Correction of Avatar Hand Movements Supports Learning of a Motor Skill

Lilija, Klemen; Kyllingsbæk, Søren; Hornbæk, Kasper

Published in:
2021 IEEE Virtual Reality and 3D User Interfaces (VR)

DOI:
10.1109/VR50410.2021.00069

Publication date:
2021

Document license:
Other

Citation for published version (APA):
Correction of Avatar Hand Movements Supports Learning of a Motor Skill

Klemen Lilija* Søren Kyllingsbæk† Kasper Hornbæk‡

University of Copenhagen

Abstract

Learning to move the hands in particular ways is essential in many training and leisure virtual reality applications, yet challenging. Existing techniques that support learning of motor movement in virtual reality rely on external cues such as arrows showing where to move or transparent hands showing the target movement. We propose a technique where the avatar’s hand movement is corrected to be closer to the target movement. This embeds guidance in the user’s avatar, instead of in external cues and minimizes visual distraction. Through two experiments, we found that such movement guidance improves the short-term retention of the target movement when compared to a control condition without guidance.

Index Terms: Human-centered computing—Empirical studies in HCI; Human-centered computing—Virtual reality

1 Introduction

Sports, games, artistic endeavors, and gestural communication all benefit from precise hand movements. However, learning complex movements is difficult and we often benefit from some type of guidance. Guidance such as playback of the target movement, or arrows rendered in virtual reality (VR), can help the user learn a movement. However, guidance can also be ineffective or distracting if the cues hide content in a scene or visually interfere with it. Further, the guidance might be confusing, or the users might get used to it to such an extent that they perform poorly once the guidance is removed. Such over-reliance on guidance can negatively affect the use of VR applications since users cannot train indefinitely. Thus, it is crucial to design appropriate guidance to benefit the user.

Many techniques have been proposed for supporting the learning of hand movements in VR. The majority of them provide augmented feedback around the user’s avatar. Such feedback includes, an extra pair of hands showing where the user should move next [8], a mirrored view of the user with cues overlaid on it [1], or abstract cues such as arrows guiding a user towards a target.

We investigate a technique that embeds the guidance directly in the avatar’s movement. The technique corrects the user’s virtual movement when it differs from the target movement (Figure 1). The user’s virtual hand location and pose are corrected to be closer to the target movement. Such correction introduces a misalignment between the user’s actual hand and its representation in virtual reality. This misalignment then serves as a guide towards the target movement. Such guidance immediately improves the user’s virtual performance, avoids introducing external visual cues, and might be used to subtly impact the user’s movement during learning.

We compare different degrees of virtual movement correction to a control condition and a conventional guidance technique. In two experiments, we found that the virtual movement correction improved the short-term retention of the trained movements. The technique outperformed the control condition in which the users did not receive any guidance and performed equally well as a conventional VR guidance technique (viz., ghosted hand). Our results suggest that the correction of virtual movement is a viable method for supporting hand movement learning. The applications that could especially benefit from it are those where minimizing visual distraction is critical (e.g., communication with gestures), as well as applications that require good virtual performance during training.

2 Related Work

Researchers have explored several VR techniques for providing guidance during a motor task. Application domains vary from learning VR painting [25], Tai-Chi practice [10], conducting [8], tennis and table tennis playing [12,26], engine assembly [11], calligraphy [30], rehabilitation and physiotherapy [23,24], as well as augmenting social interactions [20]. The review paper by Sigrist et al. gave a comprehensive overview of different feedback types that the user can receive during motor learning [21]. Apart from categorizing the types of augmented feedback, they discussed evaluation methods, motor learning theories, and the impact of different feedback strategies. According to their categorization, our technique falls within training with concurrent visual feedback in a complex task scenario. In the next section, we list the techniques and systems that are related to our work.

