 13 January 2025; accepted 13 January 2025. Date of publication 10 March 2025; date of current version 31 March 2025. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2025.3549876 provided by the authors Digital Object Identifier no. 10.1109/TVCG.2025.3549876

that the most effective way to support user learning of beyond-real action-outcome mappings remains unknown [2]. Possibly due to a lack of effective design methods, many superpower applications in MR have defaulted to using conventional input techniques to produce different superpower outcomes, such as controllers [7,33], buttons [27], and sliders [8, 32]. This risks not taking full advantage of modern MR headsets, which offer integrated hand tracking and support fully embodied interactions. There is a pressing need for well-defined design constraints for action-outcome mappings of augmented superpowers, as well as empirical evaluation of their effectiveness.

One promising candidate for such design constraints is the concept of *Sensorimotor Regularities* (SRs), defined as image schemas and the entailment between them. From early infancy, humans begin to derive a set of abstract representations (i.e., image schemas) from repeated sensorimotor experiences [36–38]. For example, we develop the image schema CONTACT from repeated experiences of observing the physical coming together of two or more things, and develop the schema COMPULSION from repeated experiences of observing an active entity forcing some passive entity to move. The entailment between different image schemas is also encoded in human cognition. For instance, we understand that a ball keeps rolling because the wind blows on it, or that a lamp topples from a table because a ball hits it [52]. We implicitly know that CONTACT is very likely to result in COMPULSION. Research in developmental psychology indicates that infants learn the mapping between CONTACT and COMPULSION within a few months [38]. There are many contingency relations between schemas, such as NEAR-CONTACT and BLOCKAGE-DIVERSION [21]. We rely on these sensorimotor regularities (image schemas and the entailment between them) to predict the outcomes of our actions, or to select actions based on our desired results. For example, we understand that placing a box in the path of a rolling ball will cause it to change direction (BLOCKAGE-DIVERSION), and if we desire to touch an object, we first need to get close to it (NEAR-CONTACT). We believe these sensorimotor regularities can be useful constraints, assisting designers in predicting appropriate actions that aligns with both general user expectations and real-world contingencies for a desired superpower outcome.

We hypothesize that implementing action-outcome mappings congruent with sensorimotor regularities in augmented superpowers will enhance task performance, user learning, and perceived agency/presence. To test these hypotheses, we designed three different control methods for a temporal manipulation superpower (i.e., being able to slow down time and reset the speed of time). The three methods include, a MID-AIR BUTTON method (Fig. 1a), an SRS-INCONGRUENT method with action-outcome mappings incongruent with sensorimotor regularities (Fig. 1b), and an SRS-CONGRUENT method with action-outcome mappings congruent with sensorimotor regularities (Fig. 1c). We compared the three superpower control methods in a set of MR object-catching games with 18 participants using a within-subject design. Results show that the SRS-CONGRUENT method enabled significantly higher success rate (the percentage of successful trials out of 32 trials of the base game), lower overall workload, greater sense of agency as well as presence compared to both MID-AIR BUTTON and the SRS-INCONGRUENT method. The SRS-CONGRUENT method also enable faster initial learning and achieves near-optimal performance (indicating that learning effects had flattened out by this point) earlier. Additionally, the SRS-CONGRUENT method was the clear favorite for most participants based on their subjective feedback. For the first time, we have implemented sensorimotor regularities as design constraints into the action-outcome mappings of temporal manipulation superpower in MR, and empirically demonstrated multiple benefits of such a design approach. In summary, this paper makes the following contributions:

1. We demonstrate a novel design approach using sensorimotor regularities as design constraints for action-outcome mappings of MR augmented superpowers, and an empirical evaluation on the benefits of this approach.
2. We present a highly effective and easy to learn gestural control method for temporal manipulation, designed based on sensorimotor regularities.
3. We present an empirical evaluation comparing the performance of the MID-AIR BUTTON, the SRS-INCONGRUENT method, and the SRS-CONGRUENT method in the context of temporal manipulation. The results indicate that the SRS-CONGRUENT method offers significant advantages in performance, user learning, and user experience.

2 RELATED WORK

2.1 Superpower Interactions

MR can enable a wide range of augmented superpowers, as it allows users to do things or experience situations that have no direct real-world equivalents [41]. One common type of augmented superpower in MR is sensory augmentation. For example, Cheng et al. [7] used MR to enable users to artificially remove or alter visual content within the environment to reduce information overload or enhance focus. Knierim et al. [27] used MR for reality slow down, allowing users to observe fast events in slow motions by pressing a button. Another type of augmented superpower in MR is motor augmentation. For instance, Irlitti et al. [23] used MR to enable remote manipulation of physical objects that are captured and volumetrically rendered in the shared space. Remixed Reality [33] enables users to dynamically alter the physical place, manipulate their viewpoint without physical movement, and manipulate time within the environment. Moreover, prior work has employed MR to augment cognitive superpowers. Lilija et al. [32] enable users to explore an item's past events by either directly dragging the item along its history movement trajectory or by using a timeline. In a similar vein, RealityReplay [8] enables users to review what happened to an real-world item when they were away using a time slider.

While most related studies focus on proposing and implementing useful augmented superpowers, there is a lack of research investigating the challenges of learnability and perceived ownership of these superpowers—which we believe are two main challenges in designing augmented superpowers in MR—and, more importantly, how to overcome them.

2.1.1 Learnability

The action-outcome mappings of augmented superpowers are initially unfamiliar to users, requiring them to learn to synthesize new control policies under novel dynamics [2]. To mitigate this unfamiliarity, researchers have suggested drawing inspiration from themes found in science fiction, superhero comics, and a wide range of narratives across books, movies, and other storytelling mediums [40, 57]. However, these superhero-inspired mappings may not be universally understood and can be challenging to generalize across a diverse group of users with varying backgrounds and cultural contexts. Another commonly adopted approach is to rely on well-established, conventional control methods when designing the action-outcome mappings for augmented superpowers, such as controllers [7, 33], buttons [27], slider [8, 32], or gaze [24]. However, although this approach might eliminate the effort required to learn the input action itself, it may still not make it easier for users to understand and adapt to the action-outcome mappings. Abtahi et al. [2] suggests that the best methods for supporting the learning of superpower interactions are still not known.

2.1.2 Perceived Ownership

Designing augmented superpowers that users can perceive as their own is no trivial task. When designing the action-outcome mappings for augmented superpowers, designers usually have to made ad-hoc decisions without being able to draw on any direct equivalents in the physical world. Due to individual differences in prior knowledge, there is a good chance designer-made mappings are not the same as what users would expect. For example, in Remixed Reality [33], pressing a button on a controller triggers changes in the user's viewpoint, while a user could expect to change their viewpoint by moving their head or eyes. According to Slater [49], if a participant carries out some physical actions which do not achieve similar outcomes as in physical reality, the sense of *presence* (i.e., *being there*) [20, 50] would be broken. In a similar vein, prior research reveals that violation of users' expectations about their actions in MR leads to reduced sense of *agency* [56], which refers to the feeling of controlling one's own actions and events in the external world through these actions (i.e., *I did this*) [18]. When there is a mismatch between the action-outcome mappings created by the designer and those anticipated by the user, feelings of "*I am not there*" and "*I did not do this*" may emerge and impair a user's perceived ownership over an augmented superpower.

Taken together, in the design of augmented superpowers, there is a lack of effective design constraints (i.e., the relations on a set of variables) [17] to assist designers in creating action-outcome mapping that are consistent with the expectations of general users, which can yield enhanced learnability and perceived ownership.

2.2 Sensorimotor Regularities as Design Constraints

2.2.1 Sensorimotor Regularities

From early infancy, humans begin to extract common patterns from recurring sensorimotor experiences, recoding them into a set of highly-abstract gestalts known as *image schemas* [38], which are defined as "recurring dynamic pattern(s) of our perceptual interactions and motor programs that gives coherence and structure to our experience" [25]. Through repeated sensorimotor experiences, we also implicitly learn the relations between these image schemas. The encoding of image schemas and lawful relations between them results in a set of sensorimotor regularities that can enable us to predict the outcomes of our actions or decide which actions to take based on the results we desire. For example, by actively moving or observing the movements of others, we develop image schemas SLOW and FAST. When we attempt to interact forcefully with objects or people, we often encounter forces that tend to impede our movement, causing our movement to slow down. Repeated experiences like these lead us to form the image schema RESISTANCE and learn the causal relationship between RESISTANCE and SLOW. In future situations, we can predict that an object or person will slow down upon encountering a resisting force like strong wind opposite to the movement direction, or we might decide to apply a resisting force when we want an object or person to slow down.

