

Motor learning in Mixed Reality

Stefan Paul Feyer
HCI Group, University of Konstanz

October 29, 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

Contents	ii
1 Introduction	1
1.1 Motivation	1
1.2 Outline	3
2 Theory and Scope	5
2.1 Visual Perspectives	5
2.2 Motor Learning	6
2.3 Mixed Reality	12
2.4 Conclusion	13
3 Related Work	15
3.1 Overview	15
3.2 Comprehension of the papers analysed	15
3.3 Analysis	19
3.4 in detail - Onebody: Remote Posture Guidance System using First Person View in Virtual Environment	27
3.5 conclusion	29
4 Proposed Study Setting	31
4.1 Hypothesises	31
4.2 Variables	31
4.3 Task	32
5 Outlook	33
5.1 Timetable	33
Bibliography	35
List of Figures	36

List of Tables	36
-----------------------	-----------

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 1.1. Microsoft also steps into this topic with partners, developing tools for apprentice, maintenance, or remote training. Eg. The Smart Glass experience Lab⁴ of the Fraunhofer Institute use the hololens for remote maintenance, compare figure 1.2.

For developing MR learning and training environments, research put much afford in developing how-to's and guidelines to ensure proper systems [5]. 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

¹TODO

²TODO

³<https://www.eonreality.com/> accessed: 14.12.2018

⁴<https://www.fit.fraunhofer.de/de/fb/cscw/smart-glasses-experience-lab.html>

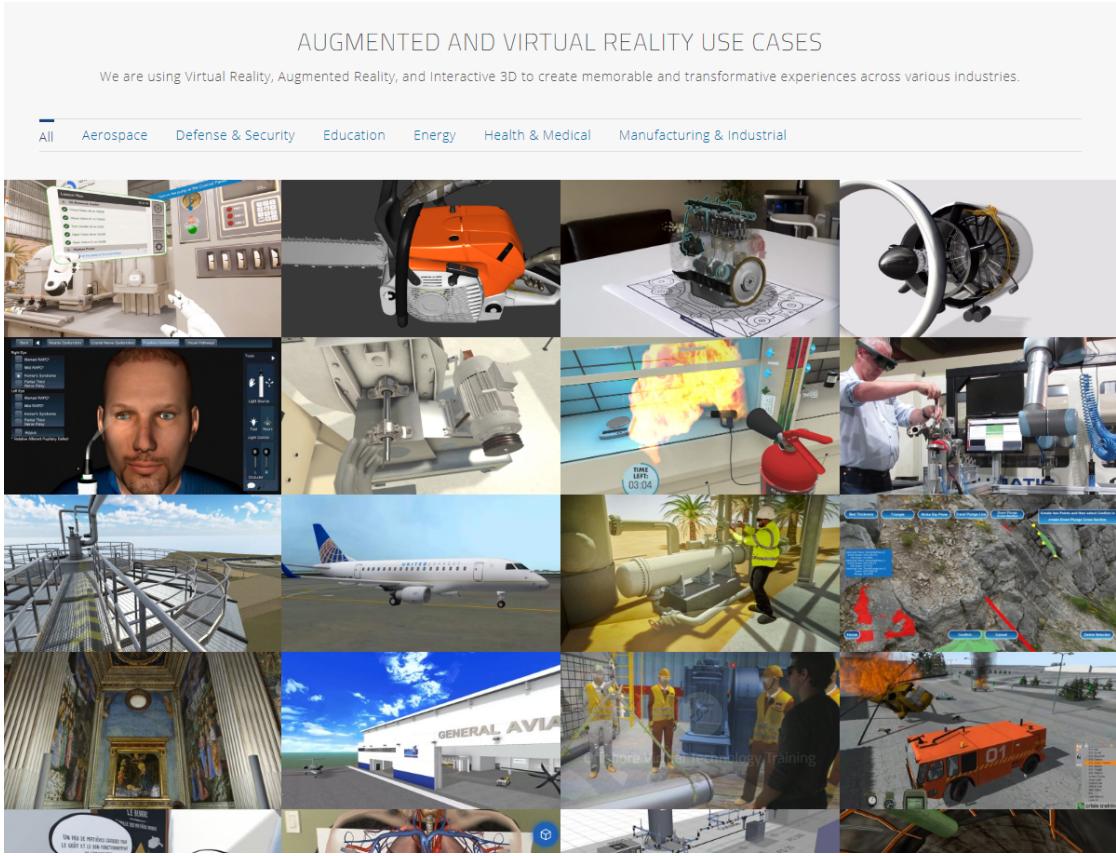


Figure 1.1: Usecases by EONReality on their AVR plattform.
<https://www.eonreality.com/use-cases/> (accessed 30.07.2019)

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 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.

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



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)

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.

- S1 Visual perspective
- S2 Motor learning
- S3 Mixed reality
- C1 Study Task
- C2 Perspective implementation
- C3 Guidance visualisation

C4 Dependent Variables

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 sets the scope of the study and provides general knowledge about the domain. The following chapter 4 analyses the work of researchers and their systems, to find suitable components for the scope that has been set in chapter 4. Here, the remaining aspects are analysed from a more practical view and it is investigated how researchers decided about the aspects and why. Whenever a decision is made to be used in the proposed study design, 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.

A*

TODOgraphical representation

2 Theory and Scope

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 transferred 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. If not other indicated, adopted from the book Motor Learning and Skills [7].

2.1 Visual Perspectives

Wang and Milgram [9] describe the perspectives on the centricity continuum see figure 2.1. 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 relationships. 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.

