Effects of Correctness and Suggestive Feedback on Learning with an Autonomous Virtual Trainer

Xiumin Shang
Computer Graphics Lab
University of California, Merced
xshang@ucmerced.edu

Marcelo Kallmann
Computer Graphics Lab
University of California, Merced
mkallmann@ucmerced.edu

Ahmed Sabbir Arif
Human-Computer Interaction Group
University of California, Merced
asarif@ucmerced.edu

ABSTRACT

In this paper we investigate interaction strategies for autonomous virtual trainers. Fourteen participants were immersed in our VR system to learn relative areas of countries by sorting virtual cubes. We evaluated two different feedback strategies used by the virtual trainer assisting participants. One provided Correctness Feedback at the end of each task, while the other provided Suggestive Feedback during the task. Correctness feedback was the most effective given that it received higher preference and led to shorter task completion time with equivalent performance outcomes.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**.

KEYWORDS

Virtual agents, virtual reality, feedback strategies.

1 INTRODUCTION

Virtual demonstrators, teammates, or tutors represent a powerful approach to address a number of applications [4], and the interaction strategies used by these virtual characters are known to impact user experience and performance. This paper investigates the use of Correctness Feedback (CF) and
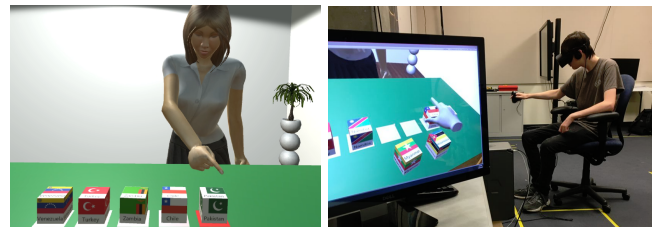
Figure 1: An autonomous virtual trainer assisting a user to learn a sorting task in our VR system (left). A user is interacting with the system using voice commands and manipulating objects (i.e., cubes) using hand controllers (right).

Suggestive Feedback (SF) strategies with virtual trainers in a VR-based application. While CF follows the strategy of informing users about their performance only at the end of a task [1], SF provides information during the task performance, by continuously providing clues or suggestions about the activity [2].

In this work, we investigate CF and SF strategies in the context of a virtual trainer providing feedback to users during a cube-sorting task (see Figure 1). Our task scenario involves a relatively short activity, each lasting from 10 to 20 minutes. Therefore, our system primarily investigates how short-term and procedural memory [3] work together to retain the information delivered by the system. Our results show that the CF strategy was the most effective given that it was preferred by the participants and it led to less time to complete the tasks while achieving similar performance levels.

2 METHOD

After running a preliminary study to validate our design, our main study evaluated the task of sorting nine countries based on their area, using cubes representing the countries. Each sorting task was divided in four stages: first, sorting three countries, then two additional countries were added for each subsequent stage. Each country was represented by a $7 \times 7 \times 7$cms cube displaying the country's flag and name.

The CF strategy was executed as follows. At each stage, after the participant manipulated and sorted the cubes on the table, the agent would correct the cubes in wrong positions

In Proceedings of IUI'19. This is the manuscript of the authors.

Shang et al.

one by one, from left to right. The participant would then observe the correct solution for as long as needed before requesting the next stage to start.

With the SF strategy, every time the participant placed all cubes on the given positions, the agent would check and point to the cubes in the wrong order, from left to right, while saying: "Attention, those two countries are in the wrong order". The procedure repeated until all cubes were sorted in order, forcing the participant to fix all mistakes.

Our system was implemented with Unity and the Oculus Rift HMD. The Windows Speech Recognition module was used to recognize voice commands given by the user; and Unity plugins SALSA with RandomEyes and FinalIK were used to solve lip syncing and Inverse Kinematics, in order to implement the virtual trainer's gaze behavior, object manipulation, and pointing actions.