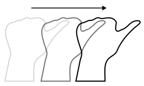
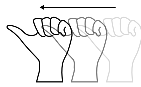
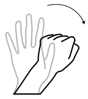
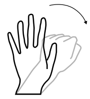
2.2.2 Hypothesized Benefits in Superpower Interactions

A substantial body of work in developmental psychology [4, 5, 13, 14, 28, 29, 34, 36–38, 59] and psycholinguistics [6, 15, 16, 46] has provided evidence supporting the psychological reality of image schemas, and further their universality [22, 30, 54]. Sensorimotor regularities (image schemas and relations between them) should be universal among people as their encoding is based on universal sensorimotor experiences (e.g., the experience of gravity and force dynamics). In the design of an MR instruction authoring system, Li and Kristensson [31] empirically demonstrated performance, learning and user experience benefits of image-schematic metaphors, the conceptual associations between a schema and a concept, (e.g., *Hidden Information* is DOWN). Taken together, we hypothesize that sensorimotor regularities (image schemas and entailment between schemas) can serve as effective design constraints to inform designers in creating action-outcome mappings for augmented superpowers that are consistent with the expectations of general users, which can potentially yield enhanced performance, learning, perceived presence, and perceived agency.

3 Designing Augmented Superpowers

To investigate the hypothesized benefits of sensorimotor regularities, we chose an augmented superpower, the ability to control the pace of time, to be the investigation context. Specifically, users can perform one action to slow down time and perform another action to reset the pace of time to normal. We designed three sets of action-outcome mappings for the proposed temporal manipulation superpower (see Tab. 1): (1) action-outcome mappings based on a well-establish MRTK3 Mid-Air button (MID-AIR BUTTON); (2) action-outcome mappings incongruent with sensorimotor regularities (SRS-INCONGRUENT); and (3) action-outcome mappings congruent with sensorimotor regularities (SRS-CONGRUENT).

Table 1: Three sets of action-outcome mappings we designed to control the pace of time in MR. All time-control actions are left-hand based as the dominant hand is tasked with game-related interactions in our experiment.

Outcome	*"Slow Down Time"*	*"Reset Time Speed"*
Action (MID-AIR BUTTON)	 (Single Tap)	 (Tap Again)
Action (SRS-INCONGRUENT)	 (Swipe Right)	 (Swipe Left)
Action (SRS-CONGRUENT)	 (Grab)	 (Release)

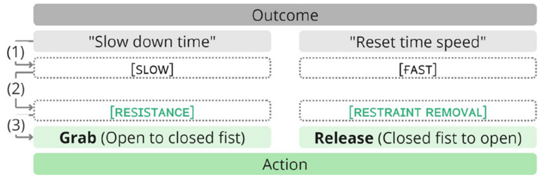


Fig. 2: A simple three-step process for the development of action-outcome mappings congruent with sensorimotor regularities. (1) Identify image schema representations for the known "Outcomes". (2) Identify an image schema (the "Action" schema) that is most likely to cause the corresponding 'Outcome' schema, based on relations between different image schemas. (3) Select a physical action that effectively represents the chosen "Action" schema.

3.1 Button-Based Action-Outcome Mappings

As discussed in Sec. 2.1.1, many augmented superpowers in MR rely on well-established input techniques such as buttons and controllers [7, 27, 33] when designing the action-outcome mappings. In the context of manipulating the pace of time, Knierim et al. [27] used a button on the HoloLens remote to trigger the slow motion effect. Another common example of using a button to control speed is the slow playback feature found in many video players and websites (e.g., YouTube, where clicking a "0.5" button will trigger slow playback). We adopt an MRTK3 Mid-Air button as a discrete binary control for two events, "slow down time" and "reset time speed to normal" (see Tab. 1). The action of tapping the button triggers the outcome of slowing down the pace of time, and tapping the same button again will reset the speed of time. We use the well-established MID-AIR BUTTON as a calibration point to investigate potential benefits of action-outcome mappings that align with sensorimotor regularities (SRS-CONGRUENT mappings), as well as the potential risks when these mappings violate those regularities (SRS-INCONGRUENT mappings).

3.2 SRs-Congruent Action-Outcome Mappings

To identify action-outcome mappings congruent with sensorimotor regularities, we propose a simple three-step process (see Fig. 2).

3.2.1 The "Outcome" Schema

Two possible outcomes can be achieved through using the temporal manipulation superpower, "slow down time" and "reset time speed". The most relevant image schema that can represent the outcome of "slow down time" is SLOW, and the most relevant image schema to represent "reset time speed" is FAST. We identify the most relevant

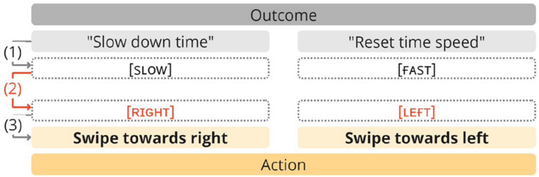


Fig. 3: Action-outcome mappings incongruent with sensorimotor regularities. The break in sensorimotor regularities was intentionally introduced in step (2). The image schema RIGHT is unlikely to yield the outcome of SLOW and similarly, LEFT is unlikely to cause FAST.

schemas based on image schema definitions provided by the ISCAT database [21].

3.2.2 The "Action" Schema

After examining the 46 image schemas as well as their relations to other schemas recorded in the ISCAT database [21], we found that the schema that will most possibly lead to the outcome of SLOW is RESISTANCE, defined as "a FORCE image schema that involves a force that tends to oppose or retard the motion of another entity". Also, we found that the schema that will most possibly lead to the outcome of FAST (following a previous state of SLOW) is RESTRAINT REMOVAL, defined as a FORCE image schema involving the physical or metaphorical removal of a barrier to motion. Now we identify two action-outcome mappings that align with sensorimotor regularities: RESISTANCE-SLOW and RESTRAINT REMOVAL-FAST.

3.2.3 Physical Action Congruent with 'Action' Schema

The schema RESISTANCE can be effectively represented by the commonly used mid-air gesture of *Grabbing*, where the hand changes from an open position to a closed fist. The schema RESTRAINT REMOVAL can be well represented by another commonly used mid-air gesture of *Releasing*, where the hand changes from a closed fist to an open hand. Therefore, we posit that a *Grab* gesture can meaningfully yield the outcome of "slow down time", and the *Release* gesture can meaningfully yield the outcome of "reset time speed".

3.3 SRs-Incongruent Action-Outcome Mappings

We often encounter arbitrary mappings between actions and outcomes in our interactions with systems. For instance, on a MacBook, the "Natural Scroll" feature means that scrolling up on a mouse or touchpad moves the window's content down. In contrast, on Windows systems, scrolling up on a mouse or touchpad moves the content up. A similar example can be found in the joystick controls for an airplane cockpit in a game: should the player pull down on the joystick to ascend, as is typical in real-world aviation, or push up to achieve the same result? To investigate potential impacts of arbitrary mappings, we design a set of action-outcome mappings incongruent with real-world sensorimotor regularities (see Fig. 3). A mistake was intentionally introduced when identifying "Action" schemas for the two known "Outcome" schemas (SLOW and FAST). We assigned the image schema RIGHT to be the cause of SLOW, and LEFT to be the cause of FAST. With the SRS-INCONGRUENT mappings, users slow down time by swiping thumb towards right, and reset the speed of time by swiping towards left.

Based on natural sensorimotor relations between schemas, the image schema RIGHT is unlikely to produce the outcome of SLOW, and similarly, LEFT is unlikely to result in FAST. However, the *Swipe Right/Left* actions are commonly used in gesture-based interactions. For instance, *Swipe Right* might represent "Execute", while *Swipe Left* might represent "Undo" [3,39,44]. Therefore, it is possible for a designer to choose a *Swipe Right* gesture to apply the time control effect and a *Swipe Left* gesture to reverse it. Despite violating sensorimotor regularities, the SRS-INCONGRUENT mappings still makes a reasonable design choice.

