# Perspective Matters: Design Implications for Motion Guidance in Mixed Reality

Xingyao Yu*
University of Stuttgart

Katrin Angerbauer†
University of Stuttgart

Peter Mohr‡
Graz University of Technology

Denis Kalkofen§
Graz University of Technology

Michael Sedlmair¶
University of Stuttgart

## Abstract

We investigate how Mixed Reality (MR) can be used to guide human body motions, such as in physiotherapy, dancing, or workout applications. While first MR prototypes have shown promising results, many dimensions of the design space behind such applications remain largely unexplored. To better understand this design space, we approach the topic from different angles by contributing three user studies. In particular, we take a closer look at the influence of the perspective, the characteristics of motions, and visual guidance on different user performance measures. Our results indicate that a first-person perspective performs best for all visible motions, whereas the type of visual instruction plays a minor role. From our results we compile a set of considerations that can guide future work on the design of instructions, evaluations, and the technical setup of MR motion guidance systems.

**Index Terms:** Information Interfaces and Presentation—User Interfaces—Evaluation/Methodology; Computer Graphics—Three-Dimensional Graphics and Realism—Virtual Reality

## 1 Introduction

Providing effective and precise motion guidance is an essential part in areas such as physiotherapy, sports, dance, and yoga. Practicing specific motions is, for instance, necessary to regain strengths after an injury or to get fit for sport competitions. The traditional way of offering motion guidance in such situations is through a physiotherapist or trainer, who can observe the motion and offer verbal or physical advice and correction if needed. In many situations, however, an external advisor might not be available. This is, for instance, the case when exercises need to be practiced alone at home, or if the physical presence of an advisor is not possible due to long distances or a pandemic. When training unsupervised, however, there is the problem of lacking feedback. Motions might be executed sloppily or even incorrectly, potentially leading to worsening health conditions or hard-wiring of bad habits.

A widely used approach to overcome this problem is to use (live or prerecorded) video tutorials for motion guidance. Video tutorials are a powerful method to convey tasks and motions effectively. However, precisely following instructions provided in a video tutorial can be difficult. Motion paths and velocity need to be inferred from a 2D screen and be translated into 3D motions by the user. This task is challenging even if multiple perspectives are given in the video [21, 37].

In this paper, we focus on the use of Mixed Reality (MR) for motion guidance. 3D Mixed Reality environments offer a viable option to overcome these issues. First, approaches in the literature have demonstrated the potential benefits of using MR for motion guidance. The application areas include physical exercises [25], martial arts [11, 16], physiotherapy and rehabilitation [31, 37], as well as repair and maintenance tasks [9, 14]. A major advantage of MR environments is their ability to change the viewpoint and to display instructions in 3D space. While these systems provide interesting point solutions in a large design space, a systematic understanding of the role of different design factors in this space is still lacking. For instance, no conclusive answers are available for the perspective in which such a system should operate (first-person, mirror-person, or third-person), how guidance should be visually encoded, and on how different motion types influence the design.

Toward filling this gap, we contribute three controlled user studies, which were run at two universities independently. In Study 1, we investigate solely the effect of perspective, while the remaining studies focus on the additional influence of motions (Study 2 and 3) and visual encoding (Study 3). To conduct these studies, we designed two prototypical MR motion guidance systems. The first prototype was designed by the team at Graz University of Technology and was used for Study 1 and 2. The second prototype was designed by the team at the University of Stuttgart and used for Study 3. For both prototypes we used Virtual Reality (VR) devices, as they did not come with the same current technical limitations as Augmented Reality (AR) devices, and as they allowed us to better control the studies. We present all technical designs and experimental results in one paper, as it allows us to reflect on the replicability of results across different systems and development teams.

Our most prominent finding points at the superiority of using a first-person view for motion guidance, leading to better performance in terms of motion accuracy and time. The first-person approach presents motion guidance in an egocentric view directly on the user's body. Here, the user's body and the guidance paths reside in the same coordinate system, avoiding additional cognitive load for mapping motions between different coordinate systems. The usefulness of the first-person perspective holds specifically in situations where the respective motion allows to easily observe such motions (single motions visible in front of the body). Motions that necessitate coordinating multiple parts of the body simultaneously, such as moving two arms, are not well supported by a first-person view. Synthesizing the results and experiences from the three studies, we offer further recommendations and considerations for the design and future studies on MR motion guidance systems.

In summary, we make the following two contributions: (1) three user studies on MR motion guidance, for which we implemented the respective prototypes, and (2) a set of synthesized design and research implications for MR motion guidance systems. We see the main novelty of our work in the careful combination of perspective, movement path visualization [2, 35, 37], and 6-DOF in-air motions [10, 16].

## 2 Background & Related Work

To set our work in a broader context, we provide a general overview of MR used for tutorials, training and assistance in Section 2.1, and we discuss related motion guidance systems in Section 2.2.

*Xingyao.Yu@visus.uni-stuttgart.de
†Katrin.Angerbauer@visus.uni-stuttgart.de
‡mohr@icg.tugraz.at
§kalkofen@icg.tugraz.at
¶Michael.Sedlmair@visus.uni-stuttgart.de


Figure 1: Prototype system for Study 1: (a) Motion guidance around the lifelike avatar. (b) Blue rubber bands indicate links between the desired posture and the user pose, the highlighted glyph indicates when posture is completed. (c) Arrows show the motion path and an animated stick figure is used for postures. (d) User during the study, wearing the Oculus Rift HMD.

### 2.1 MR-based Tutorials

MR is increasingly used in various domains for learning purposes, especially when training in the real world is safety-critical or expensive. McHenry et al. [20] tested VR training for satellite assembly tasks. To prepare for emergencies, one can also simulate the dangerous conditions in VR, for instance, the evacuation of a mine [1] or the simulation of an earthquake [17]. In medicine, there exist training programs to prepare medical students for surgery [32].

Besides upfront training, MR can also be used as on-the-spot guidance, for example, through situated visualizations. There are example use cases in assistive technology [8] to help cognitively impaired workers in assembly tasks with projections, in medicine to provide support for needle injection points [13], in music instruction to display visual guides for chords on guitar [23], and in the military to support mechanics in repair tasks [14]. With more and more handheld devices capable to run MR applications, there are also scenarios where MR is used for remote assistance, for instance, a remote worker can generate cues to help a local worker on the field [22].

MR tutorials range from projected hints, that can be prerecorded [37], drawn by a remote expert [22], extracted from videos [21, 26], or simulated scenarios [1]. As seen above, the application area of MR tutorials is vast. In this paper, we focus on the subset of explicit motion guidance, where the motion itself is the task of the tutorial.

### 2.2 Motion Guidance for Limbs

A motion guidance system can guide the user to move the limbs to a predetermined position along a preset trajectory without requiring a face-to-face tutorial from a coach, by providing hints in different modalities like visual guidance [37], haptic [24] or multi-modal feedback [18, 30]. Motion guidance systems differ in the types of motion they support. OctoPocus3D provides gesture guidance for precise in-air hand movement [6]. In the medical field, visual motion instructions were found to improve the upper motor functions for stroke survivors [19, 36]. YouMove [2] simulates an interactive floor-to-ceiling mirror for a ballet tutorial. Our setup is not modeled for a specific application scenario. We use arm motions in an abstract context to better control motion guidance factors such as visual design, motion complexity, and perspective.

The existing motion guidance systems can be divided into three categories based on personal perspective: first-person (1PP) [16], third-person (3PP) [12] and mirror-person perspective. Salamin et al. [27] proposed a series of experiments in Virtual Reality (VR), where the users were allowed to switch their viewpoint between 1PP and 3PP. Their results showed that users in 3PP could make a better prediction on the trajectory of mobile objects, while 1PP was better for delicate hand manipulations. Examples for the mirror-person perspective usually simulates surroundings for better learning efficiency, e.g. a mirror in a ballet studio [2] or a physiotherapist's office [37].