2.1 Techniques and Systems

Most existing techniques that provide visual feedback during training place the guiding cues in the world around the user. For example, EGuide rendered an extra pair of hands in the user’s periphery for egocentric guidance of the user [8]. Similarly, the Just Follow Me...
system guided the user during calligraphy by placing an additional translucent brush (i.e., ghost brush) in the virtual environment [30]. Systems such as YouMove [1], MotionMA [27], SleeveAR [23], and Physio@Home [24] placed cues in the world from an allocentric perspective. They provided a mirrored view of the user with added cues to serve as guides. The cues varied from realistic representations of the user [29], abstract representation of the user (e.g., rendering of a skeleton in YouMove [1]), to use of abstract symbols (e.g., arrows and circles in SleeveAR [23]).

Previous work has also explored placing cues on the user’s body and manipulation of the user’s avatar to help execute and learn motor movement. LightGuide projected visualizations on users’ real hands to guide their movement [22]. Visualizations such as 2D and 3D arrows on the hand’s surface helped users execute simple mid-air movements. Work on internatural skill transfer [19, 28] looked at the effects of manipulating the user’s avatar on motor learning. For example, Xiao et al. [28] showed that practicing with one hand while seeing the same movement in the other hand, helps improve performance for the unpracticed hand. Similarly, Ossmy and Mukamel showed that seeing the movement in immobile hand as well as yoking it to passively follow the practicing hand further improves the motor learning [19]. Such visual manipulations of the user’s avatar are similar to the ones we apply to the practicing hand.

2.2 Problem and Evaluation

The problem domain we focus on is how to best support motor learning by providing visual guidance during training. Sigrist et al. characterized motor learning as a lasting increase of performance after training, that is, once the guidance is removed. To evaluate the effectiveness of different guidance techniques, they suggested testing the short- and long-term retention of the practiced movement. Much of the related work performed such an evaluation (e.g., [1, 26, 30], while some did not (e.g., [6, 8, 22]). The latter techniques therefore fall in a different problem domain or need further evaluation to establish if they are suitable for motor learning.

2.3 Theoretical Background

Apart from specific techniques, previous work also investigated the general influence of avatar manipulations. Gonzalez-Franco et al. [9] reported the self-avatar follower effect. They found that, under certain conditions, participants followed their avatar when it did not overlay with their physical body, without being explicitly instructed to do so. Cohn et al. continued their work in SnapMove in which they mapped arbitrary physical reaching movement to a single virtual movement and reported on drift of participant’s hand toward the avatar’s movement [7]. The opposite effect was observed already in the 1960s by Nielsen [18]. In his experiment, the participants drifted away from their illusory hand during a motor task. This effect has been frequently exploited in motion retargeting work [2, 5, 14, 15]. For example, Azmandian et al. [2] shifted the user’s virtual arm and hand during the reaching movement to redirect the user’s reach towards a physical prop. Similarly, Bergström et al. [5] warped the user’s fingers during the reaching movement to resize the users grasp.

These contrary reactions to visual manipulation could potentially be explained by predictive coding models where the conflicts are minimized at different hierarchical levels [16]. That is, normally the conflicts between proprioception and visual feedback are minimized. However, in case of a visual task that does not allow this, the users decrease their reliance on proprioception and increase their reliance on the visual feedback. In such a task, the minimization of visual conflicts becomes more important than the minimization of proprioceptive conflicts. We designed our hand guidance technique to produce a self-avatar follower effect and support the minimization of the virtual and real hand movement’s misalignment.

3 Correction of Virtual Hand Movement

When learning a new hand movement, the user first builds up a motor program [21]. Learning is often initiated by showing the user a visual demonstration of the target movement. The movement program is then later refined through training. Initially, users make many errors and often need feedback to correct their movement. Visual guidance provided during training can help in this phase of learning. It can prevent cognitive overload, remind the users of the next movement in a sequence, and help them notice and avoid errors.

The key idea of our approach is to correct the virtual representation of the user’s hand movement to be closer to the target movement. The correction depends on the size of the movement error. When the user’s hand location and pose is far removed from the target one, the correction is large; when the subject is doing well the correction is smaller.

The implementation of our approach has two parts, we first predict where the user’s hand should be at any moment during the movement, and second, we correct the hand’s virtual representation.