3.4 Minimizing Confounding Factors in Gesture and Button-Based Interactions

Acknowledging the challenges of precise interaction with mid-air buttons in MR, we increased the size of the MRTK3 button to roughly match the dimensions of an adult's palm, and positioned this button within easy reach of the user's left arm to improve usability, based on Fitt's law [12,47] (Fig. 1a). For other aspects, such as the visual contraction and highlighting of the 3D button's front plate when pressed (to aid depth perception) and the audio feedback, we strictly followed the default settings of the MRTK3 Pressable button[1]. Moreover, we understood that performing a free-hand gesture in MR can be inherently more efficient than pressing a mid-air button, as it does not require visually locating the button. To ensure a fair comparison, we constrained the effective area for all free-hand gestures to a semi-transparent action box approximately the same size as the mid-air button and placed it at the same location (Fig. 1b and Fig. 1c).

After taking these steps to minimize confounding factors—and considering that all three actions are so simple and brief—if the level of congruence with sensorimotor regularities has minimal effect, participants should be able to learn and adapt to all of them quickly, showing no significant differences after minimal exposure. However, if congruence with sensorimotor regularities does have an impact, we would expect to observe performance differences even among these very simple and brief action-outcome mappings.

3.5 Context

To create an MR environment where participants are encouraged to use this temporal manipulation superpower, we developed a set of HoloLens2-based MR games where players were tasked to catch rapidly falling virtual objects—an almost impossible task without altering the pace of time. In all three games, players used left-hand actions for temporal manipulation and used right-hand gestures for game-related tasks.

In the first game ('CatGame' in Fig. 4), a virtual cat rapidly falls from a different position at each trial. To succeed, players slow down the pace of time for the cat, and move a virtual box to the cat's landing spot with a standard *AirTap-and-hold*[2] gesture. In the second game ('DinosaurGame' in Fig. 4), a virtual dinosaur falls from the same position but with different rotational poses at each trial. To succeed, players slow down the pace of time for the dinosaur and rotate a virtual box in the landing zone to match the dinosaur's rotational pose, by *AirTapping* a rotation button that introduces a 45-degree clockwise rotation at each tap. In the third game ('GemGame' in Fig. 4), four gems of the same color drop into a landing zone containing four boxes, each in a different color. The color of the gems changes with each trial. To succeed, players slow down the pace of time for the gems, and put on or take off the lid for each box by *AirTapping* a lid-control toggle button. The goal is to ensure that one gem falls into the color-matching box while preventing the other three gems from landing in boxes with mismatched colors. We used the above three games as the base games, and each participant completed 32 trials (i.e., repetitions) of each base game.

4 EVALUATION

We conducted a within-subject user study (3 methods $\times$ 3 games) to evaluate performance, learning, overall workload, perceived presence, and perceived agency across three different action-outcome mappings (MID-AIR BUTTON, SRS-CONGRUENT, SRS-INCONGRUENT). Each participant was exposed to all three action-outcome mappings, with each mapping paired with a different game. We used a Latin Square counterbalancing design to control for order effects across three action-outcome mappings and three games. The research was approved by the Local Research Ethics Committee.

[1] https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-uxcomponents/packages/uxcomponents/button
[2] https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-gestures-hl2

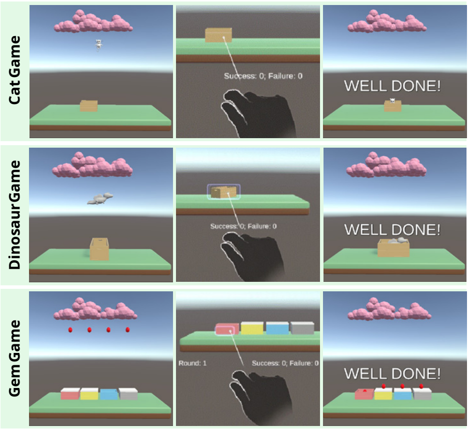


Fig. 4: Three MR games we developed as the task contexts to investigate the use of an augmented superpower. In each game, players must adjust a box's position, rotation, or lid to the correct state before the rapidly falling objects land. To do this successfully, participants must control the pace of time using one of the three sets of controls we designed.

4.1 Participants

We recruited 18 participants via convenience sampling (6 females and 12 males; 22–30 years; mean age = 26.24). Five participants identified themselves as experienced users, reporting that they use AR or VR devices from sometimes to always, while 13 participants had little to no experience, indicating they had never to occasionally used AR or VR headsets before. We distributed a follow-up questionnaire to assess participants' familiarity with superpowers from movies and books as well as their associated gestures, and we have received 16 responses. Seven participants reported high familiarity with superpowers, six reported moderate familiarity, and three reported little to no familiarity. Regarding the superpower gestures, three reported high familiarity, two reported moderate familiarity, and eleven reported little to no familiarity.

4.2 Apparatus

The experiment software was developed on the Microsoft HoloLens 2 using Unity and the Microsoft Mixed Reality Toolkit (MRTK3). The software supports bare-hand gesture recognition and no controllers were involved in the study. Throughout the experiment, the participants maintained a seated position while wearing the HoloLens 2 device. The experiment software provided discrete audio feedback for each successful gestural interaction on both hands. For the left hand, audio feedback was given for each successful *Tap* on the mid-air button or each successful gesture (*Grab*, *Release*, *Swipe Right*, and *Swipe Left*). For the right hand, audio feedback was provided for each successful *AirTap* or *AirTap-and-hold* gesture. Participants could manipulate time at any point between the appearance of "GO" in the game and the landing of the falling object. The duration of time manipulation could range from 0 to a maximum of 13.86 seconds, achieved if participants began slowing time immediately when "GO" appeared and maintained it until the object landed.

4.3 Tasks

In this user study, each participant played three different object-catching games (CatGame, DinosaurGame, GemGame), completing 32 trials of each game. Participants manipulated the pace of time using one of the three sets of left-hand actions (MID-AIR BUTTON, SRS-CONGRUENT, SRS-INCONGRUENT), giving them enough time to manipulate a virtual box with right-hand gestures including *AirTap* and *AirTap-and-hold*. Free-hand bimanual interactions were consistently employed by participants for all three methods, enabling a fair comparison. To achieve success, participants slowed down time, made the necessary box adjustments depending on the games, and then reset the time speed. The specific box adjustments varied across the games: In the CatGame, participants moved a virtual box horizontally to catch a falling cat. In the DinosaurGame, they rotated the box using a button to align it with the rotational pose of a falling dinosaur. In the GemGame, participants manipulated the lids of the boxes to ensure that color-matching gems landed in the correct boxes while preventing mismatched gems from landing incorrectly. Each trial of all three games started with "GO" displayed, signaling participants to perform left-hand gestures for time manipulation and right-hand gestures for box manipulation. The overall task completion time was measured from the appearance of "GO" to when the falling object landed, encompassing both time manipulation and box manipulation durations. Additionally, the outcome of each trial was recorded as either "success" or "failure". For each method, the average success rate (the percentage of successful trials out of 32 trials) and the average task completion time for successful trials were detailed in Sec. 5.1.

4.4 Procedure

4.4.1 Familiarization Phase

At the start of the experiment, we gave participants a brief introduction to the HoloLens2 and a tutorial of the AirTap-based right-hand gestures for virtual box manipulation. Participants were given the opportunity to practice manipulation of box's position, rotation, and lid configuration. Before starting to play each game, we introduced to participants the rules of the game and the left-hand action they used in this game to manipulation the pace of time.

4.4.2 Test Phase

After the introduction about game rules and left-hand actions, participants proceeded with the test phase, where they played 32 trials of each game. For each game, they employed different actions to control the pace of time. After playing each game, participants filled out the NASA Task Load Index (NASA-TLX), a Perceived Agency questionnaire consisting of three questions adapted from The Sense of Agency Scale (SoAS) [53] measured on a seven-point Likert scale, and a Perceived Presence questionnaire consisting of four questions adapted from Witmer and Singer's Presence Questionnaire (PQ) [58] measured on a seven-point Likert scale. At the conclusion of the experiment, participants filled out a User Preference Questionnaire, which gathered their rankings of the three methods for temporal manipulation and their feedback on what they liked and disliked about each method.

