Multiple-pose Augmented Learning System

Manuel Cardenas
Advisors: José Vega, Dr. Thomas Bohné
Chair of Decentralized Systems Engineering
https://dse.in.tum.de/

15.06.2022 – 17.10.2022

Much content, little time

- Motivation
- State of the art
- Research gap
- Solution overview
- Implementation
- User study
- Conclusion

There is no perfect system to learn a new pose

Two options are widely available for a student:

Teacher:

- The student receives guidance

But

- Availability of the teacher is limited
- A (high) price can be involved
- Teacher cannot see everything in a group setting

Online Video:

- Wide array of content
- Always available

But

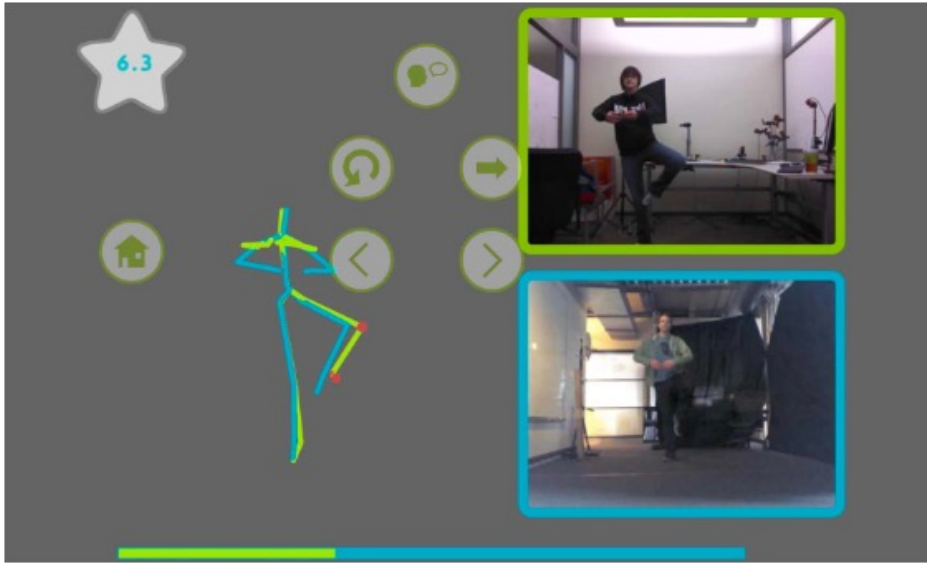
- Student receives no guidance
- Student can only correct via feeling

A new method has been under development

- Technology is being used to allow both the **availability** of videos and the **guidance** of a teacher

YouMove [1]

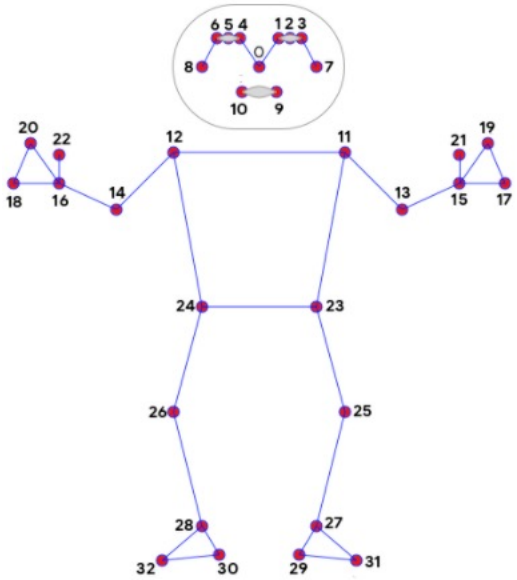
- Allows users to record and learn physical movement sequences.
- Users receive feedback on their movements that changes as the user progresses
- Pose recognition via *Kinect* [2]
- Feedback via *augmented mirror*



YouMove's terminal feedback screen. Source: [1]

New pose recognition system developed by Google

- *Google's Blaze Pose GHUM* [3]*:* With only one (or a series of) image(s) of a person doing a pose, it can detect joints



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

All joints extracted by Google Blaze Pose GHUM. Source: [10]

Result of the framework. Source: [10]

A pose teaching system that uses simple hardware

A software is needed that

- Simple hardware: just a camera, a screen and internet connection (what a normal laptop has)
- Teaches a new pose following multimedia teaching principles
- Allows change of pose with ease: not fixed to a single use case
- Is easy to use

This thesis aimed to review **the feasibility** of such a system

To do so, a prototype was developed that followed the above requirements

First, the final pose to be learned is displayed

Gustavo **guides** the trainee.