Depending on the type of the target movement, the prediction can either be simple or complex. If the target movement requires to move a hand from one point to another in a straight line, only moving forward (e.g., as in SnapMove [7]), then it is relatively simple to predict where the hand is supposed to be at any point during the movement. In contrast, if the target movement is poorly defined and requires intricate hand movements (e.g., sign language), the prediction is more complicated. For our study, we selected a target movement that requires a high degree of precision while being simple to predict. The participants had to trace an invisible line in a single movement going from left to right (see Figure 1 for an example target movement). We predicted where the hand should be, by using the participant’s hand location on the left-to-right axis and projected it on the target movement.

Once we can predict with high accuracy where the user’s hand needs to be, we can correct its virtual representation. We can either fully correct the virtual hand by moving it to the predicted location and pose, or we can partially correct it by moving it somewhere between the predicted and actual location. In our initial study, we used three variations of movement correction to investigate if they support learning of a motor movement.

4 Study 1

We conducted an exploratory study to investigate the usefulness of virtual movement correction for learning of hand movement patterns. We were interested if such guidance improves the short-term retention of the trained movement.

4.1 Participants

We recruited 30 participants via an online Oculus Quest community.

The study was administered remotely by using participants’ personal Oculus Quest headsets. We discarded data from participants that experienced poor hand tracking quality. We used jitter experienced during the measured repetitions as a proxy for the quality of hand tracking. If a participant experienced more than ten virtual hand moves larger than 20 cm from frame to frame during the experiment then their data was discarded. With this criteria, we discarded the data of 8 participants. The average age of the 22 participants left (all male) was 31.5 years (SD = 9.3).

4.2 Hand Guidance Techniques

We had three conditions of virtual movement correction: interpolation50, interpolation75 and snapping. Additionally, we had a ghost condition (Figure 2) where the participant’s movement was guided.  

---

1https://www.reddit.com/r/oculus/
We used a within-subject design and a Graeco-Latin square to balance the five conditions and the five target movements the participants needed to replicate (8). A control condition in which the participants did not receive any guidance.

The five experimental conditions were implemented as follows: control: Participant’s virtual hand was always rendered at the hand’s actual location. Participants did not receive any guidance during training.

ghost: The participant’s virtual hand was rendered at its actual location, while a ghosted hand (i.e., an additional translucent hand) indicated where to move next (Figure 2). The ghosted hand therefore moved with the same speed as the participant’s hand while always showing the next location and pose in the target movement.

interpolation50: The participant’s virtual hand was rendered 50% between the target location (i.e., where the hand should be) and the actual location of the participant’s hand. Similarly, the virtual hand’s pose was interpolated to be mid-way between the target pose and the participant’s actual hand pose.

interpolation75: The participant’s virtual hand was rendered 75% of the way between the target location and the actual location of the participants’ hand. The virtual hand’s pose was also interpolated to be three-quarters of the way between the target pose and the participant’s actual pose.

snapping: The participants’ virtual hand was rendered at (i.e., snapped to) the target location and in the target pose throughout the task.

All conditions were set in an identical virtual environment with minimal visual distractions. The virtual floor was textured with a grid pattern to aid the depth perception (see Figure 2).

4.3 Design

We used a within-subject design and a Graeco-Latin square to balance the five conditions and the five target movements the participants were instructed to repeat (see Figure 3 for target paths).

4.4 Procedure

Each experimental condition consisted of a training block and a test block. The participants started with the training block in which they needed to replicate a target movement in mid-air as many times as possible within 40 seconds. The target movement was shown to them before the training by an animated hand performing the movement three times. The index finger tip of the animated hand traced one of the target paths (see Figure 3 for the target paths); the target path was not visible during the animation. Once the participants started the training, they were guided by using one of the five techniques described in Section 4.2. After completing the training block, the participants were given a Likert-scale question about the task difficulty. Once they answered the question, they started the test block (i.e., short-term retention test), where they were instructed to repeat the movement from the training block as many times as possible within 20 seconds. Participants were not shown the target movement again before starting the test block, nor did they receive any guidance during the test block. They could only rely on what they learned in the training block.

This procedure was the same for each condition; each participant went through it a total of five times. Once the participants completed the experiment, the study application sent the logged movement data to a remote server. Finally, they were asked to fill a post-study questionnaire and receive a Steam game worth $15 as a reward.

