Modelling Visuo-Haptic Perception Change in Size Estimation Tasks

Jian Zhang
University of Melbourne
Melbourne, VIC, Australia
jianzhang10@student.unimelb.edu.au

Wafa Johal
University of Melbourne
Melbourne, VIC, Australia
wafa.johal@unimelb.edu.au

Jarrod Knibbe
The University of Queensland
St Lucia, QLD, Australia
j.knibbe@uq.edu.au

Abstract

Tangible interactions involve multiple sensory cues, enabling the accurate perception of object properties, such as size. Research has shown, however, that if we decouple these cues (for example, by altering the visual cue), then the resulting discrepancies present new opportunities for interactions. Perception over time though, not only relies on momentary sensory cues, but also on a priori beliefs about the object, implying a continuing update cycle. This cycle is poorly understood and its impact on interaction remains unknown. We study (N=80) visuo-haptic perception of size over time and (a) reveal how perception drifts, (b) examine the effects of visual priming and dead-reckoning, and (c) present a model of visuo-haptic perception as a cyclical, self-adjusting system. Our work has a direct impact on illusory perception in VR, but also sheds light on how our visual and haptic systems cooperate and diverge.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; **Empirical studies in HCI**; **Virtual reality**; **Haptic devices**.

Keywords

Perception, Visuo-Haptics, Proprioception, Illusion, Virtual Reality, Acclimation

1 Introduction

Humans integrate cues across multiple sensory modalities in order to understand the world around them [29]. When you walk into a new room, visual cues give you the rough spatial layout, thermal cues let you know the heating is on, and audio cues might inform you that there is machinery around the corner. Together, this information combines into a more detailed, perceptual model of the space.

Not only is integration important for our understanding of the world, but also for our ability to act within it. Visual cues can provide information about an object's location and size, for example, but accurate and controlled interaction also requires detailed information about weight and mass distribution (among many other things), requiring, at least, the addition of tactition and kinaesthesia [41, 43].

Research has shown that there are exciting opportunities for new interactions and experiences when designers actively introduce conflicting sensory cues. For example, by decoupling visual and haptic feedback, the rubber hand illusion can make participants believe that a fake rubber hand is their actual hand [8]. This works by simultaneously stroking the participant's real hand and a dislocated rubber hand, thereby matching the tactile stimuli while introducing an offset error through both the visual and proprioceptive stimuli. For interaction, these visuo-haptic illusions have proven popular in virtual reality, for example, to alter users' reaches through haptic retargeting [1], or to alter their grasps through resized grasping [4].

These perceptual illusions, however, only work within a fairly limited scope. As a result, much research effort is given to investigating and understanding perceptual limits for these interactions (e.g., [1, 4, 52]). For example, haptic retargeting illusions have been shown to work up to approx. 15° around an arc to the left/right of a virtual object [14], and grasping illusions work up to approx. 17% just noticeable difference for 6cm objects [52].

Work in sensorimotor neuroscience and cognitive psychology tells us about multiple important mechanisms of perception that come to inform these limits. For example, we know that the perceived reliability of sensory information changes [19], and that multi-sensory cues are integrated according to functions similar to maximum-likelihood estimators [18]. We also know that sensory memory impacts our estimations about objects (i.e., the results of our integrated perception) through processes such as cue integration theory [21]. And research has also found closed-loop processes between perception and interaction in motor adaptation [18, 49], indicating a changing process of cognition and behaviour.

Existing work in HCI has begun to discuss how different patterns of movement and interaction may impact the reliability of different perceptual channels (e.g., [14]), but little attention has been given to sensory integration and memorisation over time. Previous explorations of illusory limits were estimated either within a short time duration of use (e.g. approximately 10 minutes for each physical object in the resized grasping estimation [4]) or in an experimental context where the participants have almost no knowledge about the appearance and mechanism of the haptic proxies [4, 26], drastically impacting the available prior sensory information. We do not know, however, how continued illusory perception (i.e., perception of mismatched multi-sensory cues) compounds – *do visuo-haptic illusions*

continue to work over time? We also do not know how momentary sensory realignment impacts object perception – *what happens to an illusion when a user sees the actual proxy device?* The answers are vitally important in prolonged haptic interactions and meaningful for understanding human cognition.

We study visuo-haptic perceptual conflict resolution, specifically considering the effects of *time* (as acclimation, through prolonged use) and *knowledge of the haptic device* (through select visual priming). We conduct one-hour user studies in VR, during which we vary either the visual cue (with a passive haptic device) or both the visual and haptic cues (with an active, shape-changing device). We designed a device that dynamically alters its size between 6cm and 8cm. We ran multiple forced choice tasks to estimate the perception, alongside acclimation games to simulate continued exposure and practice with the device. We controlled the participant's view of the device, either never revealing it (non-priming), or revealing the true device (true visual priming) or a false device of a different size (misleading priming) at certain times. We recruited 80 participants and divided them into 4 groups with conditions of (1) non-priming fixed-size device, (2) non-priming size-changing device, (3) fixed-size device with correct visual priming, and (4) fixed-size device with a misleading visual priming.

Based on our empirical results, we make the following claims:

- Perception drifts over time and the size is increasingly over-estimated for passive objects, following a curve to a local asymptote
- Active, shape-changing devices skew subsequent size perception differently to passive objects. The perceived size drifts to be underestimated when the physical size actively shrinks.
- Visual priming sets a more accurate initial perception and reduces perceptual drift across a whole experience
- Intermittent visual priming serves as a form of dead-reckoning, resetting perception towards that prime
- Sensory integration can be modelled as a first-order control system, wherein the perceptive discrimination works as an amplifier and the feedback loop is influenced by confidence in the perception

These claims (as contributions) have important implications for tangible interactions. First, they begin to describe what we should expect if we deploy visuo-haptic illusions, such as resized grasping [4], over time. Second, they reveal how selective priming and dead-reckoning may enable designers to skew or reset perception at key moments. And third, they provide early insights into what users come to believe about tangible interactive objects when their vision is directed away from it.

2 Related Work

Mismatched visual and haptic cues can easily occur in scenarios with screens and head-mounted displays. This can happen accidentally, such as a result of tracking errors and latency causing spatial misalignments between visual cues and their physical, haptic counterparts. But it can also be leveraged purposefully, through the design of specific haptic devices and visuals, that induce mismatches for novel experiences and opportunities. Integral to this opportunity, however, is an understanding of how visuo-haptic cues are perceived and integrated by users.

We review recent literature on haptic devices and highlight how visuo-haptic illusions are being leveraged and understood.

2.1 Haptic Devices

Efforts to deliver high-resolution haptic feedback have led to the development of a variety of tangible devices targeting immersive interaction, especially in VR. Recent advances have largely concentrated on wearable and handheld formats. Representative examples include wearable systems for grasping rigid forms [11, 12], objects changing geometry features [26], and complex shapes via multi degree-of-freedom mechanisms [33]. However, these devices often face challenges of being complex, cumbersome, and costly.

By contrast, grounded haptic systems (particularly those of encounter-type interfaces) can offer feedback only when necessary, leaving users hands free. Commercial products such as the Touch[1] and Omega[2], as well as research prototypes like inFORM [22], ShapeShift [45], and REACH+[23], are examples of such devices enabling different interactions. These platforms are capable of delivering highly detailed feedback but similar to the wearable devices, their complexity, cost, and infrastructural requirements limit their accessibility for everyday interaction.

As a result, research has also been exploring passive haptic techniques with simple designs or daily accessible objects. For example, by leveraging only simple geometric building blocks, HapTwist demonstrated a range of complex haptic objects [55]). Another popular avenue has been to use everyday objects, in concert with visuo-haptic illusions to, e.g., haptically render objects of different sizes [4] and in different locations [1].

2.2 Visuo-Haptic Illusions

Usually, visual and haptic stimuli match and integrate with each other to form a perception of reality in interactions such as touching and grasping. In scenarios such as extended reality, however, tangible interaction often features a decouple and recouple process of the visual and haptic cues by blocking the real vision and rendering a virtual one. These visuo-haptic discrepancies, if tightly controlled, can create new experiential and interactive opportunities. The rubber hand illusion [8], for example, introduces a spatial discrepancy between the user's real hand and a fake rubber hand that, combined with synchronous visual and tactile cues, allows the user to perceive body-ownership of the fake hand.

Research in this space continues to develop and expand the boundaries of these illusions. Much of the work on visuo-haptic illusions has focused on two primary purposes: illusions of *where* the object is, or *what* the object is.

Illusions of *where* have been heavily influenced by redirected touching [35] and haptic retargeting [1]. These illusions seek to guide the user's physical hand towards a proxy object that is spatially decoupled from its virtual counterpart. Examples have demonstrated redirect controller buttons [25, 50], enabled users to grab objects placed around them [14], and attempted to retarget random, unscripted reaches [13]. From work on these illusions, we have come to understand the spatial *haptic coverage* of a physical object

---

– the area within which it can provide haptic feedback for virtual objects [15].

Illusions of *what* the object is largely explore the extent to which one physical object can feel like another. In addition to the devices that can directly alter the physical size stimuli (e.g. CLAW [12], X-Rings [26], etc.), applications of illusions aim to convince users they are interacting with an object with one property (for example, a heavy hammer), while they in fact interact with an object with a different property or, at least, with an object with a different magnitude of that property (e.g., a lightweight bottle).

In a visuo-haptic system, simple physical devices can express various shapes with edges, curves, and surfaces [2, 3]. The method was further developed by Zhao et al. [54] to extend haptic retargeting to complex, arbitrary shapes. Bickmann et al. [7] showed that haptic feedback can be created by illusion without grasping any physical prop and developed a haptic illusion glove that can provide haptic feedback for various virtual objects such as a cup, a hammer, and a water can. Yang et al. [48] demonstrated that the illusion of size change could be induced not only through direct hand interaction, but also when manipulating virtual tools such as chopsticks. In many of these applications, the illusion-based tangible interaction often relies on visual distortion to alter perception. Kim et al. [34] introduced a fixed-size haptic controller that uses finger repositioning to create the illusion of dynamic size change. This allows users to perceive changes in object size with proper visual feedback. Furthermore, illusions have been used to simulate factors in interaction such as geometry [46], weight [51], force feedback [39], stiffness [44] and texture [9]. The research community has developed the approaches described above to either induce or exploit visuo-haptic illusions. Consequently, the underlying mechanisms and characteristics of these illusions (e.g., their limits) have also attracted much attention.

2.3 The Mechanism of Visuo-haptic Illusions and Adaptation

Modern theories describe our cognition and understanding of the world around us as a predictive process. That is, our brains make predictions about the world around us, based on expectations, and our previous experience and knowledge. These predictions are then compared to the information we receive through sensory cues, and subsequently updated and corrected in a closed-loop adaptation process [16, 29].

In cognitive science, research explains this adaptation as an active process where humans adjust their perception and behaviour to improve each other. Sensory inputs, while being noisy or unreliable sometimes, are processed by the brain in a fashion similar to a maximum-likelihood integrator, which can variously weight and attenuate different cues to update our perception [6, 18, 36].

Based on this adaptation and integration of multi-sensory cues, several studies have attempted to model the patterns of the visuo-haptic illusions and accordingly proposed hypothetical theories. For the size perception, Zhang et al. [53] has described a simple linear pattern of perceived size increasing with the virtual size in augmented reality. Similarly, Bergström et al. [4] built a model indicating that illusions are more easily detected when objects are larger. Zhang et al. [52] further built a multi-dimensional model of the interaction of size and angle in visuo-haptic perception. These models focused on the impacts of the physical properties and human biological features (e.g. hand sizes) on the visuo-haptic illusions.

Furthermore, more complex models have been proposed in related tasks such as haptic retargeting. Gonzalez et al. [24] recently proposed a control process with an internal motion controller updating the hand state in a closed-loop system in haptic retargeting. Based on this model, Lebrun [38] further adapted a stochastic optimal feedback control model to describe the relationship between natural tendency with illusions when performing haptic retargeting. These models apply control theories in human perception and motor control, attempting to interpret the visuo-haptic perception from the perspective of cognitive science and neurobiology.