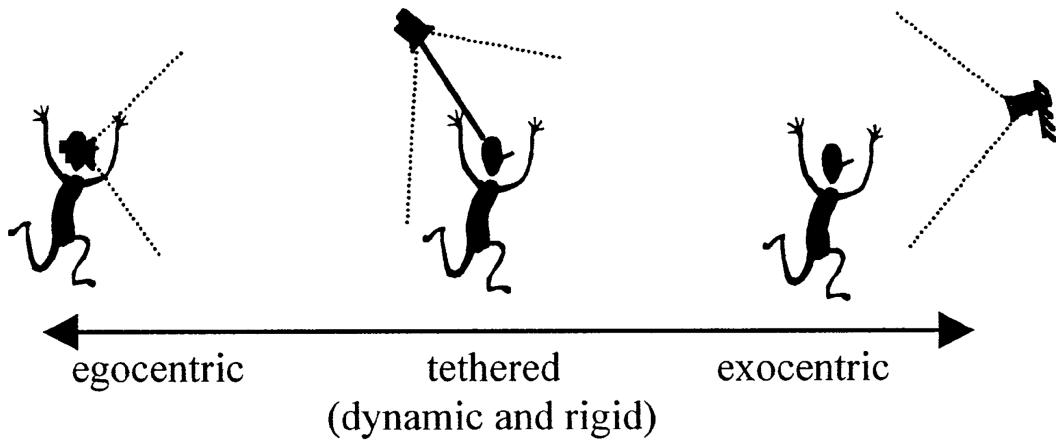


Figure 2.1: Centricity continuum by Wang and Milgram 2001 [9]

Degree of realism of teacher avatar

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 this should go in the next chapter. what would be more work.

Feedback and instruction behaviour

feedback **TODO**can be given verbally, in MR visually ... example 1,2 and 3, but in the end: we do not want to evaluate feedback so there will be none in the study . In addition, for high validity instructions must be the same, so a pre recorded instructions necessary partly answered, but lets look in next chapter about continuity in instruction.

2.2 Motor Learning

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 themselves 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" 2.2 and figure 2.4. 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 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.

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.

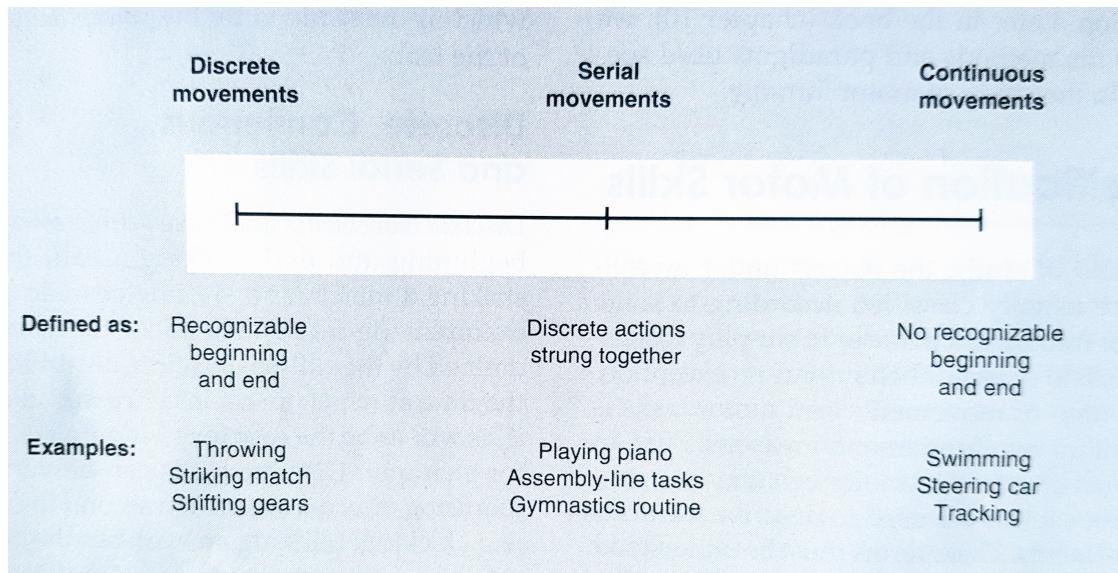


Figure 2.2: Continuum of movements [7]

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 too short for a proper task, *serial movements* are chosen for the study task .

A3

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 extent to which an 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 in-

