Motor learning in Mixed Reality

Stefan Paul Feyer
HCI Group, University of Konstanz

August 5, 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 for apprentice, maintenance, or remote training. Eg. The Smart Glass experience Lab[4] of the Fraunhofer Institute use the hololens for remote maintenance.
For developing MR learning and training environments, research put much afford in developing how-to's and guidelines to ensure proper systems. However - as we will see in Chapter 4 - there is a research gap about the perspective in these systems. For example, a student wants to learn a speacial task from a teacher. In the real world, the teacher stands in front of the student preforming the task and the student tries to mimic it. This perspective is called exocentric or 3rd person. In contrast, in MR we have the possibility to change this perspective what we cannot do in the real world. We can "step into" the teachers virtual body and see the instruction from the 1st person view of the teacher, also called egocentric view. Changing the perspective could have influence on the learning. This rises

[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

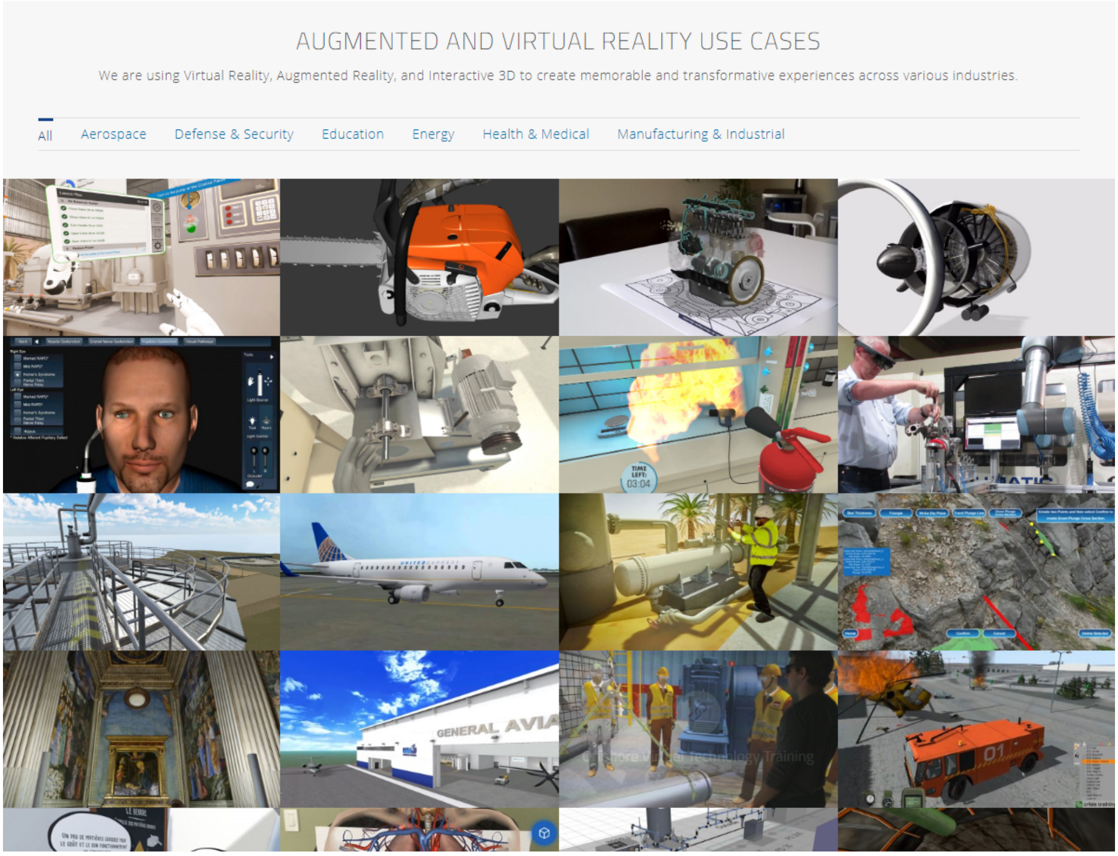

Figure 1.1: Usecases by EONReality on their AVR plattform. https://www.eonreality.com/use-cases/ (accessed 30.07.2019)

the following question: Does the perspective has influence on learning in MR environments? Furthermore, to develop guidelines for MR learning environments, it would be helpful to know if the there are conditions where either of this perspectives is more suitable than the other. Since learning and training in MR is a vast area, the topic is narrowed down to the subset of motor learning. This seminar thesis is the first out of three parts, followed by a Masters Project and a Masters thesis. The overall aim of this work is to answer to following research question:

RQ1 Does the perspective on a virtual avatar has influence on learning in MR environments.

