Motor learning in Mixed Reality

Stefan Paul Feyer
HCI Group, University of Konstanz

July 30, 2019

Abstract

- Overall aim of the Seminar thesis: how to investigate the influence of perspectives on virtual avatars in MR for motor learning
- therefore analysis of motor learning, related work, research questions
- propose study setting

Contents

1 Introduction

1.1 Motivation

In recent years, Mixed Reality (MR) devices became more affordable [1], portable [2] and usable in more conditions. Not only academic researchers are interested in this technology, commercial companies also found MR devices helpful to explore new possibilities to use it profitable. With this development, learning and training in MR became possible for many cases, too. EON [3] for example calls themself "the world leader in Virtual Reality based knowledge transfer for industry, education, and edutainment". They develop MR programs for several platforms, eg. with the aim to guide workers, reducing mistakes and thus reducing costs. These programs address a lot of usecases in the field of education, energy, health & medical, manufacturing & industrial, defence & security and aerospace. Tasks include eg. ground crew training for a Boeing 777, augmented reality (AR) assembly training, exploring or anatomy simulation to mention only a few, compare 1.2. Mircosoft also stepps into this topic with partners, developing tools apprentice, maintenance, or remote training. Eg. The Smart Glass experience Lab[4] of the Fraunhofer Institute use the hololens for remote maintenance.

Since MR learning or guiding programs reached the commercial market, many applications will be created. It is important to build these applications on well founded research.
Developing a system for MR learning can be complex and mistakes can be made. Providing a developer with guidelines to design such a program could help decreasing design faults. But before guidelines can be created, groundwork has to be done and be investigated with sophisticated research methods. This seminar thesis will take a look in the background of motor learning, perspectives on virtual avatars, VR and tracking technologies and

[1] **TODO**
[2] **TODO**
[3] https://www.eonreality.com/ accessed: 14.12.2018
[4] https://www.fit.fraunhofer.de/de/fb/cscw/smart-glasses-experience-lab.html

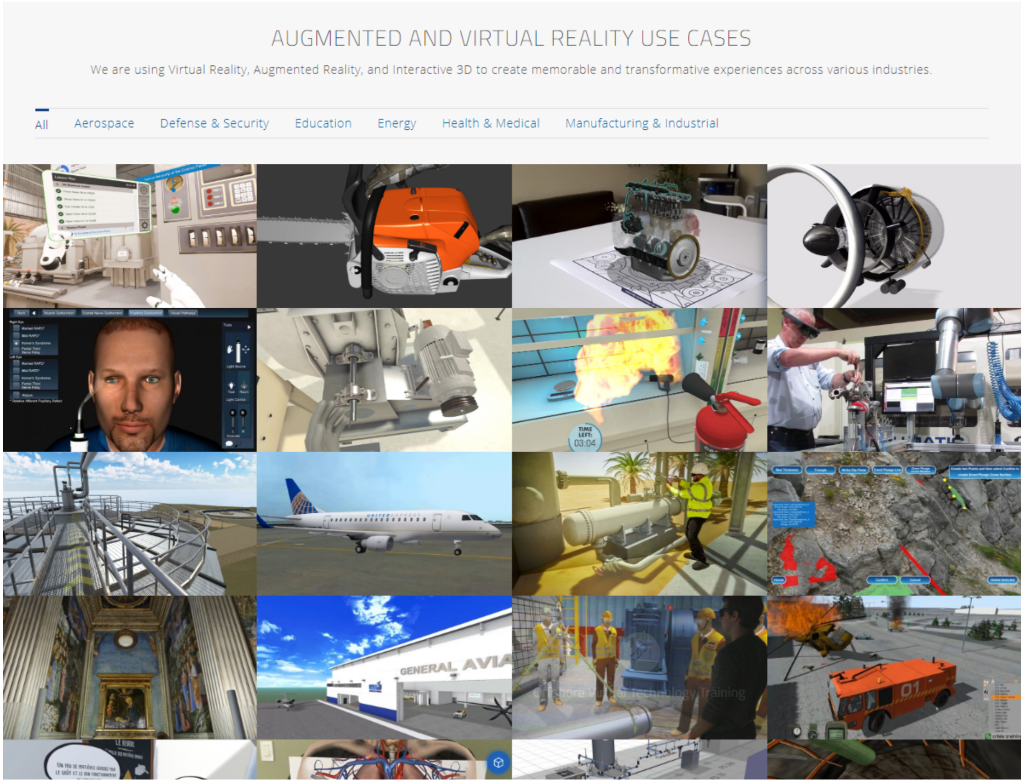

Figure 1.1: Usecases by EONReality https://www.eonreality.com/use-cases/ (accessed 30.07.2019)