5 RESULTS

For statistical analysis of the parametric data, we applied repeated measures analysis of variance (ANOVA) and post-hoc paired t-tests with Bonferroni correction. We used the Mauchly's test to check the data's sphericity. For statistical analysis of the non-parametric data, we used Friedman's test and post-hoc Wilcoxon Signed Rank test with Bonferroni correction.

5.1 Performance

Fig. 5(a) shows the distributions of participants' average success rate across 32 trials of each game for each method. A repeated measures ANOVA was conducted to assess the differences in average success rates across three methods: MID-AIR BUTTON, SRS-CONGRUENT, and SRS-INCONGRUENT. The results showed a significant effect of method on average success rates ($F(2,34) = 13.79$, $p < 0.001$, $\eta^2 = 0.33$). Post-hoc pairwise t-tests with a Bonferroni correction (corrected $\alpha = 0.0167$) revealed that the SRS-CONGRUENT method enabled significantly higher success rate compared to both MID-AIR BUTTON ($p < 0.001$) and SRS-INCONGRUENT method ($p = 0.0022$). No significant difference was found between MID-AIR BUTTON and SRS-INCONGRUENT method ($p = 0.1350$).

Fig. 5(b) shows the distributions of average time taken to complete a successful trial for each method. A Friedman test revealed a significant effect of method on the average time taken to achieve each success ($\chi^2(2) = 7.44$, $p = 0.0242$). Post-hoc Wilcoxon Signed-Rank test with Bonferroni correction (corrected $\alpha = 0.0167$) revealed that the SRS-CONGRUENT method enabled significantly shorter time to

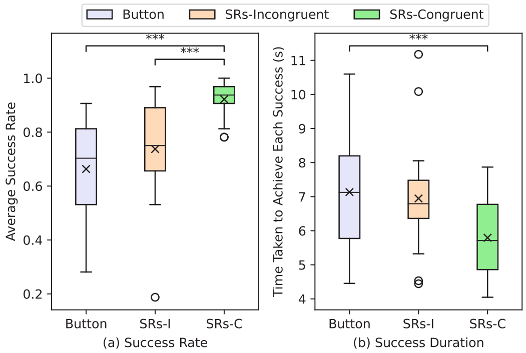


Fig. 5: (a) shows the distributions of average success rate per method. (b) shows the distributions of average time taken to achieve each success per method. The box plots show the median (the horizontal line), the mean ('x'), the first and third quartile (the box) and the minimum and maximum (the whiskers). The circle signs ('○') indicate outliers. The asterisks ('***') indicate a statistically significant difference (p<0.01) between two methods.

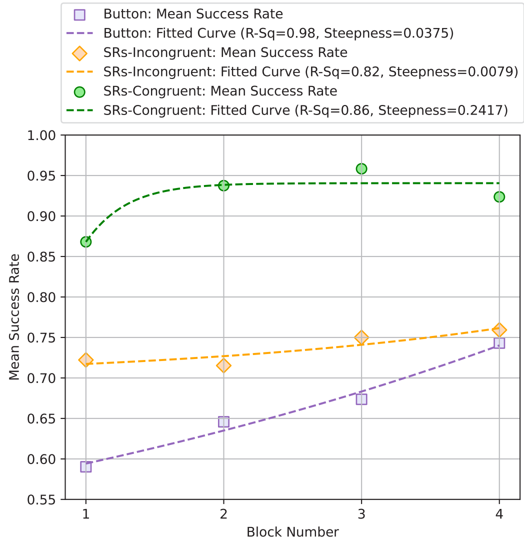


Fig. 6: Fitted learning curves across four blocks (each containing eight trials) of three methods. The square signs ('□') indicate real data points for the MID-AIR BUTTON method. The diamond signs ('◇') indicate real data points for the SRS-INCONGRUENT method. The circle signs ('○') indicate real data points for the SRS-CONGRUENT method.

achieve each success compared to MID-AIR BUTTON ($p < 0.01$). No significant difference was found between MID-AIR BUTTON and SRS-INCONGRUENT method ($p = 0.5798$) or between SRS-CONGRUENT and SRS-INCONGRUENT methods ($p = 0.0237$).

5.2 Learning

5.2.1 Learning Rate

For each method, we divided the 32 trials into four blocks and performed non-linear least squares to find the best-fitting curve for the average success rate data across the four blocks for each method (see Fig. 6). The MID-AIR BUTTON and SRS-INCONGRUENT data were effectively explained by an exponential fit, and SRS-CONGRUENT data followed a logistic fit. All methods achieved R^2 values above 0.8, indicating a good fit for each model. We performed Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) analysis to further validate model fit for each method, and the results supported the previous conclusions based on R^2 values. Specifically, for MID-AIR BUTTON, the exponential model had AIC and BIC values of 6.000058 and -34.841349, lower (which indicates a better fit) than 6.000063 and -34.551797 for the logistic model. Similarly, for SRS-INCONGRUENT data, the exponential model yielded AIC and BIC values of 6.000062 and -34.622429, lower than 6.000072 and -34.018945 for the logistic model. In contrast, for SRS-CONGRUENT data, the logistic model demonstrated a better fit, with AIC and BIC values of 6.000152 and -31.020604 compared to 6.000679 and -25.021765 for the exponential model. Using the fitted curve for each method, we calculated the steepness at the first block to represent the learning rate during the initial trials. The SRS-CONGRUENT method showed the fastest initial learning ($Steepness_{block1} = 0.2417$), followed by the MID-AIR BUTTON method in the middle ($Steepness_{block1} = 0.0375$), while the SRS-INCONGRUENT method resulted in the slowest initial learning ($Steepness_{block1} = 0.0079$).

5.2.2 Saturation Point

Since the SRS-CONGRUENT data was effectively modeled by a logistic fit, we were able to determine the inflection point x_0, representing the block where learning is at its highest (maximum learning rate). After this point, learning begins to decelerate, and performance (success rate) starts to approach its limit. With the SRS-CONGRUENT method, user learning peaked at block 0.31 ($x_0 = 0.31$, $Steepness_{block_{x_0}} = 0.8494$). Since each block consisted of eight trials, this suggests that learning rapidly reached its maximum between trials 2 and 3, after which performance began to stabilize. In contrast, both the MID-AIR BUTTON and SRS-INCONGRUENT methods followed an exponential fit, showing no clear inflection point throughout the four blocks. This suggests that participants may not have reached the learning saturation point by the end of the experiment and had not yet begun to approach optimal performance.

5.2.3 Model-Based and Model-Free Learning

Participants using the SRS-CONGRUENT method showed a learning pattern typically observed in *Model-based* learning, characterized by rapid and substantial performance improvements that gradually stabilized [19]. Planning of movement is usually involved in *Model-based* learning, which is indicated by shorter times taken to achieve each success when using the SRS-CONGRUENT method. *Model-based* learning is argued to be "the most powerful and flexible approach" to learn a good control policy in an uncertain environment [19]. In contrast, participants using both MID-AIR BUTTON and SRS-INCONGRUENT method exhibited learning patterns typically associated with *Model-free* learning, marked by a much slower performance improvement compared to the *Model-based* learning and a large number of attempts required before a good policy can be acquired [19, 45, 48]. Participants rely on remembering and repeating actions that previously led to success [19].

5.3 Perceived Agency and Presence

Fig. 7(a) presents the mean ratings on three agency questions across all participants, categorized by each method. To calculate each participant's overall perceived agency score for each method, we averaged the ratings from the three questions, resulting in a mean perceived agency score ranging from 1 to 7. A Friedman test revealed a significant effect of method on the overall perceived agency score ($\chi^2(2) = 11.37$, $p = 0.0034$). Post-hoc Wilcoxon Signed-Rank test with Bonferroni correction (corrected $\alpha = 0.0167$) revealed that the SRS-CONGRUENT method had a significantly higher overall perceived agency score compared to both MID-AIR BUTTON ($p < 0.01$) and SRS-INCONGRUENT ($p = 0.0109$) method. No significant difference was found between MID-AIR BUTTON and the SRS-INCONGRUENT method ($p = 0.2554$).