RQ1.1 Are there conditions, under which one perspective works better than the other.

The outcome in this work is a study design that will be able to address the research

Figure 1.2: Remote maintenance with Hololens by Smart Glass Experience Lab. The on site worker wears a Hololens, while the remote trainer draws green hints to resolve a miswiring, taken from https://www.youtube.com/watch?v=1QFMPo5k6p0, accessed (30.07.2019)

questions. In the Masters project the proposed study setting will be implemented, to be able to conduct the study and collecting the data necessary to answer the question. The Masters thesis itself will take the generated data to answer the research question.

1.2 Outline

For a proper study design many aspects must be taken into consideration. The main aspects this thesis will discuss are defined in the following, while further aspects like algorithms are discussed in the masters project.

A1 independent variables

A2 dependent variables

A3 movement classification

A4 open vs. closed skills

A5 task

A6 perspective

A7 MR vs. VR

A8 MR technology

A9 tracking technology

A10 synchronous vs. asynchronous instructions

A11 behaviour of instructions

A12 measures

A13 considered body parts

A14 degree of realism of the avatar

A15 feedback

Each of these aspects must be chosen wisely. In order to do so this seminar thesis will systematically go through every one of them. Chapter 3 provides insight to these aspects from a theoretical view. In addition, this chapter provides general knowledge about the domain. In contrast, chapter 4 takes the work of other researchers into consideration. Here, the aspects are analysed from a more practical view and it is investigated how these researchers decided about the aspects and why. Whenever a decision is made to be used in the proposed study design, a
A* symbol on the side of the text can be found . After all aspects are clear and reasoned, a study design is proposed in chapter 5. In the end, an outlook on further work on the Masters thesis is given and highlighted what aspects are still to decide of.

TODOgraphical representation

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

If not other indicated, adopted from the book Motor Learning and Skills [6].

Learning movements

In the real world Motor learning takes place by instruction, trying, imitation or a combination of two or all three. A learner can observe another person and imitate the movement, try to accomplish a task by themself or can follow instructions. Instructions can be written, visual or verbal. Written instructions are not bound to words solely, eg. Rudolf van Laban developed a dace notation system, compare section "quantify movements" 2.1 and figure 2.3. Visual or verbal instruction include a trainer, teaching the student movements. In this case verbal and physical feedback also plays a role in the learning process. The process of motor learning is divided in three parts. Once a student starts learning using what ever technique, it starts in the *cognitive* state. In this stage the students tries to figure out what is to be done to achieve the task. For this high cognitive activity is required, strategies are evaluated. The performance gains dramatically and is larger than in any other stage, but also inconsistent. The use of instructions and other training techniques are most effective. The next stage is the *fixation*. It begins when the student had determined the most effective way of doing the task. Performance increases more gradually but becomes more consistent. In the last *autonomous* stage, the performer gains proficiency and other tasks are less likely to interfere. Since the use of training techniques and the high performance gain in the *cognitive* state, tasks in this stage are best suited for the study. This will be considered in