previous work investigating on that topic to later conduct this groundwork that can be used for guidelines for designing a MR motor learning system.

1.2 Outline

This seminar thesis will have a look at the grounding principles to define a system that lead to the guidelines in question. Therefore we first step into the theoretical background in the next chapter. We will investigate how people achieve motor skills, how to quantify and measure movements, what hardware techniques are suitable, perspectives and mixed reality. In the following chapter we analyse how other researchers used the theory to gain insights in Motor learing in VR. In the end we propose a study setting to that can be used to investigate on this topic. **TODO**graphical representation

Figure 1.2: Remote maintenance with Hololens by Smart Glass Experience Lab, taken from https://www.youtube.com/watch?v=1QFMPo5k6p0, accessed (30.07.2019)

2 Theoretical Background

Motor learning in MR builds on many aspects, eg. suitable technology for MR representations and what perspectives to use. Furthermore, human motor learning must be suitably transfered in the digital world. And eventually, we need to match movements and measure the error correctly to derive adequate from the performance of a learner. In this chapter we take a look into these topics.

2.1 Movements

How do we learn movements

We learn by mimicking others, follow textual descriptions, videos etc. **TODO**

Movement classification

For a simplified discussion a classification of movements is provided in the following. There are two important classification schemes. The first one is based on the particular movements performed and are divided into *discrete*, *continuous* and *serial movements*. The second one is based on perceptual attributes of the task and are divided into *open* and *closed skills*. Both classification representing a continuum.

Discrete, Continuous and Serial Movements

Discrete movements are located on the one end of the continuum. These are movements with a recognisable beginning and end. The end of a discrete movement is defined by the task itself and can be very rapid like blinking or longer like making the signing. Examples are kicking a ball, shifting gears in a car or striking a match
Continuous movements are located on the other end of the continuum. These movements don't have a recognisable start and end, with behaviour

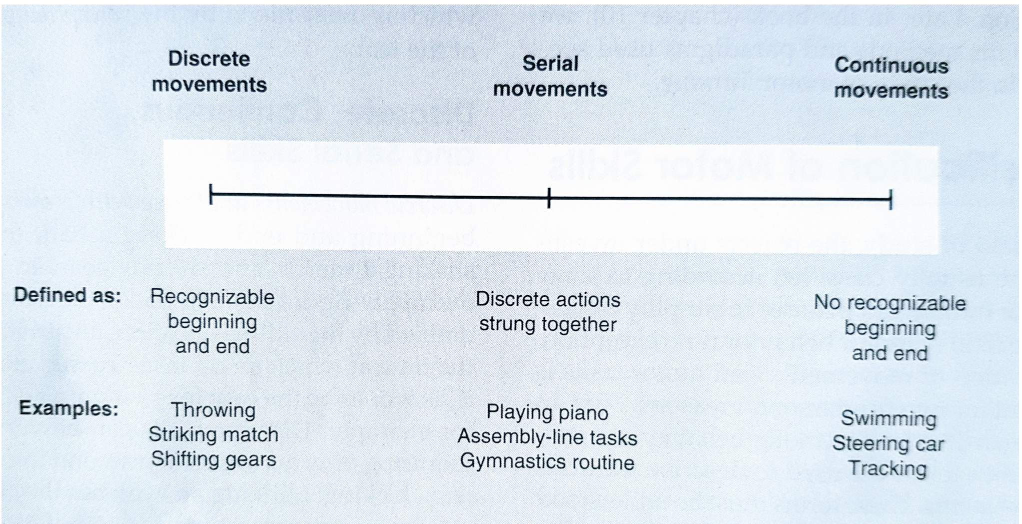

Figure 2.1: Continuum of movements buch [6]

continuing till the movement arbitrarily stopped. Continuous tasks tend to be longer than discrete tasks. Examples are swimming, running or steering a car.
Serial movements are located in the middle part of the continuum. Following the nature of a continuum these movements are neither discrete nor continuous. They can consist of smaller movements tied together. Furthermore, discrete movements can be rather long but are not stopped arbitrarily. Serial tasks can be seen as many discrete tasks strung together and the order (and sometimes timing) is important. Examples are starting a car or preparing and lighting a wood fireplace.