A personalized agent is **more effective** than just objective instruction

Goal orientation: Final pose is presented. This improves learning results by building expectations

Welcome screen recording from MPALS

The final pose is divided into sub-poses

Dividing the final pose into **small steps** makes learning easier

Sub poses help **visualize progress**

Pose selection screen recording from MPALS

Before the training, the learner can get ready

Visualizing the pose before moving helps the learner prepare

Key joints give emphasis to the training, which improve results

Training screen recording (during pre training) from MPALS

During the training, the learner receives feedback

The learner's avatar moves as the learner moves

The color coding gives **feedback** to the learner during movement

Training screen recording (during training) from MPALS

The learner can then analyse the movement

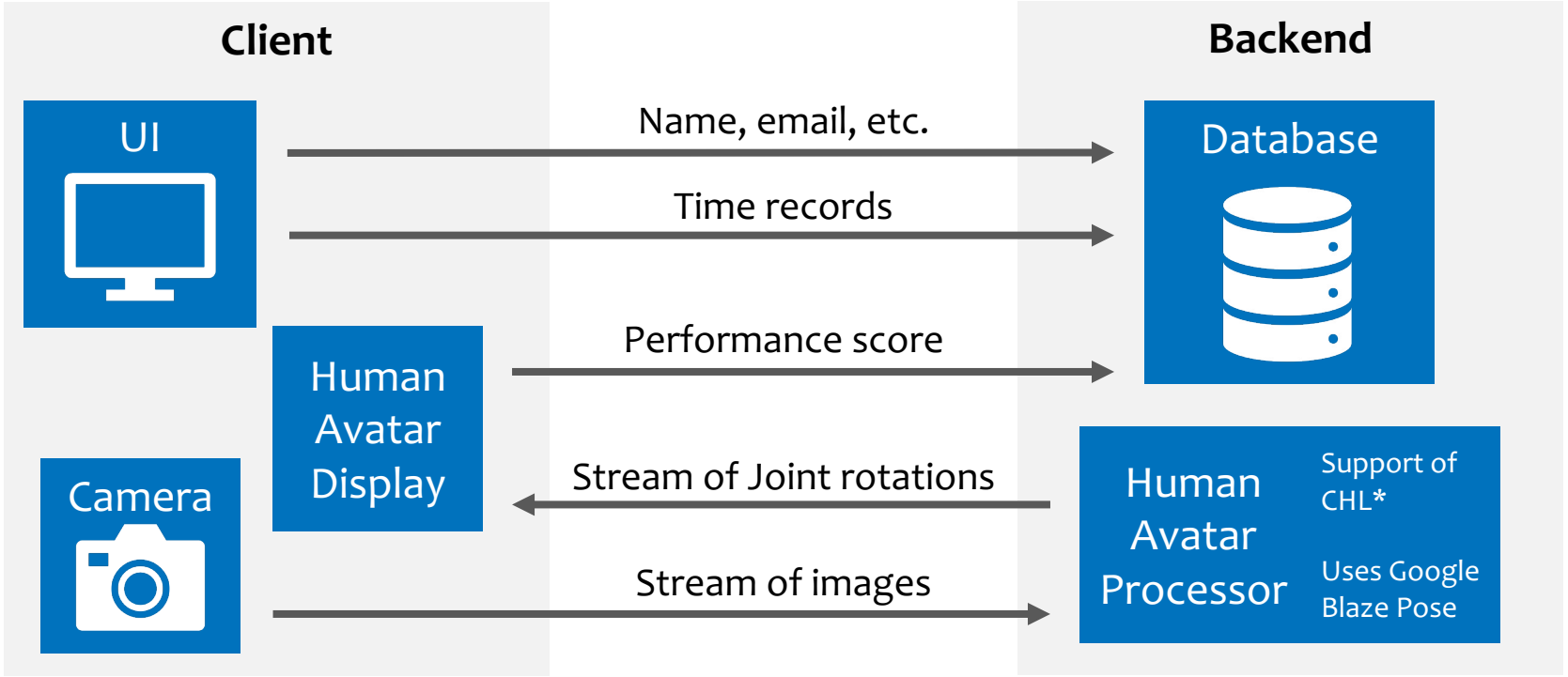
Terminal feedback screen recording from MPALS

After doing one pose, the next poses are unlocked

The learner **decides on the pace:** continue to next pose or try the same one again

Pose selection screen recording from MPALS

The software is divided into client and backend



*Cyber Human Lab – part of the Institute for Manufacturing of the University Of Cambridge

