

Assessment of Academic Performance in Free to Select Anime Avatar Educational Video for Ray Tracing Programming Class

Rex Hsieh¹, Kosei Yamamura¹, Satoshi Cho¹, Hisashi Sato¹

¹Kanagawa Institute of Technology

Abstract.

This project is a 11-week anime avatar based eLearning programming program that took place from September 2019 to February 2020. It serves as the 2nd step of vTuber based eLearning programming program where researchers used the survey results and academic performances from the spring semester's eLearning program as basis and made refinement upon it in an effort to provide better avatar based eLearning contents to students. In the previous program, students are organised into groups and assigned to 1 or 2 of 6 videos composed of 1 of 3 visuals: male styled avatar (A), female styled avatar (B), or original lecturer avatar (O) and 1 of 2 audios: original audio (O) or transformed audio (T). In this semester, transformed voice (T) is dropped after finding out they offered no benefits to students' motivation or grade performance. Furthermore, students are now free to choose from the 3 avatars offered instead of being assigned to specific videos for the entire semester. Lastly, weekly videos now encompass multiple subjects with clearly labelled titles to allow students watch the content most beneficial to them. The research team believed the increase in avatar type and subject choice will make students more motivated when viewing the contents. Currently this program is in its 7th week and at the moment, the research team has found that students enjoyed the audios much more, as compared with the previous semester. Students are also observed to be selecting videos according to their needs.

Keywords: eLearning, Avatar, User Experience

1. Introduction

With the rise in popularity of YouTube, a variety of YouTube channels and ways of broadcasting contents were developed. In Japan, where the world renowned anime industry originated from, a trend of using anime characters to broadcast contents known as vTuber was developed. As vTubers, broadcasters used 3D anime styled characters animated by motion tracking technology to show full body or upper body animated characters to act as narrators. In 2018 it was estimated the number of vTubers is in the 4,000s, which is a significant increase from the 2,000s in 2017. Currently the most popular vTuber star: Kizuna AI has over 2 million

followers and has generated a variety of contents ranging from tutorials to hosting talk shows with other internet stars.

This paper reports the findings of the 2nd stage of a vTuber based eLearning programming program where some 200 freshman students currently taking the POV-Ray Introduction class are being exposed to videos featuring male styled avatar (A), female styled avatar (B), or real lecturer's avatar (R). Students are free to choose what video they want to view in order to assess when given the freedom to choose the avatar they want, whether anime styled avatar or real lecturer's video will garner more popularity. Furthermore, videos featuring voice transform audio (T) were not used in this experiment after finding out their impacts on student's academic performance is neutral at best. By giving students more freedom to choose their educational content, the team also hoped to evaluate the effect this new method had on student motivation and academic performance.