Open and Closed Skills

Open skills: The environment is constantly, unpredictably changing, so the performer cannot plan his activity effectively in advance. Own movements depend on the environment. For example, if a ice hockey player shoots a shot in ice hockey, his own movement is dependent on the movement of the keeper. Another example is driving on a free way. The driver needs to adjust his own driving dependent on the behaviour of the other cars. Success in open skills is largely determined by the extend to which a individual can adapt the planned motor behaviour to the changing environment.
Closed skills: The environment is predictable, mainly because it is stable. This means that the performer can plan his activity in advance. Examples

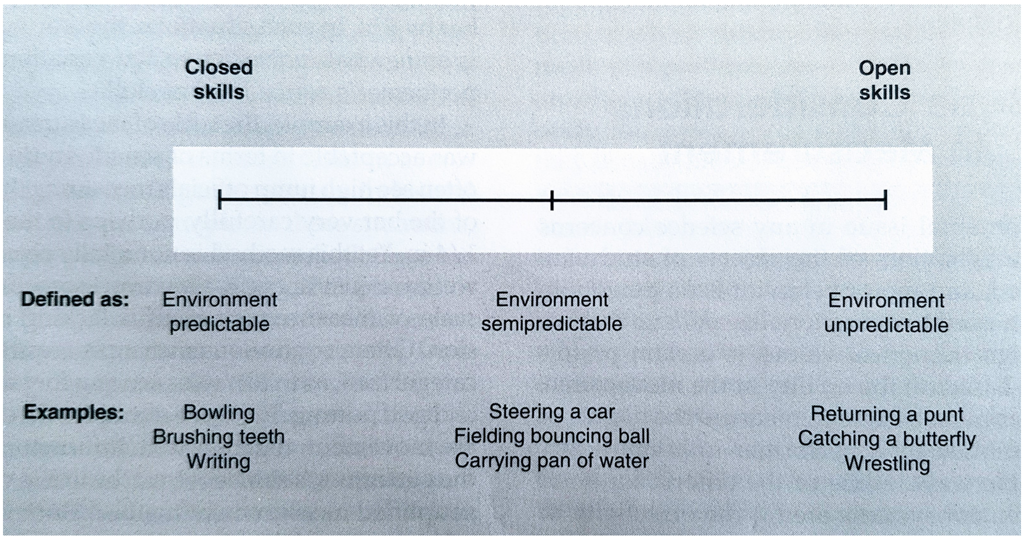

Figure 2.2: Continuum of skills [6] **TODO**seite

are bowling, archery or singing. **TODO**+ citations

How to quantify movements

Judging motions and matching them to a given motion is not a trivial task. One approach follows Rudolph von Laban - a professional dancer. Von Laban developed a broadly used dance notation. His work lead to the *Laban Movement Analysis* with which human movements could be quantized.[1] There are four main components to systematically describe movements in the *Laban Movement Analysis*: body, effort, shape and space. Each component can describe movements independently or combined. Hachimura et al. [2] used the methodology of *Laban Movement Analysis* and adopted it to for digital movements.
Yoshimura et al. [8] followed a similar approach from another dance movement description theory called *furi*. *Furi* is also described by four so called *indices*: *kamae*, *jyu-shin*, *koshi*, *uchiwa*. Yoshimura at all could map these indices to concrete markers on the body of a performer. Qian et al. [1] developed a gesture recognition system for performing arts. To match the motions ten body parts were defined: head, torso, upper arms, forearms, upper legs and lower legs. For each body part the Mahalanobis distance