The type of visualized motions usually depends on the system implementation and the feedback modality of the system. Due to the inability to visualize movement depth, traditional display techniques like TVs [37] and projectors [34, 35] can only visualize a movement trajectory on surfaces, which is likely not suitable for in-air 6-DOF motions. In contrast, MR displays such as head-mounted displays (HMD) are able to visualize coherent in-air 6-DOF movements [10]. Currently, most of the current MR motion guidance systems [10, 11, 16] visualize the postures without movement path. However, with such instructions, it is still difficult for the users to move accurately and realign themselves to the correct positions when deviating, especially for long-trajectory motions [37]. In contrast, the visualization of the movement path constantly provides the user with future movement direction and feedback regarding their movement accuracy and progress.

Our work puts a specific emphasis on the combination of different perspectives, different movement path visualizations, and different 6-DOF motions. Unlike the recent related work [7] which investigated the continuity and realism of posture animation, we focus on movement path visualizations, since constantly showing a visualization of the movement path could help learning and memorizing the motions [4, 37].

## 3 Study 1 – 3D First-person vs. 2D Mirror

While following 3D motions from 2D presentation might be very challenging, 3D MR environments have the advantage in supporting users [9, 15], due to their ability to change the viewpoint and to display instructions in 3D space. Therefore, in Study 1 we set out to understand how far a dedicated 3D MR-based guidance system situated in the users' bodies might outperform a more classical representation offered in a 2D mirror-person view. In the latter, users have to match their 3D body postures with the instructions demonstrated in the 2D mirror-person perspective. We describe below the prototype implementation and the study setup in more detail. Additional information for all our studies can also be found in our supplemental material.

### 3.1 Prototype Design

To study the question outlined above, we first developed a prototype system for MR motion guidance, in which we visualized 3D poses and motions from an arbitrary point of view. Our system generated an augmented visualization of the motion around the user by providing 3D directional arrows and an animated 3D skeleton (Figure 1c). The visualizations were registered to a 3D avatar which we derived from a rigged 3D reconstruction of the user (Figure 1a).

#### 3.1.1 Guidance

To offer visual guidance, we segmented long sequences into chunks of shorter movements in order to reduce clutter in the user's view. For each shorter segment, the motion path was visualized using an array of 3D arrows, pointing in the direction of the movement (Figure 1c). In the process of motion guidance, the arrows would vanish for segments of the path which the user has followed successfully.

There was a 3D stick figure demonstrating the postures of this segment, where bones and joints were represented by white tubes and circular glyphs respectively. Once the user reached the desired pose, the circular glyphs would be highlighted in green (Figure 1b). We furthermore connected the arm joints to the desired positions using a rubber-band visualization for the first posture, to guide the users to the correct positions.

In addition, the stick-figure depiction was animated with the velocity of the motion. However, the stick-figure depiction would only move a predefined amount of distance ahead in time to give the user time to react and reorient when deviating too far from the instruction.

#### 3.1.2 Perspective

To support a first-person perspective (*1PP*) visualization, we attached the virtual camera to the user's head position, enabling the users to change the viewpoint by natural head motion (Figure 1b). *2D-Mirror* was established by providing a presentation on a large virtual screen in front of the user (Figure 2a). Similar to Tang et al. [37], we used a split-screen setup to provide the user with two views: a top-down and a front view.

#### 3.1.3 Implementation

Our prototype requires a large field of view to visualize the 6-DOF guidance in 3D space and low latency for an accurate and reliable motion guidance. With that in mind, we opted to implement our prototype in VR, since current VR HMDs are more sophisticated concerning these technical requirements than AR systems. We tracked the user's skeleton with a Microsoft Kinect, using the open-source skeleton tracking framework NiTE[1], which provides positional and rotational data of up to 15 joints. We applied the tracking data to the rigged 3D avatar and displayed it in VR with the Oculus Rift (Figure 1d).

### 3.2 Study Design and Setup

We designed a within-subject study to compare the performance and user experience of the independent variable PERSPECTIVE in the conditions *1PP* and *2D-Mirror*. Previous works [2, 37] stated that visual AR instructions outperformed video tutorials. Consequently, we decided not to compare to video tutorials directly but to a mirror embedded in the virtual test environment. We let the users perform four motion tasks, while we measured task completion time, and movement error, i.e. positional deviation of the wrist to the given instructions computed by absolute Euclidean distance. After the performance, we gathered subjective opinions. A total of 12 participants

[1] https://structure.io/openni

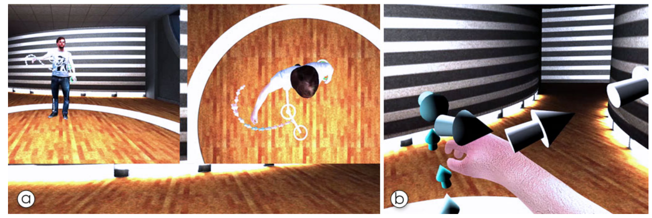


Figure 2: (a) In Study 1, we use a front view and top-down visualization of the avatar in the *2D-Mirror* condition, and (b) we compare this perspective condition to a first person view.

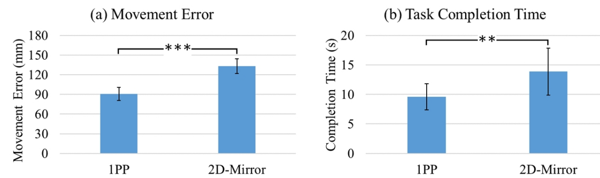


Figure 3: Overview of results of Study 1: *1PP* outperformed *2D-Mirror* in both (a) movement error and (b) task completion time. The significant differences have been marked with stars * (* for p<.05, ** for p<.01, and *** for p<.001). The error bars denote the 95% confidence intervals. Same definitions are used in the other figures.

(3 female, 9 male) who aged 22 to 32 were recruited, all of them being university students. We hypothesized that movement error and completion time would be lower in *1PP* compared to *2D-Mirror*.

**Tasks** We asked the participants to follow four different motions as precise as possible in time and distance. We designed the motions with approximately the same length and difficulty. Since we were specifically interested in testing the performance of the 3D guidance visualization we avoid motions within a single plane. To further restrict the complexity, we limit the motions to a single upper limb at a time. In order to counterbalance user preferences, we designed two of the motions to be performed with the left arm and the other two motions with the right arm.

**Procedure** We started by calibrating the motion to the body size of the participants. The participants were asked to stand in an upright T-pose while the virtual avatar was scaled to fit their height and arm length. After completing an informed consent form and a demographic questionnaire, we explained the visualizations and allowed participants to get used to the immersive environment. After they were confident with the setup, we asked them to perform the given tasks. To avoid order effect, we counterbalanced the tasks and conditions with a Latin square.

### 3.3 Results

In the following, we present our results including task movement error, completion time, and subjective feedback. For statistical analysis, we first tested for normality (Shapiro Wilk [33]). Since the data were normally distributed, we used t-tests for statistical analysis. Effect sizes are shown graphically with 95% confidence intervals [5].

Figure 3 shows an overview of the results of movement error and completion time. Overall, *1PP* had a significantly lower movement error ($|t|=7.31$, $p<.0001$) with a mean error of 90.72 (SD=25.09) compared to *2D-Mirror* with had a mean error of 133.34 (SD=27.72). The task completion time for *1PP* with M=9.58 (SD=3.98) was also lower than that of *2D-Mirror* with M=13.89 (SD=6.43). This difference was also significant ($|t|=3.41$, $p=.006$).

Regarding subjective preference, 11 out of 12 participants preferred *1PP* over *2D-Mirror*, and 1 participant was neutral. A common remark from participants was that *1PP* allows following along the path much more confidently, while *2D-Mirror* requires looking at one's own body and the mirror concurrently. The participants also mentioned jitter of the tracked skeleton but did not feel obstructed in solving the tasks.

To sum our results up, users were consistently faster and more precise using *1PP* compared to *2D-Mirror*. Subjectively, the participants also seemed to prefer *1PP* over *2D-Mirror*. The results of this study thus further support the fact that an egocentric VR guidance system outperforms mirror-based approaches.