Many of these models consider visuo-haptic illusions and integration as static. As these cues are prone to error and noise, however, this updating process is continuous [30], and human activities, in general, involve a large amount of learning and adaptation over time, such as to improve motor control [37]. Similar to a control system with feedback loops, errors are corrected and the larger the errors are, the larger the corrections are (e.g., in motor learning). As part of motor control, perception based on multisensory inputs plays an important role in the learning process. Specifically in tangible interactions, cases such as rubber hand illusion [8] have shown the effect of an acclimation process in coupling sensory stimulus and creating the illusions.

Previous studies indicate the existence of an adaptation process in motor control and perception [30, 31]. While the process of changing perception has been influencing experiences and immersion in tangible interactions, this cognitive process of fundamental properties such as size perception, however, has not been estimated or explored.

3 Methods

We ran a study to examine the impact of time and visual priming on visuo-haptic illusions. Specifically, through four conditions, we estimated (1) the impact of prolonged interaction on perception of fixed-size objects, (2) the impact of prolonged interaction on perception of size-changing devices, (3) the impact of correct visual priming and (4) the impact of incorrect visual priming on size perception. We designed a three-phase study, including (a) estimation tasks, (b) breaks, and (c) an acclimation game. The estimation task captured participants' size perception through a forced choice task, where participants were presented with haptic and/or visual stimuli and asked to choose between two alternative size options in VR. The break phase included filling out a NASA Task Load Index questionnaire on paper. The acclimation game, in VR, was designed to ensure prolonged haptic exposure to the device, making participants grasp, hold, and release the device repeatedly.

3.1 Apparatus

3.1.1 Haptic Proxy Design. We prototyped a device that can change its size (specifically, the width for grasping) between 6 cm and 8 cm, as shown in Figure 1. An Arduino UNO board controls a FT5330M aluminium servo to rotate the gear in the centre of the device, and drives the gear racks to expand/contract the sides of the cube. This servo offers up to $35kg/cm$ of torque, ensuring that the cube will

(a) Mechanical design   (b) Contraction mode: 6 cm   (c) Expansion mode: 8 cm   (d) Photo of the device in contraction mode
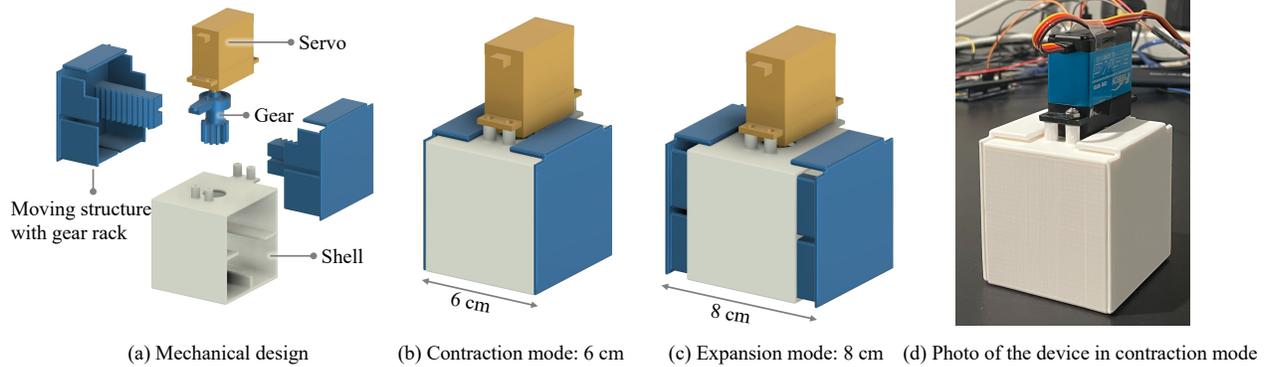
Figure 1: A device changes its size (width of grasp) between 6 cm and 8 cm. (a) Breakdown mechanical design of the device. The motor and gear in the centre drives the sides of the cube via the gear racks to expand/contract. (b) The contraction mode of the device with the appearance of a cube with the width of 6 cm. (c) The expansion mode of the device with the width of 8 cm. (d) The photo of the prototype (in contraction mode).

achieve the specified size (i.e., it is not easily overpowered by the participant). It takes 0.9±0.1 seconds to change size between 6 cm and 8 cm with the error less than 0.05 cm, and the speed change when applying a grasp force on the device is negligible according to technique estimation.

The device weighs 162±1 grams (including Optitrack markers). The device was fully 3D printed with polyethylene terephthalate glycol (PETG), except for the servo, electronic components, and fastening hardware. The device was connected with long jumper wires (approximately 40 cm), to ensure a free, unrestricted movement. When contracted, the device is a 6.0±0.05 cm cube.

3.1.2 Virtual Scene and Tracking System. The VR scene was built in Unity 2022 on a laptop PC (13th Gen Intel i9-13900HK 2.60GHz, 32.0GB RAM, NVIDIA GeForce RTX 4090 GPU, Windows 11 Pro). The scene was presented to the participants via a Meta Quest 3 head-mounted device (HMD). We used an Optitrack motion capture system for tracking the device, the hand movement, and the headset. There were 7 Prime 13W cameras, running at 240Hz, placed around the table where the participants performed the tasks (see Figure 3). The markers of the motion capture system were mounted on the participant's thumb and index finger, the device for grasping, the HMD, and the table. Markers mounted on the tracked objects did not interfere with the participants' range of motion or object interactions.

3.1.3 Finger Tips Rendering. Although only the thumb and index finger of the participant's right hand were tracked and rendered with capsule-shaped objects in the virtual scene, they were asked to hold the device with all the fingers of their right hand, also known as Large Diameter grasp in the GRASP taxonomy [20]. As our participants were grasping parallel faces of a cube, we assume their fingers are in a line opposing their thumb. To prevent tracking marker occlusion and to reduce confounding effects of hand rendering (e.g., [32, 47]), we use only a simplified and reliable rendering scheme to assist with targeting.

Similar to previous studies [4, 52], we applied a resized grasping algorithm between the fingers in real-time. The distance of the

index finger and thumb to be rendered in the virtual scene D_v was calculated based on the scaling ratio between the widths of the virtual object and the physical device S_v/S_p, where S_v is the size of the virtual object and S_p is the size of the physical object (note that the virtual object size does not always equal the physical object size, as mismatching sizes were presented for the estimation tasks and for a random visual stimuli in the acclimation game). The distance D_v between the virtual index finger and thumb was resized by the method:

$$D_v = D_p \times (S_v/S_p) \tag{1}$$

where D_p is the distance between the index finger and the thumb in the real world. Through resizing, the participant's fingers were always aligned to the virtual cube when they grasped the physical device.

3.2 Experimental Procedure

The procedure is shown in Figure 2. To evaluate the influence of time and visual priming on both fixed-size (*passive*) devices and size-changing (*active*) devices, the study was designed into 4 conditions.

3.2.1 Estimation Task (6 minutes). The Estimation Task (ET) was carried out by presenting forced choice questions to the participants and collecting their answers, similar to the task settings in previous studies [4, 17, 52]. The participants were asked to grasp the device with all the fingers of their right hands, compare the physical size in their right hands and the virtual size they saw in the VR scene, and choose between "virtual smaller" (the virtual size is smaller than the physical size they perceive in the hands) and "virtual larger" (the virtual size is larger than the physical size they perceive in the hands) within 6 seconds (visually indicated by a time bar to the participants). They input their answer with the Meta Quest 3 controller in their left hand. After answering the question, the participants were asked to hold the device in the air and wait for 2 seconds. During this wait, the virtual fingers and objects were all invisible, and, in the size-changing condition, the device would change size. At this point, the virtual object and the fingers would become visible again. Subsequently, while still holding the device,
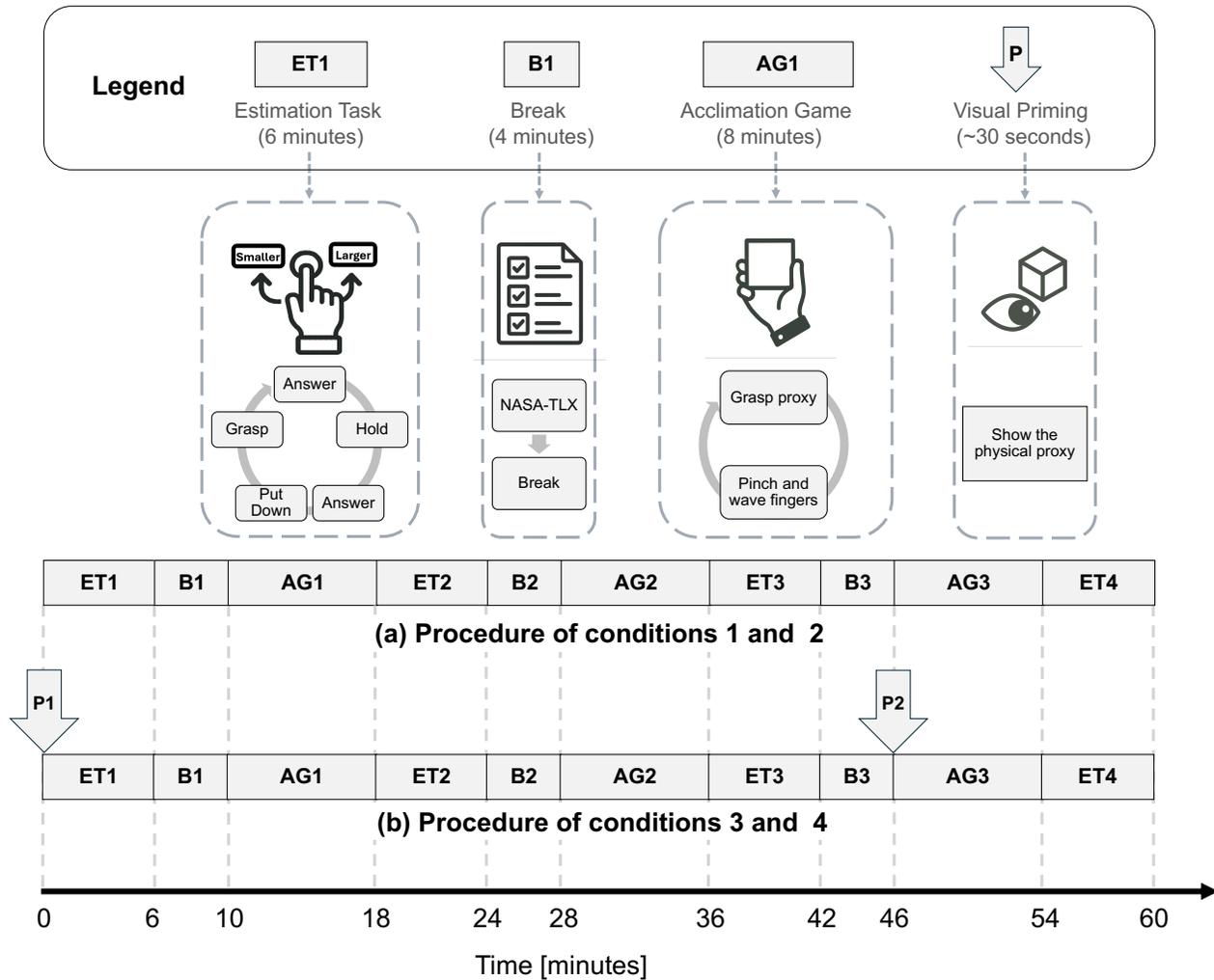
Figure 2: The procedure includes four 6-minute estimation task blocks (an initial one and three after each acclimation game block), three 4-minute break blocks (before each acclimation game block) and three 8-minute acclimation game blocks. The overall designed study time is 60 minutes. The figure shows (a) procedure of conditions 1 and 2 without any priming and (b) procedure of condition 3 and 4 with visual priming at the beginning of the study and the beginning of the last (third) acclimation game block.

the participants then compared a new virtual object in their hands with the physical one. After answering this second question, the participants were asked to put the device down on the table, rest their hand in front of their right side, and wait for the next group of questions. During this period, the virtual object was not visible. Throughout the whole study, participants would lift - answer - wait - answer - put down. This allowed us to examine the immediate impact of size-change (during the wait period) on perception. We use the same procedure when the physical size does not change, to avoid introducing time confounds between conditions.