[1]Brockhaus, Rudolf Laban. http://www.brockhaus.de/ecs/enzy/article/laban-rudolf (accessed 2018-10-25)

is calculated to an ideal point. The Mahalonobis distance describes the distance between point p and distribution D. Kwon et al. [3] **TODO**

- K. Hachimura, K. Takashina, and M. Yoshimura, "Analysis and Evaluation of Dancing Movement Based on LMA," Proc. IEEE Int'l Workshop Robots and Human Interactive Comm., pp. 294-299, 2005.
- M. Yoshimura, N. Mine, T. Kai, and L. Yoshimura, "Quantification of Characteristic Features of Japanese Dance for Individuality Recognition," Proc. IEEE Int'l Workshop Robot and Human Interactive Comm., pp. 193-199, Sept. 2001.
- G. Qian, F. Guo, T. Ingalls, L. Olson, J. James, and T. Rikakis, "A Gesture-Driven Multimodal Interactive Dance System," Proc. IEEE Int'l Conf. Multimedia and Expo (ICME '04), pp. 1579-1582, June 2004.
- D.Y. Kwon and M. Cross, "Combining Body Sensors and Visual Sensors for Motion Training," Proc. ACM SIGCHI, pp. 94-101, 2005.
- vr dance trainer

How to measure movements

In order to judge if a movement is performed correctly methods need to be applied to measure the error of a performed action. In literature, three main categories are listed: error of a single subject, measures of time and speed and measures of movement magnitude.

Measures of Error for a Single Subject

Measures of error for a single subject represent the degree to which the target was not achieved. A target can be to perform an act at a particular time (time stamp), move with a certain force (amount of force) or hit a spatial target (a point in spatial volume). The attribute of the target serves as the variable in question, see braces behind the examples. The error itself describes the distance - in regard to the dimension - from the target. The following list gives an insight to the most important error measures.

- ***Constant Error*** describes the average error between the actual accuracy and the target. Means, in average the performer missed the target by CE.
$$CE = \frac{\sum_i (x_i - T)}{n} \tag{2.1}$$

with x_i: score, n: number of values, T: target value.

- ***Variable Error*** measures the inconsistency in movements. The more consistent the movements, the smaller VE. VE does not depend on whether or not the subject was close to the target.

$$VE = \sqrt{\frac{\sum(x_i - M)^2}{n}} \tag{2.2}$$

- ***Total Variability*** describes the total variability around a target. The combination of VE and CE represents the total amount of spread about the target. It is an overall measure how successful was the subject in achieving the target.

$$E = VE^2 + CE^2 = \sqrt{\frac{\sum(x_i - T)^2}{n}} \tag{2.3}$$

with x_i: score, n: number of values, T: target value.

- ***absolute error*** is a measure of the overall accuracy in performance.

$$AE = \frac{\sum |x_i - T|}{n} \tag{2.4}$$

with x_i: score, n: number of values, T: target value.

- ***Absolute Constant Error*** is the absolute value of CE. Because of negative and positive values can cancel each other out

$$ACE = |CE| \tag{2.5}$$

these measures can be applied to other movements. like pursuit motor: TOT, Mashburn task, stabilometer, two hand coordination task.

Measures of Time and Speed

measures of time and speed: basic to this idea: performer who can accomplish more in a given amount of time or who can accomplish a given amount of behavior is more skillfull. time measure:c $\frac{time}{unit}$. speed:$\frac{units}{time}$.

reaction time (RT): can also be a performance measure. a measure of time from the arrival of a sudden and unanticipated signal to the beginning of the response.

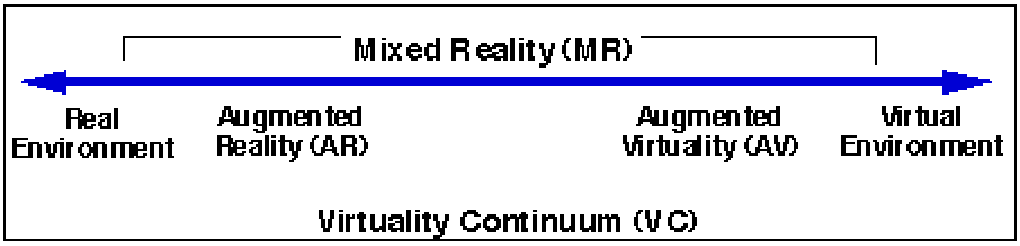

Figure 2.3: Mixed reality continuum by Milgram et al. [4]

Measures of Movement Magnitude

movement time (MT): how long does the movement last. somtimes commbined with RT: response time= $RT + MT$

Movement Classification

eg. Postural, Transport, Manipulation. p4 in motor learning, principles and practices **TODO**after analysis

2.2 Mixed Reality

Mixed reality continuum [4] and something about when AR or VR is better.

