

## RESEARCH ARTICLE

# Shared Avatar for Hand Movement Imitation: Subjective and Behavioral Analyses

YASUNOBU KATSUMATA<sup>1</sup>, YASUYUKI INOUE<sup>2</sup>, SATOSHI TORIUMI<sup>1</sup>, HIROKI ISHIMOTO<sup>1</sup>, HARIN HAPUARACHCHI<sup>1</sup>, AND MICHITERU KITAZAKI<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

<sup>2</sup>Department of Information Systems Engineering, Toyama Prefectural University, Imizu, Toyama 939-0398, Japan

Corresponding authors: Yasunobu Katsumata (yskatsumata@gmail.com) and Michiteru Kitazaki (mich@tut.jp)

This work was supported in part by the Japan Science and Technology Agency (JST) ERATO under Grant JPMJER1701; and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP22H04774, Grant JP20H04489, and Grant JP22KK0158.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee at Toyohashi University of Technology.

**ABSTRACT** Virtual co-embodiment enables sharing avatars with others in virtual environments and can be applied for training motor skills by allowing teachers to share movements with learners in first-person perspective. We conducted a task where participants were asked to imitate pre-recorded hand movements of a teacher as accurately as possible. The participant's virtual hand movements were averaged in real-time with those of the teacher (shared avatar hand). We compared their usability ratings and behavior against a controlled condition with full control of the hand (solo avatar hand). The teacher's hand was displayed facing the same or the opposite direction as the participant's hand. We hypothesized that using the shared avatar hand would improve imitation over using the solo hand, and the teacher's hand presented in the same direction is better than that in the opposite direction. Subjective ratings showed that the shared hand was easier to use than the solo hand, and the teacher's hand when presented in the same direction was easier to imitate than when presented in the opposite direction. Spatial error was less with the opposite-direction presentation than the same-direction presentation of the teacher's hand, irrespective of movement sharing. Time delay was less when the participants used the shared hand compared to when they used the solo hand, irrespective of the teacher's hand direction. These results suggest that sharing movements enhances usability and matching speed during movement imitation, and the same-direction presentation of the teacher's hand improves usability while the opposite-direction presentation improves spatial accuracy of motor imitation.

**INDEX TERMS** Avatar, co-embodiment, imitation, virtual reality, human-computer interaction.

## I. INTRODUCTION

Humans can feel illusory ownership towards bodies or body parts other than their own due to visual-tactile synchrony, or visual-proprioceptive synchrony. When a person observes a rubber hand being brushed with a brush while his or her real hand, which is out of sight, is also being brushed synchronously, an illusory feeling of owning the rubber hand occurs (Rubber hand illusion) [1]. It is known to be induced by the so-called visual-tactile synchrony. A person observing an artificial hand of which the index finger is connected to

his or her real index finger out of sight may also feel a sense of ownership for the artificial finger when she or he moves her/his finger causing the artificial finger to move along. Such sensations of illusory ownership are known to be induced by the visual-proprioceptive synchrony [2], [3], [4]. Here, the observer feels illusory ownership of the artificial index finger even when it is passively moved by the experimenter [2], [3], [4]. Virtual-reality technology can expand illusory body ownership through visual-proprioceptive synchrony by presenting virtual avatars (virtual human bodies) that move synchronously with the observers (virtual embodiment) [5], [6]. We can embody virtual avatars in different colors [7], [8], ages [9], [10], sizes [11], genders [12], and

The associate editor coordinating the review of this manuscript and approving it for publication was Giacinto Barresi<sup>1</sup>.

species [13]. Illusory ownership of invisible bodies also can be induced either by visual-tactile synchrony [14], [15] or by visual-proprioceptive synchrony [16], [17]. Visual-proprioceptive synchrony between different body parts such as a real finger and a virtual arm [18] or real legs and a virtual third and fourth arms [19] also enable us to feel those virtual body parts as our own body parts (body parts remapping). Such virtual embodiments are based on the concept that one person has one body. However, virtual co-embodiment [20], [21], [22] and virtual multibody embodiment [23], [24], [25] have further expanded the perspective of human embodiment as augmented humans [26], [27].

Virtual co-embodiment is based on the concept of sharing a virtual avatar with another entity, such as another person, robot, or an autonomous agent. It allows users to be immersed in virtual environments and have varying levels of control over a shared avatar. For example, if a person controls a shared avatar at a 75% level of control, another person can control the same shared avatar with a 25% level of control. Fribourg et al. showed that participants can estimate their actual level of control but feel the sense of agency higher than the actual controlling level of the co-embodiment avatar if the goal is common for the two participants sharing the avatar [20]. Hagiwara et al. investigated the senses of agency and ownership towards a shared body whose movement was generated by averaging the movements of two participants (control level 50% for each participant), and its motor performance. Hagiwara et al. also showed that the sense of agency was higher than the actual level of control, and the reaching movement of the shared avatar's hand is straighter and smoother (less jerk) than a solo avatar controlled completely by one person [28]. These findings suggest that humans prioritize these movement characteristics of a shared body over their own movements during body sharing.

The virtual co-embodiment is expected to be applied as a novel tool for remote training. Several embodied training systems using virtual reality have been proposed. The ghost-simulation method combined with first-person perspective visualizes a trainer's movements as a ghost superimposed on the trainee who follows the ghost image [29]. Another study developed a mixed reality system that provides a local novice user with two additional virtual arms controlled by a remote expert who uses them to train or guide the local user [30]. A study called "Fusion" proposed two wearable robotic arms and a robot head worn by a local user, while an expert remotely controls them to guide the local user from his or her viewpoint. Fusion's robotic arms can be mounted on the local user's wrists to directly force the user's arms to move [31]. These studies utilize virtual co-presence, where two or more individuals (trainer and trainee) can experience a sense of being at the same place even though they are physically at distant places. For enhancing remote training, sharing of viewpoints or first-person perspective is critical. It enables a trainee to experience a trainer's action from the same viewpoint. Virtual co-embodiment could further

enhance such aspects by directly sharing the same avatar body with the trainer.