Throughout the estimation block, participants compared physical objects against virtual objects between ± 2.0 cm in 0.5 cm steps (e.g., the 6 cm physical object was compared against nine virtual objects between 4.0 cm and 8.0cm, and the 8 cm physical object was

compared against nine virtual objects between 6 cm and 10 cm). In the size-changing condition, the physical object would change from 6 cm to 8 cm (expand) and from 8 cm to 6 cm (contract). Each physical size was compared against the nine virtual sizes, for 18 total questions. In the fixed-size condition, always 6.0 cm, to keep the duration the same, each virtual size was presented twice (1 size × 9 virtual sizes × 2 times = 18 questions).

In the fixed-size condition, we only analyse the results of the first question in each lift, to understand the impact of grasp on size perception. We believe the second response is likely biased by the first answer. Additionally, as the device is not released and grasped again, the participants receive less haptic information for consideration in their second response (in contrast to receiving new, conflicting visual information). As such, we chose not to analyse
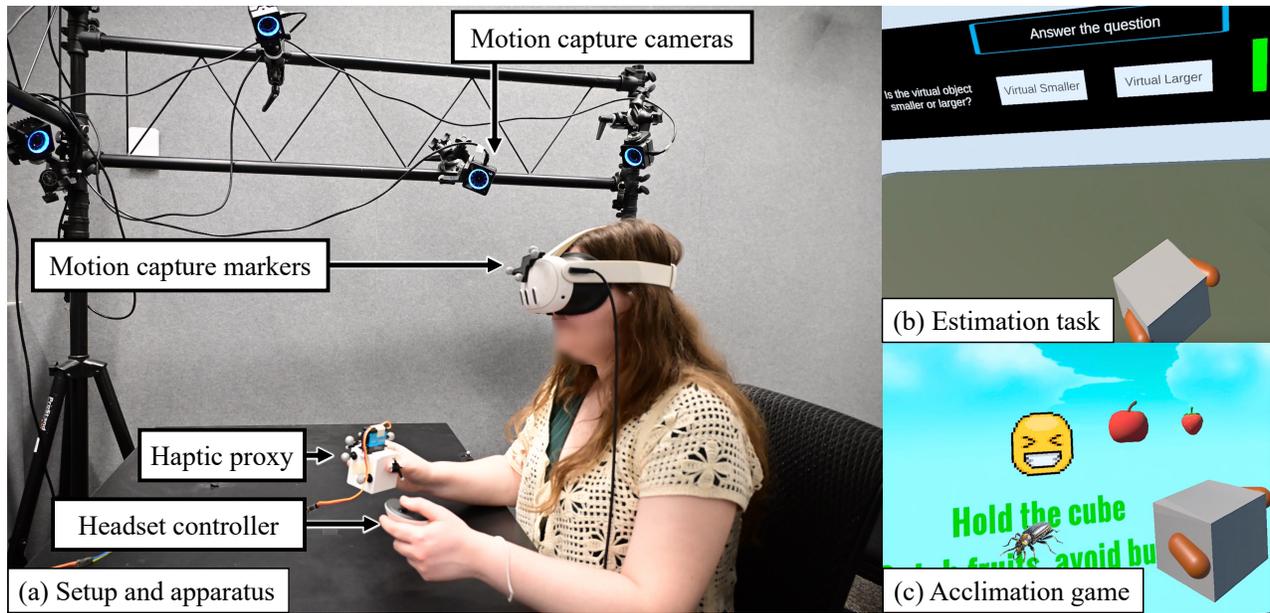
Figure 3: Experimental setup. (a) shows the apparatus and overall setup of the physical environment. (b) and (c) are accordingly the virtual scenes of the estimation task and acclimation game.

the data of the fixed-size second response. We keep the lift-respond-wait-respond format, however, to standardise task duration for considering acclimation. For transparency, we include the omitted results in the supplementary material.

As for the size-changing cube, the second answer was analysed to simulate a size-changing device performing changes in the users' hands. Although there was no "putting down" action, the haptic stimuli and visual stimuli both changed, which could serve as a reset and minimise the impact of the former question.

The order of presentation of the virtual sizes was randomised. The virtual sizes were derived from the results of prior work [4].

In the **size-changing** condition, the order of the size changes(6 cm-8 cm and 8 cm-6 cm) was also randomised.

For each question, the participants had 6 seconds to answer the question while lifting the device. The data would be invalid if they failed to answer within the time limit.

3.2.2 Break (4 minutes). During every break a NASA Task Load Index questionnaire [28] was filled out by the participants. There were 21 gradations on the scales with 0 - "Very Low" and 20 - "Very High", except the fourth question where 0 means "Perfect" and 20 means "Failure".

3.2.3 Acclimation Game (8 minutes). We designed a short acclimation game for the participants to complete with the study device. Participants used the device to move a virtual object in the scene. The game lasted for 8 minutes with 8 levels (1 level per minute). For the first 50 seconds of the game, the participants were instructed to "catch the fruits dropping from the sky with the controller in their hand, and avoid catching falling bugs". For the last 10 seconds of each level, participants were asked to "put the device down on the table and pinch the falling bugs with their thumb and index

finger". This game was designed to both acclimate participants to the device and refresh their proprioception through different hand poses.

During the game, the virtual object changed size randomly every 10 seconds. For the **fixed-size** condition, the virtual size changed randomly between 4.0 cm and 8.0cm, with a 0.5 cm step – the same as during the estimation task. In the **size-changing** condition, the virtual cube changed size between 4.0 cm and 10.0cm, at 0.5 cm steps. When the virtual cube was ≤7.0 cm, the physical cube was 6 cm, changing to 8 cm for virtual objects larger than 7.0 cm.

3.3 Experimental Conditions

We designed 4 different conditions as shown in Table 1. The participants were divided into 4 groups and the study followed a between-subject design.

3.3.1 Condition 1 - fixed-size device. Condition 1 was the simplest condition: a **fixed-size** device of 6 cm without any visual priming. In this setup, participants never saw the physical device during the entire study, allowing us to investigate the influence of interaction time on the visuo-haptic perception in VR. As shown in Fig. 2 (a), size perception was estimated four times over the course of the 1-hour session.

3.3.2 Condition 2 - size-changing device. In Condition 2, the physical device performed a **size change** in the user's hand to simulate active haptic devices in VR. The overall procedure was identical to Condition 1 (see Fig. 2 (a)). The participants never saw the physical device.

At the beginning of each estimation trial, the device was set randomly to either 6 cm or 8 cm for the participant to grasp. After the first question was answered, the device changed size while

Table 1: User study of four conditions

	No priming	Correct priming (6 cm)	Misleading priming (5 cm)
Fixed-size proxy	**Condition 1**	**Condition 3**	**Condition 4**
Size-changing proxy	**Condition 2**	-	-

still held in the hand and a forced choice question was asked again. During the size change, and in the brief intervals between the paired questions (when the device was put down and lifted again), the virtual object was not displayed to prevent additional visual cues. Therefore, the data collected includes both perception of the object after the initial grasp and the perception after a size change.

3.3.3 *Condition 3 - correct visual prime.*
In condition 1 and 2, participants were asked to finish the task without seeing the physical device. The visual priming of the physical device may pre-inform the perception of size, however [34]. As shown in Fig. 2 (b), right before the first estimation task and before the third (final) acclimation game, the physical device was shown in contracted mode (6 cm) to the participants as instructions were given (the instructions were repeated before the final acclimation game in condition 3 and 4). The participants didn't touch the device during the visual priming. In this condition, the physical device never changed size.

3.3.4 *Condition 4 - misleading visual prime.*
To explore whether a misleading visual prime can consistently skew size perception, in condition 3 we showed participants a physical device of a smaller size (5 cm). This device was produced with a corresponding motor and cable, such as to convince the participant it was the true device. This device was shown to the participants during the instructions before estimation task 1 and acclimation game 3, but was replaced by the researchers after the participants put on the headset, such that they would not notice the change of device. At no stage could they see the real physical device. The participants did not touch the device during the visual priming. In this condition, the device never changed size.

Across all conditions, the same device was used, for consistent weight and mass distribution.

3.4 Participants

We recruited a total of 80 right-handed participants for the user studies through a public university website, social media and other networks. Of these participants, 44 self-identified as female and 36 as male.

Participants ranged in age from 19 to 39, with a mean age of 25.46 (SD = 4.18). Regarding prior experience with virtual reality (VR), 31 participants reported having no experience, 41 reported some experience, and 8 reported good experience.

The study was approved by the local ethics board. Each participant received a gift card for participating in the study.

Participants were randomly assigned across the 4 study conditions, with 20 assigned to each condition.

4 Results

4.1 Condition 1 Results

For every estimation task the proportion of answering "virtual smaller" among all valid answers ("virtual smaller" and "virtual larger") was calculated for each virtual size. For example with the 6 cm devicen the proportion was very high for the 4 cm virtual size (because most of the participants felt the virtual one was smaller), and was very low for the 8 cm virtual size (because most of participants felt the virtual one was larger). The proportions of different virtual sizes were then fitted to a sigmoid function, following the procedure of prior work [4, 52] for forced choice task diagrams:

$$f(x) = \frac{1}{1 + e^{ax+b}} \qquad (2)$$

where $f(x)$ is the proportion of choosing "virtual smaller" and x is the virtual size. The parameters a and b were obtained by fitting.

Thus, each data point represents the responses from 20 participants by calculating the percentage of selecting "virtual smaller" (and/or selecting "virtual larger", which won't affect the results in the study). They were analysed later in groups of 9 comparisons (9 different virtual sizes). The results estimated represent a population visuo-haptic perception instead of individual perception, following traditional HCI practices [4, 52].

There were in total 17 (among 20 participants × 9 virtual sizes × 2 times × 4 estimation tasks = 1440 questions, 1.18%) missing answers for condition 1 because the participants did not answer the question in time.

After fitting the data to sigmoid functions of each estimation task, as shown in Fig. 4 (a), we calculate the point of subjective equality (PSE, where there's a 50% probability of selecting "virtual smaller", indicating a random choice and, thus, uncertainty about the stimuli) as well as the perception thresholds (where there's a 25% or 75% of selecting "virtual smaller", meaning the users start to be uncertain about the size, as interpreted in previous studies [4, 17, 52]). The results of condition 1 are shown in Fig. 4 (b), representing the PSE and thresholds as the perception in the middle of each estimation task (i.e. the first data point is considered to be at the moment of the middle of estimation task 1, 3 minutes from the beginning of the study). In this task with only 2 alternatives ("virtual smaller" and "virtual larger"), the calculated perceptual thresholds and PSEs are the same using the percentage of either "virtual smaller" or "virtual larger".

The upper and lower detection thresholds form an area where the users don't notice the physical and virtual size differences, while the PSE represent the absolute size perception. Both the area and PSE show a clear increasing pattern over time in condition 1, slowing down at the last estimation task (ET4). This indicates that participants perceived the physical device to be larger and larger over time, resulting in a 2.1 mm difference between the first estimation task (6.34 cm) and the last estimation task (6.55 cm).