2.3 Perspectives

Wang and Milgram [7] describe the perspectives on the centricity continuum see figure 2.4. On the most left hand side of the continuum the egocentric perspective is located. Egocentric means that the anchor of the viewport camera is located inside the object to control - for simplicity, this object in question is referred as avatar. On the left hand side the exocentric perspective is located. This viewport camera is a fixed camera in the scene not to be controllable. The exocentric perspective gives the user the possibility to examine the scene from a bird's-eye view. The movement or angle of the avatar has no influence on the cameras position or angle. So the main difference is the so called tether distance and the degree of freedom of the camera. Milgarm and Wang investigated on tethered cameras and define it as the distance between the avatar and the camera which is following the avatar. This describes the middle part of the continuum. Zero-distance camera describes the egocentric perspective. The longer the tether distance

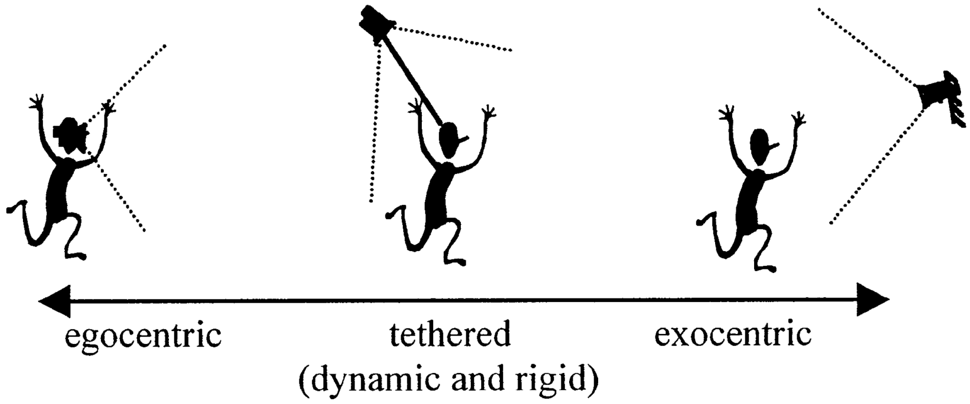

Figure 2.4: Centricity continuum by Wang and Milgram 2001 [7]

the more the perspective is located on the right of the scale to the exocentric perspective. They also distinguish between dynamic and rigid tethering relation ships. A dynamic tethered camera is controlled by the user in all six dofs (**TODO**) while a rigid stands like a pole and can only be controlled in 3 dofs. Rigid tethered cameras are common in modern 3rd person computer games.

2.4 VR Technologies

HMD,3D screens, tablets

2.5 Motion Tracking Technologies

External vs. internal tracking. Drift problem, accuracy

2.6 asynchron vs synchron

2.7 Application

Einsatzmöglichkeiten und MLSF **TODO**

- Theory practical transfere by variation of virtual content
- location independet communication and sharing of experience

- individual support of reflextion internal
- support of reflextion external

3 Scope

TODOmacht kein sinn mehr hier. soll das raus?

3.1 Motor Learning

- discrete movements
- closed skills
- at least 2 different movement categories
- how to measure movements
- posture vs movement

3.2 Mixed Reality

- Milgram
- AR or VR

3.3 Perspective

3.4 Misc

- synchron asynchron
- colocated/remote
- perspective
- hardware?
- feedback!

- real world, not abstract avatars
- only visuals - no audio or textual explanation

4 Related Work

Researchers utilised the theory of the last chapter to design MR Motor Learning systems in various ways. They differ in the technology, tasks, perspectives measures etc. In this chapter we analyse these systems to extract valuable insights to design a MR Motor learning System. First we analyse the aspects of MRML systems. After that we take a close look on some systems how they conducted their investigation and their outcome.