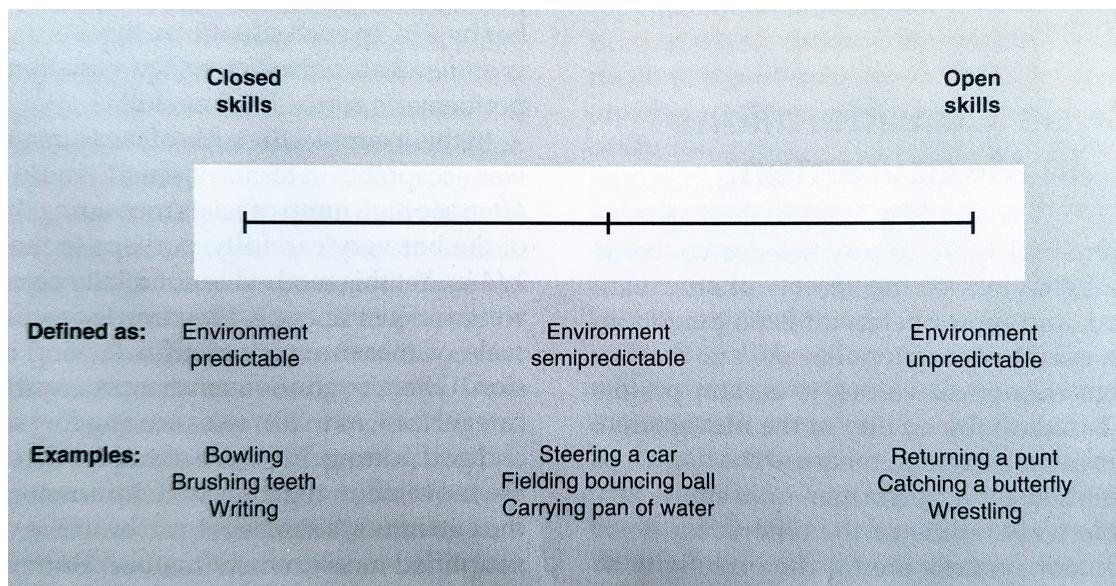


Figure 2.3: Continuum of skills [7] **TODO**seite

fluences the study is conducted in a controlled environment in a laboratory. Thus only *closed skills* are taken into consideration .

Quantify movements **TODO**in measures

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 analog descriptions for dancing and translate them in the digital world. Choensawat [?] 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 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. [4] used the methodology of *Laban Movement Analysis* and adopted it to for digital movements.

Yoshimura et al. [?] followed a similar approach from another dance movement description theory called *furi*. *Furi* is also described by four so called *indices*: *ka-mae*, *jyu-shin*, *koshi*, *uchiwa*. Yoshimura at all could map these indices to concrete markers on the body of a performer. Qian et al. [?] developed a gesture recognition

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

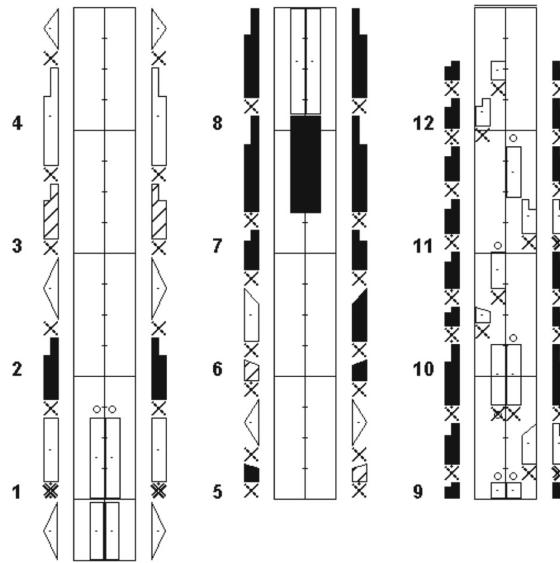


Figure 2.4: Laban notation. Generated through automatic movement interpretation by Choensawat [?]

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.

Measuring 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*. As chapter 3 will show, *error of a single object* is most commonly used, the latter two are only discussed in short.

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.

$$CE = \frac{\sum_i(x_i - T)}{n} \quad (2.1)$$

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.

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

- **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.

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

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

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

$$AE = \frac{\sum|x_i - T|}{n} \quad (2.4)$$

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

$$ACE = |CE| \quad (2.5)$$

Measures of Time and Speed

Basic to this idea is, a performer who can accomplish more in a given amount of time or who can accomplish a given amount of behaviour in less time is more skillful. Measures here are $\frac{\text{time}}{\text{unit}}$ or $\frac{\text{units}}{\text{time}}$.

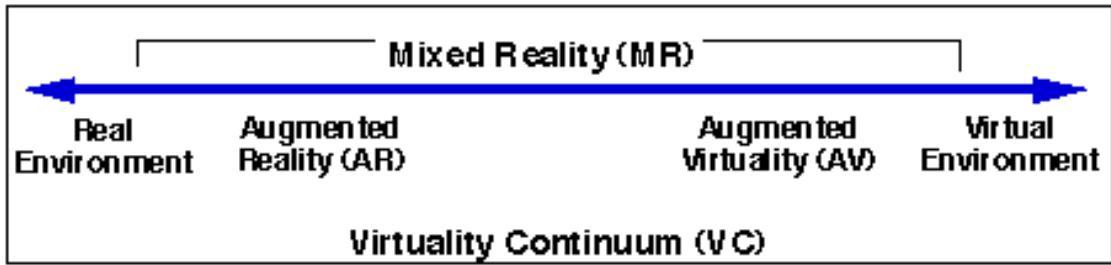


Figure 2.5: Mixed reality continuum by Milgram et al. [?]

The two most common examples are *reaction time* (RT) and *movement time* (MT). Reaction time describes the amount of time between a stimuli and the regarding start of a movement. This time span is important for two reasons. On the one hand, RT has a high validity for real-life tasks, and, on the other hand, RT measures the time taken for mental events like stimulus processing or decision making. *Movement time* is the time interval between the end of the RT phase, though the start of the response, and the completion of the movement. The sum of RT and MT is called *response time*.

Measures of Movement Magnitude

To measure a skill, the produced magnitude of a behaviour can be used. Eg. the distance a discus is thrown. A famous example is the "ski simulator". Rubber bands hold a plate centred between two poles. The magnitude in this case is the dislocation of the board from the centre by using full body movements.

2.3 Mixed Reality

Mixed reality continuum [?] and something about when AR or VR is better. 2.5

A7 **TODO**

VR Technologies

HMD, 3D screens, tablets.... **TODO** or in project

Motion Tracking Technologies

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

2.4 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

TODO: remake description!