(a) Results fitted to sigmoid function

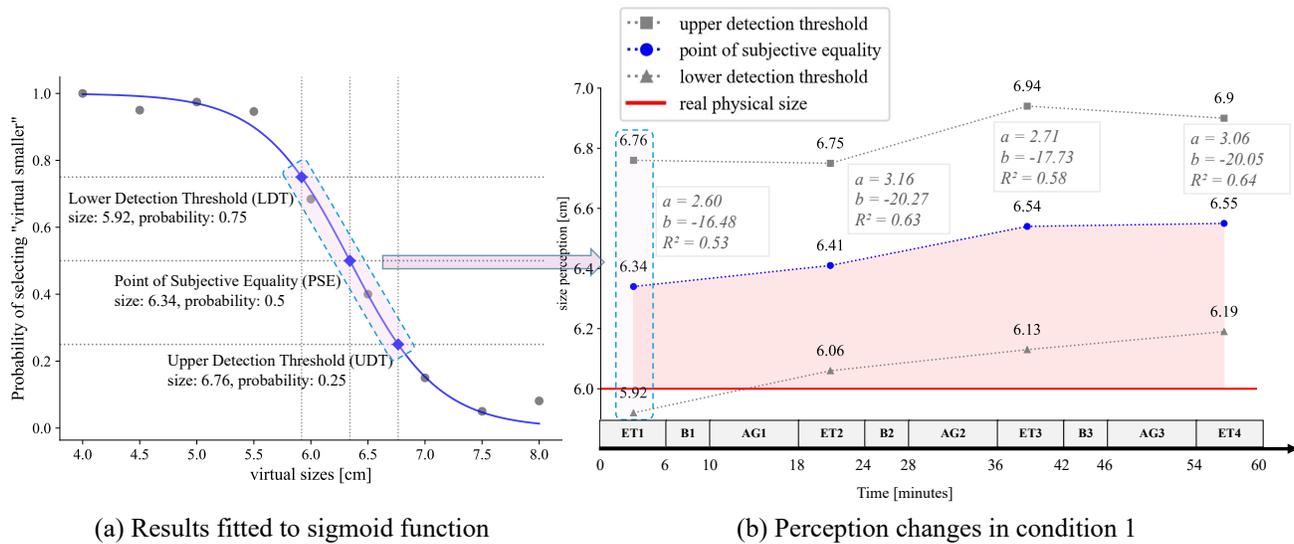(b) Perception changes in condition 1

Figure 4: Data analysis of condition 1. (a) showcases an example of how the proportions of answers are fitted to sigmoid function and how upper detection thresholds, point of subjective equality and lower detection threshold are calculated with data of condition 1, estimation task 1. (b) plots the perception thresholds and PSE of all estimation tasks in condition 1 together with time as the horizontal axis. Fitting parameters including a, b and McFadden R^2 are shown alongside.

McFadden's pseudo-R^2 for every curve shows a very strong fit (>0.5).

4.2 Condition 2 Results

In condition 2, participants interacted with a size-changing device that alternated between 6 cm and 8 cm throughout the 1-hour experiment. At no point were they shown the physical device.

Perception was estimated for two physical sizes (6 cm and 8 cm) at two different moments (immediately after grasping and immediately after the size change), resulting in four scenarios in total presented as four plots within Fig. 5. The fittings are all good or strong according to R^2 (0.23 0.50). There were 31 (among 20 participants × 9 virtual sizes × 2 physical sizes × 2 answers each lift × 4 estimation tasks = 2880 questions, 1.07%) missing answers for condition 2.

Plot (a) is comparable to condition 1, as both present the 6 cm device estimated after grasping. However, unlike condition 1, the results in condition 2 show a downward trend of size estimation. This indicates that interacting with a size-changing device-even after putting it down and grasping again- affects subsequent size estimation. Initially, the size was overestimated (6.26 cm), though still lower than condition 1 in the first estimation task (6.34 cm). Over time, the estimated size kept decreasing until it became stable in Estimation Task 4. The size perception changed approximately 5.5 mm (6.26 cm in ET1 - 5.71 cm in ET4).

Plot (b), which shows the 6 cm device after changing from 8 cm, indicates that the 6 cm size was generally underestimated in condition 2. This perception was relatively stable over time.

Plot (c) of 8 cm after grasping before size changes and plot (d) of 8 cm after changing from 6 cm show an increasing perceived size over time and tend to stabilise at the end of the session, similarly

to condition 1. The size was largely overestimated for the 8 cm physical size, ranging from 8.55 cm to 9.1 cm.

All four plots show a concentration of perception and a decrease of just noticeable differences (JNDs, half of the difference between the upper detection thresholds and lower detection thresholds, indicating a range when the users cannot notice the physical and virtual difference).

4.3 Condition 3 Results

Results from condition 1 and 2 show that the perception tended to drift over time, further from the real physical size. To examine whether a visual priming acts as a form of calibration, reducing or removing this drift, we designed condition 3 to offer a potential for visual corrections at different stages of the study.

The results of condition 3 are shown in Fig. 6 (a). Only 4 answers out of 20 participants × 9 virtual sizes × 2 times × 4 estimation tasks = 1440 questions (0.02%) were missing. The PSE ranges from 6.16 cm to 6.35 cm over time, which is notably smaller than in condition 1 (6.34 to 6.55 cm), where the only difference is the visual prime.

Similar to condition 1, an upward drift of size perception can be observed from estimation task 1 to 3. However, after receiving the second visual priming, prior to the final game, the estimation at task 4 dropped back closer to the real physical size (ET4: 6.25 cm).

4.4 Condition 4 Results

A misleading visual prime (a 5 cm physical device) was shown to the participants at the beginning of the study and again before the third acclimation game. The real 6 cm device was still used throughout the condition. The estimated perceptual thresholds and PSEs of condition 4 are shown in Fig. 6 (b). There were 8 answers out of 20 participants × 9 virtual sizes × 2 times × 4 estimation

(a) Perception for 6 cm physical proxy when grasping, in condition 2 (active device)

(b) Perception for 6 cm physical proxy after changing from 8 cm, in condition 2 (active device)

(c) Perception for 8 cm physical proxy when grasping, in condition 2 (active device)

(d) Perception for 8 cm physical proxy after changing from 6 cm, in condition 2 (active device)
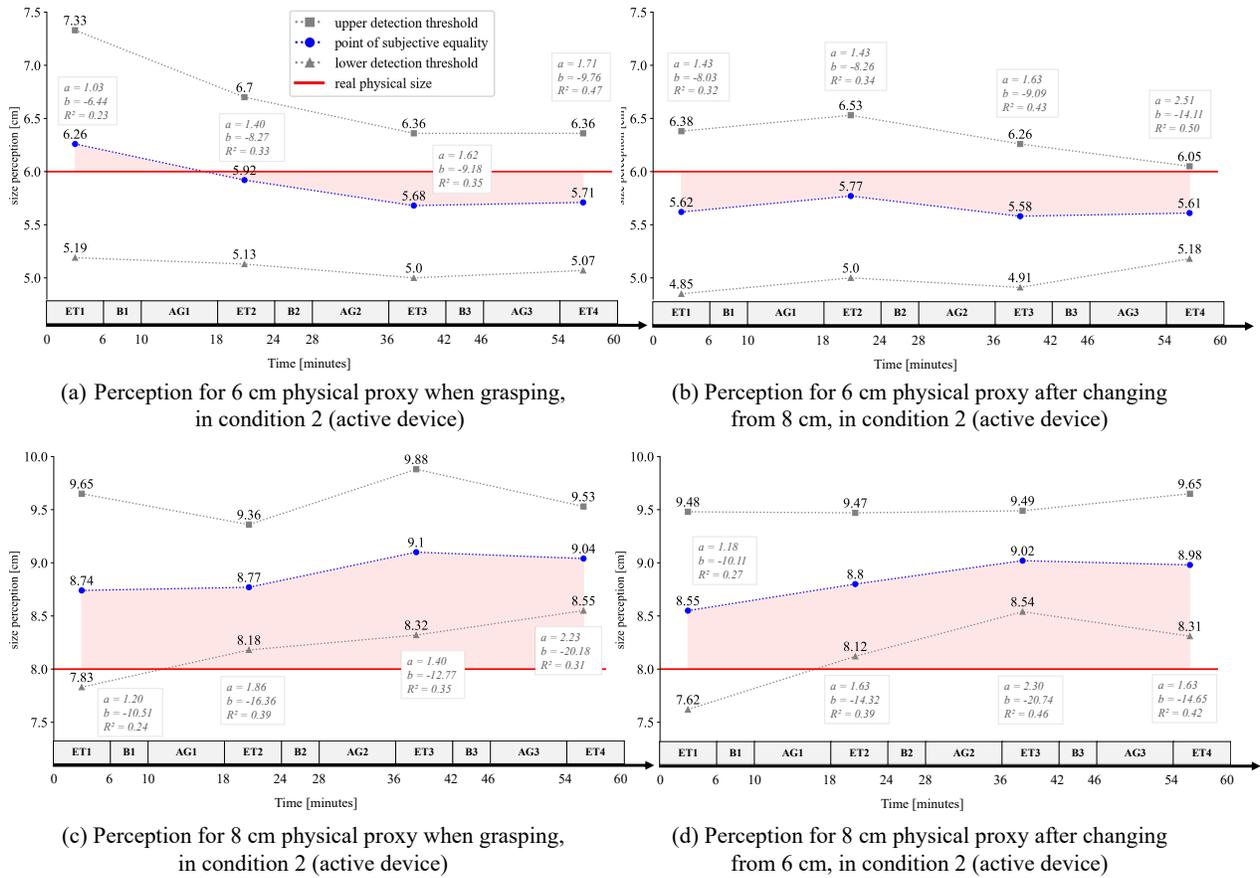
Figure 5: Data analysis of condition 2. Plots are the perception thresholds and PSE of all estimation tasks over time in condition 2. The device performed a size-changing in hand in condition 2 between 6 cm and 8 cm (orders were randomised). And therefore, we have results for both physical sizes before and after size-changing. (a) shows results of 6 cm physical device when grasping (before changing to 8 cm in hand), and (b) shows also 6 cm physical device results but after changing from 8 cm in hand. Similarly, (c) shows results of 8 cm physical device when grasping, before changing to 6 cm, and (d) shows results of 8 cm physical device after changing from 6 cm in hand. Fitting parameters including a, b and McFadden R^2 are shown alongside.

tasks = 1440 questions (0.06%) missing. Starting from 5.96 cm in estimation task 1, the perceived size is much closer to the real physical size (6 cm) when the smaller incorrect visual prime was given. Although the perceived size again drifted to be larger over time, the overall perception remained close to the real size. After the misleading small priming was given before acclimation game 3, the size perception dropped dramatically from 6.10 cm to 5.84 cm.

4.5 NASA-TLX Results

The NASA-TLX scores of the 80 participants are summarised in Fig. 7. Our intention from the NASA-TLX results was to examine whether effort and mental workload were the mechanisms behind changes in perception. The high standard deviations indicate various task load answers from different participants. We analysed the results using a Friedman test and a post-hoc Wilcoxon (Holm-corrected) test. The results indicate a significant decrease in temporal demand (Q3) and performance (Q4, note that the decrease

in this value means the performance is closer to "perfect") over time in different estimation tasks. Although not supported to be significant by the post-hoc test, the mean score and Friedman-test indicate an increase of physical demand (Q2).

For the decreasing temporal pressure (Q3), we interpret that over time the participants were getting used to the task with time limits and, subsequently, not feeling as hurried. They also seemed to increasingly believe their performance is getting better from the third estimation task (Q4).

5 Discussion

The results of the user study revealed a drift in size perception over time, and a complex interplay between visual and haptic perceptual cues. With the abundant results collected, we attempt to explain these observations, present a model of the underlying perceptual mechanisms, and examine their implications for the broader research community.