4.1 Aspects of MR Motor Learning systems

paper to come:

- Cruz: Cyclone uppercut
- Davcev: AR Environment for Dance Learning
- Chan: Immersive Performance training Tools Using Motion Capture Technology
- Han: AR-Arm: Augmented Visualization for Guiding Arm Movment in the First-Person Perspective
- han: My Tai-Chi Coaches: An Augmented-Learning Tool for Practicing Tai-Chi Chuan

Method

Jacky Chan et al. **TODO**created a VR dance training system using an optical motion capturing system to compare the movements performed by the student with movements from the avatar. These movements are presented to the student as a 3D rendering on large screen. The movements of the students are visualised on the same screen as a coloured stick figure. The student mimics these movements and gets instant feedback as well as a feedback as a summary.

In contrast, Onebody by Hoang et al. **TODO**use a VR headset for a first person remote posture guidance system.

Tasks

In Chan et al. **TODO**the dance student is presented a virtual avatar performing dance moves of A-go-go or Hip-Hop style. The avatars movement is based on the motion capturing data of a professional dancer. Onebody **TODO**is not only restricted to dance moves but also include other posture based sports or physical activities like Yoga or Mixed Martial Arts.

Onebody **TODO**uses a number of martial arts postures or stances.

- Onebody: 16 artificial postures not from but like: tai chi, matial arts
- VR Dance Trainer: dance movements, 15 min for each move
- you move: various movements to perform. using a whip, baseball, boxing, ballet, dance moves
- training archived physical skills IVE: physical skills in sport activities, especially baseball pitching

Measures and variables

Jacky Chan et al. **TODO**defined 19 body parts that are considered in the measure of the performance of the dancing student. They name three features to compare the difference between two motions common: joint position, joint velocity and joint angle. Chan investigated which of these features suits most to judge the two dancing motions. The outcome of this investigation names the joint position to have the highest discriminative power. Hence, the joint position suits them best for their evaluation, Chan et al. calculate a score of the position error for each of the defined body parts, as well as an overall score.

Onebody **TODO**uses skeletal of the instructor and the student "Posture accuracy is determined by the extend to which the student can replicate the final posture as instructed and demonstrated as by the instructor." Independent variable: mothods for posture training. dependent variables: performance factors of posture accuracy, completion time, subjective instructor rating, users preference.

- scientific work, how to measure movements: hachimura et al, yoshimura et al, qian et al, kwon et al, all use joint angles, (mentioned in: vr dance trainer)

- onebody: skeletal tracking, how much percent do the postures match? 3d positions of limbs measured: wrist, elbow, shoulder, hip knee ankle. so angle between bones is the main measure for accuracy. additionally a subjective instructor score was recorded. And, completion time, topped by 2 min.
- VR Dance trainer: there are 3 common features for measuring hte difference between movements: joint position, velocity, angle. they tested which feature descibes movements best: joint position. base line vs post training movements are compared.
- you move: in each keyframe, score based on the joint with the maximum error, measured in euclidean distance. but only "important joints" are measured. timing errors: 0.5s error on each side of the frame for matching posture. is timing important, the window is reduced to 0.25s. max eucl. distance is linear mapped to a score. 0 error is 10 (max), 10cm is 7.5 what is the score to pass. if precision is important, 10cm needed for pass.

Considered Body Parts

scientific work, how to measure movements: hachimura et al, yoshimura et al, qian et al, kwon et al, all use joint angles

error prevention **TODO**

4.2 Detailed description of 6-10 papers incl. Table

Onebody: Remote Posture Guidance System using First Person View in Virtual Environment

Hardware: Kinect, Oculus Rift

Task: sports, dance, martial arts, yoga

Perspectives: First Person of the teacher

Measures: Position, completion time, subjective score

investigation: comparing different remote guidance systems: onebody, pre recorded video, video conference, VR third person