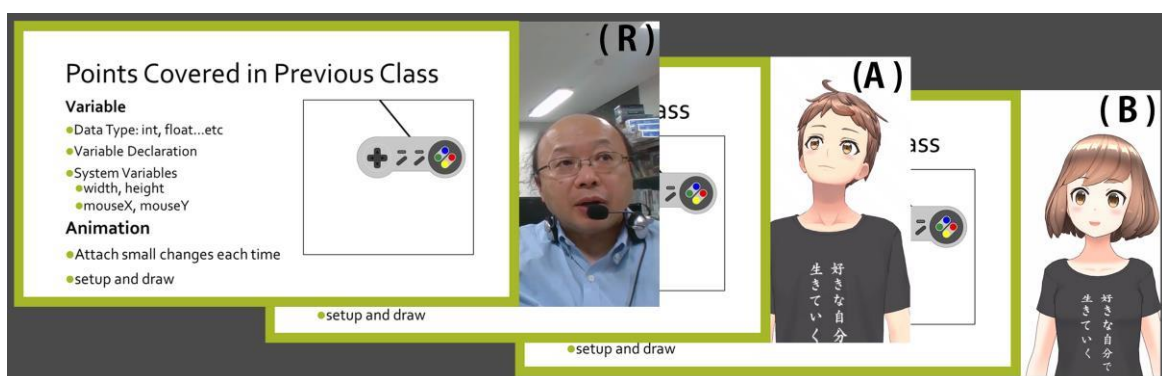
2. Past Researches

Numerous researches have been done regarding how to best utilize eLearning materials and course structures. (Zou et al., 2017) [1] created and evaluated what was called the Mulsemmedia (Multi-Sensational Media) aimed at providing supplementary video material to aid the education of STEM course works. Several other researchers such as (Giannoukos et al., 2008) [2], (Kim et al., 2009) [3], and (He & Ueno, 2012) [4] created complete online networks and attempted to promote more collaboration and sharing of contents. Studies concerning the effectiveness of E-Learning as in contrast to traditional learning have also been conducted with the majority showing a positive impact of E-Learning on promoting engagement of students. Despite of the richness in eLearning research subjects, most of the previous works merely combines traditional media together in a unified platform and do not attempt to refine the presentation of these media. This paper seeks to enhance the visual aspects of eLearning programs by finding the most preferred lecturer.

3. 2019 Spring semester vTuber based eLearning Program

The predecessor of this project is the 1st step of vTuber based eLearning programming program conducted by (Rex et al., 2019) [5] from April 2019 to September 2019 during a Processing class attended by 186 students held in Kanagawa Institute of Technology. For that experiment, there were a total of 6 types of videos composed of 3 visuals: male styled avatar (A), female styled avatar (B), real lecturer avatar (R), and 2 audios: original audio (O) and transformed audio (T). The vTuber characters A and B were generated using an application called REALITY that was created by Wright Flyer Live Entertainment, a subsidiary company of GREE. The virtual characters are given the most basic clothing and hairstyle without any decoration in order to eliminate outside factors aside from gender of avatars that can impact student learning. The transformed audio T was created using the Roland VT-4 vocoder which allows for the adjustment of a speaker's pitch, formant, balance, and reverb.

Fig 1. Real Avatar (R), Male Styled Avatar (A), and Female Styled Avatar (B) of Previous Research



3.1 Student Division

The students were divided into 15 groups labelled from A to O with 10 groups (A – I) assigned to 2 videos and alternated between watching 1 type of video between weeks while the last 6 groups (J – O) were assigned to 1 video and acted as the control group. The table below shows the videos assigned to each group and their function in the experiment.

Fig 2. Group video assignment and purpose

Groups	Odd Weeks (1, 3, 5, 7, 9, 11)	Even Weeks (2, 4, 6, 8, 10)	Purpose
A	RO	RT	Measure Audio with same visual
B	AO	AT	
C	BT	BO	
D	AO	BO	Measure Avatar Gender change
E	AT	BT	
F	RO	AO	Real Visual vs "REALITY" Generated Visual Avatars
G	BO	RO	
H	RT	AT	
I	RT	BT	Control Group: Only Watched 1 Video
J	RO	RO	
K	AO	AO	
L	BO	BO	
M	RT	RT	
N	AT	AT	
O	BT	BT	

3.2 Student Subjective Impressions

The experiment uses google form survey to gather subjective student impression regarding the videos' audio and visual qualities. In the questions, students were asked to pick from a Likert Scale of 1 to 4 with 1 being disagree and 4 being agree. The scale was made purposefully even number to get rid of a neutral option. It was discovered that while the visuals are rated 3 or above for all videos, the audio differ in quality greatly between O and T groups with O videos

performed much better compare to T. There is also much greater mismatch between visual and audio for T videos with the lecturer’s real avatar and original audio RO having the least amount of mismatch while the female styled avatar and transformed audio BT having the most amount of mismatch. The overall quality of all videos are at least a 3 or above save for the exception of AT. Surprisingly, BT, despite of its poor audio quality and highest amount of mismatch received the 4th highest score when it comes to overall impression. Below are the graphs detailing student impression of videos.