(a) Perception for 6 cm physical proxy with visual priming of the real size (6 cm)

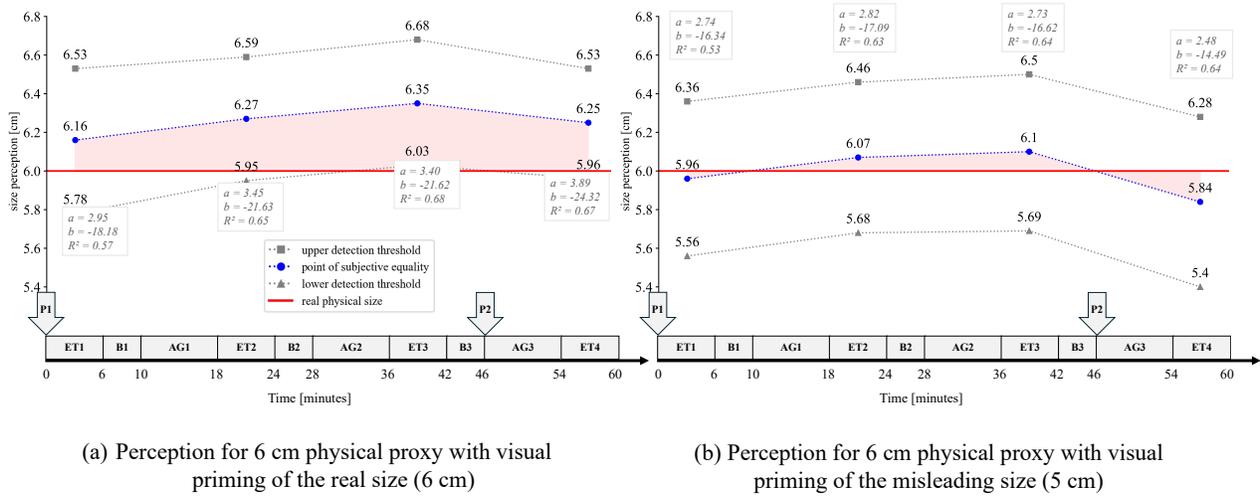(b) Perception for 6 cm physical proxy with visual priming of the misleading size (5 cm)

Figure 6: Data analysis of condition 3 and 4. Plots are the perception thresholds and PSE of all estimation tasks over time in condition 3. (a) and (b) are results of condition 3 with the real physical size visual priming (6 cm) and the results of condition 4 with the misleading physical size visual priming (5 cm), respectively. Fitting parameters including a, b and McFadden R^2 are shown alongside.
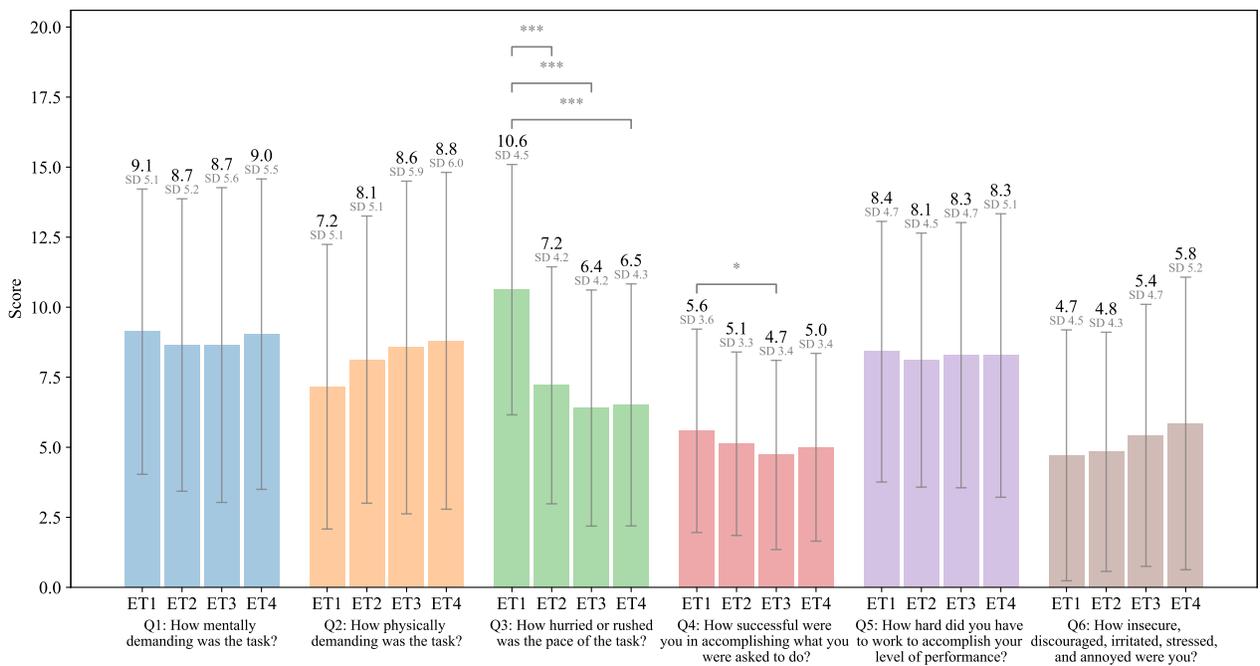

Figure 7: Mean scores and standard deviation (SD) of NASA-TLX questionnaire. ET1-ET4 represent the questionnaires being filled out after estimation task 1 to 4, respectively.

5.1 Implications

The PSE and JND results of all conditions are summarised in Fig. 8. For the fixed-size objects, all the PSEs drifted to be larger in an almost parallel pattern over time. The estimated size was smaller

following visual priming, and became even smaller when following the misleading 5 cm priming. The JNDs of the fixed-size device didn't show any unified continuous pattern.

The results of condition 2 for the 6 cm physical device, prior to it changing to 8 cm, differs from the other 6 cm results, revealing

how the PSE was increasingly underestimated over time. The JND was also the largest in this condition, showcasing how swapping sizes led to uncertainty of perception and created a larger illusory area. Over time, this JND became noticeably smaller, indicating that participants gradually became more certain about the object's size.

5.1.1 *Size is overestimated and its perception drifts larger over time for passive objects*. In the fixed-size conditions, with or without visual priming (condition 1 and 3), the size of the 6 cm physical device was generally overestimated. This overestimation of size has also been identified in previous studies [4, 17, 52]. For example, in the study of Bergström et al. [4], perception of 3 physical proxies (3 cm, 6 cm and 9 cm) was estimated 3 times over the course of approximately 30 minutes. Therefore, the interaction time for the 6 cm device was about 10 minutes, during which the PSE was estimated to be 6.36 cm, close to the PSE in our first estimation task (6.34 cm).

The upper and lower detection thresholds of the previous resized grasping study [4], however, are 5.4 cm and 7.32 cm for the 6 cm physical object, forming a much larger illusory range than we see in our results in condition 1 (5.92 and 6.76, respectively). This indicates that participants were more confident in their size perception in our study, though the basic techniques and procedures employed are similar in both. One potential reason for this is that in these conditions in our study, we only used a 6 cm physical object, where Bergstrom et al. [4] used 3 cm, 6 cm, and 9 cm objects. We hypothesise that our use of a single physical object gave participants more repeated sensory samples, consolidated in working memory, thereby strengthening their internal representation of the object.

Importantly, the overestimation becomes larger over time in fixed-size conditions (condition 1, 3 and 4), and shows a sign of stabilising towards the end of the study, from 39 minutes to 57 minutes (6.54 cm in ET3 to 6.55 cm in ET4 in condition 1). This indicates the size perception drifts towards an asymptote and any perceived size change reduces after a certain time of interacting with the object.

According to the NASA-TLX results, the participants' experience of physical demand increases throughout the study. This could also impact their perception. For example, participants could feel tired and perceive the objects to be heavier and thus larger over time. However, this does not explain the results, as (1) the score change is not significant according to the post-hoc Wilcoxon (Holm-corrected) test, (2) the perception stopped increasing at the end, and (3) the perception drifted towards smaller sizes in condition 2 with the 6 cm device, not following the tired-heavier-larger pattern.

The perceptual drift can be understood as a sensory phenomenon due to adaptation in cognitive psychology [30]. The perception is under real-time adjustment through a closed-loop process with exploration causing the perception to change over time. Beyond perceptual drift as a sensory phenomenon, our results can also be understood in light of cross-modality memorisation. While visual memory and its role in working memory have been extensively studied in HCI to inform design guidelines [29, 42], cross-modality memorisation has been largely overlooked. Our findings call for further studies in this vein, which could help assess how users build, maintain, and recalibrate their understanding of multimodal interactive systems.

5.1.2 *Grasping an object while it changes size strongly impacts subsequent size perception*. In condition 2, we examined size perception both immediately before and after the physical object changes size in each grasp-lift pair. Other than the very first trial, all of the subsequent trials were influenced by the fact that the physical device could change size. We focus on the results of the initial perception after grasping in each grasp-lift pair, rather than on the response after the size change (i.e., we concentrate on the results (a) and (c) in Fig. 5). Although the results immediately after each size change also show smaller JND (for 6 cm and 8 cm configurations) and increasing size over-estimation (for the 8 cm physical device), this is likely a result of complex effects across the virtual size size change, physical size change, visuo-haptic illusion, and a fatigue effect from holding the object in the air without refreshing by putting it down. The perceptual impact and memory of this size change remains to influence subsequent grasp-lift pairs.

Before size change, while the haptic proxy was still 6 cm, the PSE was smaller from the beginning of the study (6.26 cm in ET1), as shown in Fig. 5 (a) compared with condition 1 (6.34 cm in ET1), and more obviously, drifting towards increasing underestimation over time, which is the opposite pattern to the other conditions with the 6 cm physical device.

More similarly to conditions 1, 3, and 4, the results of the 8 cm size also drifted to be larger (Fig. 5c) and stabilised at the end. The extent of this overestimation was proportionally larger than the overestimation of 6 cm in condition 1. This is different from the conclusions of prior work [4], where sizes were underestimated for larger, 9 cm cubes.

Therefore, we observed that size perception for active, size-changing objects is different to passive, fixed-size objects. The frequent changing between physical object sizes influences the visuo-haptic perception. The smaller physical size is perceived as increasingly small, and the larger size is perceived as increasingly large. As this pattern of perception is not observed in prior work with different-sized passive objects, we conclude that the dynamic change of size, and the perceptual cues accumulated during that change, have a lingering impact on subsequent perception of the object when static. In turn, prior motion impacts subsequent perception, but only during contact with an object (at other times, this effect is not seen, so the impact of non-contact motion of perception is, seemingly, attenuated).

For HCI, this suggests that **the design of active haptic devices must account not only for the feedback they provide at a given moment, but also for the perceptual aftereffects of dynamic transformations**, since these can bias how users interpret and interact with objects long after the motion has stopped.

5.1.3 *Visual priming and its timing influences size perception*. In condition 3, participants were shown the physical device prior to the first estimation task. This resulted in their estimations being closer to the object's true size, than in condition 1 where they were given no visual prime of the actual device (6.16 cm when primed, in comparison with 6.34 cm when not primed). Though the participants did not see the physical object again for estimation tasks 2 and 3, the impact of that initial visual stimulus remained and their estimations remained lower (though continued to drift slowly

(a) PSE of all conditions
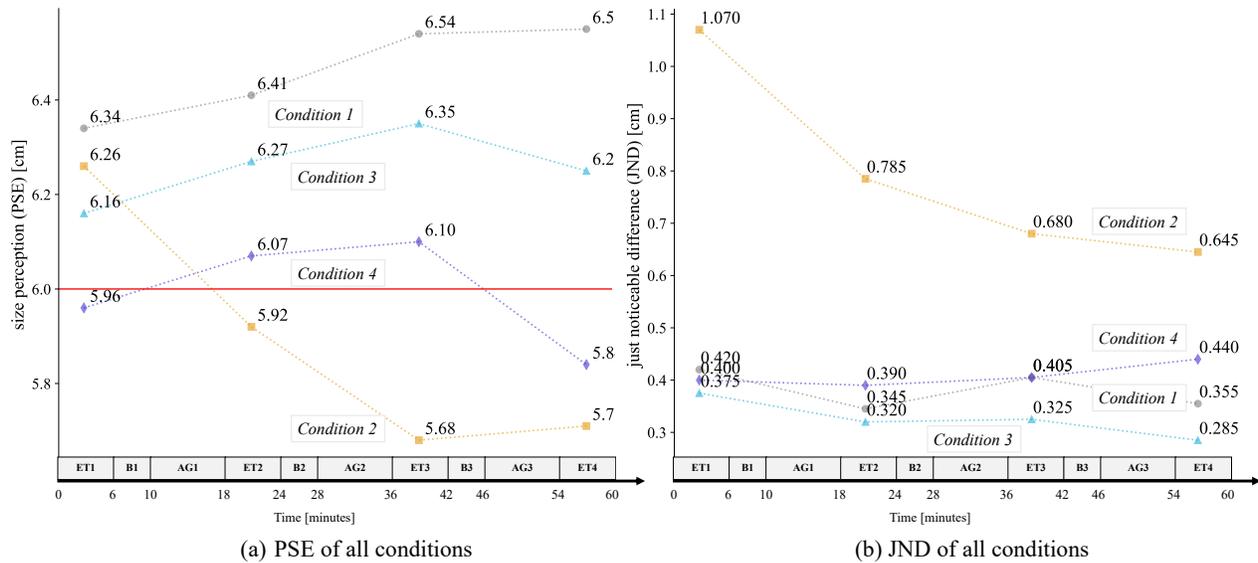
(b) JND of all conditions

Figure 8: Summary of the results of condition 1-4. Note that the results of condition 2 is represented by the perception of the 6 cm configuration after grasping and before size change. (a) summarises the points of subjective equality and (b) summarises the just noticeable difference by (upper detection threshold - lower detection threshold)/2. The timing of priming of condition 3 and 4 are also marked in the plots.

upwards, towards 6.35 cm, in comparison to 6.54 cm in condition 1).

