



Peripersonal Space Tele-Operation in Virtual Reality: The Role of Tactile - Force Feedback

Yiru Liu^{1,2}, Nicholas Katzakis³, Frank Steinicke³, and Lihan Chen^{1,2,4}(✉)

¹ School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China

CLH@pku.edu.cn

² Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China

³ Department of Informatics, Universität Hamburg, Hamburg, Germany

⁴ National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing 100871, China

Abstract. In tele-operated human-robot collaboration, a human operator typically engages with a distant physical environment through a robotic system equipped with multiple sensors and actuators, allowing for haptic-based precise manipulation. Although these technical systems have been in use for years, the connection between multisensory perception and action in peripersonal space during tele-operations remains less understood. To delve deeper into this relationship, we examined distance perception in virtual peripersonal space. Participants wore an HTC Vive head-mounted display (HMD) featuring integrated eye-tracking (SMI) and moved a comparison object (a yellow ball) towards a target object (a blue ball) using a Geomagic Touch haptic device stylus, receiving either force feedback ('closed-loop') or no force feedback ('open loop') during the operation. They were instructed to focus on fixation points while performing the task, with SMI eye-tracking monitoring their gaze. The spatial positions of the comparison and target objects were arranged in four layouts: (i) center-to-center, (ii) center-to-peripheral (20 degrees in visual eccentricity), (iii) peripheral-to-center, and (iv) peripheral-to-peripheral. We employed seven distance levels between the objects in Experiment 1 and five distance levels in Experiment 2, using consistent methods of stimuli presentation. The findings revealed that estimation errors were significantly influenced by force feedback, spatial arrangement, and distance. Crucially, the visibility of the movement trajectory enhanced the effectiveness of tactile force feedback. Overall, this study proposes a potential guideline for human-computer ergonomic design, emphasizing the importance of force feedback for accurate targeting.

Keywords: tele-operation · multisensory · virtual reality · peripersonal space · eye-tracking · tactile feedback

1 Introduction

Efficient and skillful human-robot collaboration necessitates a comprehension of how vast amounts of sensory data (e.g., visual, tactile, proprioceptive, and kinaesthetic) are combined and how the connection between perception, action planning, execution, and

learning is addressed [1–3]. This collaboration has recently seen significant advancements in teleoperation, where a distant robot is controlled remotely by a human operator. Typically, the operational scenario is simulated in virtual environments (VEs) to facilitate the manipulation of experimental variables [4–9]. However, this presents considerable challenges for empirical perceptual research. For instance, sensory feedback (including force-feedback) is diminished in VEs. The field of view and access to visual cues are significantly limited in VEs. In earlier studies, researchers discovered that users could accurately perceive the position of virtual objects, with a precision of approximately 1 mm in augmented reality (AR) environments. By utilizing consistent proprioception and corrective visual feedback, operators could achieve better matching accuracy [10, 11]. A haptic system integrated with a virtual environment engine was designed to simulate delicate multi-finger manipulation. Throughout the interaction, the most relevant associations between physiological and physical parameters involved in manipulation were well maintained in virtual operations, with fidelity largely reliant on the quality of perceived force feedback [12].

Prior research has analyzed the impact of haptic feedback and visual indicators (like stereo cues) on motor skills tasks (such as basic target selection) in simulated environments, revealing the enhancing effects of force feedback under low task difficulty scenarios [11, 13], as well as the characteristics of force-feedback and spatial references (e.g., egocentric-based operations) [14]. This line of research implies that combining sensory cues can improve human-computer interaction, but the benefits depend on functional priorities and specific task demands (like movement trajectories). Consequently, an empirical study is needed to determine how these sensory cues are combined and prioritized. Furthermore, there is limited knowledge about how visuomotor coordination is achieved when interactions primarily occur in the visual periphery, which is often the case when focusing on a primary task while simultaneously performing a secondary task, such as grabbing an object or moving an item without directly looking at the peripheral location. Previous research has explored the impact of force-feedback and target size on reaction time and movement trajectories, but these studies did not investigate the relationship between initial operation positions and final targeting, nor did they accurately record perceptual errors in spatial arrangements [15, 16]. Numerous studies have demonstrated that perception is impaired at greater visual eccentricities (i.e., peripheral conditions) [17–21]. Importantly, an accurate perception of space and a thorough understanding of actions within the given peripersonal space (including restricted vision conditions) are crucial for effectively operating remote robotic systems [5, 22–24].

Considering these outstanding issues and research objectives, our current study investigates users' depth perception during actions in peripersonal space with or without force feedback (Experiment 1). We also examine if and how visual-tactile interactions can compensate for and potentially enhance performance when reaching for target depths in virtual environments (Experiment 2).

2 Methods

2.1 Participants

Thirteen college students (with mean age of 22.3, 7 females) attended in this experiment. The experiment was performed in compliance with the institutional guidelines set by the Academic Affairs Committee, School of Psychological and Cognitive Sciences, Peking University. The protocol was approved by the Committee for Protecting Human and Animal Subjects, School of Psychological and Cognitive Sciences, Peking University. All participants gave written informed consent in accordance with the Declaration of Helsinki, and were paid for their time on a basis of 50 CNY/hour.