5 Study 1: Results

The main performance metric used to compare the conditions was the accuracy of the executed movements. We used mean squared error (MSE) as a proxy for accuracy and compared the absolute performance in the training block, the absolute performance in the test block (i.e., short-term retention test), and the performance increase from the training block to the test block. We also investigated how performance changed over time, the perceived task difficulty, and compared the total number of repetitions the participants executed per condition. For the analysis, we used linear mixed effects models (LMM) due to the advantages over more commonly used ANOVA [3, 4, 17].

5.1 Number of Repetitions

The LMM analysis did not reveal any significant differences in the number of executed repetitions among the conditions for neither the training block (F(4, 84) = 2.167, p = 0.08) nor the test block (F(4, 84) = 2.167, p = 0.08). In the training block, the participants on average executed 12.15 repetitions (SD = 3.13), while in the test block the mean number of repetitions was 5.81 (SD = 1.57). Figure 4 shows the mean number of repetitions per condition for the training and test block.

Lack of differences in the number of repetitions indicates that participants executed the mid-air movements with similar speed, independent of the guidance technique used in the training block.

https://store.steampowered.com/
Figure 4: Mean number of repetitions per condition for the training and test block in Study 1. The error bars represent the standard error of the mean.

5.2 Movement Accuracy

We compared the accuracy of the executed movements between the conditions to see if any of the guidance techniques were better at supporting learning of the target movements. The accuracy measure we used was the mean squared error (MSE) of the executed movement over the target movement. We calculated the error by squaring the difference between the index finger’s logged position and the target position (i.e., where on the target path the index finger should be) for each point on the target path. To calculate the MSE of a repetition, we summed all the errors and divided the sum with the number of them. To calculate the MSE for a specific condition we averaged the MSE across all the repetitions for a participant and then averaged again across all participants. Figure 6 shows the MSE for each of the conditions for the training and test block, and the performance improvement (i.e., decrease in MSE) from the training to test block.

To see if there are significant differences between conditions, we conducted an LMM analysis. The analysis of absolute accuracy in the training block did not reveal a main effect of condition on MSE ($F(4, 84) = 1.313, p = 0.272$). Thus, participants did not significantly benefit from any guidance during the training.

The analysis of absolute accuracy in the test block found a main effect of condition on MSE ($F(4, 84) = 3.931, p < 0.01$). A post hoc pairwise comparison test with Bonferroni adjustment showed significantly lower MSE for the snapping condition when compared to the control condition ($p = 0.046$).

The analysis of accuracy improvement from the training to test block found a main effect of condition on MSE improvement ($F(4, 84) = 5.169, p < 0.001$). The post hoc pairwise comparison with Bonferroni adjustment found that accuracy in snapping and interpolation75 conditions significantly improved ($p = 0.005$ and $p = 0.012$, respectively) when compared to the control condition.

To better understand the size of the movement errors, Figure 5 shows a movement plot and MSE of a select participant in the snapping condition. The participant’s accuracy in the training block was poor and good in the test block.

5.3 Perceived Difficulty

After the training in each experimental condition, the participants answered a Likert-scale question on task difficulty. The LMM analysis did not reveal any significant differences in perceived difficulty among conditions ($F(4, 84) = 0.773, p = 0.546$).

5.4 Learning Effect

To investigate the learning effect, we used the correlation of MSE and time for the training and test block. The Pearson correlation coefficients indicate a weak positive correlation to no correlation in MSE over time (see Table 1). Lack of negative correlation indicates that MSE did not decrease with time during the training block, meaning that participants did not get more accurate during training. Furthermore, the Pearson correlation coefficients suggest that participants were less accurate with time in most conditions (e.g., in the control condition in the training block).