3.1 Overview

Here will stand an aggregated overview about work. The systems of the next section are not mentioned here.

3.2 Comprehension of the papers analysed

because I chose to split by topics (task, measures...) here is an overview over the papers that going to be discussed.

In detail - 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

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: Ego-centric

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

In detail - Training for Physical Tasks in Virtual Environments: Tai Chi

Hardware: HMD, Optical motion tracking

Task: Tai Chi

Perspectives: Ego-centric, exo-centric & egocentric combined

Measures: position

investigation: different perspectives on virtual avatar(s)

variables: independent: perspectives, dependent: precision

Outcome: None representation proved to be significantly better than the other representations

YouMove: Enhancing Movement Training with an Augmented Reality Mirror

Hardware: Augmented reality mirror, Kinect

Task: dance moves, sport moves

Perspectives: exo-centric, ego-centric, combined

Measures: position matching

investigation: Comparing video based learning with YouMove

variables: independent: training method, dependent: precision

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

AR-Arm and tARget

Hardware:

Task:

Perspectives:

Measures:

investigation:

variables:

Outcome:

LightGuide

Hardware: Depth Camera, Projector

Task: Abstract movement figures

Perspectives: Ego.centric

Measures: Accuracy

investigation:

variables: independent: guiding technique, dependent accuracy

Outcome: 85% more accuracy compared to video

Generic Heading

Hardware:

Task:

Perspectives:

Measures:

investigation:

variables:

Outcome:

Generic Heading

Hardware:

Task:

Perspectives:

Measures:

investigation:

variables:

Outcome:

Generic Heading

Hardware:

Task:

Perspectives:

Measures:

investigation:

variables:

Outcome:

3.3 Analysis

Study task

The task in Chua's [3] **Tai Chi trainer** is – as the name indicates – a Tai Chi motion. A professional Tai Chi trainer was invited to perform a so called *Tai Chi form* and recorded offline. This *form* was segmented in four ca. 20 seconds long sequences (Motion 1-4). ”Motion 1 featured simple hand movements and a 90 degree turn to the right; motion 2 had little hand motion, a 180 degree turn of the feet, and some turning of the upper body; motion 3 featured slow hand motions, large movement of the feet, and some turning of the thorax; and motion 4 featured swift hand movement and turning of the thorax and hips but little movement of the feet.” [3]. Error measurements indicated that all motion but motion 1 had the same difficulty, being significantly easier. These movements can be classified as *sequential movements* according to chapter 2 **TODO**ref. 40 volunteers conducted the movements in a study to evaluate their system. They randomized the condition (described in next chapter, compare figure 3.2 condition a-e) and the motion to minimize learning effects. For each motion and condition pair, the Tai Chi student were asked to match the Tai Chi teachers demonstration during twelve repetitions.

YouMove by Anderson et al. [1] is a movement training system, suitable for a vast range of moves. A movement can be recorded and then edited by an authoring tool. After the editing, the movement is added to the internal library, from which it can be chosen by a student. For the study itself, an author or the paper – though no professional – recorded four movements. Two of them from ballet and two abstract movement. The authors decided to variate the difficulty of the task, namely ”the ballet movements [...] were easier to conceptualise and required only moderate movement.”, while ”the abstract movements were more difficult to perform, as they were a series of postures with no clear structure and required substantial movement.” The movements consisted of four keyframes. Keyframes are important points during the movements, determined and set by the person recording a movement. Jacky Chan et al. developed a **VR Dance trainer** [2]. For their system they invited professional Hip-Hop dancers and recorded their movements. The learner of the movement can choose a movement out of a database. One movement lasts around 2 seconds. First, the teacher appears and the student watches the demonstration, while he can adjust the demonstration speed and viewpoint. After that, the student practices the dance moves by mimicking the teachers avatar. Finally the student can see a slow motion replay of the performance. One hole session takes 15 minutes. In contrast to all other systems, Lightguide by Sodhi et al. [8] build on abstract movements to evaluate their system. Sodhi et al. focus on single arm movements, directed by an guidance visualisation on the

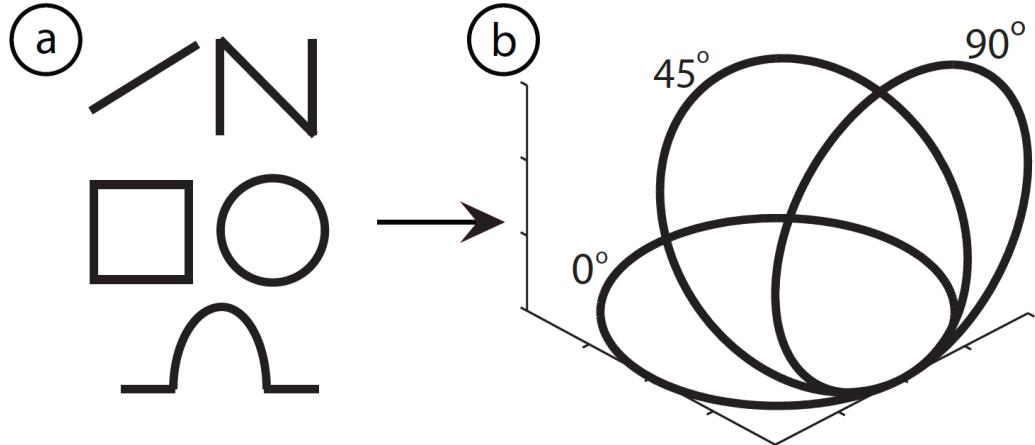


Figure 3.1: movements by lightguide [8]

hand. The participants perform five movements, namely a line, a "N", a square, a circle and a bowed line, compare figure 3.1 left. All movements were performed in three different angles, like shown in figure 3.1 right.