2.2 Stimuli and Apparatus

We composed three balls (with radius of 0.1 in Unity 3D scale) to show the correspondence between standard (target) stimulus and comparison stimulus in a given trial. The red ball indicated the manipulating hand/effector for a participant. The yellow ball showed the initial position of comparison stimulus and the blue ball was the target stimulus. Participants were encouraged to move the red ball to the position of the yellow one until both balls were overlapping (at this time point instantly the red ball disappeared), and then performed the moving task as described in the following ‘procedure’ section.

The yellow ball, as a comparison, was located either in the center (near the participants hand-homing position), or left periphery (20° to fovea) or right periphery (20° to fovea). The blue ball, as a target was placed in mirror positions to the yellow ball but was farther way from the participant. The (vertical) distance between the yellow ball (standard stimulus) and the blue one (target) was defined and picked from one of the seven levels (for the nearest to the farthest from the observers’ perspective): -1.9, -1.2, -0.5, 0.2, 0.9, 1.6, 2.3 (relative to calibrated homing point “0” in the Unity3D design environment, i.e., the middle point of 70 mm for the motion range in Z-axis for Geomagic Touch), with fully randomization and counterbalance across trials. The visual stimuli were presented in virtual reality with Unity 3D program (Unity3d.com, 2015). Participants wore a HTC Vive head-mounted display (HMD) (with a refresh rate of 90 Hz), with integrated SMI eye-tracking components. The HMD was interfaced with a LCD display (with resolution of 2160 × 1200). The force feedback, if presented, was given by Geomagic Touch device (3D systems, USA). The haptic device has 6 degrees of freedom. Its workspace is within the range of 160 × 120 × 70 mm. The maximum output force is about 1N. The moving trajectory of ‘yellow’ ball (comparison) was invisible during the trajectory towards the ‘blue’ ball (target).

The parameters for force feedback was designed as follows:

$$\text{Force} = \text{startZCallback}(\text{dis} * 25.0f, -\text{target.transform.position.z} * 25.0f) \quad (1)$$

In Formula (1), the *startZCallback* is a customized function. *Dis*25* indicates the magnitude of force behaves as a function of discrepancy between the Z- axis depths between comparison and target. *-target.transform.position.z* shows the force in negative direction with reference to the center point of the ‘blue’ ball (target) in diameter.

2.3 Experimental Design and Procedure

We adopted within-participants $2 \times 4 \times 7$ factorial design. The first independent variable is the force-feedback (no force vs. force-present). The second one is the spatial correspondence between the standard stimulus ('yellow ball') and comparison ('blue ball'): center-to-center (i.e., C-C), center-to-periphery (C-P), periphery-to-center (P-C) and periphery-to-periphery (P-P). The third one is the distance (depth) between the standard stimulus (target: blue ball) and comparison stimulus (yellow ball). During the experiment, participants were encouraged to use the stylus of Geomagic Touch device and move the standard stimulus (yellow ball) with the stylus of Touch device to the position of comparison stimulus (blue ball). In the force-feedback trial, when they approached the comparison within three radii of the target, they perceived incremental force on the end of the thumb and forefinger. However, when they moved out of the range (three radii of the target), the force (intensity) was decreased and finally disappeared. Throughout the experiment, participants fixed their eyes on the central fixation point (a red cross with 2°) and was monitored by the integrated SMI eyetracking sensors. To initiate a valid trial, the participants' eye-gaze range should be within 3° around the fixation point and maintain above 200 ms. Otherwise, the same trial would restart. We implemented eight blocks with rests in between. Each block included 112 trials (2 repetitions \times 8 locations of standard vs. comparison stimuli \times 7 levels of distances). The open-loop (no force-feedback) and closed-loop (force-present) conditions were arranged in blocks. However, within each block, the sub-conditions of the distance and position were fully randomized. Between blocks, participants could take a rest up to 5 min (Fig. 1).

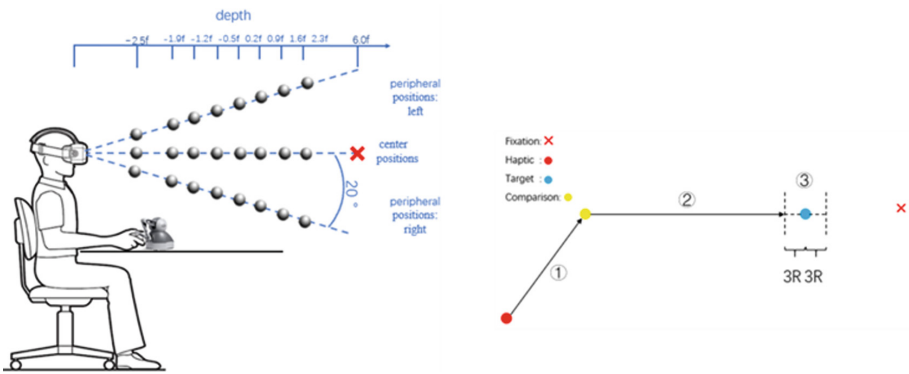


Fig. 1. Paradigm and schema for the present experiment. (A) Participants controlled the stylus of Geomagic Touch by moving the comparison to the target (location). (B) The view in head mounted display. The red ball indicates the current position of moving (right) hand. The yellow ball was the comparison and the blue one was the target. (C) The sample moving trajectory and experimental procedure. Participants initially moved the red ball to the position of yellow one, and both became invisible. Then they moved the comparison (yellow ball) to the target (blue ball). During this process, they had to fixate upon the fixation point.