Fig 3. Visual is Good Likert Scale Graph

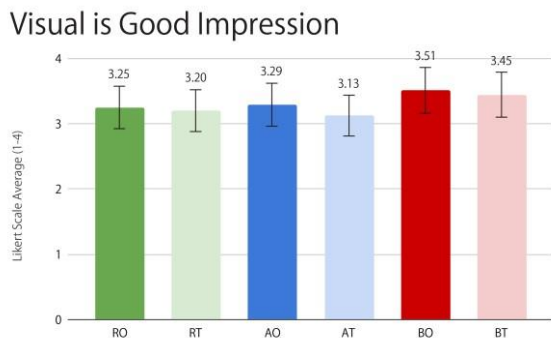


Fig 4. Audio is Good Likert Scale graph

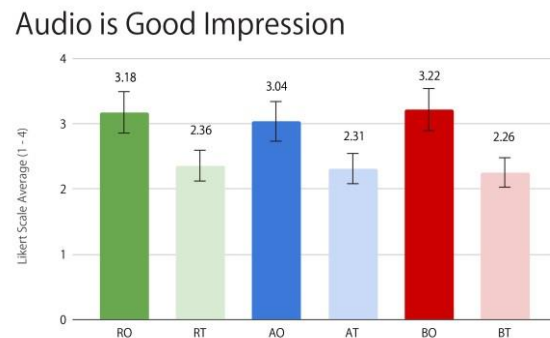


Fig 5. Visual and Audio Mismatch Likert Scale Graph

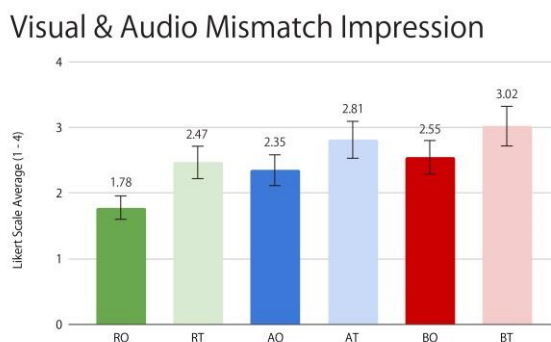
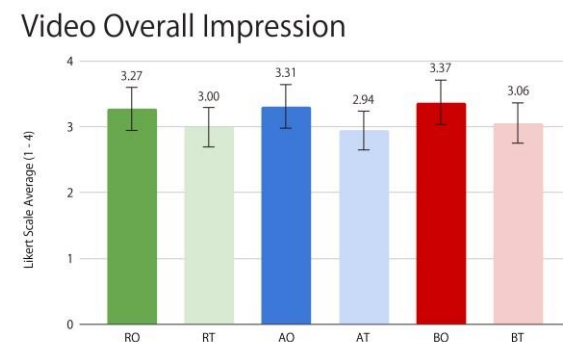


Fig 6. Overall Quality is Good Likert Scale Graph

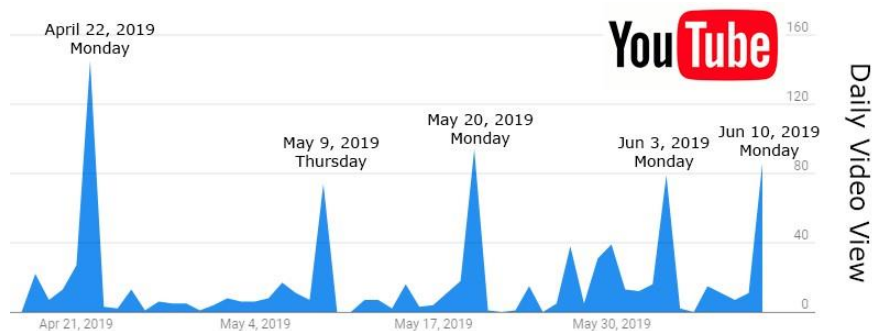


3.3 YouTube Objective Analysis

In addition to Google Form surveys, the research team also used YouTube Studio Analytics to conduct objective data analysis regarding watch duration and watch date of the videos. It was found that the majority of the views took place on Monday and Thursday, which were either the day before class lecture on Tuesday or discussion class on Friday. The research team also received data regarding students’ motivation in regards to how much time

they spent watching different types of video. Overall videos featuring original audio received longer viewing time in comparison to videos featuring transformed audio by an average of 16.87 seconds.