16 participants tested MPALS, 8 were observed

- The **Time spent** on the following activities was tracked:
  - Pose selection: selecting the pose
  - Pre teaching: after the pose is selected, before the training starts
  - Training: the participant moves
  - Analysis: the participant reviews the movement

- The **performance** of the participants doing the pose was tracked and stored every 0.2 seconds
  - Performance score is computed from the **joint rotation difference**

- The participants answered two questionnaires afterwards
  - Raw TLX: Analyses the **effort** of using the tool [4]
  - UE-SF: Analyzes how **engaged** the users were using the tool [5]

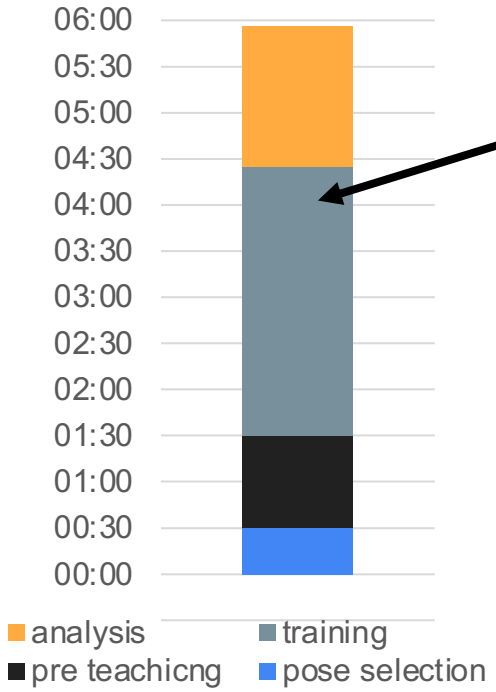
Much was learned from simply observing learners

- The dialogs are read very quickly: participants usually just want to get to the activity as quickly as possible
  - *Possible improvement*: tutorial-like explanation. Gustavo asks the learner to do something, and the learner can only continue after action is successfully performed.

- Key joint functionality is not used enough
  - One one out of the eight participants used it
  - Even after Gustavo points the user that It can help with focus
  - *Possible improvement*: tutorial-like explanation

- After finishing the pose, participants were somewhat lost
  - Even if Gustavo explains how to continue
  - *Possible improvement*: animation after successfully finishing a pose

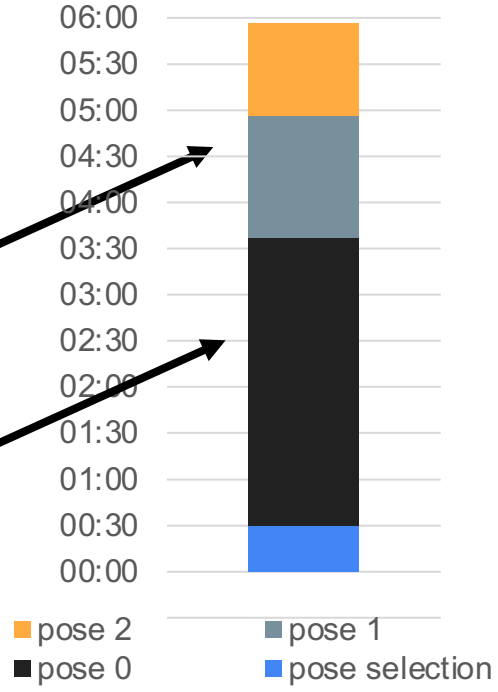
On average participants spent 5min 56s on MPALS