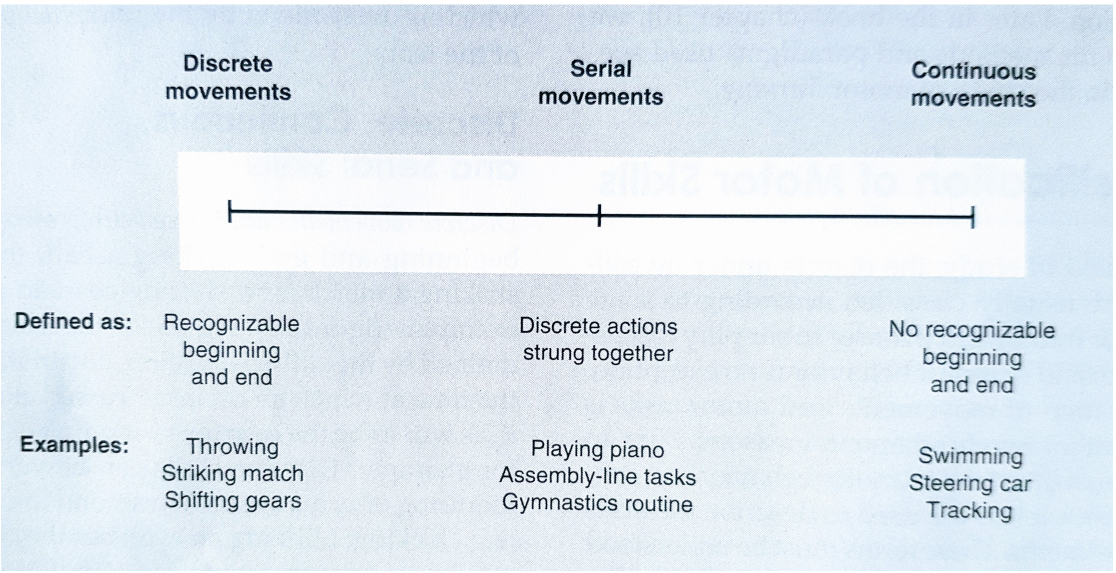

Figure 2.1: Continuum of movements buch [6]

the next chapter, too.
In the digital world most of this remains like it is in the real world. But with computational power new learning techniques can be used to teach motor learning. Instructions can be given visually by a teacher without the teacher being physically in the room, eg. by video instructions or virtual avatars. Also remote instructions become more practical eg. in video conferences. One important point is feedback. **TODO**can be given verbally, in MR visually ... exaple 1,2 and 3, but in the end:
A15 we do not want to evaluate feedback so there will be none in the study .

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 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. The nature of *Continuous movements* having no recognizable beginning and end makes it hard to describe a distinctive task for a study design while *discrete movements* are to short for a proper task, *serial movements* are choose for the study task .

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 are bowling, archery or singing. To evaluate only the motor learning and not environmental influences the study is conducted in a controlled environment in a laboratory. Thus only *closed skills* are taken into consideration .

Quantify movements TODOor out

Judging motions and matching them to a given motion is not a trivial task. Since eg. dancing is a pure physical task, movements must be recognised, digitalised and judged. One approach is to use a analog descriptions for dancing and translate them in the digital world. Choensawat [1] began with 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