Fig. 7(b) presents the mean ratings on four presence questions across all participants, categorized by each method. To calculate each participant's overall perceived presence score for each method, we averaged the ratings from the four questions, resulting in a mean perceived presence score ranging from 1 to 7. A Friedman test revealed a significant effect of method on the overall perceived presence score

($\chi^2(2) = 7.42,\ p = 0.0244$). Post-hoc Wilcoxon Signed-Rank test with Bonferroni correction (corrected $\alpha = 0.0167$) revealed that the SRs-Congruent method had a significantly higher overall perceived presence score compared to both Mid-Air Button ($p = 0.0156$) and SRs-Incongruent ($p = 0.0144$) method. No significant difference was found between Mid-Air Button and the SRs-Incongruent method ($p = 0.7768$).

5.3.1 Correlations with Performance

We also examined whether participants' perceived *Agency* and *Presence* were correlated with their performance (*SuccessRate* and *SuccessDuration*). Spearman's correlation was computed for the following variable pairs: *Agency-SuccessRate*, *Agency-SuccessDuration*, *Presence-SuccessRate*, and *Presence-SuccessDuration*, with the results summarized in Fig. 8. Across all three methods, we observed significant, strong positive correlations ($p < 0.05$) between *SuccessRate* and *Agency*, as well as significant, strong positive correlations ($p < 0.05$) between *SuccessRate* and *Presence*. However, *SuccessDuration* showed weaker correlations with subjective ratings, as the only significant result was a large negative correlation between *SuccessDuration* and *Presence* in the SRs-Congruent method.

5.4 Subjective Workload

Fig. 7(c) shows the NASA-TLX ratings of six items across all participants for each method. We averaged each participant's ratings on six items for overall workload, and performed a Friedman test to compare the overall workload across three methods. Results revealed a significant effect of method on the overall workload ($\chi^2(2) = 9.66$, $p = 0.0080$). Post-hoc Wilcoxon Signed-Rank test with Bonferroni correction (corrected $\alpha = 0.0167$) revealed that the SRs-Congruent method had a significantly lower overall workload compared to both Mid-Air Button ($p < 0.01$) and SRs-Incongruent ($p = 0.0120$) method. No significant difference was found between Mid-Air Button and the SRs-Incongruent method ($p = 0.2274$).

5.5 User Preferences and Feedback

As shown in Fig. 9, the Mid-Air Button method was generally the least preferred by participants, while SRs-Congruent emerged as the most preferred. Specifically, for the Mid-Air Button method, eleven participants rated it as least preferred, five rated it as neutral, and two rated it as most preferred. For SRs-Incongruent, seven participants ranked it as least preferred, eight as neutral, and three as most preferred. Meanwhile, for SRs-Congruent, no participants ranked it as least preferred, five ranked it neutral, and 13 rated it as most preferred.

5.5.1 Participants Feedback

The Mid-Air Button was predominantly characterized as "error-prone," with ten participants highlighting this issue. Eight participants mentioned terms like "miss-click" and "double-click". P1 mentioned that "it is more error prone as there was no haptic feedback". Two other participants attributed the difficulty to depth perception issues. P7 noted that it was "only effective at certain depth," while P18 observed that "the pushing threshold is not very clear." These comments highlight general usability challenges related to depth perception and the lack of haptic feedback, inherent to all types of 3D mid-air buttons. Notably, participants did not find the button too small or difficult to locate or reach, suggesting our adjustments improved usability, though some inherent challenges remain. The second most frequent theme was "ease of use and understanding", mentioned by five participants. One reason given was "button was a standard way of applying control" (P1). The third most frequent theme was the feeling of "triggering on/off mechanism", highlighted by four participants using terms like "...versus a button that I know it's simply triggering a mechanism" (P9) and "press a button and have a function executed" (P17).

The SRs-Incongruent method was characterized as problematic in terms of the mapping between gestures and outcomes by nine participants. P1 noted that "it is very counter-intuitive. The right swipe to me felt like speed-up, but was used as slow-down in this experiment." P9 proposed a different mapping that they think was more sensible "I might want to use 'up' as slow down (since it is the opposite direction of the falling object) and 'down' as speed up (since it's the same direction of the falling object)". And P17 mentioned that "it is not super logical as to why right swipe and not left swipe should be to slow down." Further, three participants expressed that such mappings issues led to difficulties in remembering the gestures: "had to remind myself actively which direction I needed to swipe to get what I needed" (P1), "easy to forget how to use sometimes for some reason" (P11), and "this requires some remembering" (P17). Additionally, three participants appreciated that this method had two distinct states. For example, P9 mentioned that "I like that it has two movements (left and right) that represent different things (versus a button, both speed up and slow down are the same actions)". P18 mentioned that "the threshold is super clear".

The SRs-Congruent method was primarily described as easy to learn, use, and control by seven participants. P2 commented that this method is "easy to use and to comprehend", and P4 found it "very easy and comfortable to control". Seven participants also highlighted the natural, intuitive, and logical mappings between gestures and outcomes. P14 described it as "logical as a temporal control superpower", and P17 found that "it is more natural than the directional swipe method as grabbing is a common gesture to express slowing down". P3 commented that "the gesture control is very natural". Additionally, six participants reported feeling as if they actually owned this superpower. P1 expressed that "it feels like I am actually controlling the time in my hand", and P6 commented that "it feels like I got hold of time", P9 felt that "I can really feel that it's doing something to the time and I have the superpower". Interestingly, four participants described experiencing perceptual illusions about time when using this method in the game. P9 described that "I imagine myself almost like holding a handful of sand (like in a hour glass), when I hold it, the time is still, when I let go, the sand runs through my hand and the time keeps running". P12 described that "Time is intangible, but I feel as if the grabbing gesture has transformed time into some kind of tangible material". However, it's important to note that six participants mentioned experiencing fatigue with this method. P1 explained that "It does get a little tiring after a while as I grasped really hard during the experiment. Maybe because the temporal demand of the task was high and I tried very hard to get the time to slow down".

6 Discussion

Our data analysis reveals that action-outcome mappings aligned with sensorimotor regularities (SRs-Congruent) significantly outperformed both Mid-Air Button and the SRs-Incongruent method. Statistically significant advantages were demonstrated in better performance, enhanced sense of agency and presence, and lower overall workload. The SRs-Congruent method also enabled faster initial learning and earlier achievement of near-optimal performance compared to other two methods. No significant differences in performance were found between the SRs-Incongruent and Mid-Air Button methods. Prior research [10, 11, 26, 60] suggests that the performance of mid-air buttons suffers from the lack of haptic feedback, as force or vibrotactile feedback can not be provided without extra devices. In contrast, the performance of mid-air gestures, being a more embodied interaction technique, is less dependent on haptic feedback and thus unaffected by its absence. However, despite the benefits of gestural interaction in mid-air environment, the SRs-Incongruent methods failed to yield any significant performance benefits compared to the Mid-Air Button. This could be due to the SRs-incongruence in their Action/Outcome mappings negating the advantages of their embodied and haptic-independent nature. This highlights the risks of making arbitrary design choices when designing freehand gestures in MR in general—if the action-outcome mapping of a gesture is not aligned with sensorimotor regularities (which reflect general user expectations), converting conventional mid-air button controls to freehand gestures can be an effort in vain, as such mappings are unlikely to deliver improvements in performance, learning, workload, or perceived agency/presence. Based on users' subjective feedback, the SRs-Congruent method was the clear favorite for the majority. When using this method to control time,

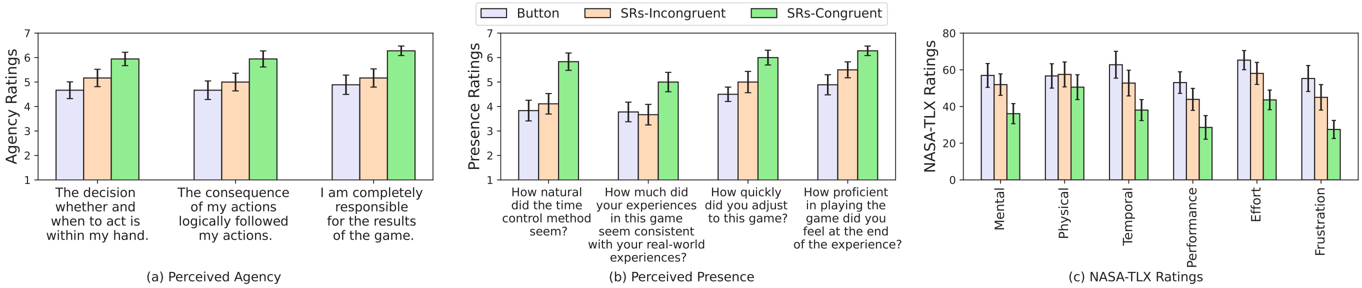


Fig. 7: (a) shows subjective ratings on the sense of agency. Each item on the x-axis represents one question adapted from the Sense of Agency Scale (SoAS) [53]. (b) shows subjective ratings on the sense of presence. Each item on the x-axis represents one question adapted from Witmer and Singer's Presence Questionnaire (PQ) [58]. (c) shows the NASA-TLX scale ratings for each method. Error bars show one standard error.