Table 1: Pearson correlation coefficients and p-values for time and MSE for the training and test block of Study 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>r</th>
<th>p</th>
<th>r</th>
<th>p</th>
<th>r</th>
<th>p</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>control</td>
<td>0.30</td>
<td>0.01</td>
<td>0.16</td>
<td>0.05</td>
<td>0.09</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>ghost</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.15</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>inter50</td>
<td>0.20</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.15</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td>inter75</td>
<td>0.01</td>
<td>0.11</td>
<td>0.71</td>
<td>0.02</td>
<td>0.16</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>snapping</td>
<td>0.02</td>
<td>0.17</td>
<td>0.12</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>control</td>
<td>0.22</td>
<td>0.04</td>
<td>0.03</td>
<td>0.00</td>
<td>0.20</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td>ghost</td>
<td>0.14</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.15</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>inter50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.15</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>inter75</td>
<td>0.01</td>
<td>0.11</td>
<td>0.71</td>
<td>0.02</td>
<td>0.16</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>snapping</td>
<td>0.02</td>
<td>0.17</td>
<td>0.12</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

5.5 Summary

The results of Study 1 suggest that correcting participants’ virtual movement helps learning of a target movement. Training with snapping guidance increased the accuracy of participants in the short-term retention test (i.e., the test block). Similarly, snapping and interpolation75 improved the most from the training to test block. While interpolation50 also performed well, the analysis did not show significant difference when compared to other conditions. Full correction of virtual movement (i.e., snapping) was better at supporting learning than partial correction (i.e., interpolation).

We expected to see a clear increase or decrease in performance from interpolation50, interpolation75 to snapping, however, there was no clear relationship between the three conditions. A more complex approach than interpolation may be needed for varying the degree of virtual movement correction.

We did not notice any clear patterns of improvement within the blocks. The correlation analysis does not indicate an improvement in accuracy during the training or test, even when there is a large improvement in accuracy from the training block to the test block. To confirm the findings from Study 1, we decided to replicate it.
while keeping only the best performing correction condition (i.e., snapping).

6 STUDY 2

Study 2 replicated the initial study while removing the interpolation50 and interpolation75 conditions. With replication, we intended to confirm the benefits of the virtual movement correction for short-term retention of the target movement and to confirm the surprising lack of improvement during training. We kept only one of the virtual movement correction conditions (snapping) as we were more interested in the general effects than the effects of specific variations (e.g., degree of correction) of the virtual movement correction technique. Furthermore, having only one virtual movement correction condition avoided any potential training effect across conditions. Our hypotheses for Study 2 were that virtual movement correction (i.e., snapping) will outperform the control and the ghost condition in the short-term retention test, and that snapping will show the largest performance improvement from the training to test block. In other words, except for snapping significantly outperforming ghost, we expected to replicate the results from the Study 1.

6.1 Participants

We recruited participants over online Oculus Quest communities until we had 36 validated data sets (1 female, 35 male, age $M = 28.7$, $SD = 8.23$). To validate the data sets, we used the same criteria as in Study 1 and discarded the data from three participants that experienced hand tracking problems during the tasks. The drastic reduction of the rejected data when compared to the the Study 1 was due to the added screening test that the participants had to take before starting the experiment. In the screening test, the participant’s hand tracking quality was evaluated in a simple target selection task. If the participant’s hand tracking quality was under a predetermined threshold, then the study application did not allow them to proceed to the experiment. In such a case, the participants were encouraged to find a room with better lighting conditions and try the screening test again. We conducted Study 2 two months after the initial study and did not limit the participation to only those who did not take part in Study 1. We assumed that after such a time span any carryover effect from Study 1 would be negligible. In the post-study questionnaire only one participant stated that he took part in Study 1.

6.2 Design

In Study 2 we used control, ghost and snapping conditions, implemented as in Study 1. We selected the three target paths from Study 1 with the least deviation from their overall mean MSE. The selected target paths were T1, T4 and T5, as seen in Figure 3. We used a within-subject design and balanced the conditions and target paths by using all 36 combinations and order permutations.

6.3 Procedure

Except for the screening test and fewer conditions, the study procedure was identical to the one of Study 1. The study took approximately 20 minutes and the participants were given a Steam game worth 15$ as a reward for their participation.

7 STUDY 2: RESULTS

We conducted an identical LMM analysis to Study 1, using MSE as our main performance metric, while also looking into the learning effect and the number of executed repetitions per condition.