These movements were performed with six conditions resulting in 90 dataset per participant (6 conditions x 5 path x 3 angles).

Conclusion

Implemented Visual Perspective

Chua et al. [?] implemented five different perspectives to teach Tai Chi, compare 3.2. There are three exo-centric perspectives (a-c, further called conditions a-c) and two ego-centric perspectives combined with exocentric (d,e, further called conditions d,e). In **One on One** the student stand next to one teacher, which is closest to a real world training scenario 3.2(a). In **con – Four Teachers** – the student is surrounded by four teachers, with the student in the middle 3.2(b). The **Side by Side** 3.2(c) condition shows four pairs of teacher and student in the same formation as in condition (b). The first ego-centric & combined condition is called **Superimposition 1** 3.2(d). Here the formation still remains, but the student in the middle is surrounded by four more students. On each of the student a red wireframe teacher is superimposed. **Superimposition 2** 3.2(e), is similar to Superimposition one and differs only in the visual representation: the student is now a green, transparent wireframe. Chua et al. use no pure ego-centric perspective, since in the latter two conditions the teacher can be seen also from the exo-centric

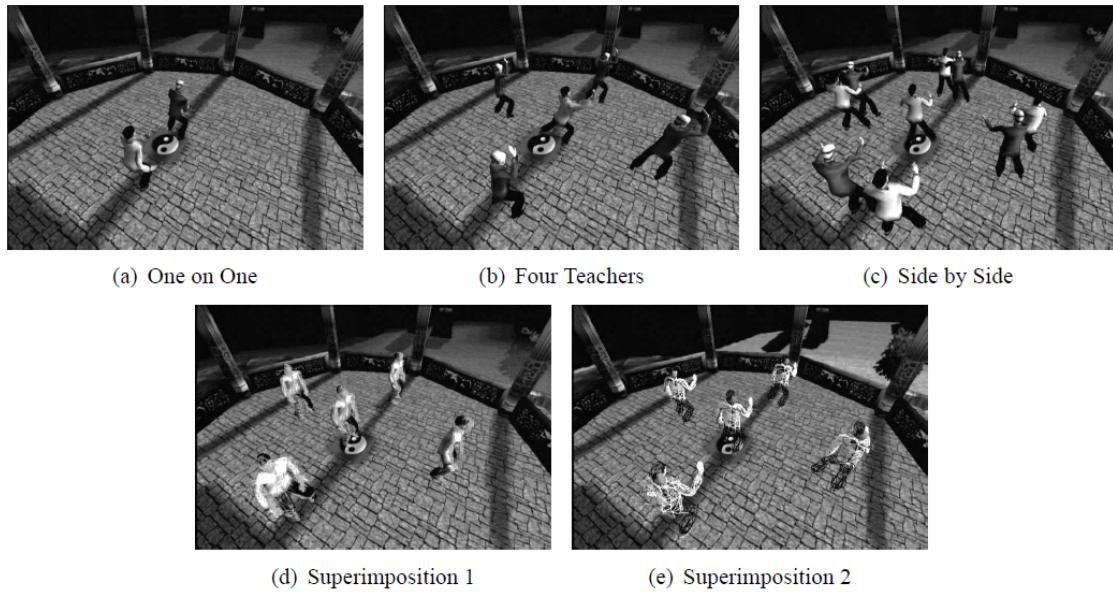


Figure 3.2: Used perspectives at [3]

perspective.

YouMove [1] provides multiple stages for the learning process: (1) Demonstration, (2) Posture Guide, (3) Movement Guide, (4) Mirror, (5) On your own. In (1) the user sees a video of the movement, though the perspective is exo-centric. In stage 2 & 3 the user is superimposed by a skeleton, though the perspective changes to ego-centric. Stage 4 & 5 does not allow a view on the teachers representation. After each stage a feedback is given. This view provides multiple perspectives for the four keyframes: a skeleton of the teacher and student superimposed (ego-centric), a video of the teachers demonstration and a video of the moves of the student (exo-centric) – this can be seen as a combination of both perspectives but only in the aftermath of the learning sequence. Chan et al. [2] facilitate a 3D screen for rendering the avatars. The student is tracked by an optical motion capturing system. In the first phase (demonstration) only the teacher is shown on the screen. In the second phase (practise) teacher and student can be seen simultaneously, standing side by side. Additionally, both avatars are mirrored, so the exo centric viewer can observe the avatars from back and the front at the same time, compare figure 3.3. In the final phase (feedback) the student sees the teacher and the students performance in slow motion side by side. LightGuide [8] achieved an ego-centric guidance without using a HMD. The student stands under a depth camera and a projector. The depth camera tracks the hand of the student. This position is utilised and the guidance visualisation is projected directly on the users hand. During the evaluation of the system, four conditions used this ego-centric

position. The remaining conditions shown exo-centric on a screen or directly on the hand.