The time distribution per activity follows expectations

As the trainees get used to MPALS, they need less time

Many recommendations by Gustavo increase time spent on the first pose



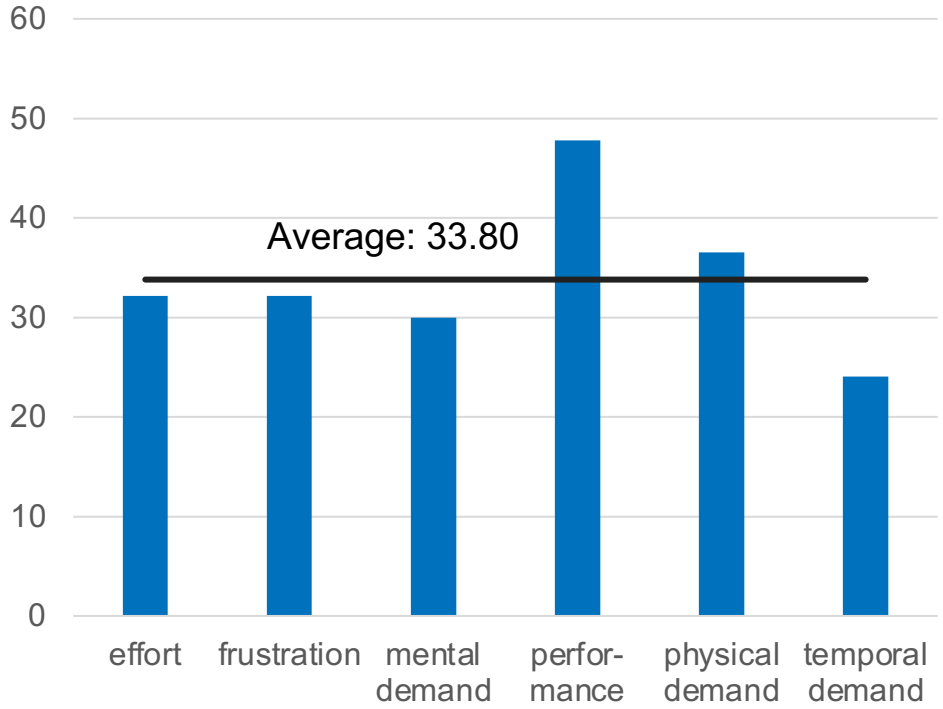
We use the score to measure the learners' performance

- Successful attempt: **the learner reached a score larger than 0.95 for at least 20 frames**
- Out of 80 training attempts, 73 were successful
- The 7 unsuccessful attempts, were not unique: the learner tried the same pose again before or after the unsuccessful attempt.

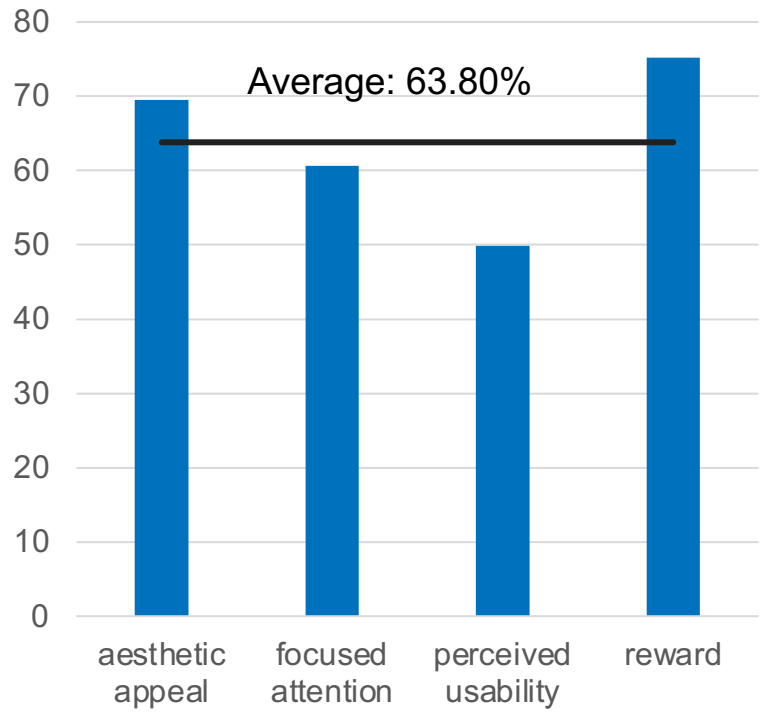
Conclusion:

- The feedback was successful in pointing the learner to an attempt that did not work
  - Important note: doing a successful training attempt is not a condition to continue. But users choose to continue only after the attempt is successful

Raw-TLX: The tool was perceived as easy to use



UE-SF: The engagement was high, but could be higher



We can answer the research question positively

Is it feasible to develop a pose teaching system that uses **simple hardware** (only a camera, screen, and an internet connection), that **follows teaching principles**, and that **users find engaging and easy to use**? → **YES**

- MPALS is a first step towards a **generalized pose teaching software**
- Big user study groups can be organized thanks to MPALS: the software can be sent over the internet, and participants' data can be gathered with ease
- The technology behind MPALS can be used for multiple use cases
  - A new service just must connect to the backend and send images, to receive a stream of joint rotations that can be used for therapy, ergonomics, etc.
  - Projects such as YouMove [1] do not need to use Kinects and expensive hardware

Backup

General teaching principles

- **Cognitive load:** Amount of working memory resources being used [6].
  - The learner's cognitive load **should be spent on the task itself, and not on the tool**

- **Guided activity principle:** learning with a personalized learning agent works better than with objective instructions (example: *Duolingo* and *Duo*) [7]