Kodama et al. proposed the application of virtual co-embodiment for motor skill learning [32]. Learner participants performed a dual motor task in which they drew different shapes with left and right hands either in the co-embodiment condition, the perspective sharing condition, or the alone condition. In the co-embodiment condition, the teacher's motion and the learner's motion were averaged into a shared avatar, and the learner participant experienced it as his or her own avatar from the first-person perspective. In the perspective sharing condition, the teacher's movements were presented as a ghost in first-person perspective of the learner participants. Participants performed a baseline session (pre-test) and a test session with their own solo avatar in addition to a learning session in one of the three conditions (co-embodiment, perspective sharing, and alone in a between-subjects design) between the baseline and test sessions. Results showed that virtual co-embodiment makes motor skill learning more efficient than that in the other two conditions. The learner's performance improved faster and higher during the co-embodiment learning session, and the performance in the test session after learning was the highest. However, the performance in the test session with the solo avatar drops compared with the learning session using the co-embodiment avatar. To prevent this, the weight adjustment method has been developed by making the level of control of the teacher greater than that of the learner in the early stages of learning and decreasing it as the learning progresses and gradually allowing the learner to move independently [33]. Virtual co-embodiment facilitates motor skill learning with declarative memory as well. The co-embodiment learning provides higher performance improvements over the solo avatar during the learning session, and higher retention of learned skills one week later [34].

Mimicry, the imitation of observed actions or movements of others, has been observed in humans in a variety of contexts from early infancy to adulthood [35], [36]. While mimicry often occurs spontaneously and has benefits for social interaction such as increasing prosocial behavior [37], the conscious and explicit imitation of an expert's movements is a basis for acquiring complex skills [38], [39]. Thus, typical motor skill learning involves observing demonstrations of an expert from a third-person perspective, and imitating its movements. It has an advantage in visually comparing one's own movements with an expert's movements simultaneously. In contrast, it is difficult to separate the expert's movements and one's own movements from a co-embodied avatar.

In this study, we aimed to combine a third-person perspective for presenting a teacher's full movements (without averaging) in addition to the first-person perspective of the shared movements (average of the learner and the teacher). The teacher's hand movements were presented in front of participants (learner) and participants observed them

in third-person perspective. The participant's hand movements were presented either as the shared avatar hand in which the participant's movements and teacher's movements were averaged or as the solo avatar hand which reflected only the participant's motion. The participants were asked to imitate or replicate the teacher's movements as accurately as possible. We hypothesized that the shared avatar hand would improve the usability (make the task easier to perform) during imitation over the solo avatar (H1). In this study, we focused on the performance of imitation as the basis for motor skill learning, not on the learning process. Therefore, we employed a within-subject design to compare conditions directly.

Previous virtual co-embodiment studies used arm and hand movements such as reaching or line drawing [20], [21], [22], [28], [32], [33], [34]. Detailed movements such as hand gesture and finger movements are required for a variety of motor tasks. Thus, we developed a hand-movement co-embodiment system with finger movements (a shared avatar hand) and employed hand signs from American sign language as stimuli for participants to imitate. Teachers of sign language usually show their hand face-to-face with students so that the direction of the learner's hand is opposite to the student's hand. If the teacher's and student's hands are in the same direction, it may be easier to make comparisons. In this study, we compared how usability, matching time delay, and spatial error are affected when the presentation of the teacher's hand is in the opposite direction to the learner compared to when it is in the same direction as the learner. We hypothesized that the same-direction presentation could provide better performance than the opposite-direction presentation (H2).

## II. METHODS

### A. PREPARING TEACHER DATA

Twenty nonsense syllables/words consisting of three letters chosen from the evaluated list [40] were prepared for the stimuli (for example, YOS, WEF, HUI). The nonsense syllables were chosen instead of meaningful words to eliminate the effects of the meaning of the words and individual knowledge/experience with known words affecting the learning results. We captured and recorded three-dimensional (3-D) hand movements of forty candidate words represented with American manual alphabet (American sign language) from a person with American sign language skill using a motion capture system (Manus Prime II). An expert evaluated the captured hand motions of sign languages, and we chose the best twenty words for the experiments based on the evaluation.

### B. PARTICIPANTS

Twenty-six healthy adults (three females, mean 21.69, SD 1.44 years old, all right-handed) participated in the experiment. The sample size was determined by a power analysis: minimum twenty-four participants with a medium effect size  $f = 0.25$ ,  $\alpha = 0.05$ , power = 0.8, and repeated

measures of analysis of variance (ANOVA) — two avatar conditions  $\times$  two direction conditions — using G\*Power 3.1 [41], [42]. All participants had normal binocular vision and physical abilities. None of them had learned American Sign Language, and two participants had some experience with Japanese Sign Language. They provided written informed consent before the experiment. The methods of the experiment were approved by the Ethical Committee at Toyohashi University of Technology, and all methods were carried out in accordance with the relevant guidelines and regulations.

### C. APPARATUS

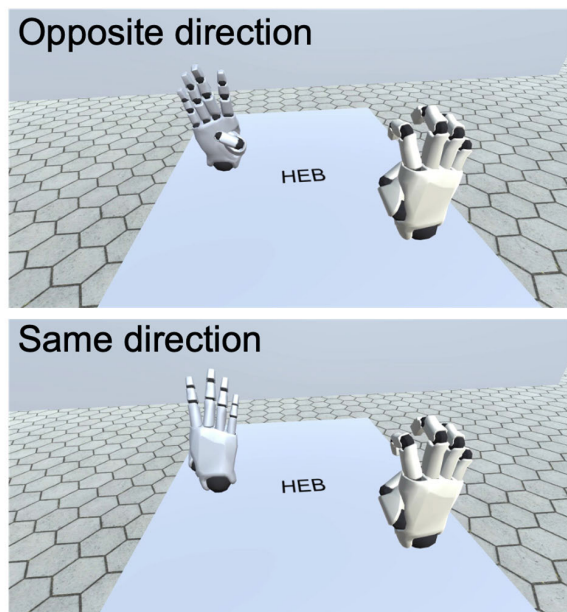
Participants wore a head-mounted display (HMD, HTC Vive Pro EYE, 1440  $\times$  1600 pixels per eye, 90  $\times$  110deg, 90 Hz refresh) and a glove with a hand motion capture system (Manus Prime II, ManusCore v1.9.0, sampling at 30 Hz) with a tracker (HTC Vive Tracker 3.0) on the right hand. The experiment was controlled by a computer (Intel Core i7 10700 2.9GHz CPU, NVIDIA GeForce RTX 3060 Graphics, DDR4 32GB memory). The virtual environment and experiment task were created using Unity (2020.3.9f1).