Conclusion

Guidance visualisation

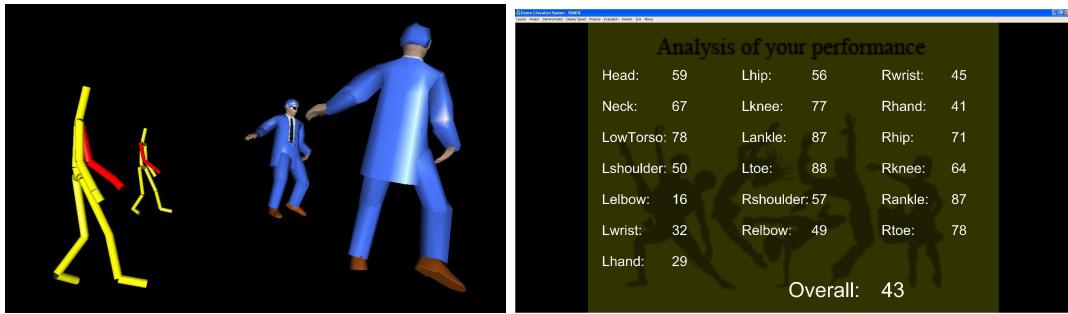
Chua's **Tai Chi training** scenario takes place in pure VR using a HMD. The lessons are conducted in a virtual pergola standing in a nature or park environment. In condition a to c (compare 3.2), the student and the teacher are rendered as non transparent, high realism degree avatars. However, in condition d and e, the visual appearance changes. In condition (d) the student is still rendered normally but the teacher is now represented as a red wireframe. Condition (e) was introduced late in the development of the system based on early subject feedback. The students here are rendered as green wireframe avatars and the teacher as red wireframe avatars.

In terms of guiding technique, the student is presented the pre recorded motion of the teacher. The teacher avatar performs the motion in question and the student tries to mimic this motion. During the mimicking process no feedback is provided to the student. But in condition (d) and (e) the teacher is superimposed on the students avatar. This can be interpreted as feedback since the difference between the students e.g. arm position and the teachers arm position can be seen easily.

YouMove uses Microsoft Kinect ([TODOref](#)) to record motions and to track the student. The recorded instruction as well as the students movements are projected on an AR mirror. The special about this mirror is, that the degree of reflection can be changed by simply adjusting the light in the room. With this, the student can see his own real body as reflection as well as the projection of a beamer simultaneously. If the room is bright, the student sees only his reflection and in a dark room only the projection is visible.

The visualisation of the teacher itself is rendered in a low degree of realism – only a stick figure in yellow. During the feedback screen after each stage, the student is also rendered as a stick figure but in blue.

For teaching, YouMove utilise the above mentioned five training stages. In the first stage (Demonstration), a video of the teacher performing the movements is shown. The student only sees the video and not the reflection of the own body. The second stage (Posture guide) a video and the stick figure of the student are shown, but still not the own reflection. The student is asked to match they postures at specific keyframes, where the demonstration stops. After matching the posture of a keyframe, the demonstration moves on to the next keyframe. In the third stage (Movement Guide) the demonstration no longer stops at the keyframes.

Figure 3.3: **TODO** [2]

Furthermore, the reflection and the projection are visible simultaneously, the student is superimposed by the teachers stick figure. (**TODO** change skeleton to stick figure). Stage four (Mirror) is like the name indicates: the student stands in front of the mirror seeing only his own reflection but additionally audio queues are provided. In the last stage (On you own) even the mirror is removed and the student performs the movements without any guidance.

In terms of feedback, Anderson et al. state "[...] the availability and modality of feedback can greatly impact skill acquisition". To match this, feedback is a part of YouMove. During the second stage (Demonstration), red circles indicate misplaced limbs. The bigger the circle the bigger the misplacement. Additionally, a side view on the scene is faded in. At the end of each stage, a so called "summary feedback" is provided. Here the both stick figures of the teacher and student are superimposed and a video of both can be seen. Eventually a score is provided, which will be explained in the next chapter. The visualisations in Chan et al. VR Dance Trainer choose two different visual representation. The teacher is shown as a high realism degree avatar, while the student is shown as a low realism degree stick figure. The students stick figure serves additionally as feedback regarding the students performance. The body parts are coloured from green (perfect motion match) over yellow (acceptable motion match) to red (poor motion match). In all phases these visual representations stay consistent. In the first phase (demonstration) no feedback is given, in the second phase (practise) the student gets immediate feedback in the colour code described above. The last phase provides a slow motion replay showing with feedback in the same colour code. Additionally, the student can get feedback by the score board see figure 3.3 right. Here all body parts are shown with a numeric indicator (0-100) how good the motion was performed on the specific body part. LightGuide [8] compare six guidance visualisations:

- Follow Spot: Moving light spot with elevation information. System sets pace.

- 3D F-Arrow: 3D arrow indicates the direction to move. System sets pace.
- 3D SG-Arrow: 3D arrow indicates the direction to move. User sets pace.
- 3D Pathlet: a line indicates the direction to move next. Current position indicated by a red spot. System sets pace.
- Video on hand: Instruction video projected on the hand. User sets pace.
- Video on screen: Traditional instruction video .User sets pace.

The main idea behind a guidance visualisation directly on the students body itself, is that the student can concentrate on the body part in question and not share the with a instruction medium. The last condition proofed to be better than the second last, supporting this theses. The student gets instant feedback by comparing the indicators with the hand. A larger offset results in increased indication in real time.