Fig 7. YouTube Daily View Count Graph



3.4 Student Academic Performance

A total of 3 assessments took place during the Processing course with a mid-term on July 2nd, final exam at the end of July, and final assignment with a due date at the beginning of August. As we can see from the graphs, the control groups (J – O) all performed consistently well throughout all assessments with RO constantly placed in the upper half of the class. The groups that were assigned to 2 videos alternate between original voice and transformed voice (A – C) did much worse in comparison with B and C group being the bottom 2 performers in 3 out of 4 assessments. While transformed audio alone does not contribute to worse grade performance, it is the switching of original and transformed audio between weeks that contributes to poor grades. Other aspects of the video, such as switching avatars in between weeks does not impact the academic performances of students although it is still apparent that maintaining the same visual and audio yields the best outcome.

Fig 8. Mid-Term Grades

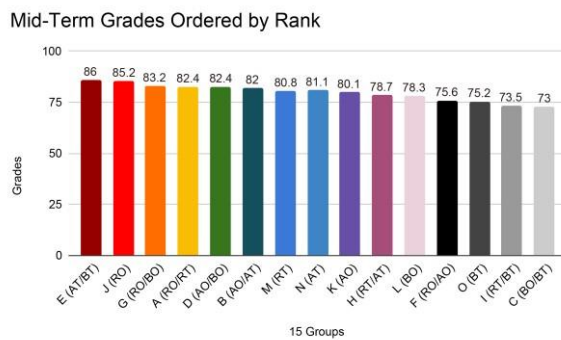


Fig 10. Final Assignment Grades

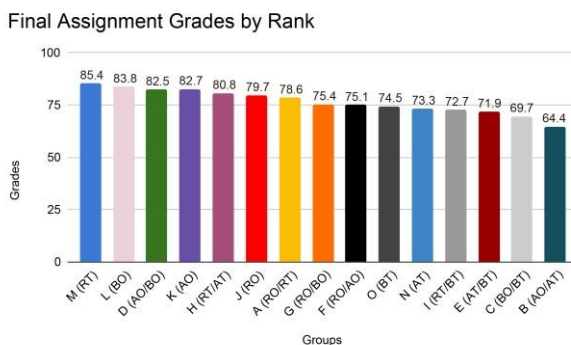


Fig 9. Final Exam Grades

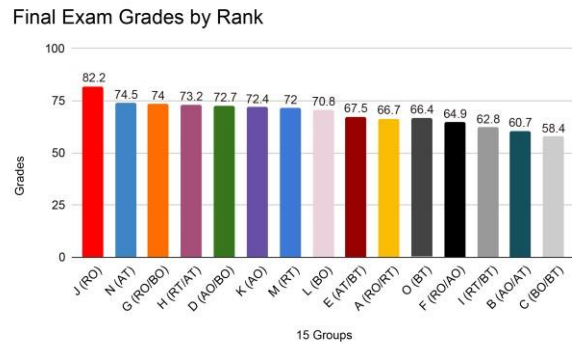
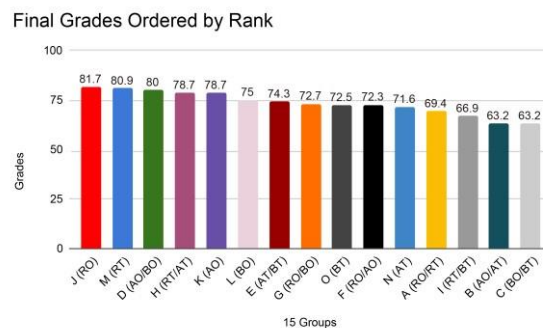