### D. STIMULI AND CONDITIONS

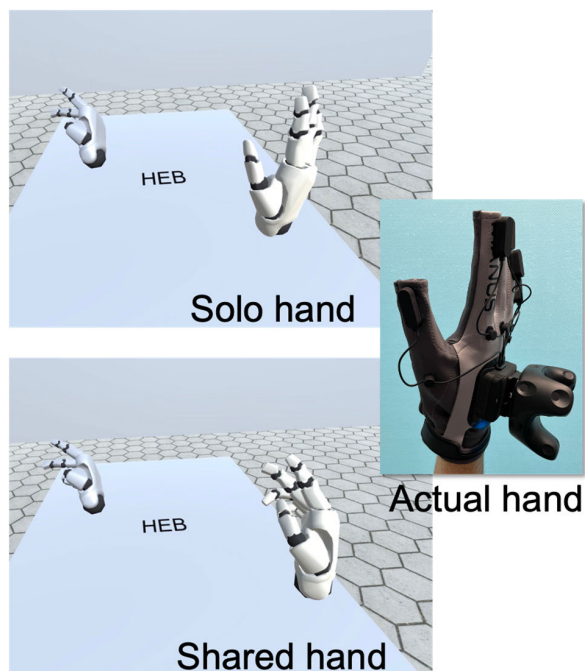
Twenty nonsense syllables/words represented by American manual alphabet were used as stimuli. The teacher's hand automatically moved based on pre-recorded data. Their duration was approximately 13 s (13.27 s in average, SD 1.97, Min 10.07, Max 15.53). The participant's virtual hand and the teacher's hand were presented side-by-side. The teacher's hand appeared either in the opposite (face-to-face; Figure 1 Top) or same direction (Figure 1 Bottom) as the participant. The participant's virtual hand was either moving with full control of the participant (solo avatar hand) or with the shared avatar hand, which was made by averaging the participant's hand movements with the teacher's pre-recorded hand movements in real-time (Figure 2). The teacher's virtual hand was not affected by the participant's hand movement in either condition. Thus, the conditions were  $2 \times 2$  (solo/shared avatar hand  $\times$  opposite/same direction).

### E. PROCEDURE

The participants were asked to keep their hand at an initial position with the initial posture. After maintaining at the initial position for 3 s, a teacher's hand appeared towards the left side of the participant's hand, and a stimulus word (three letters) was presented visually on a table. After 1 s, the teacher's hand moved to show the word with American manual alphabet. The participants were asked to imitate the movements of the teacher's hand as accurately as possible while observing their own avatar hand. A blocked design was used. Five randomly chosen words were used for each block in one of four conditions. In the end of each block/condition, the participants rated how easy it was to imitate in 7 levels (1: very difficult, 7: very easy). All conditions were repeated



**FIGURE 1.** Avatar-hand direction conditions (Top: opposite, Bottom: same directions). In scenes, the left one is the teacher's hand and the right one is participant's avatar hand. These are examples of scenes from the first-person perspective of participants who used their right hand.



**FIGURE 2.** Example scenes of the solo hand (top) and the shared hand (bottom) conditions. The right picture represents the actual hand of the participant. In the solo hand condition, the participant's avatar hand posture (right in scenes) was the same as the actual hand's posture, while in the shared hand condition, the posture of the participant's avatar hand was the average of the teacher's hand (left in scenes) and the actual hand postures.

5 times (sessions) in random order. The participants' hand motion was measured during experiments.

### III. RESULTS

#### A. SUBJECTIVE SCORES ON USABILITY

We firstly tested the normality of subjective data (7 levels discrete data) using Shapiro-Wilk test, and found that no dataset deviated from the normality (same direction with solo hand:  $W=0.960$ ,  $p=0.389$ , same direction with shared hand:  $W=0.954$ ,  $p=0.287$ , opposite direction with solo hand:  $W=0.963$ ,  $p=0.457$ , opposite direction with shared hand:  $W=0.958$ ,  $p=0.348$ ). Thus, we performed two-way repeated-measures ANOVA (solo/shared hand  $\times$  opposite/same direction) with rated scores. The interaction between avatar hand type and direction was not significant ( $F(1,25)=0.01$ ,  $p=0.99$ ,  $\eta_p^2=0.001$ ). We found main effects of solo/shared avatar hand ( $F(1,25)=30.21$ ,  $p<0.001$ ,  $\eta_p^2=0.547$ ) and hand directions ( $F(1,25)=10.44$ ,  $p=0.003$ ,  $\eta_p^2=0.295$ ). The post-hoc analyses showed that the score of the shared hand was higher than the solo hand ( $t(25)=5.496$ ,  $p<0.001$ ), and the score of the same direction was higher than the opposite direction ( $t(25)=3.231$ ,  $p=0.003$ ).

The easiness of imitation was significantly higher with the shared avatar hand than the solo hand, and higher with the same-direction presentation than the opposite-direction presentation (Figure 3). Thus, the shared avatar hand and the same-direction presentation were better than the own solo hand and the opposite-direction presentation, respectively.