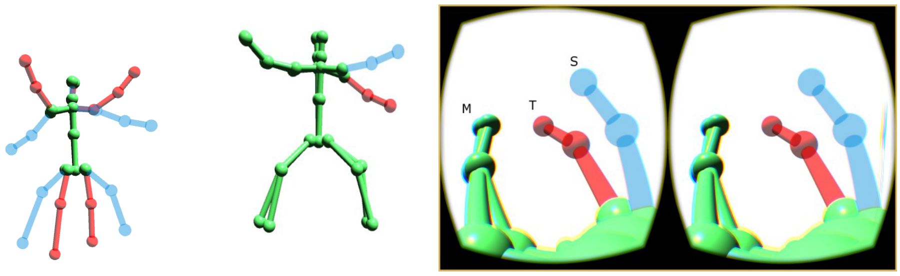

Figure 4.1: Left: student avatar (blue) and teacher avatar (red). Green limbs are matching limbs. Right: students view on the scene. [5]

variables: Independent: training method, dependent: performance

Outcome: Onebody offers better posture accuracy in delivering movement instructions

Onebody by Hoang et al. [5] is a VR system for remote posture guidance. Onebody is designed for sports or physical activity training like yoga, dance or martial arts. The student and the teacher are both tracked by skeletal tracking. The visualisation of this tracking are shown via a VR headset, allowing the student to follow the instruction of the teacher in first person view of the teacher - which means the student "stands inside the body of the teacher". Both, the student and the teacher are visualised by stick figures. The teachers avatar is red, the students blue and matching joints are green like shown in figure 4.1 left. Figure 4.1 right shows the scene from the first person perspective. When the teacher moves his limbs, the student can see the movement emerging from himself. Now the student can move his own limbs to mimic the movement till the teachers posture is matched. The teacher sees the students limbs likewise allowing him to give instant feedback to the student. Thus, "Onebody provides a medium to deliver body movement instructions for non-collocated instructor and learner." **TODO**. The visualisations are attached to the hip but keeps the mapping between the user and corresponding avatar. To overcome different body sizes, the avatars are normalised and scaled to the size of the person seeing the avatars.
For transfering data, both the teacher and the student are clients in a server-client system. The clients are sending the their tracking data to the server which is broadcasting it to the clients. The comparison of the limbs for colour coding is performed on the client side. The matching of the limbs is calculated by the position of the single limbs (see equation (2.1)) with a threshold of 5cm to reduce jitter and tracking errors. Limbs in question

	Video	*Skype*	*VR-3PP*	*Onebody*
Synchronous Interaction	No	Yes	Yes	Yes
VR Medium	No	No	Yes	Yes
1st Person View	No	No	No	Yes

Figure 4.2: Training methods and their differences used in the study to evaluate Onebody [5]

are wrist, elbow, shoulder, hip, knee and ankle. The feedback with colour codes is provided in realtime.
With this system hoang et al. designed a user study to evaluate the performance of posture accuracy and user's preference. Their main hypotheses is "*Onebody delivers better posture accuracy than existing remote movement instruction methods*". "Posture accuracy is determined by the extend to which the student can replicate the final posture as instructed and demonstrated by the instructor." In addition, completion time and a subjective score of the instructor are considered. To test the hypothesis, Onebody was compared with three other remote posture training methods (independent variables): pre recorded video, video conference (Skype), VR 3rd person perspective. Each of the systems differs to Onebody in terms of synchronous interaction, VR medium and perspective see figure 4.2. The study was a 4x4 within subject. Each participant stated with a training session in which the not collocated instructor teach a posture physically and verbally. Verbal feedback was given the training repeated until the student was confident. After that the final posture was recorded. A set of four of postures with every system were performed with different complexities.
The results show a significant difference in accuracy. Onebody performed significantly better in over video conference, 3rd person VR and pre recorded video. Furthermore, the completion time was significantly higher with Onebody as in the other three systems. The subjective score of the instructor showed no significant differences. A post questionnaire indicated that Onebody is harder to understand and use than the other systems, but at the same time it also indicated that Onebody was perceived to be more exact. Participants rated video conference as their most preferred system over Onebody and 3rd person VR.