Before the formal test, participants received a practice in which the conditions and trials were reduced. It included 32 trials in which we manipulated the depth (4 levels:

−2.2, −0.7, 0.8 and 2.3), the correspondence of locations between comparison and the target (4 levels) and the visibility of the red ball (visible vs. invisible). In the first part of practice (16 trials), participants got familiar with the sensation of force-feedback along the trajectory of moving red ball (being visible). In the second part of the practice (another 16 trials), participants were requested to fixate on the central fixation point and the red ball was invisible. After issuing each response, participants obtained feedback of either ‘correct’, ‘underestimation’ or ‘overestimation’ by the text message that lasted about 1 s on the screen. However, during the formal test no feedback was given. The response time window for a given trial was three seconds.

2.4 Results for Experiment 1

The deviation of estimated depths from the physically target positions and the reaction times were obtained. We then performed repeated measures analysis of variance (ANOVA). Moreover, we used MATLAB *grpstats* function to sort out the data by averaging the of reaction time and deviation for the depth judgments, and their associated errors of means. We obtained the proportions of ‘overestimation’ and then used SPSS 16.0 (Chicago, SPSS Inc.) to conduct repeated measures ANOVA. We also sorted out the reaction times under each experimental condition and did repeated measures ANOVA.

Reaction Time Under Force and Spatial Correspondence Factors

The repeated measures of ANOVA showed that the reaction time under force-feedback condition was longer (1370 ± 61.5 ms) than the one under no-force condition (1243 ± 59.4 ms), $F(1,12) = 23.441$, $p < 0.01$, $\eta^2 = 0.054$. The finding was counterintuitive but suggested that participants might purposely exploit the depth by relying on the ‘force-feedback’ information, this waiting strategy could extend the response time. The main effect of the spatial correspondences was significant, $F(3,36) = 23.212$, $p < 0.01$, $\eta^2 = 0.145$. The mean RTs for the center-to-center (‘c-c’), center-to-peripheral (‘c-p’), peripheral to center (‘p-c’) and peripheral to peripheral (‘p-p’) conditions were 1272 ± 56.6 ms, 1317 ± 55.2 ms, 1299 ± 59 ms, and 1335 ± 60.3 ms. Bonferroni-corrected comparisons showed the RT for ‘c-c’ was shorter than those in ‘c-p’ and ‘p-p’ conditions, $p < 0.001$. With the increased depths, the RTs increased as well, $F(6,72) = 204.751$, $p < 0.01$, $\eta^2 = 0.750$. The interaction between the factors of force condition and spatial correspondences was not significant, $F(3,36) = 1.181$, $p = 0.331$, $\eta^2 = 0.090$ (Fig. 2).

Reaction Time Under Factors of Force and Depth Levels

The repeated measures of ANOVA showed that the reaction time under force-feedback condition was longer (1370 ± 61.5 ms) than the one under no-force condition (1243 ± 59.4 ms), $F(1,12) = 11.272$, $p < 0.001$, $\eta^2 = 0.484$. The main effect of the depth was significant, $F(6,72) = 61.749$, $p < 0.001$, $\eta^2 = 0.837$. The mean RTs for the seven depths (from near to far, labeled as ‘D1’ to ‘D7’) were 1060 ± 52.3 ms, 1176 ± 53.4 ms, 1250 ± 57 ms, 1329 ± 58.8 ms, 1392 ± 64.2 ms and 1501 ± 67.7 ms. Bonferroni-corrected comparisons showed except for the comparisons between D2 and D3 ($p = 0.115$) and between D5 and D6 ($p = 0.392$), the other cohorts for comparisons were significant, $p < 0.05$. The two-way interaction between force and depth conditions was significant, $F(6,72) = 7.074$, $p < 0.001$, $\eta^2 = 0.371$. Further simple effects analysis indicated that