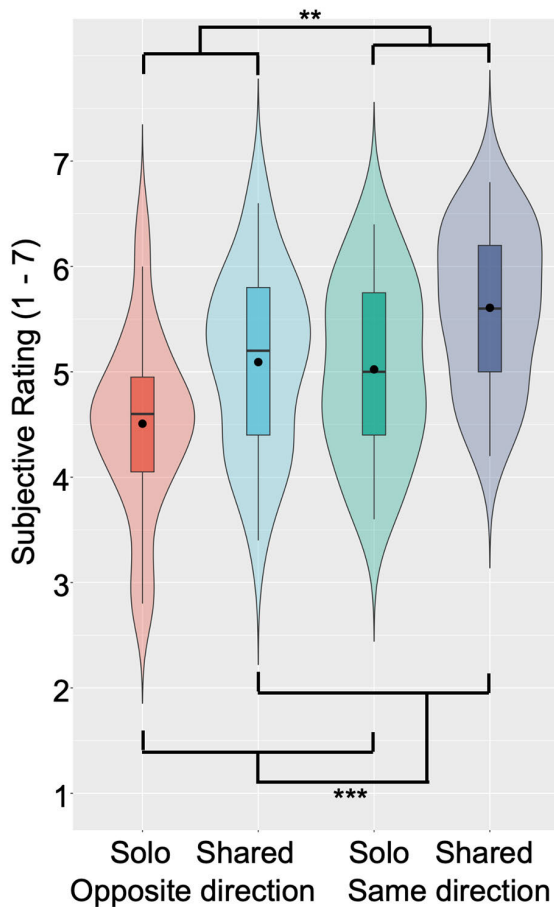
#### B. SPATIAL ERROR OF HAND MOVEMENTS

As a behavioral performance measure, we calculated the root mean square error (RMSE) by comparing the teacher's hand movement and the participant's (real) hand movement. Data of three joints of 5 fingers (15 joints in total) were sampled at 30 Hz, and the bending angle of each joint was calculated. Those data were averaged for each trial. We performed two-way repeated-measures ANOVA (solo/shared hand  $\times$  opposite/same direction) with RMSE. The interaction was not significant ( $F(1,25)=0.17$ ,  $p=0.69$ ,  $\eta_p^2=0.007$ ). We found a main effect of hand directions ( $F(1,25)=4.86$ ,  $p=0.037$ ,  $\eta_p^2=0.163$ ), but no main effect of avatar hand was observed ( $F(1,25)=0.47$ ,  $p=0.50$ ,  $\eta_p^2=0.018$ ). The post-hoc analysis showed that the RMSE of the same direction condition was larger than the opposite direction condition ( $t(25)=2.205$ ,  $p=0.037$ ).

The RMSE was significantly less with the opposite direction presentation than the same direction presentation (Figure 4). Thus, the opposite direction presentation of the teacher's hand was better than the same direction presentation, irrespective of the avatar type (solo or shared avatar hand). This is controversial compared with the subjective ratings.

#### C. TIME DELAY OF HAND MOVEMENTS

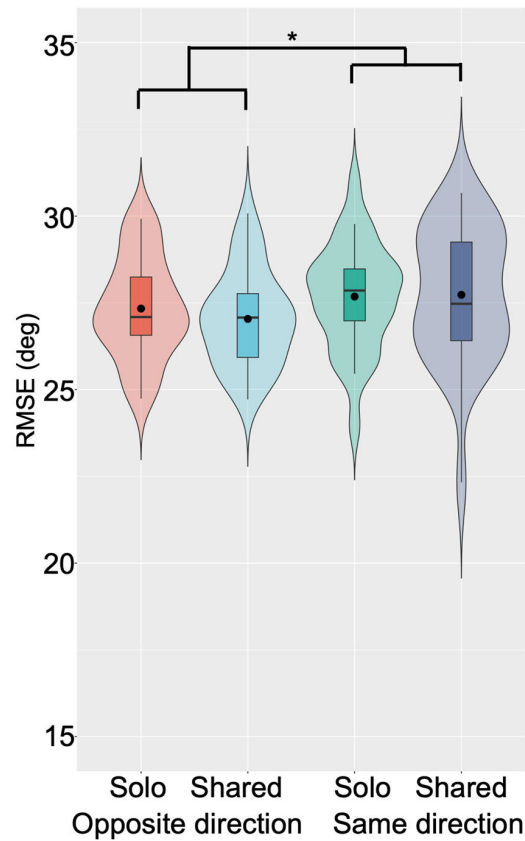
As another behavioral performance measure, we calculated the time-lag of which cross-correlation between the teacher's hand motion and the participant's hand motion shows the maximum value with limiting the lag range from -2 to +2 s.



**FIGURE 3.** Results of subjective ratings of usability. Data were plotted in box plots with violin plots. A box indicates the range between Q1 (25%) and Q3 (75%), the thick line in the box indicates the median (Q2), and a set of whiskers indicates the maximum and minimum values. The solid circle in the box indicates the average. A wide shape like a violin represents probability distribution by a kernel density estimator. \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

Data of three joints of 5 fingers (15 joints in total) were sampled at 30 Hz, the time-lag for the maximum correlation was calculated for each joint and each trial, and averaged with all joints for each trial. Positive time-lag shows that the participant’s hand motion was delayed compared to the teacher’s hand. We performed two-way repeated-measures ANOVA (solo/shared hand x opposite/same direction) with the time lag. The interaction was not significant ( $F(1,25)=0.04$ ,  $p=0.85$ ,  $\eta_p^2 = 0.002$ ). We found a significant main effect of avatar types ( $F(1,25)=7.50$ ,  $p=0.011$ ,  $\eta_p^2 = 0.231$ ), but no main effect of hand direction was observed ( $F(1,25)=1.30$ ,  $p=0.266$ ,  $\eta_p^2 = 0.049$ ). The post-hoc analysis showed that the time lag with the shared hand was less than the solo hand ( $t(25)=2.738$ ,  $p=0.011$ ).