Study Design A within-subjects design was used with 14 volunteers from the university community. Their age ranged from 17 to 25 years, with an average of 20.14 years (SD = 1.99). Each participant experienced both feedback strategies with two different material sets. Each material set included a group of nine different country pictures displaying the country's flag and name. The order of feedback strategies and material sets was counterbalanced. The overall within-subjects design was: 14 participants × 2 conditions (with different learning material sets) × 4 task stages × {3, 5, 7, 9} country cubes. Feedback was the independent variable, while learning outcome was the dependent variable. Feedback consists of CF and SF. Learning outcome was quantified by a sorting score defined as the combination of all cube pairs in correct order divided by all possible combinations of pairs (in total 36). Sorting scores were measured as pre-test and post-test questionnaires.

3 RESULTS

Quantitative We used a repeated-measures ANOVA with an alpha of .05 for analysis. ANOVA failed to identify a significant effect of feedback on learning outcome ($F_{1,13}$ = 1.66, p = .22), but identified a significant impact of pre-test and post-test questionnaires test order on user's learning outcome, the result is: $F_{1,13}$ = 49.44, p = .000009. In SF, the average suggestion times increased as the stage number increased, as shown in Figure 2. The average total time needed for CF was 339 seconds, while for SF was 559 seconds, which is 65% higher than the time needed for CF. This shows that, although there was no performance difference between CF and SF, there was a substantial difference in task execution time, and both feedback strategies can improve learning outcome.

Qualitative A Wilcoxon Signed-Rank test was used to analyze the post-questionnaire data with respect to the preferred feedback strategy. It failed to identify a significant impact of feedback strategy on user preference. However, comparing
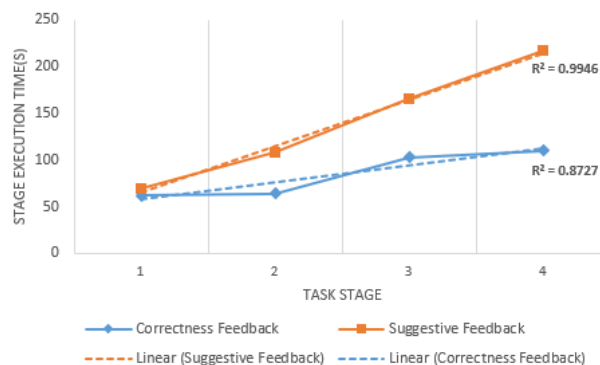

Figure 2: Average task execution time (in seconds) at each stage for CF and SF.

the median values of CF and SF, more participants cited CF as preferable and as more effective. In addition, more participants responded that they would prefer using CF in our system instead of learning from a real human. Participants also rated interaction with speech commands unfavorably and rated HMD visual resolution and quality as not interfering or distracting from the task execution. The Simulator Sickness Questionnaire (SSQ) [5] score was 9.82 (the maximum possible SSQ is 300), which indicates that our system was suitable for conducting the user study.

4 CONCLUSIONS

While both strategies were effective, CF was found to be most preferred and more efficient time-wise. A linear relationship between SF execution time per stage and stage completion times was found. In general, interactions with virtual trainers were rated as comparable in preference to hypothetically performing the same task with human interactions.

REFERENCES

[1] Ahmed Sabbir Arif, Cristina Sylla, and Ali Mazalek. 2017. Effects of different types of correctness feedback on children's performance with a mobile math app. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2844–2849.

[2] Kristen Pilner Blair and Kristen Pilner. 2013. Learning in critter corral. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*. ACM Press, New York, New York, USA, 372–375.

[3] Anna-Katharine Brem, Kathy Ran, and Alvaro Pascual-Leone. 2013. Learning and memory. *Handbook of clinical neurology* 116 (2013), 693–737.

[4] W. Lewis Johnson and Jeff Rickel. 1997. Steve: an animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin* 8, 1-4 (12 1997), 16–21.

[5] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.