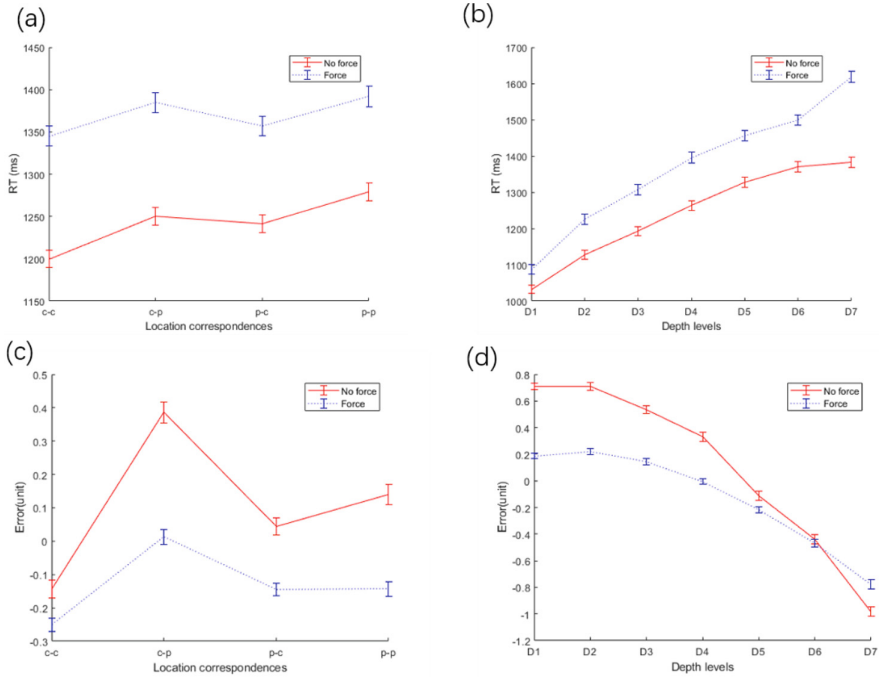


Fig. 2. Results for Experiment 1. (a) Reaction Time as function of the four types of location correspondences; (b) Reaction Time as a function of the seven levels of depths; (c) The Error (deviation) as a function of the location correspondences; (d) Error (deviation) as a function of the seven levels of depths. Error bars denoted the standard errors.

except for D1, the RTs in D2-D7 were longer for force-feedback conditions than those for force-absent conditions, $ps < 0.05$.

Deviation Analysis in Experiment 1

The deviation (error) was defined as the difference between the recorded depth estimation and the given depth. The positive deviation indicated over-estimation while the negative one under-estimation. We implemented repeated measures of ANOVA on the deviation. The repeated measures of ANOVA showed that the deviation under force-feedback condition was nearly the same (0.107 ± 0.150) as the one under no-force condition (-0.131 ± 0.108), $F(1,12) = 3.312$, $p = 0.094$, $\eta^2 = 0.216$. The main effect of the spatial correspondences was significant, $F(3,36) = 27.288$, $p < 0.001$, $\eta^2 = 0.695$. The mean deviations for the ‘c-c’, ‘c-p’, ‘p-c’ and ‘p-p’ conditions were -0.197 ± 0.109 , 0.2 ± 0.126 , -0.05 ± 0.103 and -0.001 ± 0.126 . Bonferroni-corrected comparisons showed the deviation for ‘c-c’ was smaller than the ones in ‘c-p’, ‘p-c’ and ‘p-p’, $ps < 0.05$. Moreover, the deviation in ‘c-p’ was larger than the one in ‘p-c’, $ps < 0.001$. There were no differences of deviations in ‘p-c’ and ‘p-p’. The two-way interaction between force condition and spatial correspondences was significant, $F(3,36) = 13.593$, $p < 0.001$, $\eta^2 = 0.531$. Further simple effects analysis indicated that in ‘c-p’ condition, the deviation without force (0.387 ± 0.166) was larger than the one in force-present

situation (0.013 ± 0.114), $p = 0.016$. Also, in ‘p-p’ condition, the deviation without force (0.140 ± 0.157) was larger than the one in force-present condition (-0.143 ± 0.128), $p = 0.061$.

The repeated measures of ANOVA showed that main effect the depth was significant, $F(6,72) = 57.897$, $p < 0.001$, $\eta^2 = 0.828$. The mean deviations for the seven depths (from near to far, D1-D7) were 0.448 ± 0.076 , 0.466 ± 0.106 , 0.338 ± 0.107 , 0.162 ± 0.118 , -0.165 ± 0.128 , -0.454 ± 0.156 and -0.880 ± 0.183 . Bonferroni-corrected comparisons showed except for the comparisons of deviations within D5-D7 ($ps > 0.37$), for the other cohorts, the deviations under force-absent were larger than those in force-present conditions, $ps < 0.01$. The two-way interaction of force and depth factors was significant, $F(6,72) = 11.256$, $p < 0.001$, $\eta^2 = 0.484$. Therefore, the current findings suggested that an overall over-estimation for near depths but under-estimation for far depths, i.e., with the increased depth, participants tended to shift from over-estimation to under-estimation.

3 Experiment 2

3.1 Participants

Forty college students (with mean age of 22.2, 18 females) attended in this experiment, however, 4 of them could not complete the task, and 6 of them failed to maintain the fixation as required, so that finally 30 participants fulfilled the task and their data were valid for subsequent analysis. All the participants had normal or corrected-to-normal vision. The experiment was performed in compliance with the institutional guidelines set by the Academic Affairs Committee, School of Psychological and Cognitive Sciences, Peking University.

3.2 Experimental Design and Procedure

The general design was similar to that in Experiment 1 but we made the moving trajectory visible. To balance the effectiveness of the depths as well as to reduce the number of trials (preventing fatigue), in Experiment 2 we reduced the depths as 5 levels: $-1.9f$, $-0.85f$, $0.2f$, $1.25f$, $2.3f$. The spatial correspondences were the same as in Experiment 1 and the periphery area was defined as 20° eccentricity. The side length for fixation was given by the following formula:

$$\begin{aligned} \text{fixationRange_length} = & 2 * ((\text{fixationRange.transform.position.z} - \text{camera_z}) \\ & * (\text{float})\text{Math.Tan}((\text{float})\text{Math.PI}/(180/\text{fixationAngle}))) \end{aligned} \quad (2)$$