Fig 11. Final Grades

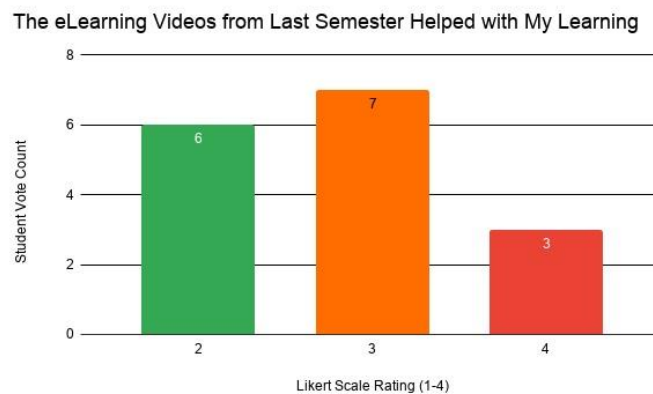


4. Post Spring 2019 Class Impression

After the Spring 2019 Processing class ended, a post class survey was distributed in

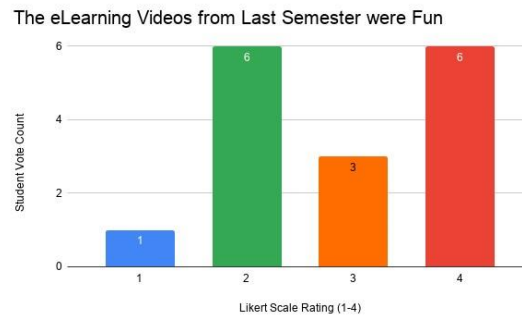
September to inquire about the students’ impression regarding the eLearning videos used during class and which areas do they think should be improved. 16 students have since answered the survey and their feedback compiled. According to the students’ feedback, when asked the question: “The eLearning Videos from Last Semester Helped with My Learning “, 62.5% of students gave either a 3 or a 4 with 18.8% voted 4 meaning they strongly agree the videos helped and 43.8% voted 3 weakly agreeing the videos helped. Of the 37.5% who disagree with the notion that the videos helped, all gave the question a 2 rating meaning their disagreement is leaning towards the weaker side.

Fig 12. Student Notion on how eLearning Videos Helped



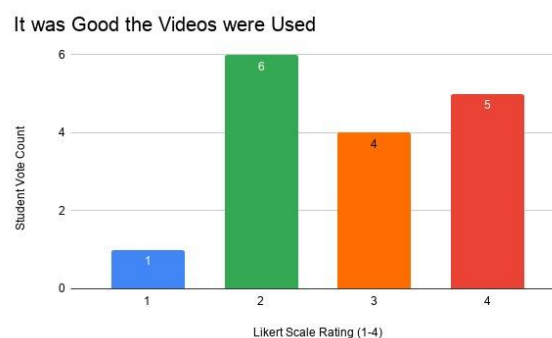
When it comes to the question: “The eLearning Videos from Last Semester were Fun”, the results is much less positive with 6 students voted 4, 3 voted 3, 6 voted 2, and 1 student voted 1 in the Likert Scale.

Fig 13. Students Notion on whether the eLearning Videos were Fun



The question of “It was Good the Videos were Used” would also generate mixed reaction with 5 voting 4, 3 voting 3, 6 voting 2, and 1 voting 1, creating a picture of 56.3% students believe it was good the videos were used and 43.7% disagreeing with the notion that videos being used in class is good. Despite of the seemingly equal split of opinion, 6 out of 7 who disagreed voted 2 standing for weak disagree so the overall opinion is still leaning towards favouring the videos.

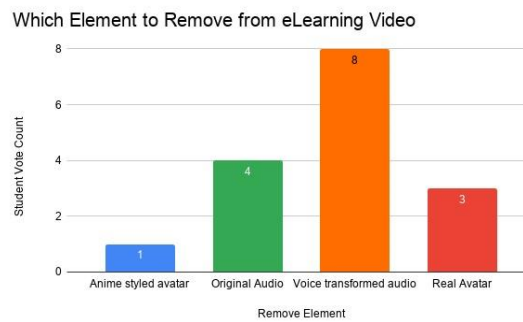
Fig 14. Students Notion on whether eLearning Video Inclusion is Good