The time delay of the participant’s movements was significantly less with the shared avatar hand than with the solo avatar hand (Figure 5). Thus, the shared avatar hand was better than the solo avatar hand, irrespective of the teacher’s hand direction. This is basically consistent with the subjective ratings for the shared avatar hand.

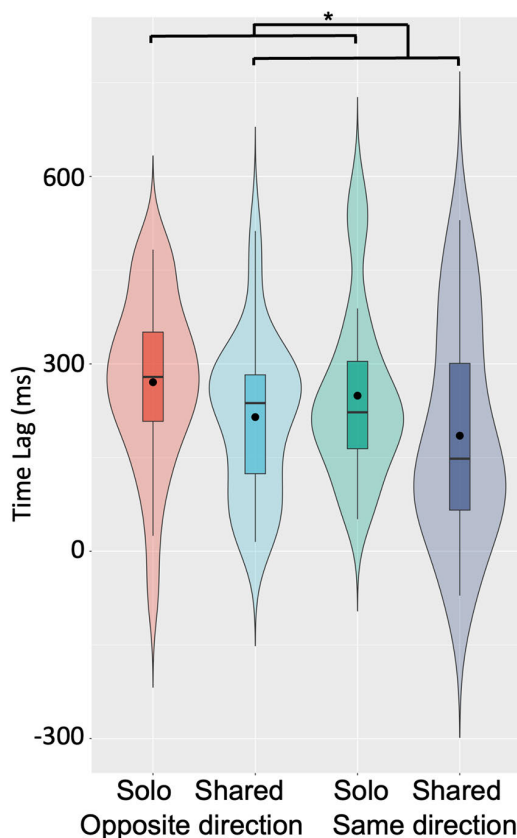


**FIGURE 4.** Results of spatial error of hand movements. Data were plotted in box plots with violin plots as same as in Figure 3. \* $p < 0.05$ .

#### IV. DISCUSSION

We developed a virtual hand imitation system combining the virtual shared (or co-embodiment) hand from the first-person perspective with the teacher’s hand presented in front of the learner and investigated the effects of the shared avatar hand and the presentation direction of the teacher’s hand on motor imitation performance and subjective impression with regards to usability (easiness in imitation). The subjective evaluation showed that the impression was better with the shared avatar hand with the teacher’s hand presented in the same direction compared to the own solo hand with the teacher’s hand presented in the opposite direction. The spatial error data showed that the opposite-direction presentation of the teacher’s hand was better than the same direction presentation, irrespective of the avatar hand type. The temporal delay data showed that the shared avatar hand was better than the own solo avatar hand, irrespective of the hand direction of the teacher.

The results of the subjective ratings and the temporal aspect of motor performance supported the hypothesis, H1 (the shared avatar hand could improve the imitation over the solo avatar hand). However, the spatial aspect of motor performance was not significantly improved by the shared avatar hand. According to Hagiwara et al., hand movement of the participants deviates from each other as the shared



**FIGURE 5.** Results of time delay of hand movements. Data were plotted in box plots with violin plots as same as in Figure 3 and 4. \* $p < 0.05$ .

avatar's hand moves to a goal [28]. The same effect is shown during motor learning [32]. Participants prioritize the motion of the shared avatar over their own movements. Therefore, the shared hand method might not be good for tasks requiring spatial accuracy of one's own (real) hand.

It should be noted that the fact that the delay is smaller in the shared hand condition than in the solo hand condition is not due to the methodological effect. The teacher's virtual hand was not affected by the participant's hand movement, and the movement of the teacher's hand was identical in both the shared and solo hand conditions. Moreover, we compared the participant's real hand movement, not the shared or solo avatar hand movement, with the teacher's hand movement.

As for H2 (the same direction presentation of the teacher's hand could provide better performance than the opposite direction presentation), it was supported by the results of the subjective ratings. However, the temporal aspect of motor performance was not significantly improved by the same-direction presentation and the spatial aspect of motor performance was better with the opposite-direction presentation inconsistently with the hypothesis. In the same-direction presentation, the detailed posture of the teacher's fingers may not be well observed by participants due to occlusion of the back of teacher's hand.

We employed sign languages gestures as stimuli for motor imitation because they include detailed movements with finger movements. Sign languages are visual languages with manual articulation. Many students learn American Sign Language in high schools, colleges, and universities [43]. Video materials are often used in addition to textbooks and in-person lectures. A sign language learning system based on two-dimensional image sampling with a convolutional neural network has been developed [44]. A game-based sign-language learning system has been developed, and it is shown that the learning system is better than the traditional face-to-face learning method [45]. It captures students' hand movements as three-dimensional data using Kinect motion capture, compares them with teachers' stored data, and provides feedback on similarity scores. With these methods, students can learn sign language alone with appropriate feedback. Our shared avatar hand can be a potential application for a sign-language learning system because the shared avatar hand could improve subjective usability and time delay of the motor imitation of detailed hand movements.

However, the focus of our current study was on the performance of imitation, not on the motor skill learning process. This is a limitation of the current study. In the future, we should investigate the motor skill learning process of detailed hand movements, which could contribute to developing a novel learning system for sign language or a skill transfer system of manual arts and crafts.

The sample of participants in this study had a gender and age bias, with the majority of participants being male university students. Conducting future research with a more diverse population, both in terms of gender and age, may yield results that are more universally applicable [46].

In our study, movements of the participants were averaged with pre-recorded movements of the teacher. Thus, the teacher's movements were not affected by the participant's movement. In contrast, previous studies on virtual co-embodiment involved two persons interacting with each other through the shared avatar [20], [21], [22], [28], [32], [33], [34]. In our experiment, we pre-recorded the teacher's movements aiming for ideal movements that remain independent from the participant's performance. Our method has an advantage in motor imitation when ideal movements are well defined. However, it is not clear how mutual interaction between a teacher and a learner in virtual co-embodiment affects motor imitation. This is another limitation of our study.