## 4 Study 2 – 3D First-person vs. 3D Third-Person

Our previous experiment explores 3D motions of a single upper limb. In such a scenario, the user can easily observe the current instruction from a first-person view at any point in time. However, more complex motions might involve moving multiple limbs at a time, preventing the user to observe all instructions at once from a first-person perspective. In contrast to an egocentric perspective, in a third-person perspective one does not necessarily need to move one's head to get an overview of more complex motion. As seen in Study 1, a mirror-person view had a negative impact on performance. Therefore, instead of *2D-Mirror*, we selected a third-person perspective (*3PP*) following game design guidelines [29]. This resulted in a camera placement from behind and above the avatar (Figure 4a), which represents the user.

In Study 2, we were thus interested in the effectiveness of 3D guidance for motions that require head rotation to be observed from a first-person view.

### 4.1 Study Design and Setup

We again used a within-subject with repeated measures design to study the above mentioned *1PP* (Figure 4b) and *3PP* (Figure 4a) for the independent variable PERSPECTIVE. Additionally, we investigated the influence of the MOTION, by considering motions in front and in the periphery of the user as described below.

In this experiment, we measured the movement error for both wrists, task completion time, the head yaw rotation in degrees, usability with the Single Ease Question (SEQ) *"How difficult or easy was the task to complete?"* [28], and overall preference.

For the technical setup, we used the same prototype as before, with participants wearing an HTC Vive. Since participants commented on the jitter of the skeleton tracking in the previous experiment, we combined the Lighthouse tracking system provided by HTC Vive and Kinect tracking. The users, therefore, wore an HTC Vive as HMD and held two input controllers in their hands to track their wrists. A total of 12 subjects aged 21-38 years (2 female, 10 male) participated in this study. None of them had participated in Study 1. We hypothesized that *1PP* was faster and more accurate for motions in front of the user based on our previous findings in Study 1. However, for motions in the periphery of the user, *3PP* should yield faster and more accurate results. Additionally, *3PP* should reduce the amount of head rotation.

**Tasks** Similar to the previous experiment, we asked participants to follow a set of predefined motions as precise as possible. To cause head rotation, we added a *Peripheral-Motion* that requires to move both upper limbs simultaneously. Figure 5b shows a yoga motion as an example, which consists of a sweeping gesture keeping both arms on the sides. Due to the limited field of view of the current generation of VR HMDs (110° in our setup) the instructions were not visible simultaneously from *1PP*. *Frontal-Motion* (Figure 5a) was inspired by physio-therapeutic instructions for diagonal shoulder flexion and extension patterns. It was extended to involve both arms simultaneously in front of the user, so that all instructions were visible from the first-person point of view. Thus, the motions were designed to take advantage of each of the viewing conditions. The *Frontal-Motion* was designed to be best visible from the *1PP* and to include changes in depth, while the *Peripheral-Motion* was designed to require *3PP* for an overview and to include fewer changes in depth.

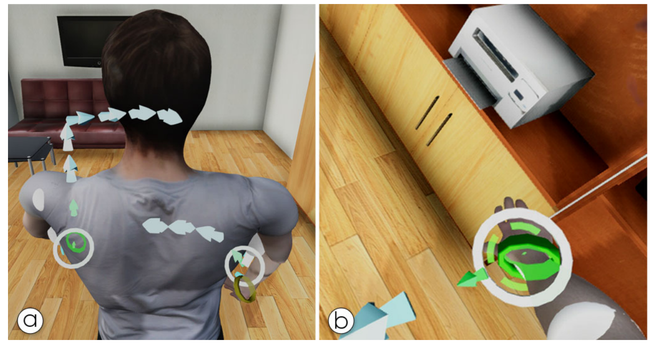


Figure 4: Conditions in Study 2. (a) The user's perspective from within the VR environment during the *Frontal-Motion* instruction from the *3PP* perspective. (b) The VR perspective of the participant as seen in *1PP*.

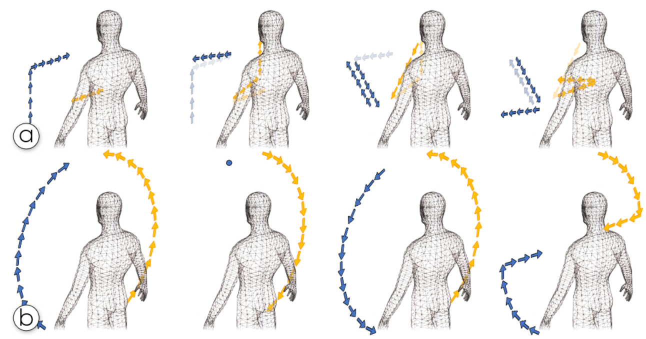


Figure 5: Motions in Study 2: (a) *Frontal-Motion* depicted with our arrow glyphs, split into four steps for better visual clarity (b) *Peripheral-Motion* depicted with our arrow glyphs, split into four steps for better visual clarity.

**Procedure** As in the previous experiment, we started by introducing and calibrating the system for each participant. We allowed participants to get familiar with the environment and to test all conditions. After they were confident, we started the first task. The conditions were counterbalanced with a Latin square.

### 4.2 Results

We used a Two-Way ANOVA for statistical analysis of movement error, task completion time, and head rotation, with PERSPECTIVE and MOTION as factors. We used Greenhouse-Geisser corrections to adjust the lack of sphericity and Bonferroni corrections for multiple comparisons in post-hoc tests. The results of the Likert-scales from the subjective questions were analyzed with a Friedman test. Effect sizes were reported with generalized Eta squared ($\eta_G^2$) [3] and Kendall's W for Friedman tests, and are shown graphically together with 95% confidence intervals.

Figure 6 shows an overview of movement error, task completion time, and head rotation data for different tasks. To look at movement error, we summed up the left and right hand error into total error and then averaged it per participant. As seen in Figure 6a, *Frontal-Motion* caused a lower movement error than *Peripheral-Motion*, which formed a significant main effect of MOTION ($F_{1,11}$=35.223, p<.0001, $\eta_G^2$=.317). As for PERSPECTIVE, the overall movement

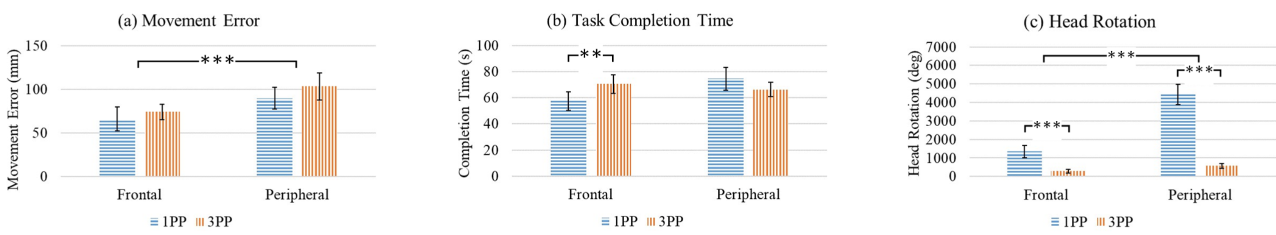


Figure 6: Overview of the results of Study 2. (a) Movement error: *1PP* significantly outperformed *3PP* in both motions. (b) Completion time: *1PP* was lower in *Frontal* and *3PP* was lower in *Peripheral*. (c) Head rotation: *3PP* significantly outperformed *1PP*, and *Frontal* significantly outperformed *Peripheral*.

error was lower in *1PP* compared to *3PP* for all motions, yet not significantly ($F_{1,11}$=4.750, p=.052, $\eta^2_G$=.074). It is interesting, that even for *Peripheral-Motion*, *1PP* slightly outperformed *3PP*, but not significantly ($|t|$=1.55, p=.150). When looking at this result more closely, we observed an interesting correlation between the accuracy of individual arms in *1PP*. A common pattern is shown in Figure 7, which presents the error of the left and right hands as line-plots for one exemplary user. The plot shows that the error increases on one side when it decreases on the other side. The plot of the head yaw angle clearly shows that the users are switching their focus between the left and right hand, as the error of the focused side decreases.