Towards the end of the condition, and prior to the final acclimation game and estimation task, the visual prime was given to the participant again. After this, the estimated size returned to a more accurate value (∼6.25 cm). This was still not as accurate as the first estimation (6.16 cm). This might result from the 8-minute acclimation game between the visual priming and estimation task 4. If we were to assume that the moment the participants saw the physical device their perception was calibrated to that ground truth (6 cm), then the perceptual drift over time that we see across estimation task 1, 2, and 3, would also explain the drift that occurs across the acclimation game into estimation task 4, yielding the 6.25 cm results. This potentially shows a mathematical relationship between perceptual drift and time.

When a misleading visual prime (5 cm, in condition 4) was presented, the resultant size perception was dragged towards that prime (i.e., initial size perception was 5.96 cm after a 5 cm visual prime, as opposed to 6.16 cm after a 6 cm visual prime). Proportionally, however, this misleading visual prime had a lesser impact on size perception – the size perception dropped, but did not approach 5 cm. This supports existing theories about optimisation and reliability across different sensory cues [18] and how humans do not simply default to trusting vision over haptic cues.

When the misleading visual prime was presented again, prior to the acclimation game before estimation task 4, the size perception dropped to even smaller than the first estimation (5.96 cm), showing that the misleading prime had a larger effect on perception after interaction, than before interaction. The correction and learning between sensory inputs described by Berkeley [5] changes

at different timings. We posit that this is due to the decreasing reliability of the haptic cues over time, during the interactions, resulting in a subsequent visual cue having a larger impact on sensory integration.

5.2 Haptic Drift Model

It is still unclear how to mathematically describe the drift pattern and why the drift exists in perception. To better describe and interpret the results, we propose a theory to fit and explain the phenomenon. The hypothesis proposed offers one possible perspective and we expect relevant studies to further develop the theory.

Given the fact that the results drift towards one direction and tend to stabilise, an exponential function is the most direct and simple approach to describe the data, and allows parameters to alter the function to represent the impact of priming.

We then propose a first-order control system model whose nature is an adaptation process to describe our results. Firstly the exponential mathematical expression can typically describe a first-order control system, and secondly several previous studies in relevant fields, such as haptic retargeting [24, 38] and motor adaptation [18, 49], proposed similar closed-loop processes/control theories to our theory. The model is intended to be descriptive, as specific parameters may change under different procedures, though we believe the overall flow would remain the same.

The results show the following patterns that can be observed across more than one condition.

- The visuo-haptic perception drifts towards a certain direction over time (in all conditions).
- The haptic drift slows down towards an asymptote and becomes stable (condition 1 and 2)

- The initial priming has an impact on subsequent perception throughout the whole study (condition 3 and 4)
- The priming in the middle changes the perception directly (condition 3 and 4).

First of all, the results of the first condition with 6 cm passive object and no priming can be described as

$$P = c_1 - c_2 \cdot e^{-c_3 \cdot t} \tag{3}$$

where P is the perceived size and t is the time since the beginning of the study. The parameters c_1, c_2 and c_3 describe the initial gap, asymptote, and rate of time accordingly. The exponential approach is chosen because it describes how the perceptual results increase, the rate of increase slows and eventually flattens out, approaching a constant value.

Note that the fit can be transformed into a first-order control system expression mathematically. Therefore, we propose a visuo-haptic perception drift model, with the expression:

$$P(t) = P_\infty + (P_0 - P_\infty)e^{-\frac{t}{\tau}} \tag{4}$$

where P_0 is the initial value of the model, showing the initial perception at the beginning of interaction, P_∞ is the steady state value, meaning the final perception formed after infinite time of interaction and τ is the time constant (τ is the time when the value is about 63% of the way to P_∞ and 3τ is the time when the value is 95% of the way to P_∞).

When modelling perception, there are usually three components involved: *sensory information*, *expectation*, and *attention* [29]. A first-order control system is characterised by a differential equation relating inputs and outputs. If we assume human perception shares the same pattern with the control system, the haptic and visual stimuli are the sensory input (*sensory information*) and the perceived size is the output of the system (estimated by our forced choice task) as shown in Fig. 9. Humans dynamically perceive the haptic stimuli across the one-hour duration of the study, and adjust the output based on their best estimation, derived from their perception (*prediction/expectation*). Over time, this estimation stabilises.

During the active process of the visuo-haptic perception, people adopt exploratory behaviours to gain sensory stimulus, to inform their understanding of objects and the surrounding environment. When these different signals are processed, however, noise is introduced into our understanding (which we term 'discrimination of reality' in our model). This has been observed previously, such as with resized grasping [4, 17, 52], retargeting [1], and the rubber hand illusion [8], where the perception didn't match the reality accurately. As a result of this noise, humans amplify the momentary sensory cues in the output perception. Their outcome understanding, as a prediction of reality, is repeatedly assessed against new incoming sensory cues in an attempt to reduce the errors.

These predictions formed in the perceptual system are partially corrected by incoming sensory signals via recurrent feedback loops, through which humans maximise the information and attempt to reduce perceptual errors. However, the extent to which predictive signals are weighted within the brain remains unclear. Confidence in internal predictions appears to bias this weighting of reliability, thereby constraining the accuracy of error correction. As a result, residual error persists even after the system reaches stability. This confidence in comparison is also related with *attention* and sometimes is noisy and lossy. Bayesian models for sensory integration are used to explain the weights of prior information and prediction [6, 36]. Over time, without continuous calibration, perceptual experience becomes amplified, leading to a measurable mismatch between sensory input and perceptual output.

The result of fitting the data points of the 6 cm physical device without visual priming is shown in Fig. 10 and the fitting parameters are $c_1 = 6.71$, $c_2 = 0.39$ and $c_3 = 0.02$, accordingly, with $R^2 = 0.94$ indicating a good fit. This model reveals an initial perception $P_0 = 6.31$, steady state perception (the asymptote) $P_\infty = 6.71$, and time constant $\tau = 56.32$ (Fig.10). This means the initial perception at the beginning of interaction, although not estimated in the experiment, is 6.31 cm according to the model. After enough time, the perception will finally approach 6.71 cm. The time constant τ shows that after 56.32 minutes of interaction, the perception is around 63% of the way to P_∞. When the interaction time is longer than 3τ the perception will be 95% of P_∞.

Therefore, while it is almost impossible to estimate the perception over such a long duration with user studies, the model predicts the starting point and ending state of the perception. More importantly, this reveals that human cognition acts in a similar way to a control system - a procedure of initial guess, correction, and approaching a steady state.

Similarly, the size-changing device perception can be fitted to the model as shown in Fig. 11 with R^2 being 0.97 and 0.76 for the small and larger configurations accordingly. Again, we focus on the perception after grasping and before size change, because the perception after size changing is a complex procedure with multiple changing sensory stimuli.

For the 6 cm configuration, the initial perception P_0 is 6.37 cm, larger than the physical size. The steady state perception P_∞, however, drops to 5.64 cm. The time constant τ is 19.69 minutes, showing that the perception drift is 95% done within 1 hour (3τ).

The initial perception P_0 (8.68 cm) and steady state perception P_∞ (9.45 cm) are both much larger than the physical size 8 cm in the larger configuration of the device. However, this procedure takes longer time to complete with $\tau = 76.79$ minutes and $3\tau = 230.37$ minutes.

In condition 3, we presented the correct visual prime again at minute 46 and, thus, we assume the perception is corrected to P_0 again from this point (through a process akin to sensory dead-reckoning). From this assumption, the perception is modelled as shown in Fig. 12 (a) with $R^2 = 0.96$. The initial perception is $P_0 = 6.14$ cm, but then drifts towards $P_\infty = 6.40$ cm. At minute 46, we propose another theoretical hypothesis that the perception is refreshed to P_0 and the procedure of drift starts over again. With an initial correct visual priming, both the initial perception (6.14 cm) and steady state perception (6.40 cm) are more accurate compared with non-priming condition 1 (initial perception 6.31 cm, steady state perception 6.71 cm). Also the time constant τ (24.49 minutes) is much smaller in value than condition 1 ($\tau = 56.32$), indicating the perception gets to steady state much faster with visual priming.

As the second visual priming was presented in condition 4 with the misleading stimuli (5 cm), a sudden drop in perception was observed and the mechanism is complex. Thus, the final data point of condition 4 is excluded from the model. With the other 3 data
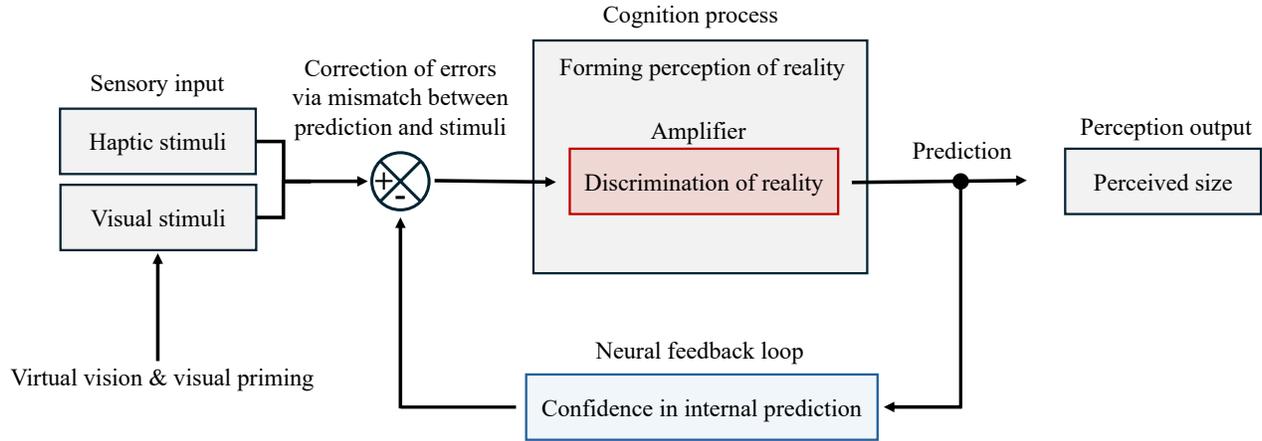
Figure 9: A model to describe the perception process as a first-order control system. The sensory input corrects the perceptual errors through a feedback loop over time, and discrimination in the cognitive process results in an amplified perception at last.
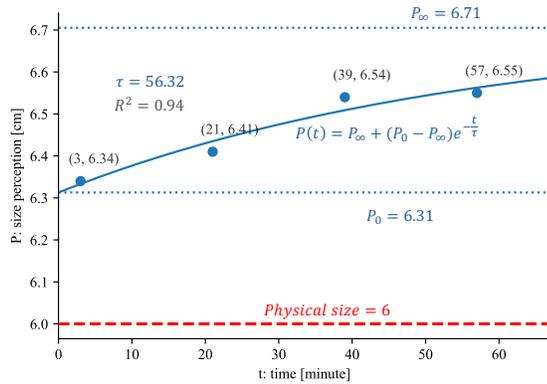

Figure 10: Haptic drift model of fixed-size device without any visual priming (condition 1).

points, the initial perception, steady state perception and time constant are calculated to be $P_0 = 5.92$ cm, $P_\infty = 6.11$ cm and $\tau = 13.85$ minutes. With the misleading smaller visual priming of 5 cm, the perception will still eventually drift to 6.11 cm to be larger than the physical size 6 cm, and the procedure is fast with $\tau = 13.85$ minutes and $3\tau = 41.55$ minutes.