in which *fixationRange_length* represents the size length of the fixation, *fixationRange.transform.position.z* shows the depth where the fixation point lies. *Camera_z* indicates the position of HTC_Vive in the Unity virtual environment. $(\text{float})\text{Math.Tan}((\text{float})\text{Math.PI}/(180/\text{fixationAngle}))$ was equal to size corresponding to $\tan 20^\circ$ (with the default 20° of *fixationAngle*). The fixation cross had a size of 2° and was located at a distance of $6.0f$ in Unity environment. After the practice, participants

received formal test with 8 blocks, in which 4 blocks with force-feedback and 4 without force-feedback. For the force-absent condition, each sub-condition had 4 trials (5 depth levels and 4 spatial correspondences), totally 80 trials were included. For the force-present condition, in addition to the 80 trials with congruent visual-tactile moving information (i.e., the maximum force was coincident with the contact of blue/target ball), we inserted randomly 20 trials as fillers in which the maximum force was given beyond the 0.5f of the very depth for the blue ball:

$$\text{startZCallback}(\text{dis} * 25.0f, -\text{target.transform.position.z} * 25.0f + 0.5f) \quad (3)$$

In formula (3) the *startZCallback* is customized function to regulate the magnitude of force. ‘-target.transform.position.z * 25.0f + 0.5f’ indicates that the (max) force is given when the stylus has been moved to the point of 0.5f further away from the blue ball. The parameters of a multiply (*25.0f) means that the distance has been transformed to meet the counterpart distance in real space.

Participants could start a trial when they met two mandatory requirements: the gaze duration upon the fixation should be above 300 ms and the response time should be within 4 s. If they failed to meet one of the two constraints, they had to redo the current trial. Before the formal test, we calibrated the eye-tracking equipment with the established 5-point protocol. After the calibration, they received a practice session of 40 trials, one half with force-feedback and the other without force. We defined the ‘overestimation’ as the deviation from the produced depth to the exact depth was above one radius of the blue ball and ‘underestimation’ as the deviation was below one radius. Participants received due verbal feedback appearing the screen, immediately after the issued the response. The response modes were the same as in Experiment 1 except that the moving trajectory was visible throughout the experiment.

3.3 Results for Experiment 2

Reaction Time

Contrary to the finding in Experiment 1, the repeated measures of ANOVA showed that the reaction time under force-feedback condition was shorter (1793 ± 82.1 ms) than the one under no-force condition (1979 ± 74.5 ms), $F(1,29) = 23.462, p < 0.001, \eta^2 = 0.447$. This suggests that with the visual feedback, the inputs from the force (tactile) facilitated the depth discrimination in virtual space. The main effect of the spatial correspondences was not significant, $F(3,87) = 1.921, p = 0.132, \eta^2 = 0.062$. The mean RTs for the ‘c-c’, ‘c-p’, ‘p-c’ and ‘p-p’ conditions were 1853 ± 75 ms, 1893 ± 80.1 ms, 1896 ± 73.8 ms, and 1902 ± 79.9 ms. The interaction between force condition and spatial correspondences was significant, $F(3,87) = 4.929, p = 0.003, \eta^2 = 0.145$. Further simple effects analysis indicated that for each spatial layout, the RTs in force-feedback condition were faster than those in force-absent condition, $ps < 0.01$. On the other hand, for no-force condition, the RTs were statistically not different among the four spatial correspondence; while for force-present condition, the RT in ‘c-c’ condition was the shortest, $ps < 0.05$.

As shown above, the main effect of force factor was significant, $F(1,29) = 23.462, p < 0.001, \eta^2 = 0.447$. The main effect of depth was significant, $F(4,116) = 176.909,$

$p < 0.001$, $\eta^2 = 0.859$. The RT increased linearly as a function of the depth. The RTs were 1544 ± 68.8 ms, 1764 ± 73.4 ms, 1909 ± 78.8 ms, 2039 ± 82.3 ms and 2174 ± 84.4 ms for D1-D5 respectively. However, the two-way interaction between force and depth factors was not significant, $F(4,116) = 2.214$, $p = 0.272$, $\eta^2 = 0.071$.

Results of Deviations for Experiment 2

The main effect of the force factor was significant, $F(1,29) = 48.277$, $p < 0.001$, $\eta^2 = 0.625$. The deviation in no-force condition (0.230 ± 0.034) was larger than the one with force-feedback (0.042 ± 0.028). The main effect of spatial correspondences was significant, $F(3,87) = 25.995$, $p < 0.001$, $\eta^2 = 0.473$. The mean deviations for the 'c-c', 'c-p', 'p-c' and 'p-p' conditions were -0.021 ± 0.009 , 0.328 ± 0.047 , 0.118 ± 0.024 and 0.116 ± 0.051 respectively. Bonferroni-corrected comparisons indicated that except for the cohort of 'c-c' and 'p-p' ($p = 0.059$), the cohort of 'p-c' and 'p-p' ($p = 1$), the other cohorts were significantly different, $ps < 0.001$. The interaction between force and spatial correspondences was significant, $F(3,87) = 17.175$, $p < 0.001$, $\eta^2 = 0.372$. Further simple effects analysis indicated that on each spatial layout, the deviations in no-force were larger than those in force-feedback conditions, $ps < 0.001$. However, on the other hand, for 'no-force' condition, the comparison between deviations in 'p-c' and 'p-p' was not significant, $p = 1$, the comparisons in other cohorts were significant, $ps < 0.01$. For 'force-present' condition, the comparison between 'c-c' and 'p-p', and the comparison between 'p-c' and 'p-p' were not significant, $ps = 1$. The other cohorts were significantly differed in the deviations, $ps < 0.001$.