In terms of completion time, we saw the influence stemming from motion more clearly. The Two-Way ANOVA found neither significant main effect of PERSPECTIVE nor MOTION, but a significant interaction effect between PERSPECTIVE and MOTION ($F_{1,11}$=14.847, p=.003, $\eta^2_G$=.194). While for *Frontal-Motion*, *1PP* was clearly and significantly better ($|t|$=3.86, p=.003), for *Peripheral-Motion* there was no significant difference found ($|t|$=1.75, p=.108), as illustrated in Figure 6b. Besides, *Frontal-Motion* had significantly less completion time than *Peripheral-Motion* when guided in *1PP* ($|t|$=2.99, p=.012), while *Peripheral-Motion* was slightly better when compared in *3PP* ($|t|$=1.60, p=.137).

Additionally, we observed that head rotation was significantly higher for *1PP* (M=2883.65 °, SD=1717.19°) than *3PP* (M=431.57°, SD= 229.74), which formed a main effect of PERSPECTIVE ($F_{1,11}$=280.273, p<.0001, $\eta^2_G$=.861). The MOTION had a significant main effect as well ($F_{1,11}$=142.469, p<.0001, $\eta^2_G$=.745), where the *Peripheral-Motion* caused higher head rotation than *Frontal-Motion*. Although there was a significant interaction effect between PERSPECTIVE and Motion (F(1,11)=107.859, p<.0001, $\eta^2_G$=.666), *1PP* caused a significantly lower head rotation than *3PP* for either motion (both p<.0001). The head rotations of the specific motions are shown in Figure 6c.

Regarding subjective feedback, the SEQ revealed a difference in subjective usability between *1PP* and *3PP* for the specific motions ($\chi^2(3) = 13.037, p = .005, W = .362$). For the *Peripheral-Motion* we measured a mean of 3.83 (SD= 1.27) in *1PP* and a mean of 5.5 (SD= 1.17) in *3PP*, which was significant with p=.034. For the *Frontal-Motion* we measured a mean of 5.42 (SD= 1.44) for 1PP and a mean of 4.41 (SD = 1.16) for *3PP*. The difference here, however, was not significant. 8 of 12 participants preferred *3PP* over *1PP*. Most participants commented on the higher amount of head rotations in *1PP* for preferring *3PP*. People who preferred *1PP* over *3PP* mentioned the visibility problems due to the occlusions of the 3D avatar, especially for the motion guidance for *Frontal-Motion* and different scale due to the distance. One participant mentioned that "it feels like remote controlling a puppet".

To summarize our results, *1PP* was beneficial for all motions in terms of movement accuracy, but *3PP* lead to faster executions of the *Peripheral-Motion*. Our assumptions were confirmed by the head rotation data. The *Peripheral-Motion* caused significantly more head

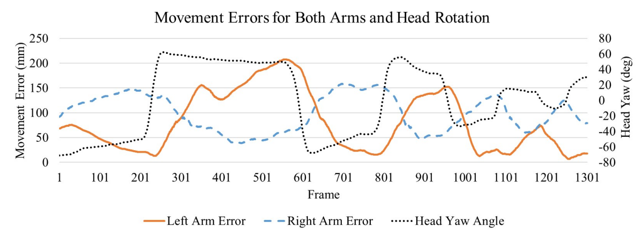


Figure 7: Error correlation between left and right hand depending on the peripheral the user is concentrating on in first-person. The data is taken from a single user during the 1PP *Peripheral-Motion* condition.

rotations in *1PP* compared to *3PP*. Interestingly, head rotations in *1PP* were also significantly higher compared to *3PP* in the *Frontal-Motion*. We believe this was caused by the depth of the motion that can be explored in *1PP* by natural head rotations, while the viewpoint in *3PP* was too far away to look at the motion path "from the side".

## 5 STUDY 3 – VISUAL ENCODING OF MOTIONS

After the comparison between *1PP*, *2D-Mirror* and *3PP*, it remained unclear how guidance visualizations influence the performance of MR motion guidance systems. To better understand that, we designed a prototype with new guidance visualizations and motions. In this study, we were primarily interested in the effect of the continuity and visual encoding of a guidance visualization for single-arm motions. We wanted to take a closer look at the interaction among perspective, guidance and motion characteristics. In terms of perspective, we opted to compare *1PP* with *3D-Mirror*, which complements the previous two studies.

### 5.1 Prototype Design

With a more focused representation and motion range, we sought to reduce the potential distraction of other components in our system. Therefore, we used a stick figure arm token rather than the entire avatar in this study. An arm token consists of three spheres representing wrist, elbow, and shoulder joints, and two sticks for lower and upper arms. The visual instructions take the user's shoulder as the origin. We developed two conditions for visual guidance with the arm token: *Streamer*, providing continuous guidance, and *GhostArm* with discrete guidance.

#### 5.1.1 Continuous Guidance: Streamer

We developed *Streamer* as continuous arm motion guidance. The main metaphor behind *Streamer* are dancing ribbons that in our case continuously precede the motion of the arm to visualize its trajectories. To create the *Streamer* we simplified the lower and upper

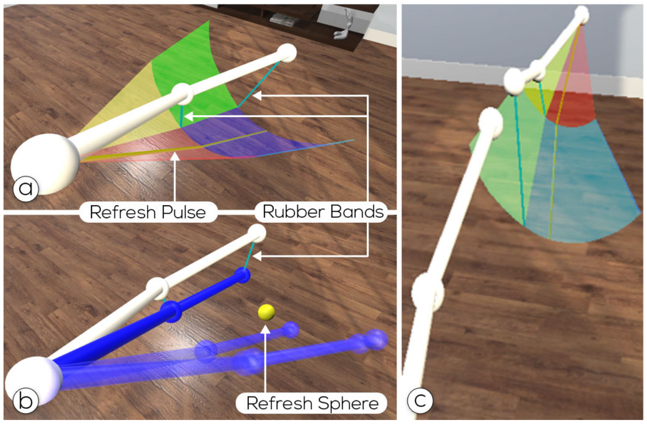


Figure 8: Instruction design for Study 3: There are a white arm token and two blue rubber bands for both guidance visualizations. A yellow refresh pulse for *Streamer* and a yellow refresh sphere for *GhostArm* serve as speed visualization. (a) On *Streamer*, blue and red represent the lower and upper arm respectively in the uncompleted part. Their counterparts in the completed part are in green and yellow. The boundary between completed and uncompleted parts represents the current target posture, where the rubber bands start. (b) The *GhostArm* was represented with blue tokens with increasing transparency along the movement path. (c) In *3D-Mirror* the guidance is visualized on the mirror-arm.

arm into two line segments, each getting their own streamer ribbon. We considered the fact that the trajectory of a mobile line segment is a curved surface in 3D space. Thus, the combined trajectories of the lower and upper arm are two curved surfaces that intersect each other, and the curve at the intersection should represent the trajectory of the elbow.

In the *Streamer* visualization, the instructions of the lower and upper arm are distinguished by different colors (Figure 8a). To represent the movement progress, the visualized trajectories of lower and upper arm change colors. Then the boundary between the completed and uncompleted parts represents the current target posture. When the user's arm deviates from the trajectory, there are dynamic rubber bands like the ones described in Study 1, aiming to guide all user's arm joints back to the desired positions. We used a refresh pulse on stream trajectory to signal the ideal velocity of motion, which continuously moves from the current target position (where the rubber bands start) to the end of motion at the same speed with preset movement.

#### 5.1.2 Discrete Guidance: GhostArm

The *GhostArm* condition provides discrete motion guidance, consisting of a set of arm tokens at certain frames of the movement. First, we sample from a preset arm motion at regular intervals to obtain arm postures at certain frames. We then use arm tokens to present these target postures, which form a set of GhostArms with the shoulder as the origin. As seen in Figure 8b, the GhosrArm guidance was represented with blue arm tokens while the user's arm was in white.

To visualize the direction of movement, all existent GhostArm tokens have the same color, but their transparency increases along the movement path. The current target posture is represented by the most opaque token (*transparency* = 0). Once the users have successfully reached the current target posture, its token will vanish immediately, and the remaining tokens will redistribute the transparency by the rules described above.