7.1 Number of Repetitions

As in Study 1, there were no significant differences ($F(2, 70) = 0.236, p = 0.79$) in the number of repetitions between the conditions in the test block ($M = 6.15, SD = 2.05$). We found a main effect of condition on the number of repetitions in the training block ($F(2, 70) = 9.555, p < 0.001$). Tukey’s post hoc test revealed that in the training block participants executed significantly more repetitions ($p = 0.001$) in the ghost condition ($M = 13.72, SD = 3.54$) than in the control condition ($M = 11.58, SD = 3.5$). There were no interaction effects between the snapping and ghost, and snapping and control condition. Figure 8 shows the mean number of repetitions per condition for the training and test block.

7.2 Movement Accuracy

Study 2 confirmed our hypothesis of snapping performing better than the control condition, and rejected our hypothesis of ghost outperforming ghost as they did not differ significantly (see Figure 7).

The analysis with LMM revealed a main effect of condition on MSE in the test block ($F(2, 70) = 6.082, p < 0.01$). Tukey’s post hoc test showed that snapping ($p = 0.005$) and ghost ($p = 0.025$) were significantly more accurate than the control condition (Figure 7, middle). We also found a main effect of condition on improvement in accuracy from training to test block ($F(2, 70) = 11.647, p < 0.001$). Tukey’s post hoc test showed that the improvement in accuracy from training to test block was significantly larger for snapping ($p < 0.0001$) and ghost ($p < 0.01$) when compared to the control condition (Figure 7, right). In the control condition, the participants were less accurate in the test block than in the training block.

We did not find any significant differences between the conditions in the training phase ($F(2, 70) = 0.428, p = 0.654$), indicating that
guidance used in snapping and ghost did not impact the accuracy during training (Figure 7, left).

7.3 Perceived Difficulty
As in Study 1, the LMM analysis of the Likert-scale answers on task difficulty did not find a main effect of condition ($F(2, 70) = 0.843, p = 0.435$). This means that all the conditions were judged as equally difficult, suggesting that the participants were not negatively impacted by the misalignment between the virtual and the real hand that the snapping introduced.

7.4 Learning Effect
Similarly as in Study 1, we found weak to no correlations between MSE and time, which indicates no improvement in accuracy during the training block. Table 2 shows Pearson correlation coefficients and p-values for each of the conditions per block.

7.5 Summary
Study 2 confirmed the benefits of virtual movement correction for the learning of hand movement while not finding any significant differences between the snapping and ghost conditions. Furthermore, Study 2 confirmed the lack of clear improvement in accuracy during the training.

8 DISCUSSION
We have demonstrated that training in which the participant’s virtual movement is corrected, significantly improves performance in the short-term retention test when compared to training without the correction. The effectiveness of such training was comparable to training with a conventional VR guidance technique (i.e., ghosted hand).

The results of Study 1 show that full correction of virtual movement (i.e., snapping) supported learning better than partial correction (i.e., interpolation50 and interpolation75). Nevertheless, there is a noticeable positive trend in improvement from the training block to the test block for interpolation conditions (Figure 6). The benefit of only partially correcting the virtual movement is that it introduces less misalignment between the location of participant’s real and virtual hand than the full correction of movement. For example, interpolation50 displaced the participant’s virtual hand only halfway towards the target location compared to snapping. In our task, the smaller misalignment of the interpolation conditions did not outweigh the benefits of full correction. We speculate that snapping might have supported learning better because it showed the participants the exact target movement. Seeing the exact movement could be especially crucial in the early phases of motor learning, when the participant is not yet familiar with the movement. In later phases of learning, perhaps the partial correction of movement would be more beneficial.

Surprisingly, we did not observe an improvement in accuracy over time during training. This could be due to many reasons. For instance, participants might have become progressively more tired and consequently less accurate with time; they might have forgotten the movement’s exact details when training without guidance; or they might have felt rushed towards the end by the training block countdown timer. Furthermore, we noticed that two participants had little regard for accuracy during the training block of the snapping condition. They executed the movement almost in a straight line...
from the marked starting point to the ending point. These participants still benefited from training with snapping and performed better in the test block than in the control condition. We did not instruct the participants in any specific strategy they should or should not take during the task. Therefore the performance of the mentioned two participants was valid and analyzed the same as others.