The main effect of depth was significant, $F(4,116) = 38.501$, $p < 0.001$, $\eta^2 = 0.570$. The mean deviations across D1-D5 were 0.126 ± 0.018 , 0.238 ± 0.032 , 0.241 ± 0.034 , 0.139 ± 0.037 and -0.066 ± 0.040 respectively. Bonferroni-corrected comparison indicated that except that there was no statistical difference between the deviations in D1 and D4 ($p = 1$), there remained significant differences in other cohorts, $ps < 0.05$. Typically, the deviations were larger with mid-range of depths. The two-way interaction between the factors of force and depth was significant, $F(4,116) = 22.595$, $p < 0.001$, $\eta^2 = 0.438$. Further simple effects analysis indicated that across each depth level, the deviation in no-force condition was larger than the one with force-feedback, $ps < 0.05$. On the other hand, in the no-force dimension, the comparisons between D1 and D4 ($p = 0.875$), between D2 and D3 ($p = 1$), between D2 and D4 ($p = 0.106$) were not significant, while the other cohorts were significantly different in the deviations, $ps < 0.001$. For the force-present condition, except that there were no differences in the cohorts of D1 vs. D3 ($p = 0.872$), D2 vs. D3 ($p = 1$), the comparisons in other cohorts showed significantly differences, $ps < 0.05$ (Fig. 3).

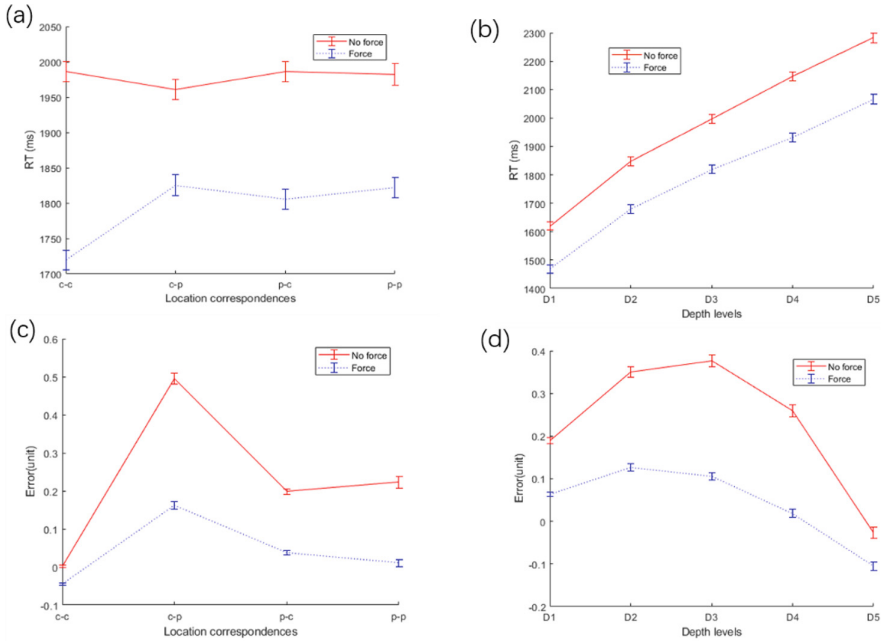


Fig. 3. Results for Experiment 2. (a) Reaction Time as function of the four types of location correspondences; (b) Reaction Time as a function of the seven levels of depths; (c) The Error (deviation) as a function of the location correspondences; (d) Error (deviation) as a function of the seven levels of depths.

4 Discussion

In this study, we explored 3D depth comparison in a desktop virtual reality system, focusing on the interaction between two essential sensory inputs: visual stimulation and force feedback. In the first experiment, we utilized force feedback while keeping the visual trajectory hidden. The second experiment involved a cross-modal design, where depth discrimination was enhanced by force feedback in conjunction with visible moving balls (specifically, the “yellow” ball used for comparison). In both experiments, we assessed participants’ depth perception performance in virtual peripersonal space, using reaction time and deviation indices for depth perceptual judgments, and analyzing the differences between the two experiments’ critical manipulations.

We discovered that estimation errors were significantly influenced by force feedback, spatial arrangement, and depth. Both deviation analysis and (over)estimation proportion analysis revealed that the smallest estimation errors occurred in the closed-loop condition (with appropriate force feedback) and congruent spatial layouts (center-to-center and peripheral-to-peripheral). Estimation errors were larger in the open-loop condition (without force feedback), incongruent spatial layouts (c-p and p-c), and the peripheral condition (p-p).

Overall, we observed an overestimation of depth perception, with participants overestimating the target’s depth (indicated by the blue ball). Short depths were overestimated,

while far depths were underestimated. Overestimation proportions were inversely proportional to actual distances, adhering to Vierordt's law as outlined in the contrast effects of temporal perceptual studies. Force inputs reduced depth estimation errors and resulted in a significant underestimation of depth compared to conditions without force feedback. This finding aligns with the notion that near-body touch/force enhances performance and fosters near-space perception.