Conclusion

Dependent variables

In the Tai Chi trainer by Chua et al. [3], the students task is to mimic the motion of a pre recorded teacher. While the independent variables are the above mentioned perspectives (conditions a-e) and the motions, the independent variable is the precision of the performed movements. To measure this precision, twelve bones of the students were tracked, namely: upper and lower arms, hands, upper and lower legs, and feet. The bones are hierarchical structured with a parent and child end. After a normalization of the parents position and bone length, an error was calculated:

$$\text{TODO} \quad (3.1)$$

This error can be seen as the euclidean distance between the teacher's and the student's position of a bone. This error was calculated for every frame of the roughly 20 seconds with a rate of 60 fps:

$$\text{TODO} \quad (3.2)$$

The last four out of twelve trials were considered. So the measure of precision is described as:

$$\text{TODO} \quad (3.3)$$

This measure is an implementation of chapter 2 ref **TODO**. Chua et al. found two major faults in this method of precision determination: yaw shift (**TODOadd**

in chapter 2 a definition of dofs) and time shift. To overcome the first, the initial position of the student was taken into account, the latter was fixed with a time frame comparison of 120 frames. Eventually this result was normalized with the difficulty (average number of errors per motion) of the motion task. This will be discussed in detail in the Projects Report. In addition, a post questionnaire was conducted where the students were asked to rate the difficulty of the representations. The results are discussed in the next section.

While Chua et al. calculate the error of the performed movement over all tracked limbs, Anderson et al. take one single joint with the greatest error and – even more constricting – only the keyframe joints (important joints, specified by the teacher) are taken into consideration. The dependent variable is a score between 0 and 10. An offset 15 cm results in a score of 7.5 and no error results in a score of 10. The offset is imply the euclidean distance. This error measurement corresponds to chapter 2 **TODO**. To overcome time shift errors, a window of 0.5 seconds is added. If the teacher specified that timing is important, this window is halved, if precision is set as important, 15cm offset results in a score of 7.5. ”This values are determined by experimentation” Chan et al. evaluate the students performance with the VR Dance Trainer. Therefore they specified 19 body parts (compare 3.3 right) and calculate a score between 0 and 100. The average over the 19 numeric indicators results in the overall performance. Before and after the training session of one move this score was calculated. Additionally, a postcourse survey asked the student specific questions about the system. This survey had the aim to evaluate if the ”[...] system is interesting and able to motivate subjects to learn.” [2] and if the ”[...] the system can provide them an easy way to learn” ibidem. Sodhi et al [8] compare in LightGuides evaluation the five conditions where the instruction is directly on the body of the user with a baseline condition where the instruction is on a screen. To measure the performance of the student, Sodh et al. developed two measures: movement accuracy and movement times. Movement accuracy describes the absolute euclidean distance from the closest point. Movement times is devided in two sections by the reason that half of the conditions were self timed and the half the pace was set by the system. For the self timed conditions the completion time is taken as measure, for system paced conditions the time before or after is taken as measure. Though, independent variable are the conditions, dependet variables are performance.

Conclusion

Results

Chua et al. compare different perspectives (conditions a-e) on the teacher and the student. The independent variables are the above mentioned precision of the students performance. By comparing the results they found condition a (One on One), b (Four Teachers), c (Side by Side) and e (Superimposition 2) aim the same precision. Only e (Superimposition 1) aimed significantly worse than the others. At the same time, the questionnaire indicated that the subjective difficulty of condition d and e (Superimposition 1 and 2) was the highest. "In fact, all of the subjects who tried Superimposition 2 thought it was the most difficult. Interestingly, although subjects considered Superimposition 2 very difficult compared to the other layouts, average error on that layout was not significantly greater than the other non-superimposed layouts." [3]. The Authors argue this result as following:

- simultaneously watching the teacher and performing own movements could interfere each other.

Chua et al. suggests to choose wisely for the task to suit into VR training.

- latency and performance correlates strongly. A lower latency could lead to better performance.

Since Chua et al. [3] was developed in 2003, there is a large improvement in latency nowadays.

- To reduce latency a low polygon count on the high realism degree avatars was used. More polygons could lead to better performance.

The system was run on a Pentium 3 processor. Today's graphic cards and processors are way above this mark, a higher polygon count is easily achievable.

- The field of view was very small.

Today's VR HMDs probably provide a higher field of view (e.g. HTC Vive, 110°).

Anderson et al. compared YouMove with traditional video training, resulting in the independent variables *YouMove* and *Video*. The study was conducted with eight participants in a two factor repeated-measures design. Each participant had one ballet and one abstract task with both conditions. *YouMove* scored significantly better than *Video* by a factor of 2. **TODO**: elaborate a little bit more. Chan et al. investigated three topics with the VR Dance Trainer. First the learning outcome, to proof if the student really got better with the system. Second,

the "Arousing Interest" to investigate whether the system motivates the students to learn. Eventually they compared the system with a traditional Self-learning method. Comparing the baseline score before a training with the VR Dance Trainer with the score after the training session proved a significant better performance. The post survey is interpreted by the authors with "Overall speaking, the subjects enjoy learning dance with our proposed system." ibidem. To compare the system with a traditional learning method, a control group conducted the same study only with the demonstration and no feedback. The baseline scores showed no significant difference between the two groups but the post training scores did. Questionable remains, if a recording of a professional dancer rendered as an high realism degree avatar on a 3D screen can be called a traditional dance learning method. Main finding of Sodhi et al. while evaluating LightGuide is an 85% higher accuracy of ego-centric on body projections compared to a exo-centric video instructions. Per student 90 datasets are generated (6 conditions x 5 path x 3 angles). The performance measure they applied, the conditions scored (best to worst): Follow Spot < 3D F-Arrow < 3d G-Arrow < 3D Pathlet < Video Hand < Video Screen. Especially the relation between video on screen and video hand shows the importance of attention during guidance instruction. The instruction on the body part itself scored better than seeing the instruction on a screen. In terms of *movement times*, both video conditions lead to lower movement times. Sodhi et al. see the reason therefore, that it makes a difference if the student sees the whole path before hand than figuring out the movement as they moved along. Additionally, in an interview, participants of the study stated that self paced guidance are subjectively like more than system paced. For system paced movement speed 30mm per second scored best in a pilot test. Further they suggest to plan regular recovery rests to exclude fatigue effects.