## V. CONCLUSION

This is the first study to apply virtual co-embodiment to hand and finger movements. Movements of a participant were averaged with a teacher's pre-recorded movements and presented to the participant as a shared avatar hand. Participants were asked to imitate the teacher's hand that appeared in front of them while observing their own avatar hand, which was either a shared avatar hand (moving with average movements of the participant and the teacher) or a solo avatar hand (full control

for the participant), as accurately as possible. The subjective usability and the spatial and temporal performances of motor imitation were measured. The shared avatar hand improved subjective usability and temporal delay of motor imitation. When the teacher's hand was presented in the same direction as the participant's own avatar hand, the subjective usability improved. On the other hand, when the teacher's hand was presented in the opposite direction to the participant's own avatar hand (like face-to-face), the spatial accuracy of motor imitation improved. These findings support the advantage of using shared avatar hands for motor imitation tasks. The shared avatar hand concept may contribute to developing efficient learning systems for sign language and other skill transfer systems for manual arts and crafts in the future. Further research is needed to explore the effects of the shared avatar hand on the motor learning process.

## ACKNOWLEDGMENT

An earlier version of this paper was presented in part at the 2022 International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments (ICAT-EGVE) [47].

## REFERENCES

- [1] M. Botvinick and J. Cohen, "Rubber hands 'feel' touch that eyes see," *Nature*, vol. 391, no. 6669, p. 756, Feb. 1998, doi: [10.1038/35784](https://doi.org/10.1038/35784).
- [2] T. Dummer, A. Picot-Annand, T. Neal, and C. Moore, "Movement and the rubber hand illusion," *Perception*, vol. 38, no. 2, pp. 271–280, Feb. 2009, doi: [10.1068/p5921](https://doi.org/10.1068/p5921).
- [3] A. Kalckert and H. H. Ehrsson, "Moving a rubber hand that feels like your own: A dissociation of ownership and agency," *Frontiers Hum. Neurosci.*, vol. 6, p. 40, Mar. 2012, doi: [10.3389/fnhum.2012.00040](https://doi.org/10.3389/fnhum.2012.00040).
- [4] A. Kalckert and H. H. Ehrsson, "The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership," *Consciousness Cogn.*, vol. 26, pp. 117–132, May 2014, doi: [10.1016/j.concog.2014.02.003](https://doi.org/10.1016/j.concog.2014.02.003).
- [5] K. Kiltner, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence, Teleoperators Virtual Environ.*, vol. 21, no. 4, pp. 373–387, Nov. 2012, doi: [10.1162/PRES\\_a\\_00124](https://doi.org/10.1162/PRES_a_00124).
- [6] T. C. Peck and M. Gonzalez-Franco, "Avatar embodiment. A standardized questionnaire," *Frontiers Virtual Reality*, vol. 1, Feb. 2021, Art. no. 575943, doi: [10.3389/frvir.2020.575943](https://doi.org/10.3389/frvir.2020.575943).
- [7] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater, "Putting yourself in the skin of a black avatar reduces implicit racial bias," *Consciousness Cogn.*, vol. 22, no. 3, pp. 779–787, Sep. 2013, doi: [10.1016/j.concog.2013.04.016](https://doi.org/10.1016/j.concog.2013.04.016).
- [8] M. Martini, D. Perez-Marcos, and M. V. Sanchez-Vives, "What color is my arm? Changes in skin color of an embodied virtual arm modulates pain threshold," *Frontiers Hum. Neurosci.*, vol. 7, p. 438, Jul. 2013, doi: [10.3389/fnhum.2013.00438](https://doi.org/10.3389/fnhum.2013.00438).
- [9] D. Banakou, R. Groten, and M. Slater, "Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 31, pp. 12846–12851, Jul. 2013, doi: [10.1073/pnas.1306779110](https://doi.org/10.1073/pnas.1306779110).
- [10] A. Tajadura-Jiménez, D. Banakou, N. Bianchi-Berthouze, and M. Slater, "Embodiment in a child-like virtual body influences object size perception, self-identification, and subsequent real speaking," *Sci. Rep.*, vol. 7, no. 1, p. 9637, Aug. 2017, doi: [10.1038/s41598-017-09497-3](https://doi.org/10.1038/s41598-017-09497-3).
- [11] B. van der Hoort, A. Guterstam, and H. H. Ehrsson, "Being Barbie: The size of one's own body determines the perceived size of the world," *PLoS ONE*, vol. 6, no. 5, May 2011, Art. no. e20195, doi: [10.1371/journal.pone.0020195](https://doi.org/10.1371/journal.pone.0020195).
- [12] S. Schulze, T. Pence, N. Irvine, and C. Guinn, "The effects of embodiment in virtual reality on implicit gender bias," in *Virtual, Augmented and Mixed Reality. Multimodal Interaction*. Cham, Switzerland: Springer, Jun. 2019, pp. 361–374, doi: [10.1007/978-3-030-21607-8\\_28](https://doi.org/10.1007/978-3-030-21607-8_28).
- [13] A. Krekhov, S. Cmentowski, and J. Krüger, "The illusion of animal body ownership and its potential for virtual reality games," in *Proc. IEEE Conf. Games (CoG)*, Aug. 2019, pp. 1–8, doi: [10.1109/CIG.2019.8848005](https://doi.org/10.1109/CIG.2019.8848005).