Contrary to expectations, the force inputs increased reaction time in general when the moving trajectory was invisible rather than decreasing it. We speculate that, within the three-second response window, observers had enough time to utilize the force cue and deliberately move the standard stimulus (yellow ball) as close as possible to the comparison stimulus (blue ball), which may have led to increased reaction times. However, when the visual trajectory was present, depth perceptual uncertainty was greatly reduced, and force inputs as a closed-loop condition indeed facilitated responses. This finding suggests that effective cross-modal integration operates in accordance with the domain recruitment hypothesis, where force information can flexibly integrate with visual information only when the latter is highly functional with a visible trajectory. This facilitation effect was more stable when the trajectory was visible. In Experiment 1, the deviation was smaller for the force-feedback condition in "c-p" and "p-p"; however, in Experiment 2, the deviation was consistently smaller under force-feedback conditions across all spatial layouts.

In conclusion, this study illuminated key perceptual principles of multisensory inputs (i.e., force feedback) in enhancing teleoperation in peripersonal space using depth perception and manual operation in virtual reality (VR), as well as the limiting factors of spatial correspondence and distances between effectors/controllers and targets in depth perception and precise designated actions (e.g., pointing). A significant implication from this study is that observers may assign different weights to the estimation of physical distances in virtual environments based on the construal level theory. For example, due to the immediate and direct interaction nature of the tactile modality, people may rely more on force feedback under uncertain visual conditions and form a general schema for depth estimation in those situations (including Vierordt's law). Moreover, when visual information was more reliable (with visible trajectory), force cues maximized their role in enhancing target-reaching accuracy, and potential contamination from response strategy (such as purposefully exploiting force inputs) was largely reduced. These findings offer valuable insights for designing perceptually-inspired visuo-haptic interactions in areas related to redirected touching, haptic retargeting, and passive haptic feedback in visually disrupted environments. These results have important implications for the development of more effective virtual reality systems and applications. By understanding how multisensory inputs, particularly force feedback, can enhance depth perception and accuracy in teleoperation, we can create more immersive and realistic virtual experiences. This knowledge can be applied to various fields, such as remote surgery, telepresence robotics, training simulations, and entertainment.

Moreover, our findings suggest that the integration of force feedback with visual cues should be carefully considered and optimized to maximize its benefits. In situations where visual information is less reliable, the use of force feedback can be crucial for enhancing depth perception accuracy. On the other hand, when visual information is

more reliable (e.g., with visible trajectories), force feedback can still play an important role in improving the user's overall performance.

Lastly, it is essential to consider the spatial correspondence and distances between effectors/controllers and targets in the design of VR systems, as these factors can significantly impact depth perception and the accuracy of designated actions. By addressing these constraints and leveraging the insights gained from our study, developers can create more efficient and user-friendly virtual environments that facilitate accurate depth perception and interaction.

In summary, this research contributes valuable knowledge to the field of virtual reality and visuo-haptic interactions. It highlights the importance of multisensory inputs, particularly force feedback, and the need to carefully integrate these elements within virtual environments to maximize user performance and the overall VR experience.

Acknowledgements. This study was funded by STI2030-Major Projects 2021ZD0202600 and Sino-German Crossmodal Learning Project from Natural Science Foundation of China (Grant No. 62061136001).

References

1. Ernst, M.O., Banks, M.S.: Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**(6870), 429–433 (2002)
2. Hillis, J.M., Ernst, M.O., Banks, M.S., Landy, M.S.: Combining sensory information: mandatory fusion within, but not between, senses. *Science* **298**(5598), 1627–1630 (2002)
3. Kording, K.P., Wolpert, D.M.: Bayesian integration in sensorimotor learning. *Nature* **427**(6971), 244–247 (2004)
4. Popescu, V.G., Burdea, G.C., Bouzit, M., Hentz, V.R.: A virtual-reality-based telerehabilitation system with force feedback. *IEEE Trans. Inf Technol. Biomed.* **4**(1), 45–51 (2000)
5. Pacchierotti, C., Tirmizi, A., Bianchini, G., Prattichizzo, D.: Enhancing the performance of passive teleoperation systems via cutaneous feedback. *IEEE Trans. Haptics* **8**(4), 397–409 (2015)
6. Neupert, C., Matich, S., Scherping, N., Kupnik, M., Werthschützky, R., Hatzfeld, C.: Pseudo-haptic feedback in teleoperation. *IEEE Trans. Haptics* **9**(3), 397–408 (2016)
7. Panzirsch, M., Balachandran, R., Weber, B., Ferre, M., Artigas, J.: Haptic augmentation for teleoperation through virtual grasping points. *IEEE Trans. Haptics* **11**(3), 400–416 (2018)
8. Bugdadi, A., et al.: Is virtual reality surgical performance influenced by force feedback device utilized? *J. Surg. Educ.* **76**(1), 262–273 (2019)
9. Wang, Z., Sun, Y., Liang, B.: Synchronization control for bilateral teleoperation system with position error constraints: a fixed-time approach. *ISA Trans.* **93**, 125–136 (2019)
10. Swan, J., Singh, G., Ellis, S.: Matching and reaching depth judgments with real and augmented reality targets. *IEEE Trans. Visual Comput. Graphics* **21**(11), 1289–1298 (2015)
11. Wall, S.A., Harwerth, R.S.: Quantification of the effects of haptic feedback during a motor skills task in a simulated environment. In: *Proceedings at Phantom User Research Symposium*, Zurich, Switzerland, pp. 61–69 (2000)
12. Bergamasco, M., Avizzano, C.A., Frisoli, A., Ruffaldi, E., Marcheschi, S.: (2006) Design and validation of a complete haptic system for manipulative tasks. *Adv. Robot.* **20**(3), 367–389 (2006)