As in *Streamer*, there are also dynamic rubber bands guiding the users back to the target positions in case of spacial deviation. Also, there is a visualization for the target speed in the form of a refresh sphere moving from the current target GhostArm token to the end of motion along the path of the user's wrist repeatedly.

#### 5.1.3 Perspective

*1PP* here is the same as in Study 1 and Study 2, where the guidance is visualized on users' arm tokens. In contrast to the virtual screen of *2D-Mirror* in Study 1, *3D-Mirror* in Study 3 provides visual instructions on a mirror-arm token in front of the user (Figure 8c). The mirror-arm token performed mirror-symmetric movements relative to users' arms. To find the optimal distance between the user's arm token and the mirrored one we empirically tested different distances. As a result, 2.2 times of the user's arm length seemed to be a good compromise between the visibility of the mirror-arm token and the overlap of both tokens.

To avoid mismatch of the preset motions due to different body sizes, we used Tang et al.'s method [37] to scale the preset movements to fit the users' bodies, which transforms an arm posture in the preset record into two normalized vectors representing the orientation of lower and upper arms, multiplies these vectors by the length of the user's lower and upper arm forming user-fitted posture, and then registers it to the user's shoulder.

#### 5.1.4 Implementation

In this study, we used an HTC VIVE Pro set, which consists of an HMD, two controllers, two base stations, and three 6-DOF trackers. To track three arm joints, users wear three 6-DOF trackers mounted on shoulder, elbow, and wrist.

### 5.2 Study Design and Setup

Like in the previous studies, we used a within-subject repeated measures design. We wanted to investigate the independent variables GUIDANCE (*Streamer* vs. *GhostArm*), PERSPECTIVE (*1PP* vs. *3D-Mirror*), and MOTION (*Ring*, *G Clef*, *Cake Piece*, and *Combo* as seen in Figure 9). Thus, we had a 2x2x4 design with 16 conditions in total. A total of 17 participants (5 female, 12 male), aged between 21 and 33, were recruited for the study. None of them had participated in Study 1 or 2.

We gathered data on movement error, relative completion time, and subjective feedback. For movement error, we accounted the positional deviation of not only *wrist* but also *elbow* into the overall movement error, as *shoulder* was the origin of visual instructions. In contrast to the other studies we did not control the trajectory length of a motion, but each movement had its own unique trajectory length. In order to get a comparable measure for the completion time, we divide the participants' completion time by the duration of a preset record of the motion, to get the normalized completion time. Due to our implementation of the guidance visualization, the user could not be faster than the preset record. Consequently, the normalized completion time would never be less than 1. The closer this value is to 1, the more accurate it is.

In this study, we chose again motions in front of the user, thus we expected *1PP* to outperform *3D-Mirror* in terms of position and time accuracy, as seen already in Study 1, Study 2, and related work [27]. In contrast to the other studies, we chose motions different in their complexity, to see how that influences the user's performance as well. However, our main focus was the comparison of our discrete and continuous modes of guidance. Here we hypothesized that the discrete *GhostArm* guidance would be faster to follow, while the *Streamer* should have a lower movement error.

**Tasks** After initial testing, we determined the motions *Ring*, *G Clef*, *Cake Piece*, and *Combo* with different characteristics and difficulty. *Ring* and *G Clef* have wrist trajectories parallel to the user's body and only require the movement of the arm as a whole. *Cake Piece* consists of five consecutive motion fragments that require the abduction of the arm, which looks like a piece of cake. *Combo*

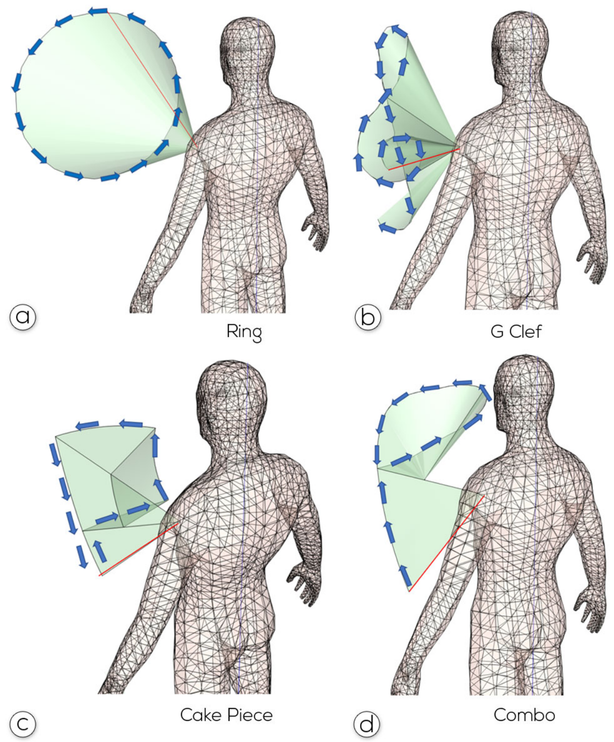


Figure 9: Motions in Study 3: the start positions are represented by red lines and the blue arrows stand for the trajectory of wrist.

includes a twisty Lasso movement of the lower arm, while the upper arm has to remain still.

**Procedure** Before starting VR exposure, participants read the study description and then signed the form of consent. After fitted with the HMD and trackers, participants were introduced to the system with a short demonstration and completed a trial run of each condition with a sample motion. For each condition, the participants were allowed to watch the tutorial animations if they felt confused about the visualization. Each participant recorded 48 trials: 2 guidance x 2 perspectives x 4 motions x 3 trials. As in the previous studies, a Latin square was used to counterbalance the order of conditions.

### 5.3 Results

Performance data were analyzed using 2 x 2 x 4 ANOVA, with PERSPECTIVE, GUIDANCE, and MOTION as factors. For statistical analyses of movement error, we averaged the error of the 3 trials in each condition for all participants. For the normalized completion time, we picked the trial with the minimum value among all 3 trials, to avoid bias from potential outliers through incorrect movements. We applied the same statistical correction and analysis methods as in Study 2.

**Perspective and Guidance** Figure 10 provides an overview of the measurements of movement error, normalized completion time, and subjective feedback depending on PERSPECTIVE and GUIDANCE used.

Statistical analysis revealed that there was a significant main effect of PERSPECTIVE ($F_{1,16}$=63.415, p<.0001, $\eta_G^2$=.198) and GUIDANCE ($F_{1,16}$=7.09, p=.017, $\eta_G^2$=.017) on movement error. Overall, *1PP* (M=347.56, SD=102.57) had a lower movement error than *3D-Mirror* (M=435.8, SD=99.98). When comparing *GhostArm* and *Streamer*, the movement error for *GhostArm* (M=380.01, SD=105.61) was significantly lower than the error for *Streamer* (M=403.44, SD=114.09). Figure 10a illustrates the findings above by showing the differences between the guidance and perspective conditions.

The normalized completion time was affected significantly as well by the factors PERSPECTIVE ($F_{1,16}$=24.215, p=.0002, $\eta_G^2$=.060) and GUIDANCE ($F_{1,16}$=8.735, p=.009, $\eta_G^2$=.065). Pairwise comparison indicated that the *1PP* had a significantly lower normalized completion time than *3D-Mirror* (Figure 10b). Furthermore, the normalized completion time for *GhostArm* guidance was lower than *Streamer*.

As shown in Figure 10c, the measurements above are also reflected in the subjective feedback. Participants found *1PP* significantly easier to follow than *3D-Mirror* ($\chi^2(1) = 9, p = .003, W = .529$). Participants also rated *GhostArm* better then *Streamer* in that aspect, but this difference was not significant ($\chi^2(1) = 3.27, p = .071$). When asked to rank the combination of GUIDANCE and PERSPECTIVE, 9 out of 17 participants ranked *GhostArm* in *1PP* first, while both forms of guidance in *3D-Mirror* were ranked last by 8 participants (Figure 10c).