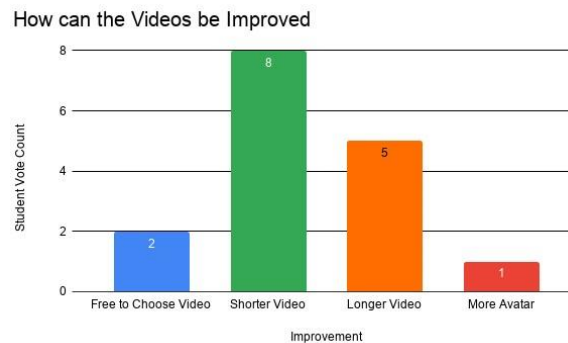
Students were also asked if given the choice, which element do they wish removed from the videos and it was revealed that 8 students (50%) believed the transformed voice (T) should be

removed, 3 students (18.8%) believing the lecturer’s original avatar (RO) should be removed, 1 student (6.3%) believing the anime styled avatar should be removed, and surprisingly 4 students (25%) believing the lecturer’s original voice (O) should be removed.

Fig 15. Students Notion on Which Element should be Removed

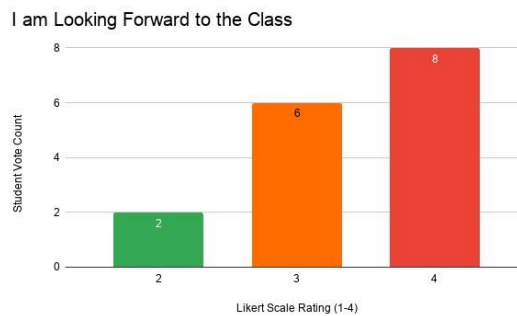


When asked “How can the Videos be Improved”, 8 students (50%) believed the videos should be shorter, 5 students (31.3%) believed videos should be longer, 1 student (6.3%) believe there should be more avatar, and 2 students (12.5%) believe they should be free to choose their own video.



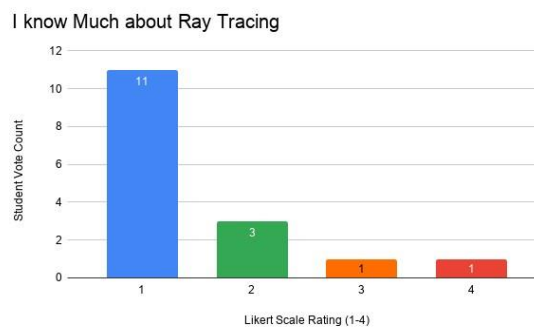
5. Fall 2019 Semester POV-Ray Before Class Survey

The same 16 students who filled the survey for how they believe the Spring Semester Processing Class can be improved also filled the survey regarding their impression on the Fall 2019 POV-Ray Introduction class. Overall, most students are looking forward to the ray tracing class with 8 students (50%) voted 4, 6 students (37.5%) voted 3, and only 2 students (12.5%) voted 2 when asked the question “I am Looking Forward to the Class”. Fig 16. “I am Looking Forward to the Class” Result



Most students also voted that they had little experience with ray tracing with 11 students (68.8%) voting 1, 3 students (18.8%) voting 2, while 1 student (6.3%) voted 3 and another 1 student (6.3%) voted 4 regarding to the question “I Know Much about Ray Tracing”.

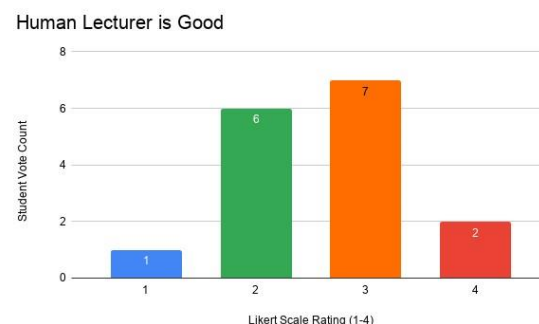
Fig 17. “I Know Much about Ray Tracing” Result



Most students also demonstrated their motivation for learning when 8 students (50%) voted 3 and the same amount of students voted 4 when asked the question “I will Take this class Seriously”. When it comes to the question “Human Lecturer is Good”, 2 students voted 4 (12.5%), 7 voted 3 (43.8%), and 2 voted 4 (12.5%), hinting at a general neutral feeling towards a human lecturer.