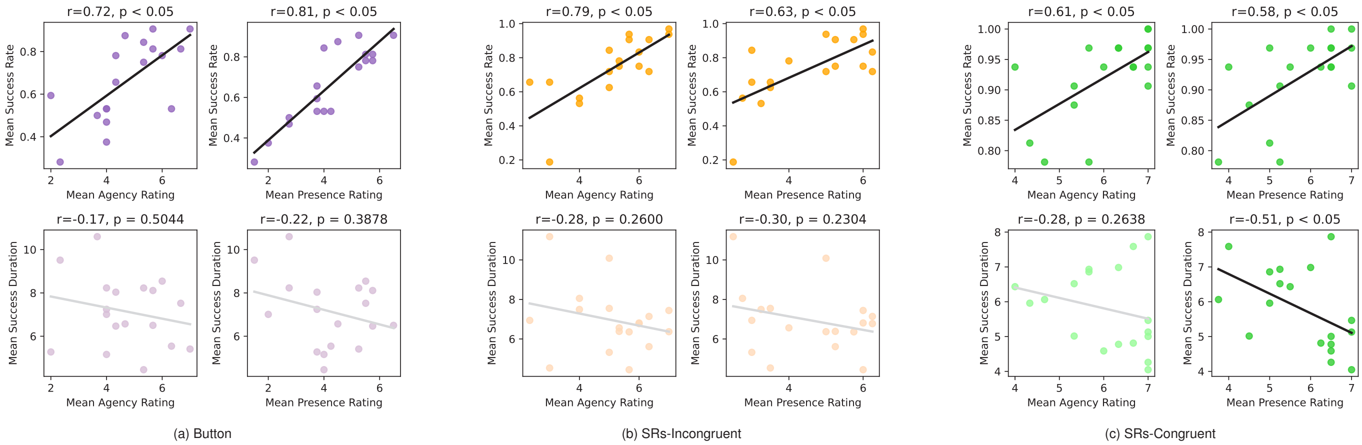


Fig. 8: Fitted line plots show the correlations between (1) *SuccessRate-Agency*, (2) *SuccessRate-Presence*, (3) *SuccessDuration-Agency*, and (4) *SuccessDuration-Presence* in each method. $r > 0.5$ indicate a strong correlation, $p < 0.05$ indicates significance of correlation.

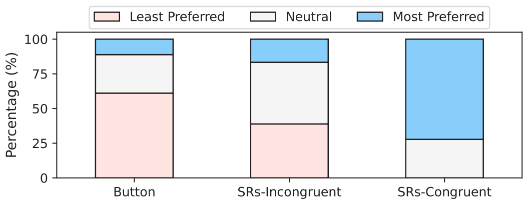


Fig. 9: The percentage of participants who ranked each method as most preferred, neutral, or least preferred.

participants uniquely reported a strengthened sense of ownership over the augmented superpower, along with perceptual illusions related to time. These findings empirically demonstrated the benefits of using sensorimotor regularities as constraints to design the action-outcome mappings of augmented superpowers in MR.

6.1 Performance Gain

Using the MID-AIR BUTTON method as a reference point, we found that applying SRS-CONGRUENT action-outcome mappings for the time control superpower yielded a significant performance gain. This was evident from the significantly higher success rate and the reduced time required for each success. A high success rate indicates that participants developed effective policies under this superpower dynamics. Moreover, shorter time required for each success reflects a refinement of these policies, with movement planning involved (a characteristic of *Model-based* learning [19]). Initially, participants would slow down time, adjust the box, and wait for the object to fall slowly. As they gained clearer understanding of different world states and causalities between them, their control policies evolved: they slowed down time, adjusted the box, and then immediately reset the time speed, resulting in shorter times to achieve success (as shown by the SRS-CONGRUENT method). In contrast, with the MID-AIR BUTTON method, P17 noted in the post-experiment survey, "Due to the low success rate, I didn't bother to resume the time flow after completing the other game action, leaving a period at the end waiting for the balls to drop slowly." Ineffective action-outcome mappings can demotivate efforts to refine control strategies and plan movement. We suggest that using SRs as design constraints for action-outcome mappings can motivate the refinement of formed control policies, apart from just enabling better performance, and encourage future research to further validate this hypothesis.

6.2 Enhanced Learning

Compared to MID-AIR BUTTON and the SRS-INCONGRUENT method, we found that the SRS-CONGRUENT method enabled faster learning during initial trials and an earlier achievement of approaching-optimal performance. The results indicate that SRS-CONGRUENT action-outcome mappings can yield faster mastery of augmented superpower.

Moreover, as discussed in Sec. 5.2.3, participants using the SRS-CONGRUENT method showed a learning pattern commonly observed in *Model-based* learning, while participants using both MID-AIR BUTTON and the SRS-INCONGRUENT method exhibited learning patterns typically associated with *Model-free* learning. One of the main differences between *Model-based* learning and *Model-free* learning is that the former involves developing an internal world model that encompasses different states, the casual relationships between actions and outcomes in state transitions, and the rewards of different outcomes; while the later lacks such a world model and is solely driven by the difference between actual and expected outcome [9, 19]. The varia-

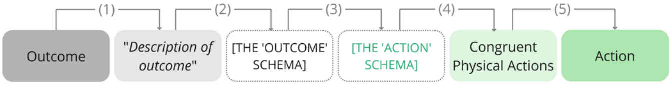


Fig. 10: SRs-constrained design pathway for action-outcome mappings in MR augmented superpowers.

tions in learning patterns observed among the three methods may be attributed to whether participants developed an internal model of the MR environment. In the SRs-Congruent condition, the physical and virtual worlds share the same contingency relationships between actions and outcomes. We conjecture that this alignment makes it easier for users to leverage their existing world model to form a unified internal representation that accurately describes both realities they simultaneously experience. The formation of an internal world model of the MR environment in the SRs-Congruent condition could also explain the significantly greater sense of presence it yielded—by keeping track of the world states and casual relations between states transitions, users could feel more "being there" [49] in the MR environment. In contrast, in the other two conditions, the virtual world operates on a different set of contingency relations from the real world. This discrepancy can hinder the formation of a unified world model for the MR environment. An inherent challenge imposed by the nature of MR is the how to naturally blend the virtual and physical realities that users simultaneously experience. Our results provide primary evidence that applying SRs-congruent mappings to MR interactions could better blend the two realities and facilitate the formation of a unified world model that users can leverage to efficiently describe both realities, improving performance, user learning, and perceived presence. Since SRs originate in the physical world, applying SRs-congruent mappings to VR interactions could help users adapt their physical world model to navigate the virtual environment, enhancing performance and learning. This presents an intriguing avenue for future research.

6.3 Strengthened Ownership

After using the SRs-Congruent method, six participants reported feeling as if they actually owned this superpower when answering the question "What do you like about this time control method", without being explicitly asked about their sense of ownership. This did not happen for the other two methods. Four of the participants who reported enhanced perceived ownership also reported experiencing perceptual illusions mainly related to the feelings of time being transformed into something tangible and grab-able. We believe such perceptual illusions could originate from their strengthened sense of ownership over the superpower—they really felt like they were "doing something to the time" (P9), which led to its transformation. Apart from users' subjective feedback, the strengthened sense of ownership enabled by the SRs-Congruent method is also evident from significantly higher ratings on perceived agency and presence compared to other two methods.

6.4 SRs-Constrained Superpower Design Pathway

This study has empirically demonstrated multiple benefits of using SRs as constraints in the design of action-outcome mappings of one augmented superpower. We encourage further comparative studies across different augmented superpowers and designers, while also considering different input modalities beyond gestures (e.g., gaze and head movement).