**Motion** The factor of MOTION also significantly influenced movement error ($F_{3,48}$=13.615, p<.0001, $\eta_G^2$=.198) and normalized completion time ($F_{1.59,25.52}$=11.336, p=.0006, $\eta_G^2$=.111). Pairwise comparisons between the motions (Figure 11a) indicated that *Ring* cause a consistently lower error than *Cake Piece* ($|t|$=7.82, p<.0001) and *Combo* ($|t|$=6.40, p<.0001), and that the error for *G Clef* was consistently lower than *Cake Piece* ($|t|$=5.30, p<.0001) and *Combo* ($|t|$=4.57, p<.0001). And *Cake Piece* caused a higher error than *Combo* ($|t|$=2.54, p=.014). As shown in Figure 11b, the *Combo* motion had a consistently higher normalized completion time than *Cake Piece* ($|t|$=3.22, p=.002), *G Clef* ($|t|$=3.76, p=.0004) and *Ring* ($|t|$=3.60, p=.0006), and the normalized completion time for *Cake Piece* was consistently higher than *Ring* ($|t|$=2.31, p=.024) and *G Clef* ($|t|$=2.69, p=.009).

Furthermore, according to Three-Way ANOVA, there was a significant interaction effect among PERSPECTIVE, GUIDANCE and MOTION on normalized completion time ($F_{1.29,20.72}$=5.413, p=.023, $\eta_G^2$=.033). *3D-Mirror* had a slightly lower normalized time than *1PP* for motion *G Clef* (p=.898), while *1PP* was significantly lower than *3D-Mirror* (p<.0001) overall. Motion *Ring* tended to have a lower normalized time than *G Clef* when guided with *Streamer* (p=.999), while it had a significantly higher value than *G Clef* when guided with *GhostArm* (p=.046). However, no significant difference overall was found between these two motions (p=.999). An overview of the normalized completion time of the motion can be seen in Figure 11b. Subjectively, the *Combo* was considered hardest by 9 participants and the *Ring* easiest by 10. A complete overview of the ranking of the motions can be seen in Figure 11c.

**Summary of Results** We replicated the findings of our previous studies that *1PP* seems fast and accurate for motions in front of the user. With *GhostArm*, the participants were faster compared to *Streamer*, as expected. However, *GhostArm* also was more positionally accurate than *Streamer*, a result which we did not anticipate. Participants' oral comments indicated that *GhostArm* provided more details on joints of correct postures than *Streamer*. Additionally, there was less visual clutter in *GhostArm*, which made it easier for them to follow the rubber bands to the desired positions. Regarding motion design, we found that the Combo Motion, which includes the twist of an arm might be the most difficult to learn in both *1PP* and *3D-Mirror*.

## 6 IMPLICATIONS FOR DESIGN

Table 1 provides an overview of our three studies and their results. The top rows summarize the design decisions we made for the prototypes used in the studies. The prototypes for Study 1 and 2 were quite similar, while for Study 3 we added major changes.

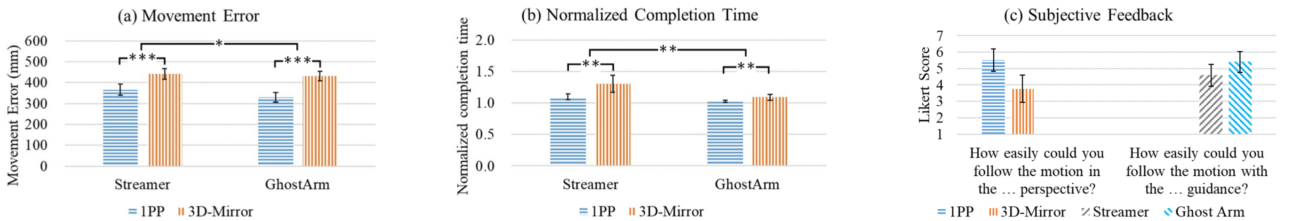


Figure 10: Overview of the results of perspective and guidance in Study 3. *1PP* outperformed *3D-Mirror* and *GhostArm* outperformed *Streamer*, in (a) movement error, (b) normalized completion time, and (c) subjective feedback.

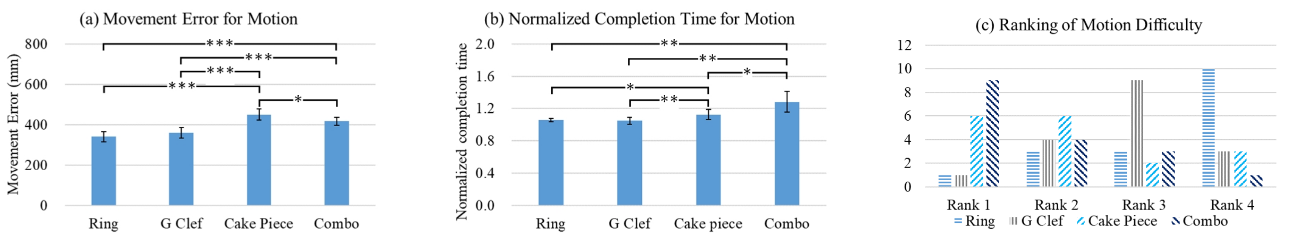


Figure 11: Overview of the results of motions in Study 3 (a) movement error: *Ring* caused a consistently lower error than *Cake Piece* and *Combo*, and that the error for *G Clef* was consistently lower than *Cake Piece* and *Combo*. (b) normalized completion time: The *Combo* motion had consistently higher normalized completion time than *Cake Piece*, *G Clef*, and *Ring*. (c) motion difficulty: *Combo* got the most votes in motion difficulty, *Ring* was considered least difficult.

Table 1: Overview of our three studies regarding the prototype design, the experiment setup, as well as the results we obtained in the studies. The results show that we could replicate the effects of perspective. The "greater" markers (>) represent "outperform".

| Study | I | II | III | |
|---|---|---|---|---|
| **Prototype** | 3D arrow & stick figure | | continuous Streamer | discrete GhostArm |
| Embodiment | realistic avatar | | abstract arm token | |
| Movement direction | 3D arrows | | dancing ribbons | increasing transparency |
| Posture maintenance | 3D stick figure | | junction in Streamer | GhostArm tokens |
| Movement progress | arrows vanish | | changes of color | tokens vanish |
| Movement speed | 3D stick figure | | refresh pulse | refresh sphere |
| Position alignment | rubber bands | | | |
| Tracking method | Kinect | Kinect+VIVE controllers | VIVE trackers | |
| **Experiment** | | | | |
| Perspective | 1PP vs. 2D-Mirror | 1PP vs. 3PP | 1PP vs. 3D-Mirror | |
| Motion Type | single arm frontal motion | peripheral vs.frontal | single arm frontal motion | |
| Position | 1PP>2D-Mirror | 1PP>3PP | 1PP>3D-Mirror<br>GhostArm>Streamer | |
| Time | | frontal:1PP>3PP<br>peripheral:3PP>1PP | | |
| Subjective feedback | | 1PP>3PP | | |

At the bottom of the table, the main findings are summarized. In the following, we jointly reflect on the three studies and derive implications for design and future work.

### 6.1 Perspective – use 1PP if the type of motion allows

The perspective in which the motion guidance was shown, could explain the largest effects of the studies. If a motion was suitable for *1PP*, showed by far the best performance and led to the best subjective user feedback.

We see multiple reasons that can explain this result. First, *1PP* provides an in-situ registration of guidance visualizations and thus enables seeing the motion in real-world scale. Thus, most participants who preferred *1PP*, considered it as *"natural"* and *"intuitive"*. Second, *1PP* also allows better leverage of human stereo perception in the near field of sight. By slightly rotating the head the users were able to see the motion paths from different angles and better explore the instructions in all three dimensions.

The other perspectives that we tested were consistently less precise and slower for motions in front of the user. Again, there are several reasons that can explain these results. In the *Mirror* perspectives, both 2D and 3D, but also in the third-person perspective (*3PP*), movements need to be translated from another coordinate-system to be applicable for the user. Furthermore, in 2D the motion path need to be scaled due to the projection in the mirror perspective. Note, that unless the user moves the head, the added value of 3D over 2D in these situations is also minimal [38]. These mental translations add cognitive load and were often the source of confusion. In Study 3, for instance, participants had problems *"judging the depth of the movement" (Participant 5)* or were confused about the direction of the movement (*"I mixed up left and right and forward and backward" (Participant 15)*).