By applying our visuo-haptic perception drift model, whose nature is mathematically a first-order control system, we interpret the procedure of human visuo-haptic perception and how it changes in different priming conditions, for different types of devices. The model can also calculate and predict perception unable, or not available any more, to be estimated.

Although many previous studies on perception and motor control indicate an improvement in accuracy over time[29, 40], the model shows the accuracy in our task only gets stabilised but is lower in general.

While it's hard to quantify the impact of visual priming in the model, the model indicates a "calibration" function of visual priming during interaction, which may relate to previous theories of visual and haptic integration models, that a maximum-likelihood integrator decides which sensory information should take domain in perception [18].

The model aligns with the perception theory in HCI with *sensory information*, *expectation* and *attention* [29]. Our model also aligns with historical and contemporary accounts of cross-sensory calibration. Berkeley's Essay Towards a New Theory of Vision (1709) [5] argued that vision is calibrated by touch; a view that has since been revisited in multisensory research [10, 27], and shows that visual input often anchors haptic estimates. In our experiments, intermittent visual priming acted precisely as such a calibration signal, resetting accumulated haptic drift. Within our study, a correct visual prime operates as an external corrective input that reduces error and stabilises the feedback loop. Moreover, because recalibrated estimates persisted after priming, we speculate that working memory retains this "reset" state, delaying the re-emergence of drift. This highlights how cross-modal calibration and working memory jointly contribute to the dynamics of perceptual stability in tangible interaction.

5.3 Application

Perception changes over time. As shown in all the results, perception changes during interaction. Our proposed model reveals how this perception changes and, thus, what users perceive about objects they interact with over time. This model gives interaction designers clearer insights, without having to repeatedly run psychometric tests to understand perception change during use.

Passive objects represent larger virtual objects. As shown in Fig. 4 (b), a passive object of 6 cm is generally overestimated in interaction. After about 21 minutes of estimation, the downscaling perceptual threshold (6.06 cm) will be larger than the physical object itself (6 cm). Given prior claims around how experience degrades when objects fall outside of this range [52], this means that, for
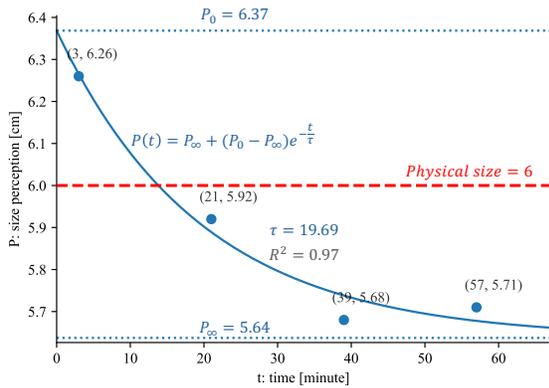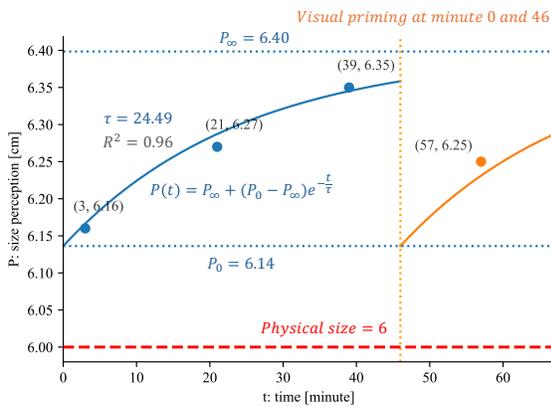
(a) Haptic drift model for size-changing proxy (6 cm before size change)
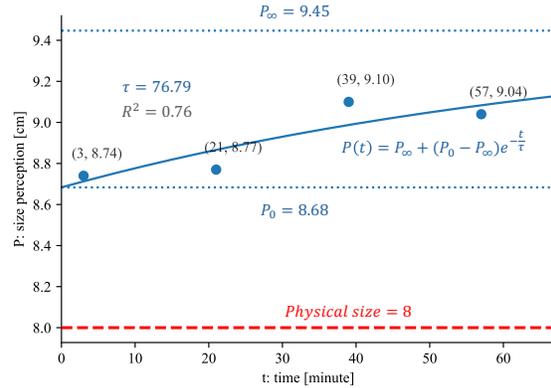
(b) Haptic drift model for size-changing proxy (8 cm before size change)

Figure 11: Haptic drift model of size-changing device (condition 2). (a) shows the perception of 6 cm configuration before size change and (b) shows the perception of 8 cm configuration before size change.

(a) Perception with 6 cm correct visual priming
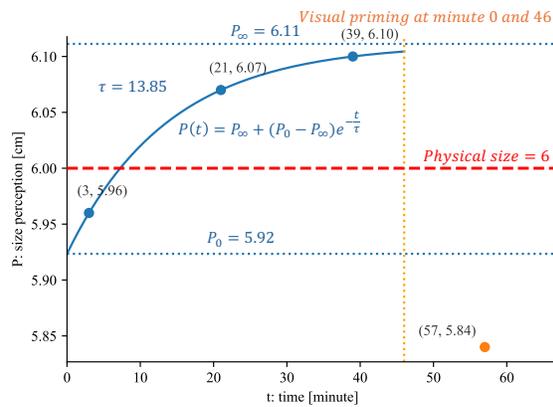
(b) Perception with 5 cm misleading visual priming

Figure 12: Haptic drift model of fixed-size device with visual priming (condition 3 and 4). (a) shows the perception with the real physical size (6 cm) visual priming and (b) shows the perception with the misleading physical size (5 cm) visual priming.

example, the physical object could no longer be used to represent itself in a virtual reality experience. This effect can be mitigated with visual priming.

Active devices impact perception differently to multiple passive devices. Many previous studies and designs assumed that passive and active devices shared the same patterns (e.g. Zhang et al. [52] applied the illusion spaces model to both passive and active devices in discussion). However, our results warn the research and design community to be more cautious on their perceptual conclusions, as the active devices yield meaningfully different results to the fixed-size objects. The size change provides an obvious contrasting stimulus between configurations. After a while of interaction (approximately minute 39 to 57 including breaks) the small configurations (6 cm) represents even smaller virtual objects (5.68 5.71 cm) and accordingly, the large configuration (8 cm) represents even

larger virtual objects (9.04 9.1 cm). This shows that in simple active devices, smaller dynamic size changes result in larger perceived size changes. In more complex designs with multiple configurations, the size should be carefully designed to reach the target effect.

Seeing the physical object affects its ability to represent objects. Many works in the research and design community ignored the influence of revealing the object to the participants during studies. While in most applications, the users would know the visual appearance of the object, the results estimated when the objects were not shown to the participants can be inaccurate in these scenarios. However, the results can be corrected, or at least predicted closer to reality, by utilising the haptic drift model and applying visual priming parameters.

Presenting the visual priming seems to calibrate the results and reset the sensory errors. This means that well-designed breaks, for

example, during which participants can see the physical device, have a meaningful impact on their ongoing perception. This is even more important when the appearance of the device is designed to mislead the users into believing something else, because showing the misleading stimuli again after some interaction has an even larger effect.

The results also suggest that potentially concealing or hiding the mechanism of the physical device may affect user's perception, leaving them with an incomplete or misleading visual reference for the object they are interacting with.

5.4 Future Works

In our study a visuo-haptic perception drift was observed in VR. It is still to investigated how this drift happens, and if this drift applies to a much larger range of interaction, even including the most common touch and grasp interactions of all objects in the physical world. We propose a first-order control system theory and an exponential mathematical model. Future related studies in human-computer interaction, psychology, and cognitive science are needed to further develop the theory.

While a prolonged interaction causes drifts in perception, it is worth exploring if there can be approaches to accelerate this procedure for effective interaction (e.g., training and education), or to slow down the drift to reduce perception errors.

As this work primarily concentrated on the fundamental factors with a specific setup, several future works are required to generalise the conclusion and theory. For example, **interaction effect** (visual priming and active device) or other factors may affect this process. **Individual perception** might also be different from our population conclusion we obtained in this study. However, it is still a challenge to enhance the generalisability in these perceptual findings to become effective both for the population and the individuals. The **grasping behaviour and type** is also a common limitation in the visuo-haptic research community. In our study we examined the moment right after grasping and lifting, and a certain grasp type. The impacts of the timing and types are still underexplored. While it's obviously tedious work to estimate on all timings and grasp types in user studies, some biomechanical/neurological interpretations behind them may form some theories, patterns or models for the community to expand our conclusions in the future. Similarly, all the **interactive objects properties** can not be included in a single study. While we employed a specific device, it's potentially predictable how the results may change when the properties differ. For example, Bergström et al. [4] and Zhang et al. [52] modelled the perceptions when the properties change. Yet it is not verified if all conclusions, including those in this paper, will follow the proposed model in previous studies and thus need future verifications.

While the visualisation of the users' hands impacts tangible interactions [32, 47], technical challenges persist for precisely tracking the whole-hand movement. Therefore, only simplified indicators were rendered in our user study to help grasp and avoid any tracking marker occlusion or inaccurate hand rendering which would have largely impacted the results. Although the absolute values of the perceived size may slightly change with different hand rendering techniques in practice, we believe the major conclusions,

including the drift patterns and influence of priming still hold true, which may be verified in future work.

Future work could also investigate how entrenched mental representations of everyday objects—formed as sensory traces consolidate into long-term memory, shape current perception in VR. This raises important HCI questions about how novel haptic devices can either exploit or must work against such perceptual baselines when simulating familiar objects.

While our study focused on size perception, it is also worth investigating whether similar forms of haptic drift occur in other object properties, such as weight, shape, texture, temperature, and stiffness. Exploring these dimensions would help determine whether drift is a generalisable feature of visuo-haptic perception or one specific to size. If perceptual drift reflects a more fundamental neurological process of forming reality in the brain, then comparable shifts might emerge across nearly all sensory modalities, not only in haptics. For HCI, this suggests that prolonged interaction with haptic devices could gradually alter how users perceive familiar material properties, potentially enhancing immersion but also introducing subtle mismatches when devices fail to recalibrate perception effectively. Studying these broader visuo-haptic properties would therefore deepen our understanding of perceptual stability and guide the design of haptic systems that remain convincing over extended use.

6 Conclusion

Understanding how people form and adjust visuo-haptic perception is critical for HCI, as it shapes the design of tangible interfaces and haptic systems. To investigate this process, we conducted a user study across four conditions, using forced choice tasks to examine how perception evolves over time and how it is influenced by visual priming. Our findings show that size perception drifts in a consistent direction before flattening out during prolonged interaction, and that both correct and misleading visual primes significantly alter how participants interpret the object's property. Based on these results, we modelled visuo-haptic perception as a closed feedback loop: sensory signals are integrated, calibrated against prior expectations, and gradually stabilise into cross-model perceptual representations. Taken together, these findings highlight that visuo-haptic perception is not static but dynamically shaped by time, calibration, and cross-modal interaction, offering both new opportunities and important challenges for the design of tangible interactive technologies.

References

[1] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2016. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*. 1968–1979. doi:10.1145/2858036.2858226

[2] Yuki Ban, Takashi Kajinami, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2012. Modifying an identified curved surface shape using pseudo-haptic effect. In *2012 IEEE Haptics Symposium (HAPTICS)*. IEEE, 211–216. doi:10.1109/HAPTIC.2012.6183793

[3] Yuki Ban, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2014. Displaying shapes with various types of surfaces using visuo-haptic interaction. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. 191–196. doi:10.1145/2671015.2671028

[4] Joanna Bergström, Aske Mottelson, and Jarrod Knibbe. 2019. Resized grasping in vr: Estimating thresholds for object discrimination. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 1175–1183. doi:10.1145/3332165.3347939

[5] George Berkeley. 1709. *An essay towards a new theory of vision.* IndyPublish. com.