To facilitate such explorations in the future, we proposed an SRs-Constrained Superpower Design Pathway (see Fig. 10). We recommend that designers (1) generate an accurate description of the expected superpower outcome(s); (2) identify the 'Outcome' schema—the most relevant image schema(s) that could best represent the expected outcome; (3) based on lawful relations between different schemas, identify the 'Action' schema—the one that is most likely to result in the 'Outcome' schema; (4) determine physical actions congruent with the 'Action' schema; and (5) choose the most appropriate physical action, taking into account practical considerations like distinguishing it from null gestures. For example, for an outcome of "blending two effects into one", the most relevant 'Outcome' schema are Merging, the 'Action' schemas that are most likely to yield the Merging schema is Compulsion and Near. Therefore, one gesture that can properly represent Compulsion and Near can be a bimanual gesture with both hands moving closer until contact.

7 Conclusion

This work empirically demonstrated the benefits of sensorimotor regularities as design constraints for action-outcome mappings of augmented superpower in MR. This is supported by the evaluation between a Mid-Air Button, SRs-Incongruent gestures and SRs-Congruent gestures in the context of temporal manipulation, in a within-subject user study with 18 participants. Using the well-established Mid-Air Button as a baseline, the results showed that SRs-Congruent gestures led to significant improvements in performance (both in terms of success rate and time to achieve each success), significantly reduced overall workload, as well as higher perceived agency and presence. Additionally, SRs-Congruent gestures resulted in faster initial learning and earlier achievement of near-optimal performance. On the other hand, SRs-Incongruent gestures, despite being a more embodied and less haptic-dependent interaction technique compared to the Mid-Air Button, did not provide any of these benefits. Additionally, SRs-Congruent gestures were overall the most preferred across all participants, followed by SRs-Incongruent gestures, and thereafter Mid-Air Button. Participants noted that the SRs-Incongruent gestures had a perceived arbitrary mapping between actions and outcomes, making them difficult to remember. In contrast, they appreciated the SRs-Congruent gestures for their natural and logical mapping, ease of learning and use, and the sense of ownership they fostered.

Recognizing the performance challenges of mid-air buttons, we used a standard MRTK3 mid-air button and attempted to improve its usability through Fitts' law-based adjustments, such as increasing its size and optimizing its position. However, these enhancements still did not fully resolve the error issues. Future research should explore better-performing mid-air button designs, such as introducing delays for consecutive clicks, and compare user performance and experience enabled by mid-air buttons with improved usability with SRs-Incongruent and SRs-Congruent methods.

It is important to emphasize that the benefits of using sensorimotor regularities as design constraints were observed in one possible implementation of SRs for one superpower interaction. Future research should further validate these findings by replicating similar experiments, applying different SRs to different types of superpower interactions. To support such validations, this paper demonstrated a SRs-Constrained Superpower Design Pathway (see Fig. 10), which can assist future studies in exploring the wider applicability of SRs as action-outcome mappings constraints.

We conclude that sensorimotor regularities offer a promising method for designers and developers to create highly usable, easy-to-learn action-outcome mappings that resonate with the mental models of a wide range of users. Beyond the context of augmented superpowers, we believe that this method could also yield similar benefits for any types of MR interactions with clear states of action and outcome.

Figure Credits

Fig. 4 includes the following 3D models from Sketchfab, used to build the MR game environment. Free Standard License: Cat Plush[3]. CC BY 4.0 license: Wooden Crate[4], Cute Spino[5], Low Poly Clouds[6].

[3] https://sketchfab.com/3d-models/cat-plush-no-texture-86857aec65d741938bdcf56bc1d01981

[4] https://sketchfab.com/3d-models/wooden-crate-eac1f95a64f5400ebf67d2129b1414d8

[5] https://sketchfab.com/3d-models/cute-spino-701f055847224f7c928787733254ae2c

[6] https://sketchfab.com/3d-models/low-poly-clouds-f653f67c8ea34b7b8f3dae162b7fff0f

REFERENCES

[1] P. Abtahi. From illusions to beyond-real interactions in virtual reality. In *Adjunct Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21 Adjunct, 5 pages, p. 153–157. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3474349.3477586 1

[2] P. Abtahi, S. Q. Hough, J. A. Landay, and S. Follmer. Beyond being real: A sensorimotor control perspective on interactions in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, article no. 358, 17 pages. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517706 1, 2

[3] I. Aslan, T. Schmidt, J. Woehrle, L. Vogel, and E. André. Pen + mid-air gestures: Eliciting contextual gestures. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ICMI '18, 10 pages, p. 135–144. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3242969.3242979 4

[4] R. Baillargeon. *The Acquisition of Physical Knowledge in Infancy: A Summary in Eight Lessons*, chap. 3, pp. 47–83. John Wiley & Sons, Ltd, 2002. doi: 10.1002/9780470996652.ch3 3

[5] R. Baillargeon and S. hua Wang. Event categorization in infancy. *Trends in Cognitive Sciences*, 6(2):85–93, 2002. doi: 10.1016/S1364-6613(00)01836-2 3

[6] L. Boroditsky. Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75(1):1–28, 2000. doi: 10.1016/S0010-0277(99)00073-6 3

[7] Y. F. Cheng, H. Yin, Y. Yan, J. Gugenheimer, and D. Lindlbauer. Towards understanding diminished reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, article no. 549, 16 pages. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517452 2, 3

[8] H. Cho, M. L. Komar, and D. Lindlbauer. Realityreplay: Detecting and replaying temporal changes in situ using mixed reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(3), article no. 90, 25 pages, Sept. 2023. doi: 10.1145/3610888 2

[9] N. Drummond and Y. Niv. Model-based decision making and model-free learning. *Current Biology*, 30(15):R860–R865, 2020. doi: 10.1016/j.cub.2020.06.051 8

[10] J. Dudley, H. Benko, D. Wigdor, and P. O. Kristensson. Performance envelopes of virtual keyboard text input strategies in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 289–300, 2019. doi: 10.1109/ISMAR.2019.00027 7

[11] A. Faeth and C. Harding. Emergent effects in multimodal feedback from virtual buttons. *ACM Trans. Comput.-Hum. Interact.*, 21(1), article no. 3, 23 pages, Feb. 2014. doi: 10.1145/2535923 7

[12] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954. doi: 10.1037/h0055392 4

[13] N. Freeman, S. Lloyd, and C. Sinha. Infant search tasks reveal early concepts of containment and canonical usage of objects. *Cognition*, 8(3):243–262, 1980. doi: 10.1016/0010-0277(80)90007-4 3

[14] D. Frye, P. Rawling, C. Moore, and I. Myers. Object–person discrimination and communication at 3 and 10 months. *Developmental Psychology*, 19(3):303–309, 1983. doi: 10.1037/0012-1649.19.3.303 3

[15] J. GIBBS, RAYMOND W., D. A. BEITEL, M. HARRINGTON, and P. E. SANDERS. Taking a stand on the meanings of stand: Bodily experience as motivation for polysemy. *Journal of Semantics*, 11(4):231–251, 11 1994. doi: 10.1093/jos/11.4.231 3

[16] R. W. Gibbs. The psychological status of image schemas. *From perception to meaning: Image schemas in cognitive linguistics*, 29:113–136, 2005. doi: doi:10.1515/9783110197532.2.113 3

[17] M. D. Gross. *Design as exploring constraints*. PhD thesis, Massachusetts Institute of Technology, 1985. 3

[18] P. Haggard. Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4):196–207, 2017. doi: 10.1038/nrn.2017.14 3

[19] A. M. Haith and J. W. Krakauer. Model-based and model-free mechanisms of human motor learning. *Advances in Experimental Medicine and Biology*, 782:1–21, 2013. doi: 10.1007/978-1-4614-5465-6_1 6, 8

[20] C. Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1(2):262–271, 05 1992. doi: 10.1162/pres.1992.1.2.262 2

[21] J. Hurtienne, S. Huber, and C. Baur. Supporting user interface design with image schemas: The iscat database as a research tool. In *CEUR Workshop Proceedngs (EUR-WS. org)*. Jönköping Sweden, 2022. 2, 4