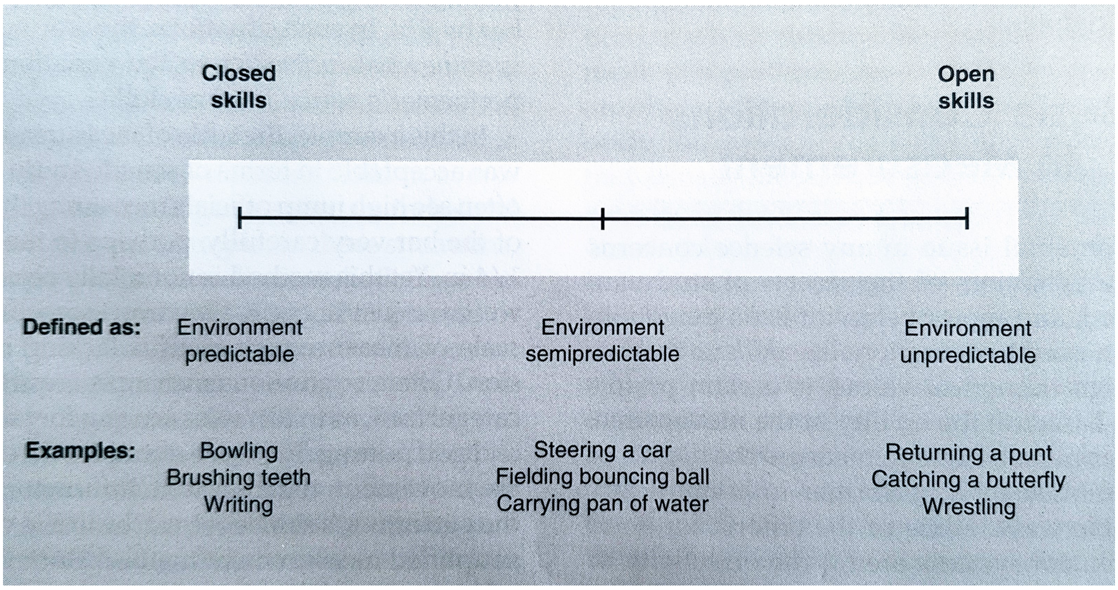

Figure 2.2: Continuum of skills [6] TODOseite

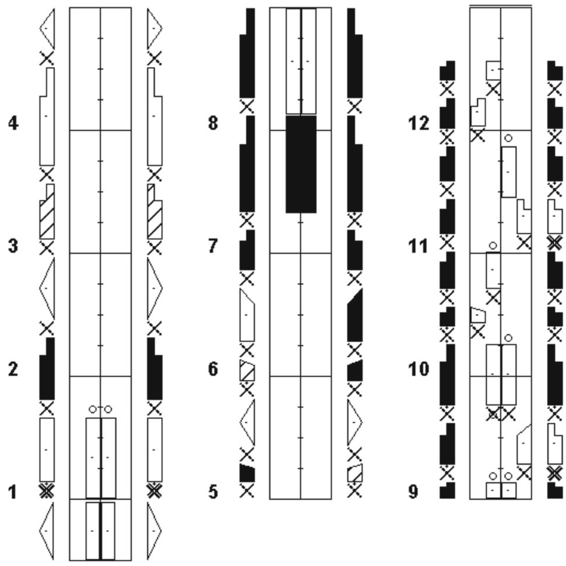

Figure 2.3: Laban notation. Generated through automatic movement interpretation by Choensawat [1]

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. [3] 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. [2] 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 is calculated to an ideal point. The Mahalonobis distance describes the distance between point p and distribution D.

So the takeaway message is, there is a need to digitalise movement and then apply suitable algorithms to judge these movements. The algorithms will be discussed in the Masters project.

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.

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

- ***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.

Measures of Movement Magnitude

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

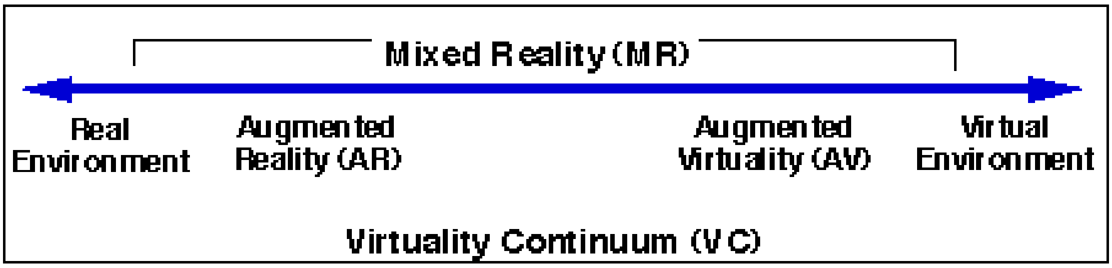

Figure 2.4: Mixed reality continuum by Milgram et al. [4]

2.2 Mixed Reality

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

2.3 Perspectives

Wang and Milgram [7] describe the perspectives on the centricity continuum see figure 2.5. 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 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 or in project

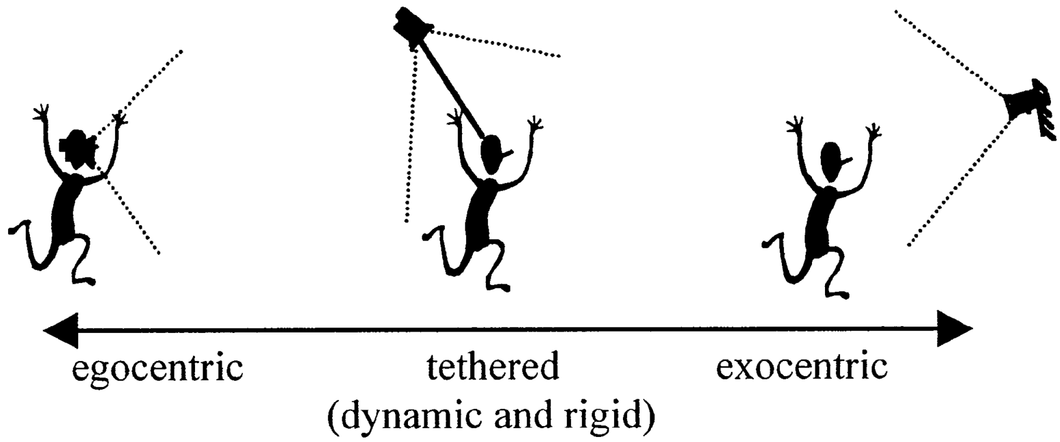

Figure 2.5: Centricity continuum by Wang and Milgram 2001 [7]

2.5 Motion Tracking Technologies

External vs. internal tracking. Drift problem, accuracy or in project.