Training for Physical Tasks in Virtual Environments: Tai Chi

Hardware: HMD, Optical motion tracking

Task: Tai Chi

Perspectives: Ego-centric, exo-centric, combined

Measures: position

investigation: different perspectives on virtual avatar(s)

variables: independent: perspectives, dependent: precision

Outcome: None representation proved to be significantly better than traditional virtual Tai Chi teacher

A VR Dance Training System Using Motion Capture Technology

Hardware: Optical Motion tracking, 3D screen

Task: Dance moves

Perspectives: exo-centric

Measures: position

investigation: Comparing video based learning with VR learning with feedback

variables: independent: training method, dependent: precision

Outcome: better assists in learning compared to traditional video approach, as well as more motivation an fun

YouMove: Enhancing Movement Training with an Augmented Reality Mirror

Hardware: Augmented reality mirror, Kinect

Task: dance moves, sport moves

Perspectives: exo-centric

Measures: postition

investigation: Comparing video based learning with YouMove

variables: independent: training method, dependent: precision

Outcome: learning and short-term retention better than traditional video representation

Generic Heading, input? TODO

Hardware:

Task:

Perspectives:

Measures:

investigation:

variables:

Outcome:

Generic Heading, any ideas? TODO

Hardware:

Task:

Perspectives:

Measures:

investigation:

variables:

Outcome:

4.3 Research Gap

lücken in der aktuellen forschung die sich lohnen zu untersuchen. **TODO**

Name	Hardware	Perspectives	Task	Measures	Variables
Onebody					
VR dance trainer					

Table 4.1: Summary of proposed MR learning systems

Conclusion of VR motor learning systems

to conclude we have a look a Table 4.1

Research questions?

Hypothesis?

5 Outlook

In the last chapter we concluded that good is ... and bad is ... so we propse the following study design:
Research questions?
hypothesis? **TODO**

5.1 Preliminary Study design

Aim of the Study

The aim of the study is to investigate the influence of egocentric and exocentric perspectives on a virtual avatar during motor learning tasks.

process

There are two groups: one learn only with the egocentric perspective, the other one with the exocentric perspective on the virtual avatar.
To derive conclusions on body regions, every participant learns movements for three different body parts. The body parts are:

- *upper body*(UB)
- *lower body*(LB)
- *full body*(FB)

To derive conclusion on movement types, two different movements per body part is learned. The two movement types are:

- mirrored movements
- independent movements

	UB	LB	FB
Ego	1 mirrored and 1 asynchronous movement	1 mirrored and 1 independent movement	1 mirrored and 1 independent movement
Exo	1 mirrored and 1 independent movement	1 mirrored and 1 independent movement	1 mirrored and 1 independent movement
Ego/Exo	1 mirrored and 1 independent movement	1 mirrored and 1 independent movement	1 mirrored and 1 independent movement

Independent variables

- perspective on the avatar (Ego/Exo centric)
- body parts (*upper body*, *lower body*, *full body*)
- movement types (mirrored/independent movements)

measures

TBA

5.2 Timetable

Bibliography

[1] Gang Qian, Feng Guo, T. Ingalls, L. Olson, J. James, and T. Rikakis. A gesture-driven multimodal interactive dance system. (January):1579–1582, 2005.

[2] Kozaburo Hachimura, Katsumi Takashina, and Mitsu Yoshimura. Analysis and evaluation of dancing movement based on LMA. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2005:294–299, 2005.

[3] Doo Young Kwon and Markus Gross. Combining body sensors and visual sensors for motion training. *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '05*, pages 94–101, 2005.

[4] Paul Milgram and Fumio Kishino. A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS. IEICE Transactions on Information Systems, Vol E77-D, No.12. *IEICE Transactions on Information Systems*, E77-D(12):1–15, 1994.

[5] Martin Reinoso. Onebody : Remote Posture Guidance System using First Person View in Virtual Environment. 2016.

[6] Richard Schmidt and Tim Lee. *Motor Control and Learning: A Behavioral Approach*. 2011.

[7] Wenbi Wang and Paul Milgram. Dynamic Viewpoint Tethering for Navigation in Large-scale Virtual Environments. *Human Factors*, pages 1862–1866, 2001.

[8] Mitsu Yoshimura, Norio Mine, Tamiko Kai, and Yoshimura Isao. Quantification of Characteristic Features of Japanese Dance for. pages 188–193, 2005.

List of Figures

List of Tables