- [14] A. Guterstam, Z. Abdulkarim, and H. H. Ehrsson, "Illusory ownership of an invisible body reduces autonomic and subjective social anxiety responses," *Sci. Rep.*, vol. 5, no. 1, p. 9831, Apr. 2015, doi: [10.1038/srep09831](https://doi.org/10.1038/srep09831).
- [15] B. van der Hoort and H. H. Ehrsson, "Illusions of having small or large invisible bodies influence visual perception of object size," *Sci. Rep.*, vol. 6, no. 1, p. 34530, Oct. 2016, doi: [10.1038/srep34530](https://doi.org/10.1038/srep34530).
- [16] R. Kondo, M. Sugimoto, K. Minamizawa, T. Hoshi, M. Inami, and M. Kitazaki, "Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity," *Sci. Rep.*, vol. 8, no. 1, p. 7541, May 2018, doi: [10.1038/s41598-018-25951-2](https://doi.org/10.1038/s41598-018-25951-2).
- [17] R. Kondo, Y. Tani, M. Sugimoto, M. Inami, and M. Kitazaki, "Scrambled body differentiates body part ownership from the full body illusion," *Sci. Rep.*, vol. 10, no. 1, p. 5274, Mar. 2020, doi: [10.1038/s41598-020-62121-9](https://doi.org/10.1038/s41598-020-62121-9).
- [18] R. Kondo, Y. Tani, M. Sugimoto, K. Minamizawa, M. Inami, and M. Kitazaki, "Re-association of body parts: Illusory ownership of a virtual arm associated with the contralateral real finger by visuo-motor synchrony," *Frontiers Robot. AI*, vol. 7, p. 26, Mar. 2020, doi: [10.3389/frobt.2020.00026](https://doi.org/10.3389/frobt.2020.00026).
- [19] K. Arai, H. Saito, M. Fukuoka, S. Ueda, M. Sugimoto, M. Kitazaki, and M. Inami, "Embodiment of supernumerary robotic limbs in virtual reality," *Sci. Rep.*, vol. 12, no. 1, p. 9769, Jun. 2022, doi: [10.1038/s41598-022-13981-w](https://doi.org/10.1038/s41598-022-13981-w).
- [20] R. Fribourg, N. Ogawa, L. Hoyet, F. Argelaguet, T. Narumi, M. Hirose, and A. Lécuyer, "Virtual co-embodiment: Evaluation of the sense of agency while sharing the control of a virtual body among two individuals," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 10, pp. 4023–4038, Oct. 2021, doi: [10.1109/TVCG.2020.2999197](https://doi.org/10.1109/TVCG.2020.2999197).
- [21] H. Hapuarachchi, T. Hagiwara, G. Ganesh, and M. Kitazaki, "Effect of connection induced upper body movements on embodiment towards a limb controlled by another during virtual co-embodiment," *PLoS ONE*, vol. 18, no. 1, Jan. 2023, Art. no. e0278022, doi: [10.1371/journal.pone.0278022](https://doi.org/10.1371/journal.pone.0278022).
- [22] H. Hapuarachchi and M. Kitazaki, "Knowing the intention behind limb movements of a partner increases embodiment towards the limb of joint avatar," *Sci. Rep.*, vol. 12, no. 1, p. 11453, Jul. 2022, doi: [10.1038/s41598-022-15932-x](https://doi.org/10.1038/s41598-022-15932-x).
- [23] R. Miura, S. Kasahara, M. Kitazaki, A. Verhulst, M. Inami, and M. Sugimoto, "MultiSoma: Motor and gaze analysis on distributed embodiment with synchronized behavior and perception," *Frontiers Comput. Sci.*, vol. 4, May 2022, Art. no. 788014, doi: [10.3389/fcomp.2022.788014](https://doi.org/10.3389/fcomp.2022.788014).
- [24] A. Verhulst, Y. Namikawa, and S. Kasahara, "Parallel adaptation: Switching between two virtual bodies with different perspectives enables dual motor adaptation," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2022, pp. 169–177, doi: [10.1109/ISMAR55827.2022.00031](https://doi.org/10.1109/ISMAR55827.2022.00031).
- [25] K. Takada, M. Kawaguchi, A. Uehara, Y. Nakanishi, M. Armstrong, A. Verhulst, K. Minamizawa, and S. Kasahara, "Parallel ping-pong: Exploring parallel embodiment through multiple bodies by a single user," in *Proc. Augmented Hum.* New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 121–130, doi: [10.1145/3519391.3519408](https://doi.org/10.1145/3519391.3519408).
- [26] M. Inami, D. Uriu, Z. Kashino, S. Yoshida, H. Saito, A. Maekawa, and M. Kitazaki, "Cyborgs, human augmentation, cybernetics, and JIZAI body," in *Proc. Augmented Hum.* New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 230–242, doi: [10.1145/3519391.3519401](https://doi.org/10.1145/3519391.3519401).
- [27] M. Inami, Ed., *Theory of JIZAI Body: Towards Mastery Over the Extended Self*. Singapore: Springer, 2023, doi: [10.1007/978-981-99-2003-7](https://doi.org/10.1007/978-981-99-2003-7).
- [28] T. Hagiwara, G. Ganesh, M. Sugimoto, M. Inami, and M. Kitazaki, "Individuals prioritize the reach straightness and hand jerk of a shared avatar over their own," *iScience*, vol. 23, no. 12, Dec. 2020, Art. no. 101732, doi: [10.1016/j.isci.2020.101732](https://doi.org/10.1016/j.isci.2020.101732).
- [29] U. Yang and G. J. Kim, "Implementation and evaluation of 'just follow me': An immersive, VR-based, motion-training system," *Presence, Teleoperators Virtual Environ.*, vol. 11, no. 3, pp. 304–323, Jun. 2002, doi: [10.1162/105474602317473240](https://doi.org/10.1162/105474602317473240).
- [30] M. L. Chenechal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "Vishnu: Virtual immersive support for HelpiNg users an interaction paradigm for collaborative remote guiding in mixed reality," in *Proc. IEEE 3rd VR Int. Workshop Collaborative Virtual Environ. (3DCVE)*, Mar. 2016, pp. 9–12, doi: [10.1109/3DCVE.2016.7563559](https://doi.org/10.1109/3DCVE.2016.7563559).