In summary, the findings in all three studies further support previous findings on the advantages of in-situ instructions in (*1PP*) [38].

### 6.2 Motion – consider alternatives for other types of motions.

The above findings, however, only hold when the motions are suitable for *1PP*. For fully leveraging the benefits of *1PP*, motions need to be (i) fully visible in *1PP*, and (ii) should not involve the concurrent movement of several body parts, specifically if they do not reside in the same field of view. We explicitly studied such motions in Study 2 (*Peripheral-Motion*). We were surprised that even for the *Peripheral-Motion*, *1PP* showed slight benefits in terms of accuracy, further hinting at the prevalence of this perspective (see above). Nonetheless, in terms of time and subjective feedback *1PP* could not compete as the *Peripheral-Motion* simply was outside the field of view. On the other hand, the *3PP* we did not fully resolve this issue either due to the aforementioned weaknesses. This is an interesting area for future work. Such motions might, for instance, be better supported by overall and viewpoint-independent accuracy indicators, or by a combination of multiple perspectives such as discussed by Tang et al. [37]. Such ideas, however, add a higher cognitive load and further work is needed to understand the design space here.

Even though motions might fulfill the criteria of fitting into a *1PP* approach, they naturally still can differ in terms of complexity. While in Study 1 and Study 2 we strove to create motions of approximately uniform complexity, in Study 3 we explicitly tested the influence of motion complexity and designed four motions with increasing complexity. To this end, we considered the different possible characteristics that might have an influence on motion complexity:

**i. Trajectory length**. There is a higher chance for users to get tired and make mistakes in a longer trajectory.

**ii. Number of intersections in a trajectory**. An intersection means the limb goes through a position for more than once.

**iii. Number of joints involved**. The more arm joints are involved, the harder it is to focus on each of them.

**iv. Number of 2D segments**. An in-air motion, for instance, used in physiotherapy, usually can be decomposed into consecutive 2D fragments, which is a dimensionality reduction. The higher number of 2D segments makes it more difficult to understand and follow a motion.

Among all four motions in Study 3, *G Clef* had the longest trajectory (i) and the most intersections (ii). Both *Cake Piece* and *Combo* involved all three arm joints (iii), but the former consists of five consecutive 2D motion fragments, while the twisting of the arm required for *Combo* cannot be decomposed into finite 2D segments (iv). As expected, the motion *Combo* had the significantly highest normalized completion time among all motions, and a higher movement error than *Ring* and *G Clef*. Additionally, *Combo* was considered the most difficult motion by 9 participants. We thus specifically believe that (iv) the number of 2D segments could play a role in motion complexity. Future studies, however, are needed to further support our observations. While it is interesting to consider the complexity of motions, we did not find any interesting interaction effects with the design parameters.

### 6.3 Guidance – visual design might have a minor effect.

While most of the effects we observed stemmed from perspective and motion type, we also observed differences caused by the visual presentation of the guidance. The main design factor that we studied in this respect was the difference between discrete and continuous visual guidance. Of course, many other factors might influence the performance as well [7].

In Study 1 and 2, we visualized motion paths "intuitively" with discrete arrows. This choice was inspired by classical flow visualization approaches. In Study 3, we wanted to take a closer look and compared a discrete *GhostArm* visualization with a continuous *Streamer* visualization. We found that indeed, the intuitive choice for a discrete representation worked better: *GhostArm* guidance performed better than *Streamer*. However, it is important to note that the effect size of GUIDANCE compared to that of PERSPECTIVE was much smaller for movement error.

There are many other design decisions in terms of the visual encoding of motion paths that we made on the way of creating our prototypes (see top of table 1). Specifically, we used dynamic *rubber bands* to visualize when the user's arm deviated from the guidance. While we had the impression that those worked well in general, in the continuous *Streamer* condition, the rubber bands were confused with movement paths themselves by four participants. One participant also mentioned that replacing the rubber bands with targeted arrows might be more helpful to align the positions. Besides, our subjective results indicated that an explicit representation of arm joints, such as the *GhostArm* tokens in Study 3, was beneficial to precisely maintain the postures. Other design considerations of guidance visualizations, which we did not explicitly study are listed at the top of table 1.

## 7 LIMITATIONS AND FUTURE WORK

As all empirical work, our studies come with a set of limitations. The performance of motions is inherently connected to a certain degree of fitness and physical characteristics of a user. To account for that, we normalized the motions according to the body size of the single individual participants in all our studies. This easy approach seemed to have worked well in most situations, however, a few participants also reported on discomfort due to their physical characteristics. Specifically, two tall participants (>185cm) and one small one (165 cm) reported difficulties when performing some of the tasks.

Naturally, as mentioned above, for our studies we also had to focus on certain factors. For many others, we could only scratch the surface through our design process. For instance, although we used different levels of realism with respect to the embodiment in our studies, we did not investigate the effect of realism. As mentioned in the recent related work [7], abstract discrete posture animation seems to be subjectively preferred by the users. For the visualization of the movement path, we had similar findings, where the users seemed to prefer *GhostArm* over *Streamer*. However, more work is needed to understand these trade-offs.

In this paper, we used precise positional and temporal metrics, which may not be optimally suited to actually learn motions in the long run. There could be a technical approach to this goal. When the user's arm deviates from the desired path, our current method directs the user back to the absolute postures with blue rubber bands. But theoretically, we can also create an asymptotic path based on the current deviation and the original path, which might help the users to learn in a more intuitive, natural, and holistic way, as proposed in LightGuide [34]. Such a solution involves path-planning of human limbs in 3D space though, which we deemed as beyond the scope of this paper.

Besides, we are aware that the motions designed in our studies may not cover all real-life scenarios. For those, we provide authoring modules in our prototype systems that allow users to record and edit their own motion instructions. For physiotherapy, as an example, motions could thus be specifically modified for individual patients, with the help of a physiotherapist. The patients could then practice motions of increasing difficulty and detail levels. In the future, we plan to add more complex motions to expand our user studies toward an increased level of ecological validity.

Due to technical reasons, we have studied our approaches in VR, while in the longer run optical see-through AR might be preferable. The latest generation of Augmented Reality (AR) headsets such as the HoloLens 2 already allow posture tracking to a certain extent, and a 3PP view in AR can be visualized using a world-in-miniature metaphor. We are confident that our results also generalize across the technical differences between AR and VR, but we have not explicitly tested for that. For similar reasons, we have opted to select a straight forward approach to motion tracking. In Study 1, we used Kinect but did not receive the accuracy we wanted. In Study 2, we thus combined it with HTC Lighthouse and we abandoned the additional Kinect tracking altogether in Study 3, as it did not add any noticeable benefit in our case. A drawback of this approach is, however, that the wearable trackers may have negatively impacted the user performance. Future work could look into using alternative tracking approaches such as professional OptiTrack or Vicon motion tracking systems.

Finally, the three experiments arose from different projects at two universities. The projects were conducted independently, and we learned of their similarity only after the data were captured. Since each participant only participated in one experiment, we decided to report them together in a paper because we believe that the benefits outweigh the drawbacks. While there are certain drawbacks in terms of the consistency between prototypes and study designs, putting them together allowed us to reflect on the replicability of our results across different systems and setups.

## 8 Conclusion

In this paper, we present three studies, where we investigated the effect of perspective, visual encoding, and motion characteristics on MR motion guidance. Specifically, we compared first-person, 2D/3D mirror-person, and third-person perspectives. Furthermore, we also explored the performance of continuous and discrete guidance on the motions with different characteristics. A total of 41 participants were recruited to accomplish the motion guidance tasks. We measured the users' performance of the motion guidance based on objective metrics regarding position and time, together with subjective questionnaires. The results of all three independent studies indicated that the first-person perspective outperformed mirror-person and third-person perspective with respect to position and time, especially for motions in front of the users. For the motions in the periphery of users, however, perspectives providing an overview like third-person view might be better for timing. It could also be concluded from the results of Study 3 that the apparent visualizations of arm joints are more important than the continuity of guidance. From our results, we propose a set of design implications regarding the perspective selection and its interactions with the motion characteristics, as well as visual design, which aim to give pointers for the design of motion guidance systems. Despite the discussed limitations, we believe that our results concerning the perspectives have a relatively certain effect, as they could be replicated throughout three studies and two different prototypes.