13. Wall, S.A., Paynter, K., Shillito, A.M., Wright, M., Scali, S.: The effect of haptic feedback and stereo graphics in a 3D target acquisition task. In: *Proceedings of Eurohaptics*, Edinburgh, UK, pp. 23–29 (2002)
14. Pawar, V.M., Steed, A.: Evaluating the influence of haptic force-feedback on 3D selection tasks using natural egocentric gestures. In: *IEEE Virtual Reality*, Lafayette, Louisiana, USA, pp. 11–18 (2009)
15. Pawar, V.M., Steed, A.: Profiling the behaviour of 3D selection tasks on movement time when using natural haptic pointing gestures. In: *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, New York, United States, pp. 79–82 (2009)
16. Pawar, V.M., Steed, A.: Poster: the effect of target size and force feedback on 3D selection within a co-located visual-haptic immersive virtual environment. In: *2013 IEEE Symposium on 3D User Interfaces (3DUI) 2013*, Orlando, FL, USA, pp. 169–170 (2013)
17. Xing, J., Heeger, D.J.: Center-surround interactions in foveal and peripheral vision. *Vision. Res.* **40**(22), 3065–3072 (2000)
18. Latham, K., Whitaker, D.: Relative roles of resolution and spatial interference in foveal and peripheral vision. *Ophthalmic Physiol. Opt.* **16**(1), 49–57 (1996)
19. Siderov, J., Harwerth, R.S.: Stereopsis, spatial frequency and retinal eccentricity. *Vision. Res.* **35**(16), 2329–2337 (1995)
20. Thibos, L.N., Walsh, D.J., Cheney, F.E.: Vision beyond the resolution limit: aliasing in the periphery. *Vision. Res.* **27**(12), 2193–2197 (1987)
21. Katzakis, N., Chen, L., Ariza, O., Teather, R.J., Steinicke, F.: Evaluation of 3D pointing accuracy in the fovea and periphery in immersive head-mounted display environments. *IEEE Trans. Visual Comput. Graphics* **27**(3), 1929–1936 (2021)
22. de Haan, A.M., Smit, M., Van der Stigchel, S., Dijkerman, H.C.: Approaching threat modulates visuotactile interactions in peripersonal space. *Exp. Brain Res.* **234**, 1875–1884 (2016)
23. Chen, Y.C., Maurer, D., Lewis, T.L., Spence, C., Shore, D.I.: Central–peripheral differences in audiovisual and visuotactile event perception. *Atten. Percept. Psychophys.* **79**, 2552–2563 (2017)
24. Grechuta, K., Guga, J., Maffei, G., Rubio Ballester, B., Verschure, P.F.M.J.: Visuotactile integration modulates motor performance in a perceptual decision-making task. *Sci. Reports* **7**(1), 3333 (2017)
25. Foley, A.J., Michaluk, L.M., Thomas, D.G.: Pace alteration and estimation of time intervals. *Percept. Mot. Skills* **98**(1), 291–298 (2004)
26. Rammsayer, T., Wittkowski, K.M.: Time order error and position effect of a standardized stimulus in discrimination of short time duration. *Arch. Psychol. (Frankf)* **142**(2), 81–89 (1990)
27. Roy, M.M., Christenfeld, N.J.: Effect of task length on remembered and predicted duration. *Psychon. Bull. Rev.* **15**(1), 202–207 (2008)
28. Ryan, L.J., Havens, A.: Responses contribute to context effects on ratio-setting timing tasks. *Perception* **42**(5), 537–550 (2013)
29. Schiffman, H.R., Bobko, D.J.: The role of number and familiarity of stimuli in the perception of brief temporal intervals. *Am. J. Psychol.* **90**(1), 85–93 (1977)
30. Vatakis, A., Ulrich, R.: Temporal processing within and across senses. *Acta Physiol. (Oxf)* **147**, 1 (2014)
31. Serino, A., Haggard, P.: Touch and the body. *Neurosci. Biobehav. Rev.* **34**(2), 224–236 (2010)
32. Hara, M., et al.: Voluntary self-touch increases body ownership. *Front. Psychol.* **6**, 1509 (2015)
33. Michalka, S.W., Kong, L., Rosen, M.L., Shinn-Cunningham, B.G., Somers, D.C.: Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. *Neuron* **87**(4), 882–892 (2015)

34. Rim, S., Uleman, J.S., Trope, Y.: Spontaneous trait inference and construal level theory: psychological distance increases nonconscious trait thinking. *J. Exp. Soc. Psychol.* **45**(5), 1088–1097 (2009)
35. Trope, Y., Liberman, N.: Construal-level theory of psychological distance. *Psychol. Rev.* **117**(2), 440–463 (2010)
36. Trautmann, S.T., van de Kuilen, G.: Prospect theory or construal level theory? Diminishing sensitivity vs. psychological distance in risky decisions. *Acta Psychol.* **139**(1), 254–260 (2012)