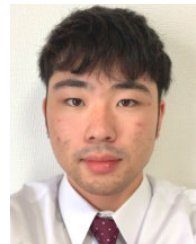
- [31] M. Y. Saraiji, T. Sasaki, R. Matsumura, K. Minamizawa, and M. Inami, "Fusion: Full body surrogacy for collaborative communication," in *Proc. ACM SIGGRAPH Emerg. Technol.* New York, NY, USA: Association for Computing Machinery, Aug. 2018, pp. 1–2, doi: [10.1145/3214907.3214912](https://doi.org/10.1145/3214907.3214912).
- [32] D. Kodama, T. Mizuho, Y. Hatada, T. Narumi, and M. Hirose, "Effects of collaborative training using virtual co-embodiment on motor skill learning," *IEEE Trans. Vis. Comput. Graphics*, vol. 29, no. 5, pp. 2304–2314, May 2023, doi: [10.1109/TVCG.2023.3247112](https://doi.org/10.1109/TVCG.2023.3247112).
- [33] D. Kodama, T. Mizuho, Y. Hatada, T. Narumi, and M. Hirose, "Effect of weight adjustment in virtual co-embodiment during collaborative training," in *Proc. Augmented Hum. Conf.* New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 86–97, doi: [10.1145/3582700.3582703](https://doi.org/10.1145/3582700.3582703).
- [34] H. Takita, Y. Hatada, T. Narumi, and M. Hirose, "Effects of virtual co-embodiment on declarative memory-based motor skill learning," in *Proc. ACM Symp. Appl. Perception*. New York, NY, USA: Association for Computing Machinery, Aug. 2023, pp. 1–10, doi: [10.1145/3605495.3605797](https://doi.org/10.1145/3605495.3605797).
- [35] S. S. Jones, "Imitation in infancy: The development of mimicry," *Psychol. Sci.*, vol. 18, no. 7, pp. 593–599, Jul. 2007.
- [36] K. A. Duffy and T. L. Chartrand, "Mimicry: Causes and consequences," *Current Opinion Behav. Sci.*, vol. 3, pp. 112–116, Jun. 2015.
- [37] R. B. van Baaren, R. W. Holland, K. Kawakami, and A. van Knippenberg, "Mimicry and prosocial behavior," *Psychol. Sci.*, vol. 15, no. 1, pp. 71–74, Jan. 2004.
- [38] G. F. Darden, "Demonstrating motor skills—Rethinking that expert demonstration," *J. Phys. Educ., Recreation Dance*, vol. 68, no. 6, pp. 31–35, Aug. 1997.
- [39] G. Goldenberg and H.-O. Karnath, "The neural basis of imitation is body part specific," *J. Neurosci.*, vol. 26, no. 23, pp. 6282–6287, Jun. 2006.
- [40] C. L. Hull, "The meaningfulness of 320 selected nonsense syllables," *Amer. J. Psychol.*, vol. 45, no. 4, pp. 730–734, 1933.
- [41] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, May 2007, doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146).
- [42] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses," *Behav. Res. Methods*, vol. 41, no. 4, pp. 1149–1160, Nov. 2009.
- [43] D. Quinto-Pozos, "Teaching American sign language to hearing adult learners," *Annu. Rev. Appl. Linguistics*, vol. 31, pp. 137–158, Mar. 2011.
- [44] Y. Ji, S. Kim, and K.-B. Lee, "Sign language learning system with image sampling and convolutional neural network," in *Proc. 1st IEEE Int. Conf. Robot. Comput. (IRC)*, Apr. 2017, pp. 371–375.
- [45] T. Kamnardsiri, L. O. Hongsit, P. Khuwuthyakorn, and N. Wongta, "The effectiveness of the game-based learning system for the improvement of American sign language using Kinect," *Electron. J. e-Learn.*, vol. 15, no. 4, pp. 283–296, 2017.
- [46] T. C. Peck, L. E. Sockol, and S. M. Hancock, "Mind the gap: The underrepresentation of female participants and authors in virtual reality research," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 1945–1954, May 2020.
- [47] Y. Katsumata, H. Ishimoto, Y. Inoue, and M. Kitazaki, "Sign language learning system with concurrent shared avatar hand in a virtual environment: Psychological evaluation," in *Proc. Int. Conf. Artif. Reality Telexistence Eurograph. Symp. Virtual Environ. (ICAT-EGVE)*, Nov. 2022, pp. 27–28, doi: [10.2312/egve.20221300](https://doi.org/10.2312/egve.20221300).



**YASUNOBU KATSUMATA** received the B.E. and M.E. degrees in knowledge-based information engineering from the Toyohashi University of Technology, Aichi, Japan, in 2008 and 2010, respectively, where he is currently pursuing the Ph.D. degree in computer science and engineering.



**YASUYUKI INOUE** received the B.E., M.E., and Ph.D. degrees from the Graduate School of Engineering, Toyohashi University of Technology, in 2003, 2005, and 2010, respectively. He was a Project Assistant Professor with The University of Electro-Communications, a Postdoctoral Fellow with Mie University and The University of Tokyo, and a Project Assistant Professor of computer science and engineering with the Toyohashi University of Technology. Since 2023, he has been a Lecturer with the Faculty of Engineering, Department of Information Systems Engineering, Toyama Prefectural University.



**SATOSHI TORIUMI** received the degree from the National Institute of Technology, Ichinoseki College, in 2022. He is currently pursuing the B.E. degree in computer science and engineering with the Toyohashi University of Technology, Aichi, Japan.



**HIROKI ISHIMOTO** received the B.E. and M.E. degrees in computer science and engineering from the Toyohashi University of Technology, Aichi, Japan, in 2020 and 2022, respectively.



**HARIN HAPUARACHCHI** received the B.E. and M.E. degrees in computer science and engineering from the Toyohashi University of Technology, Aichi, Japan, where he is currently pursuing the Ph.D. degree in computer science and engineering.



**MICHIATERU KITAZAKI** received the B.A., M.A., and Ph.D. degrees in experimental psychology from The University of Tokyo, Tokyo, Japan, in 1992, 1994, and 1997, respectively.

From 1997 to 2000, he was a Research Associate with the Department of Psychology, The University of Tokyo. From 2000 to 2003, he was a Lecturer with the Toyohashi University of Technology, Aichi, Japan, where he was an Associate Professor, from 2003 to 2016, and has been a Professor with the Department of Computer Science and Engineering, since 2017.

...