2.6 asynchron vs synchron

we must decide on one, so we do same place with pre recorded movements different
A10 time

2.7 realismusgrad

first some paper collection about how the virtual avatar looked like, and then
decide for what rebeccas work proposes. its a lot about related work, so maybe
A14 this should go in the next chapter. what would be more work.

2.8 Conclusion

we dicided for A1 A2 ... will be used in study ... some aspects missing, lets have a look how other researchers choose there, what we do in the next chapter.

3 Related Work

In chapter 2 A1, A3, A4, A6, A7, A10, A14 and 15 are already discussed. In this chapter the remaining aspects are analysed by investigating how other researchers decided on them. The remaining aspects are dependent variables (A2), the task (A5), the MR technology to use (A8), the tracking technology (A9), the behaviour of the instructions (A11), the measures to apply (A12) and the considered body parts (A14). These aspects are better discussed at existing systems to make sure the study having a high validity. This chapter is structured by means of the perspective though the independent variables, namely ego-centric, exo-centric and the combination of both. In this subchapters the aspects are discussed and concluded. In the overall conclusion a decision is made for the aspects by comparing them to each other. eg, the task must be suitable for all perspectives.

3.1 Exo-Centric

In the following systems utilise the exo-centric perspective. The remaining aspects are analysed in respect to this perspective.

Task

papers xyz used task abc

MR tech

papers xyz used abc mr tech

tracking tech

papers xyz used abc track tech

Task	MR Tech	tracking tech	instruction	measures	body parts

behaviour of instruction

instructions were given in this way

measures (for dependent vars)

paper xyz used abc as measures

considered body parts

paper xyz used bodyparts abc for the measurement

In detail - A VR Dance Training System Using Motion Capture Technology

desctiption of one paper in detail

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

conclusion

conclusion

3.2 Ego-Centric

beschr

Task

task

MR tech

mr

tracking tech

track

behaviour of instruction

behav

measures (for dependent vars)

measures

considered body parts

bp

in detail - 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

variables: Independent: training method, dependent: performance

Outcome: Onebody offers better posture accuracy in delivering movement instructions

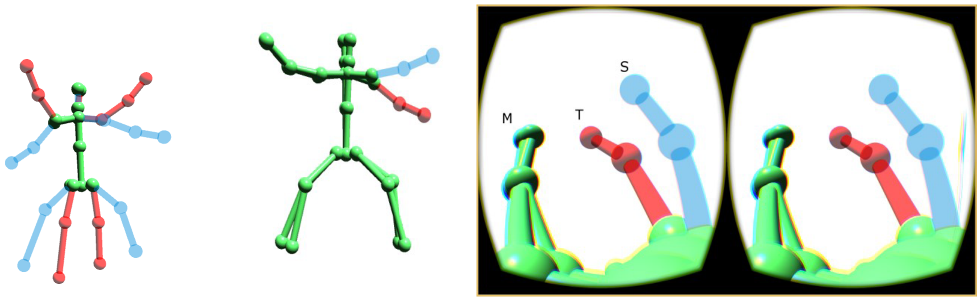

Figure 3.1: Left: student avatar (blue) and teacher avatar (red). Green limbs are matching limbs. Right: students view on the scene. [5]

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 3.1 left. Figure 3.1 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 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

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

Figure 3.2: Training methods and their differences used in the study to evaluate Onebody [5]

Task	MR Tech	tracking tech	instruction	measures	body parts

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 3.2. 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.

conclusion

conclusion

3.3 Ego-Exo Combined

beschr

Task

task

MR tech

mr

tracking tech

track

behaviour of instruction

behav

measures (for dependent vars)

measures

considered body parts

bp

In detail - 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

Task	MR Tech	tracking tech	instruction	measures	body parts

Task	MR Tech	tracking tech	instruction	measures	body parts

conclusion

conclusion

3.4 Conclusion

decide for aspects based on the tables before.
A2 dependent
A5 task
A8 MR tech
A9 track tech
A11 behav
A12 measures
A13 BP

4 Proposed Study Setting

4.1 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 Outlook

Outlook about what is next, open questions like, tracking algorithms, filters, error handling...

5.1 Timetable

Bibliography

[1] Worawat Choensawat, Minako Nakamura, and Kozaburo Hachimura. GenLaban: A tool for generating Labanotation from motion capture data. *Multimedia Tools and Applications*, 74(23):10823–10846, 2015.

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

[3] 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.

[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