[6] Max Berniker and Konrad Kording. 2011. Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science* 2, 4 (2011), 419–428. doi:10.1002/wcs.125

[7] Raoul Bickmann, Celine Tran, Ninja Ruesch, and Katrin Wolf. 2019. Haptic illusion glove: a glove for illusionary touch feedback when grasping virtual objects. In *Proceedings of Mensch und Computer 2019.* 565–569. doi:10.1145/3340764.3344459

[8] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391, 6669 (1998), 756–756. doi:10.1038/35784

[9] Detjon Brahimaj, Florent Berthaut, Frédéric Giraud, and Betty Semail. 2023. Crossmodal interaction of stereoscopy, surface deformation and tactile feedback on the perception of texture roughness in an active touch condition: Interaction intermodale de la stéréoscopie, de la déformation de surface et de la retour tactile sur la perception de la rugosité de la texture dans un état tactile actif. In *Proceedings of the 34th Conference on l'Interaction Humain-Machine.* 1–12. doi:10.1145/3583961.3583967

[10] Robert Briscoe, Rick Grush, and Alison Springle. 2015. Action-based theories of perception. (2015).

[11] Inrak Choi, Elliot W Hawkes, David L Christensen, Christopher J Ploch, and Sean Follmer. 2016. Wolverine: A wearable haptic interface for grasping in virtual reality. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, 986–993. doi:10.1109/iros.2016.7759169

[12] Inrak Choi, Eyal Ofek, Hrvoje Benko, Mike Sinclair, and Christian Holz. 2018. Claw: A multifunctional handheld haptic controller for grasping, touching, and triggering in virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems.* 1–13. doi:10.1145/3173574.3174228

[13] Aldrich Clarence, Jarrod Knibbe, Maxime Cordeil, and Michael Wybrow. 2021. Unscripted retargeting: Reach prediction for haptic retargeting in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR).* IEEE, 150–159. doi:10.1109/vr50410.2021.00036

[14] Aldrich Clarence, Jarrod Knibbe, Maxime Cordeil, and Michael Wybrow. 2022. Investigating the effect of direction on the limits of haptic retargeting. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* IEEE, 612–621. doi:10.1109/ismar55827.2022.00078

[15] Aldrich Clarence, Jarrod Knibbe, Maxime Cordeil, and Michael Wybrow. 2024. Stacked Retargeting: Combining Redirected Walking and Hand Redirection to Expand Haptic Retargeting's Coverage. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–13. doi:10.1145/3613904.3642228

[16] Andy Clark. 2024. *The experience machine: How our minds predict and shape reality.* Random House.

[17] Xavier de Tinguy, Claudio Pacchierotti, Mathieu Emily, Mathilde Chevalier, Aurélie Guignardat, Morgan Guillaudeux, Chloé Six, Anatole Lécuyer, and Maud Marchal. 2019. How different tangible and virtual objects can be while still feeling the same?. In *2019 ieee world haptics conference (whc).* IEEE, 580–585. doi:10.1109/whc.2019.8816164

[18] Marc O Ernst and Martin S Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415 (2002), 429–433. doi:10.1038/415429a

[19] Marc O. Ernst and Heinrich H. Bülthoff. 2004. Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8, 4 (2004), 162–169. doi:10.1016/j.tics.2004.02.002

[20] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M Dollar, and Danica Kragic. 2015. The grasp taxonomy of human grasp types. *IEEE Transactions on human-machine systems* 46, 1 (2015), 66–77. doi:10.1109/thms.2015.2470657

[21] Christopher R Fetsch, Gregory C DeAngelis, and Dora E Angelaki. 2013. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience* 14, 6 (2013), 429–442. doi:10.1038/nrn3503

[22] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM: dynamic physical affordances and constraints through shape and object actuation.. In *Uist,* Vol. 13. Citeseer, 2501–988. doi:10.1145/2501988.2502032

[23] Eric J Gonzalez, Parastoo Abtahi, and Sean Follmer. 2020. Reach+ extending the reachability of encountered-type haptics devices through dynamic redirection in vr. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology.* 236–248. doi:10.1145/3379337.3415870

[24] Eric J Gonzalez, Elyse D. Z. Chase, Pramod Kotipalli, and Sean Follmer. 2022. A Model Predictive Control Approach for Reach Redirection in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22).* Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. doi:10.1145/3491102.3501907

[25] Eric J Gonzalez and Sean Follmer. 2019. Investigating the detection of bimanual haptic retargeting in virtual reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology.* 1–5. doi:10.1145/3359996.3364248

[26] Eric J Gonzalez, Eyal Ofek, Mar Gonzalez-Franco, and Mike Sinclair. 2021. X-Rings: A Hand-mounted 360° Shape Display for Grasping in Virtual Reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology*

(Virtual Event, USA) *(UIST '21).* Association for Computing Machinery, New York, NY, USA, 732–742. doi:10.1145/3472749.3474782

[27] Monica Gori, Luana Giuliana, Giulio Sandini, and David Burr. 2012. Visual size perception and haptic calibration during development. *Developmental science* 15, 6 (2012), 854–862. doi:10.1111/j.1467-7687.2012.01183.x

[28] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology.* Vol. 52. Elsevier, 139–183. doi:10.1016/S0166-4115(08)62386-9

[29] Kasper Hornbæk, Per Ola Kristensson, and Antti Oulasvirta. 2025. *Introduction to Human-Computer Interaction.* Oxford University Press.

[30] Michaela Jeschke, Aaron C Zoeller, and Knut Drewing. 2024. Humans flexibly use visual priors to optimize their haptic exploratory behavior. *Scientific Reports* 14, 1 (2024), 14906. doi:10.1038/s41598-024-65958-6

[31] Roland S Johansson and Kelly J Cole. 1992. Sensory-motor coordination during grasping and manipulative actions. *Current opinion in neurobiology* 2, 6 (1992), 815–823. doi:10.1016/0959-4388(92)90139-C

[32] Sophie Jörg, Yuting Ye, Franziska Mueller, Michael Neff, and Victor Zordan. 2020. Virtual hands in VR: Motion capture, synthesis, and perception. In *SIGGRAPH Asia 2020 Courses.* 1–32. doi:10.1145/3415263.3419155

[33] Ulan Kelesbekov, Gabriele Marini, Bai Zhongyi, Wafa Johal, Eduardo Velloso, and Jarrod Knibbe. 2024. Stuet: Dual Stewart Platforms for Pinch Grasping Objects in VR. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* IEEE. doi:10.1109/ISMAR62088.2024.00045

[34] Myung Jin Kim, Eyal Ofek, Michel Pahud, Mike J Sinclair, and Andrea Bianchi. 2024. Big or Small, It's All in Your Head: Visuo-Haptic Illusion of Size-Change Using Finger-Repositioning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–15. doi:10.1145/3613904.3642254

[35] Luv Kohli. 2010. Redirected touching: Warping space to remap passive haptics. In *2010 IEEE Symposium on 3D User Interfaces (3DUI).* IEEE, 129–130. doi:10.1109/3dui.2010.5444703

[36] Konrad P Körding and Daniel M Wolpert. 2004. Bayesian integration in sensorimotor learning. *Nature* 427, 6971 (2004), 244–247. doi:10.1038/nature02169

[37] John W Krakauer, Alkis M Hadjiosif, Jing Xu, Aaron L Wong, and Adrian M Haith. 2019. Motor learning. *Comprehensive physiology* 9, 2 (2019), 613–663. doi:10.1002/j.2040-4603.2019.tb00069.x

[38] Flavien Lebrun. 2022. *Study of Visuo-Haptic Illusions in Virtual Reality : understanding and Predicting Illusion Detection.* Theses. Sorbonne Université. doi:10.1145/3649792.3649803

[39] Anatole Lécuyer, Sabine Coquillart, Abderrahmane Kheddar, Paul Richard, and Philippe Coiffet. 2000. Pseudo-haptic feedback: Can isometric input devices simulate force feedback?. In *Proceedings IEEE Virtual Reality 2000 (Cat. No. 00CB37048).* IEEE, 83–90. doi:10.1109/vr.2000.840369

[40] Byungjoo Lee, Sunjun Kim, Antti Oulasvirta, Jong-In Lee, and Eunji Park. 2018. Moving target selection: A cue integration model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–12. doi:10.1145/3173574.3173804

[41] Yuan Liu, Bo Zeng, Li Jiang, Hong Liu, and Dong Ming. 2021. Quantitative investigation of hand grasp functionality: Thumb grasping behavior adapting to different object shapes, sizes, and relative positions. *Applied Bionics and Biomechanics* 2021, 1 (2021), 2640422. doi:10.1155/2021/2640422

[42] Taku Morimoto. 2020. The nature of haptic working memory capacity and its relation to visual working memory. *Multisensory Research* 33, 8 (2020), 837–864. doi:10.1163/22134808-bja10007

[43] John R Napier. 1956. The prehensile movements of the human hand. *The Journal of Bone & Joint Surgery British Volume* 38, 4 (1956), 902–913. doi:10.1302/0301-620X.38B4.902

[44] Ferran Argelaguet Sanz, David Antonio Gómez Jáuregui, Maud Marchal, and Anatole Lécuyer. 2013. Elastic images: Perceiving local elasticity of images through a novel pseudo-haptic deformation effect. *ACM Transactions on Applied Perception* 10, 3 (2013), 17–1. doi:10.1145/2501599

[45] Alexa F Siu, Eric J Gonzalez, Shenli Yuan, Jason B Ginsberg, and Sean Follmer. 2018. Shapeshift: 2D spatial manipulation and self-actuation of tabletop shape displays for tangible and haptic interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–13. doi:10.1145/3173574.3173865

[46] Luca Turchet, Maud Marchal, Anatole Lécuyer, Rolf Nordahl, and Stefania Serafin. 2010. Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology.* 139–142. doi:10.1145/1889863.1889893

[47] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Balagopal Raveendranath, Christopher C. Pagano, Andrew C. Robb, Wen-Chieh Lin, and Sabarish V. Babu. 2023. How Virtual Hand Representations Affect the Perceptions of Dynamic Affordances in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2258–2268. doi:10.1109/TVCG.2023.3247041

[48] Jackie Yang, Hiroshi Horii, Alexander Thayer, and Rafael Ballagas. 2018. VR Grabbers: Ungrounded haptic retargeting for precision grabbing tools. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology.* 889–899. doi:10.1145/3242587.3242643

[49] Difeng Yu, Mantas Cibulskis, Erik Skjoldan Mortensen, Mark Schram Christensen, and Joanna Bergström. 2024. Metrics of motor learning for analyzing movement mapping in virtual reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18. doi:10.1145/3613904.3642354

[50] André Zenner and Antonio Krüger. 2019. Estimating detection thresholds for desktop-scale hand redirection in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 47–55. doi:10.1109/vr.2019.8798143

[51] Jian Zhang, Gavin Buckingham, Wafa Johal, and Jarrod Knibbe. 2025. ThumbShift: Modulating Perceived Object Properties Through Dynamic Thumb Repositioning. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 571–581. doi:10.1109/ISMAR67309.2025.00067

[52] Jian Zhang, Jarrod Knibbe, and Wafa Johal. 2025. Illusion Spaces in VR: The Interplay Between Size and Taper Angle Perception in Grasping. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)* (Yokohama,

Japan). ACM. doi:10.1145/3706598.3714162

[53] Li Zhang, Weiping He, Yupeng Hu, Shuxia Wang, Huidong Bai, and Mark Billinghurst. 2021. Using the Visuo-Haptic Illusion to Perceive and Manipulate Different Virtual Objects in Augmented Reality. *IEEE Access* 9 (2021), 143782–143794. doi:10.1109/ACCESS.2021.3121390

[54] Yiwei Zhao and Sean Follmer. 2018. A functional optimization based approach for continuous 3d retargeted touch of arbitrary, complex boundaries in haptic virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12. doi:10.1145/3173574.3174118

[55] Kening Zhu, Taizhou Chen, Feng Han, and Yi-Shiun Wu. 2019. HapTwist: Creating Interactive Haptic Proxies in Virtual Reality Using Low-cost Twistable Artefacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300923