We believe that VR/AR motion guidance could potentially have a broad impact in many application scenarios, such as physiotherapy, sports, or remote collaboration. Thus, we argue that a better understanding of suitable interfaces, as well as the compatibility between interface and motions, is of high importance. We hope that our work will help to further build up the foundations for future research in these directions.


## Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2075 – 390740016, by the Competence Center VRVis, and the Austrian Science Fund (grant P30694). VRVis is funded by BMVIT, BMWFW, Styria, SFG and Vienna Business Agency in the scope of COMET, Competence Centers for Excellent Technologies (854174), which is managed by FFG. We would like to thank Laurenz Theuerkauf for helping us with user study one and two and the corresponding Figures.


## References


[1] K. Andersen, S. J. Gaab, J. Sattarvand, and F. C. Harris Jr. METS VR: Mining evacuation training simulator in virtual reality for underground mines. In *Proceedings of the International Conference on Information Technology - New Generations (ITNG)*, pp. 325–332, 2020.

[2] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 311–320, 2013.

[3] R. Bakeman. Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–384, 2005.

[4] O. Bau and W. E. Mackay. Octopocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the ACM symposium on User interface software and technology (UIST)*, pp. 37–46, 2008.

[5] G. Cumming. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, 2013.

[6] W. Delamare, T. Janssoone, C. Coutrix, and L. Nigay. Designing 3d gesture guidance: visual feedback and feedforward design options. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 152–159, 2016.

[7] M. Dürr, R. Weber, U. Pfeil, and H. Reiterer. EGuide: Investigating different visual appearances and guidance techniques for egocentric guidance visualizations. In *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction (TEI)*, pp. 311–322, 2020.

[8] M. Funk, S. Mayer, and A. Schmidt. Using in-situ projection to support cognitively impaired workers at the workplace. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 185–192, 2015.

[9] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6):778–798, 2015.

[10] P.-H. Han, K.-W. Chen, C.-H. Hsieh, Y.-J. Huang, and Y.-P. Hung. Ar-arm: Augmented visualization for guiding arm movement in the first-person perspective. In *Proceedings of the Augmented Human International Conference (AH*, pp. 1–4, 2016.

[11] P.-H. Han, Y.-S. Chen, Y. Zhong, H.-L. Wang, and Y.-P. Hung. My taichi coaches: an augmented-learning tool for practicing tai-chi chuan. In *Proceedings of the Augmented Human International Conference (AH)*, pp. 1–4, 2017.

[12] T. He, X. Chen, Z. Chen, Y. Li, S. Liu, J. Hou, and Y. He. Immersive and collaborative taichi motion learning in various VR environments. In *Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*, pp. 307–308. IEEE, 2017.

[13] F. Heinrich, F. Joeres, K. Lawonn, and C. Hansen. Comparison of projective augmented reality concepts to support medical needle insertion. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 25(6):2157–2167, 2019.

[14] S. Henderson and S. Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(10):1355–1368, 2010.

[15] S. J. Henderson and S. K. Feiner. Augmented reality in the psychomotor phase of a procedural task. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 191–200. IEEE, 2011.

[16] T. N. Hoang, M. Reinoso, F. Vetere, and E. Tanin. Onebody: remote posture guidance system using first person view in virtual environment. In *Proceedings of the Nordic Conference on Human-Computer Interaction (Nordic CHI)*, pp. 1–10, 2016.

[17] K. Iguchi, H. Mitsuhara, and M. Shishibori. Evacuation instruction training system using augmented reality and a smartphone-based head mounted display. In *Proceedings of International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pp. 1–6, 2016.

[18] N. Katzakis, J. Tong, O. Ariza, G. Klinker, B. Roeder, and F. Steinicke. Stylo and handifact: Modulating haptic perception through visualizations for posture training in augmented reality. In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*.

[19] J. Kim, M. Lee, Y. Kim, S.-D. Eun, and B. Yoon. Feasibility of an individually tailored virtual reality program for improving upper motor functions and activities of daily living in chronic stroke survivors: A

case series. *European Journal of Integrative Medicine*, 8(5):731–737, 2016.

[20] N. McHenry, T. Hunt, W. Young, A. Gardner, U. Bhagavatula, B. Bontz, J. Chiu, G. Chamitoff, and A. Diaz-Artiles. Evaluation of pre-flight and on orbit training methods utilizing virtual reality. In *Proceedings of the AIAA Scitech 2020 Forum*, 2020.

[21] P. Mohr, D. Mandl, M. Tatzgern, E. Veas, D. Schmalstieg, and D. Kalkofen. Retargeting video tutorials showing tools with surface contact to augmented reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 6547–6558, 2017.

[22] P. Mohr, S. Mori, T. Langlotz, B. H. Thomas, D. Schmalstieg, and D. Kalkofen. Mixed reality light fields for interactive remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 1–12, 2020.

[23] Y. Motokawa and H. Saito. Support system for guitar playing using augmented reality display. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 243–244, 2006.

[24] S. L. Norman, A. J. Doxon, B. T. Gleeson, and W. R. Provancher. Planar hand motion guidance using fingertip skin-stretch feedback. *IEEE transactions on haptics (TOH)*, 7(2):121–130, 2014.

[25] T. Nozawa, E. Wu, F. Perteneder, and H. Koike. Visualizing expert motion for guidance in a vr ski simulator. In *ACM SIGGRAPH Posters*, pp. 1–2. 2019.

[26] N. Petersen and D. Stricker. Learning task structure from video examples for workflow tracking and authoring. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 237–246, 2012.

[27] P. Salamin, D. Thalmann, and F. Vexo. The benefits of third-person perspective in virtual and augmented reality? In *Proceedings of the ACM symposium on Virtual reality software and technology (VRST)*, pp. 27–30, 2006.

[28] J. Sauro and J. S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 1599–1608, 2009.

[29] A.-M. Schleiner. Does lara croft wear fake polygons? gender and gender-role subversion in computer adventure games. *Leonardo Music Journal*, 34(3):221–226, 2001.

[30] C. Schönauer, K. Fukushi, A. Olwal, H. Kaufmann, and R. Raskar. Multimodal motion guidance: techniques for adaptive and dynamic feedback. In *Proceedings of ACM international conference on Multimodal interaction (ICMI)*, pp. 133–140, 2012.

[31] Y. A. Sekhavat and M. S. Namani. Projection-based ar: Effective visual feedback in gait rehabilitation. *IEEE Transactions on Human-Machine Systems*, 48(6):626–636, 2018.

[32] X. Shao, Q. Yuan, D. Qian, Z. Ye, G. Chen, K. le Zhuang, X. Jiang, Y. Jin, and D. Qiang. Virtual reality technology for teaching neurosurgery of skull base tumor. *BMC Medical Education*, 20(1):1–7, 2020.

[33] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[34] R. Sodhi, H. Benko, and A. Wilson. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 179–188, 2012.

[35] M. Sousa, J. Vieira, D. Medeiros, A. Arsenio, and J. Jorge. SleeveAR: Augmented reality for rehabilitation using realtime feedback. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI)*, pp. 175–185, 2016.

[36] W. Stryla and A. Banas. The use of virtual reality technologies during physiotherapy of the paretic upper limb in patients after ischemic stroke. *J. Neurology and Neuroscience*, 6(3):33–38, 2015.

[37] R. Tang, X.-D. Yang, S. Bateman, J. Jorge, and A. Tang. Physio@home: Exploring visual guidance and feedback techniques for physiotherapy exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 4123–4132, 2015.

[38] C. Ware. *Information Visualization: Perception for Design*. 2004.