Conclusion

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

(I tried both: split by topic and by paper. This paper is described in one, not split by the aspects above. I will split that soon.) Onebody by Hoang et al. [6] 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

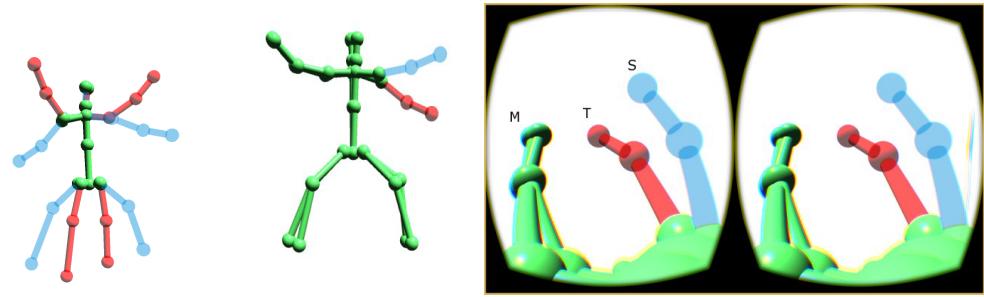


Figure 3.4: Left: student avatar (blue) and teacher avatar (red). Green limbs are matching limbs. Right: students view on the scene. [6]

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 3.4 left. Figure 3.4 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 transferring data, both the teacher and the student are clients in a server-client system. The clients are sending 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 (2.2)) 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 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

	<i>Video</i>	<i>Skype</i>	<i>VR-3PP</i>	<i>Onebody</i>
<i>Synchronous Interaction</i>	No	Yes	Yes	Yes
<i>VR Medium</i>	No	No	Yes	Yes
<i>1st Person View</i>	No	No	No	Yes

Figure 3.5: Training methods and their differences used in the study to evaluate Onebody [6]

(Skype), VR 3rd person perspective. Each of the systems differs to Onebody in terms of synchronous interaction, VR medium and perspective see figure 3.5. 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.

3.5 conclusion

Here will stand an overall conclusion

Study Task	Implemented Perspective	Guidance Visualisation	Dependent Variables	Results

Table 3.1: Table of the conclusion

4 Proposed Study Setting

4.1 Hypotheses

4.2 Variables

Independent Variables

- ego-centric
- exo-centric
- combined
 - variation of combined: sequential - parallel

further variations that could be interesting:

- body parts
- sitting/standing
- visual representation
- degree of realism of avatar
- guidance techniques: stopping at keyframes vs. fluent instructions
- audio queues
- feedback
- fixed position of teacher and avatar, vs walking around

Dependent variables

Score. Combination of objective and subjective measurements.

- precision
- subjective opinion of participant
- stress level (EKG, HRV)
- cognitive load
- retention
- reaction time

4.3 Task

- Tai Chi form, split in subtasks
- Dance moves from single dance

variations:

- difficulty
- complexity
- abstract vs. real world
- operate a control panel
- escape the room
- game

5 Outlook

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

5.1 Timetable

Bibliography

- [1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. YouMove : Enhancing Movement Training with an Augmented Reality Mirror. pages 311–320, 2013.
- [2] Jacky C P Chan, Howard Leung, Jeff K T Tang, and Taku Komura. A Virtual Reality Dance Training System Using Motion Capture Technology. 4(2):187–195, 2011.
- [3] Philo Tan Chua and Russ Schaaf. Training for Physical Tasks in Virtual Environments : Tai Chi.
- [4] 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.
- [5] Joseph J LaViola. 3D user interfaces: theory and practice , 2017.
- [6] Martin Reinoso. Onebody : Remote Posture Guidance System using First Person View in Virtual Environment. 2016.
- [7] Richard Schmidt and Tim Lee. *Motor Control and Learning: A Behavioral Approach*. 2011.
- [8] Rajinder Sodhi. LightGuide : Projected Visualizations for Hand Movement Guidance. pages 179–188, 2012.
- [9] Wenbi Wang and Paul Milgram. Dynamic Viewpoint Tethering for Navigation in Large-scale Virtual Environments. *Human Factors*, pages 1862–1866, 2001.

List of Figures

1.1	Usecases by EONReality on their AVR plattform. https://www.eonreality.com/use-cases/ (accessed 30.07.2019)	2
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=1QFMP accessed (30.07.2019)	3
2.1	Centricity continuum by Wang and Milgram 2001 [9]	6
2.2	Continuum of movements buch [7]	8
2.3	Continuum of skills [7] TODO seite	9
2.4	Laban notation. Generated through automatic movement interpreta- tion by Choensawat [?]	10
2.5	Mixed reality continuum by Milgram et al. [?]	12
3.1	movements by lightguide [8]	20
3.2	Used perspectives at [3]	21
3.3	TODO [2]	23
3.4	Left: student avatar (blue) and teacher avatar (red). Green limbs are matching limbs. Right: students view on the scene. [6]	28
3.5	Training methods and their differences used in the study to evaluate Onebody [6]	29

List of Tables

3.1	Table of the conclusion	29
-----	-----------------------------------	----