The field of human-computer interaction (HCI) only rarely discusses the importance, or lack thereof, of immediate improvement during training for motor learning. One exception is a paper by Kirsh, which describes how marking, a simplified or abstracted form of movement, is used in dance during practice [13]. The dancers execute a partial or a substitute movement while imagining the full movement. Kirsh listed general benefits of marking, such as requiring less energy, and benefits of it over pure mental simulation in which there is no movement at all. The performance of two participants, that were moving their hand almost in a straight line during training in snapping condition, could be interpreted as marking the movement. Such marking could have additional benefits when done in VR, since the users can see the full practiced movement instead of just imagining it as in Kirsh’s description of marking practice [13].

The self-avatar follower effect [9] is another mechanism that might have influenced our results. The effect suggests that users will start following their avatar’s movement to minimize the mis-alignment between the virtual and the real hand. The effect differs from marking and might be at work during training in the snapping condition for most of the participants. Participants were as accurate at executing the target movement in the snapping condition as they were in the ghost condition (Figure 6 left, and 7 left). This is interesting when considering that participant’s virtual performance (i.e., the movement the participant saw in VR) in the snapping condition would be perfectly accurate irrespective of trying to align the physical hand with their virtual hand or not. The fact that the participants did align their hand with their avatar’s hand might therefore be described as self-avatar follower effect. The participants’ accuracy did not improve over time during the training (see Table 1 and 2), however, this could be explained by the self-avatar follower effect appearing immediately after starting the training. This speculation is also supported by results of immediate avatar correction (when contrasted to gradual) in the work of Gonzalez-Franco et al. [9].

9 Limitations and Future Work

It is difficult to predict how well our findings generalize to motor tasks of different complexity and difficulty. Based on the review paper by Sigrist et al. [21], our experimental task falls within the complex type, as it has several degrees of freedom and cannot be mastered in a single practice session. The participants needed to trace an invisible path with their index finger, moving and rotating their hand and fingers freely. The target paths were multi-segment curves generated in a plane, however, to trace them accurately the participants needed to move their hand within a 3D space. Movements of similar complexity could, for example, be used as a control gesture in VR applications or for drawing in mid-air. However, it is unclear how well the virtual movement correction would work in motor tasks that, for example, required dexterous movement of all the fingers, higher degree of accuracy, or memorization of longer movement sequences. Future work should therefore investigate motor tasks of different complexity. One application we find especially interesting is learning of sign language. A sign correction system, conceptually similar to one used for text input correction, might prove immediately useful for communication while possibly have lasting long term improvements.

Apart from investigating other motor task of different complexities, future work should also collect data over longer periods of time. It is unclear how virtual movement correction affects long-term retention of the practiced movement as well as what are the benefits of such training in later stages of motor learning. In their review paper of augmented feedback, Sigrist et al. [21] suggested that concurrent feedback is most efficient in early phases of motor learning. Investigating how to integrate correction techniques with other types of feedback (e.g., terminal feedback or feedback in other modalities) is therefore another interesting research direction.

10 Conclusion

We investigated learning of a motor skill with a help of VR hand guidance technique. The technique corrected the participant’s virtual hand movement to be more accurate than the the participant’s actual movement. In the first experiment, we compared three levels of correction (interpolation50, interpolation75 and snapping) to a control condition and a conventional hand guidance technique (viz., ghost). In the second experiment, we replicated the initial experiment with only one of the correction conditions (viz., snapping). The two experiments showed that the correction of virtual movement supported motor learning by improving the short-term retention of the practiced movement. Apart from supporting motor learning, the investigated technique has other benefits, such as immediate improvement of the virtual performance during training. By embedding the guidance in the user’s avatar it also minimizes visual distractions. We connect the technique to past work on manipulating visual feedback for the purpose of user guidance and interpret the results of our study through self-avatar follower effect [9] and marking [13] theory.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement 648785).

References