- **Pacing principle:** Allow students to process **smaller pieces of information** in working memory by **letting them decide the pace** of their own learning [7]

- **Pre-training principle:** Students learn better when there is guidance before the training [7]

Pre-teaching principles

How the pretraining should be designed depends on the task, but pre training principles exist to design it well [8]:

- **Mastery goal orientation:** Goal that is focused on increasing the competence of a task works better than orienting with the goal of reaching **a specific metric** [8]

- **Attentional advice:** Administer information about the process or strategy needed to learn most. Proven to improve learners' performance [8]
  - Gustavo gives this type of advice

Score calculation

Key question for the study: did the user manage to perform the pose right?

→ A metric is defined to quantify the error of the learner

- The rotations of each joint of the learner's avatar is compared with the rotations of each joint of the avatar doing the model pose
- 3D rotations are represented in Unity with Quaternions
- We compute the distance of two quaternions using their dot product [9]:
  - $Distance = Dot(q_1, q_2)^2$
- The rotation distance of all joints is averaged, weighing the key joints more
  - Key extremities = 75% of final score
  - Remaining extremities = 25% of final score

Raw-TLX survey in detail: rate ease-of-use of MPALS

- Participants rank each subscale between with a number between 5 and 100 [4]

Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
Mental Demand	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Performance	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Physical Demand	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

UE-SF survey: how engaging is MPALS?

- Participants rank each question in a seven-point scale [5]

Aesthetic Appeal	• The application was attractive. • The application was aesthetically appealing. • The application appealed to my senses.
Focused Attention	• I lost myself in this experience. • The time I spent using the application just slipped away. • I was absorbed in this experience.
Perceived Usability	• I felt frustrated while using the application • I found the application confusing to use. • Using the application was demanding
Reward	• Using the application was worthwhile. • My experience was rewarding. • I felt interested in this experience.

Bibliography

Bibliography

[1]: Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. "YouMove: Enhancing Movement Training with an Augmented Reality Mirror". In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. UIST'13. St. Andrews, Scotland, United Kingdom: Association for Computing Machinery, 2013, pp. 311–320. isbn: 9781450322683. doi: 10.1145/2501988.2502045.

[2]: Zhengyou Zhang. "Microsoft Kinect Sensor and Its Effect". In: *IEEE MultiMedia* 19.2 (2012), pp. 4–10. doi: 10.1109/MMUL.2012.24.

[3]: Ivan Grishchenko, Valentin Bazarevsky, Andrei Zanfir, Eduard Gabriel Bazavan, Mihai Zanfir, Richard Yee, Karthik Raveendran, Matsvei Zhdanovich, Matthias Grundmann, and Cristian Sminchisescu. *BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation*. 2022. doi: 10.48550/ARXIV.2206.11678.

[4]: Sandra G. Hart. "Nasa-Task Load Index (NASA-TLX); 20 Years Later". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50.9 (Oct. 2006), pp. 904–908. issn: 1071-1813. doi: 10.1177/154193120605000909.

[5]: Heather L. O'Brien, Paul Cairns, and Mark Hall. "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form". In: *International Journal of Human-Computer Studies* 112 (Apr. 2018), pp. 28–39. issn: 1071-5819. doi: 10.1016/j.ijhcs.2018.01.004.

Bibliography

[6]: Jeroen J. G. van Merriënboer and John Sweller. “Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions”. In: *Educational Psychology Review* 17.2 (June 2005), pp. 147–177. issn: 1573-336X. doi: 10.1007/s10648-005-3951- 0.

[7]: Roxana Moreno and Richard Mayer. “Interactive Multimodal Learning Environments”. In: *Educational Psychology Review* 19.3 (June 2007), pp. 309–326. issn: 1573-336X. doi: 10.1007/s10648-007-9047-2.

[8]: Jessica Mesmer-Magnus and Chockalingam Viswesvaran. “The role of pre-training interventions in learning: A meta-analysis and integrative review”. In: *Human Resource Management Review* 20.4 (Dec. 2010), pp. 261–282. issn: 1053-4822. doi: 10.1016/j.hrmr. 2010.05.001.

[9]: Du Q. Huynh. “Metrics for 3D Rotations: Comparison and Analysis”. In: *Journal of Mathematical Imaging and Vision* 35.2 (June 2009), pp. 155–164. issn: 1573-7683. doi: 10.1007/s10851-009-0161-2.

[10]: Google LLC. *Mediapipe Pose Documentation*. Oct. 11, 2022. url: https://google.github. io/mediapipe/solutions/pose.html.