[22] J. Hurtienne, C. Stößel, C. Sturm, A. Maus, M. Rötting, P. Langdon, and J. Clarkson. Physical gestures for abstract concepts: Inclusive design with primary metaphors. *Interacting with Computers*, 22(6):475–484, 2010. Special Issue on Inclusion and Interaction: Designing Interaction for Inclusive Populations. doi: 10.1016/j.intcom.2010.08.009 3

[23] A. Irlitti, M. Latifoglu, T. Hoang, B. V. Syiem, and F. Vetere. Volumetric hybrid workspaces: Interactions with objects in remote and co-located telepresence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, article no. 802, 16 pages. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642814 2

[24] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 10 pages, p. 201–210. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357089 1, 2

[25] M. Johnson. *The Body in the Mind. The Bodily Basis of Meaning, Reason and Imagination*. Chicago University Press, Chicago, 1987. 3

[26] W. Kim and S. Xiong. Pseudo-haptic button for improving user experience of mid-air interaction in VR. *International Journal of Human-Computer Studies*, 168:102907, 2022. doi: 10.1016/j.ijhcs.2022.102907 7

[27] P. Knierim, T. Kosch, G. LaBorwit, and A. Schmidt. Altering the speed of reality? exploring visual slow-motion to amplify human perception using augmented reality. In *Proceedings of the Augmented Humans International Conference*, AHs '20, article no. 2, 5 pages. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3384657.3384659 2, 3

[28] A. M. Leslie. The perception of causality in infants. *Perception*, 11(2):173–186, 1982. PMID: 7155770. doi: 10.1068/p110173 3

[29] A. M. Leslie. The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz, ed., *Thought Without Language*, pp. 185–210. Clarendon Press/Oxford University Press, 1988. A Fyssen Foundation Symposium. 3

[30] J. Li, N. Crilly, and P. O. Kristensson. Guiding the design of inclusive interactive systems: Do younger and older adults use the same image-schematic metaphors? *ACM Trans. Comput.-Hum. Interact.*, 31(4), article no. 47, 44 pages, Sept. 2024. doi: 10.1145/3648618 3

[31] J. Li and P. O. Kristensson. On the benefits of image-schematic metaphors when designing mixed reality systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, article no. 953, 20 pages. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642925 3

[32] K. Lilija, H. Pohl, and K. Hornbæk. Who put that there? temporal navigation of spatial recordings by direct manipulation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 11 pages, p. 1–11. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376604 2

[33] D. Lindlbauer and A. D. Wilson. Remixed reality: Manipulating space and time in augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 13 pages, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173703 2, 3

[34] S. Lloyd, C. Sinha, and N. Freeman. Spatial reference systems and the canonicality effect in infant search. *Journal of Experimental Child Psychology*, 32(1):1–10, 1981. doi: 10.1016/0022-0965(81)90088-6 3

[35] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):JCMC321, 09 1997. doi: 10.1111/j.1083-6101.1997.tb00072.x 1

[36] J. M. Mandler. How to build a baby: On the development of an accessible representational system. *Cognitive Development*, 3(2):113–136, 1988. doi: 10.1016/0885-2014(88)90015-9 2, 3

[37] J. M. Mandler. How to build a baby: II. conceptual primitives. *Psychological Review*, 99(4):587–604, Oct. 1992. doi: 10.1037/0033-295x.99.4.587 2, 3

[38] J. M. Mandler. *How to build a baby: III. Image schemas and the transition to verbal thought*, pp. 137–164. De Gruyter Mouton, Berlin, New York, 2005. doi: doi:10.1515/9783110197532.2.137 2, 3

[39] D. Mauney, J. Howarth, A. Wirtanen, and M. Capra. Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, 6 pages, p. 4015–4020. Association for Computing

Machinery, New York, NY, USA, 2010. doi: 10.1145/1753846.1754095 4

[40] A. E. Mostafa, E. Sharlin, and M. C. Sousa. Poster: Superhumans: A 3dui design metaphor. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 143–144, 2014. doi: 10.1109/3DUI.2014.6798861 1, 2

[41] R. Neuhaus, R. Ringfort-Felner, S. Sadeghian, and M. Hassenzahl. To mimic reality or to go beyond? "superpowers" in virtual reality, the experience of augmentation and its consequences. *International Journal of Human-Computer Studies*, 181:103165, 2024. doi: 10.1016/j.ijhcs.2023.103165 2

[42] N. Nouri, V. Biener, and J. Grubert. Experiences with off-the-shelf solutions for xr-supported knowledge work. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 670–672, 2024. doi: 10.1109/VRW62533.2024.00132 1

[43] L. Pavanatto, V. Biener, J. Chandran, S. Kalamkar, F. Lu, J. J. Dudley, J. Hu, G. N. Ramirez-Saffy, P. O. Kristensson, A. Giovannelli, L. Schlueter, J. Müller, J. Grubert, and D. A. Bowman. Working in extended reality in the wild: Worker and bystander experiences of xr virtual displays in real-world settings, 2024. 1

[44] T. Pham, J. Vermeulen, A. Tang, and L. MacDonald Vermeulen. Scale impacts elicited gestures for manipulating holograms: Implications for ar gesture design. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, 14 pages, p. 227–240. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3196709.3196719 4

[45] J. Reis, H. M. Schambra, L. G. Cohen, E. R. Buch, B. Fritsch, E. Zarahn, P. A. Celnik, and J. W. Krakauer. Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences*, 106(5):1590–1595, 2009. doi: 10.1073/pnas.0805413106 6

[46] D. Richardson, M. Spivey, S. Edelman, and A. Naples. Language is spatial: Experimental evidence for image schemas of concrete and abstract verbs. *Twentythird Meeting of the Cognitive Science Society*, pp. 873–878, 01 2001. 3

[47] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x 4

[48] L. Shmuelof, J. W. Krakauer, and P. Mazzoni. How is a motor skill learned? change and invariance at the levels of task success and trajectory control. *Journal of Neurophysiology*, 108(2):578–594, July 2012. doi: 10.1152/jn.00856.2011 6

[49] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009. doi: 10.1098/rstb.2009.0138 2, 9

[50] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 12 1997. doi: 10.1162/pres.1997.6.6.603 2

[51] Z. Song, J. J. Dudley, and P. O. Kristensson. Hotgestures: Complementing command selection and use with delimiter-free gesture-based shortcuts in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(11):4600–4610, 2023. doi: 10.1109/TVCG.2023.3320257 1

[52] L. Talmy. Force dynamics in language and cognition. *Cognitive Science*, 12(1):49–100, 1988. doi: 10.1016/0364-0213(88)90008-0 2

[53] A. Tapal, E. Oren, R. Dar, and B. Eitam. The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology*, 8:1552, 2017. doi: 10.3389/fpsyg.2017.01552 5, 8

[54] R. Tscharn. *Innovative and age-inclusive interaction design with image-schematic metaphors*. PhD thesis, Bayerische Julius-Maximilians-Universitaet Wuerzburg (Germany), 2018. 3

[55] C. Velasco, F. Barbosa Escobar, O. Petit, and Q. J. Wang. Impossible (food) experiences in extended reality. *Frontiers in Computer Science*, 3:716846, 2021. doi: 10.3389/fcomp.2021.716846 1

[56] R. Villa, E. Tidoni, G. Porciello, and S. M. Aglioti. Violation of expectations about movement and goal achievement leads to sense of agency reduction. *Experimental Brain Research*, 236(7):2123–2135, 2018. doi: 10.1007/s00221-018-5286-3 3

[57] W. Willett, B. A. Aseniero, S. Carpendale, P. Dragicevic, Y. Jansen, L. Oehlberg, and P. Isenberg. Perception! Immersion! Empowerment! Superpowers as Inspiration for Visualization . *IEEE Transactions on Visualization & Computer Graphics*, 28(01):22–32, Jan. 2022. doi: 10.1109/TVCG.2021.3114844 1, 2

[58] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, 16 pages, June 1998. doi: 10.1162/105474698565686 5, 8

[59] A. L. Woodward and J. A. Sommerville. Twelve-month-old infants interpret action in context. *Psychological Science*, 11(1):73–77, 2000. 3

[60] X. Zhao, T. Niikura, and T. Komuro. Evaluation of visuo-haptic feedback in a 3d touch panel interface. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, 6 pages, p. 299–304. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2669485.2669536 7