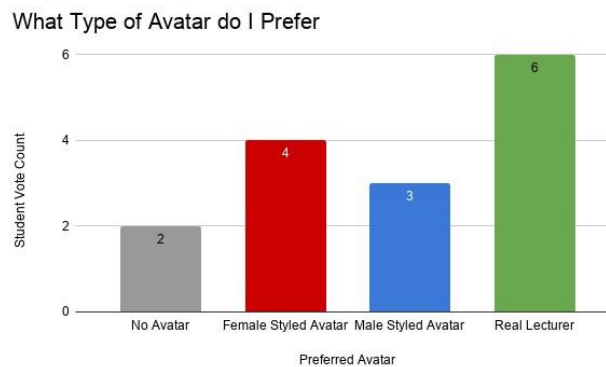
Fig 18. “I will take this Class Seriously” Result

Fig 19. “Human Lecturer is Good” Result



One last question asked to students during the survey is “What Type of Avatar do I Prefer”. For this question, 6 students (37.5%) voted real avatar (R), 3 students (18.8%) voted male styled avatar, 4 students (25%) voted female styled avatar, and 2 students (12.5%) voted for no avatar). Based on the votes, the ranking of avatars based on their popularity is: Real Lecturer (R) > Female Styled Avatar > Male Styled Avatar > No Avatar.

Fig 20. "What Type of Avatar do I Prefer" Result



6. Fall 2019 Semester POV-Ray eLearning Video Creation

The POV-Ray eLearning videos were created with the shortcomings of previous experiment in mind. The usage of transformed videos was stopped due to their inclusion having a potential negative impact on the students' learning. Furthermore, students are now free to choose the videos they want to watch and has no limit on how many videos they want to watch per week. Also in contrast to the previous semester where all of the course contents were compiled into one single video, the learning material was separated into different videos this semester and the contents covered in each labelled clearly.

The virtual avatars A and B are created using a popular 3D avatar tool named VRoid, a popular 3D model creation software in Japan and was then animated using 3tene, which allows for the animation of 3D models using previously recorded motion tracking data. This ensures that the videos can be of the same length. This also allows the lecturer to make 1 video per subject only and have the same facial data used to create 3 separate videos.

7. Subjective Evaluation Survey

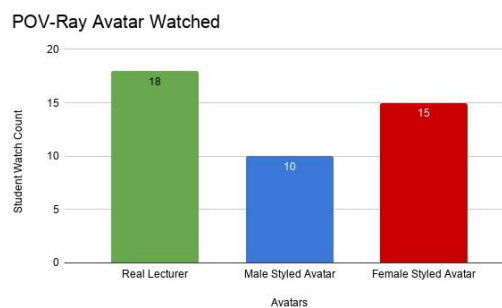
The students in this class were asked to fill out weekly surveys aimed at gathering students' subjective impression regarding the weekly videos. The surveys are structured in multiple choice (MCQ) or Likert Scale (LS) in a scale of 4 from Disagree (1) to Agree (4). The research team purposefully made the scale an even number to eliminate the neutral option. There is also a Short Answer Question (SAQ) inquiring about additional opinion. Unlike the previous semester's processing eLearning course, students have the option of going back to watch any video from previous week and are free to fill out surveys for any week at any given time. The survey questions for the weekly videos are listed below.

- The subject of this video is (MCQ/Number of choice depends on number of subjects)
- The video's avatar I watched is (MCQ of 3)
- How focused are you when watching the video? (LS Scale of 4)
- This week's visual is good (LS Scale of 4)

- This week’s audio is good (LS Scale of 4)
- There is a feeling of mismatch between the visual and audio (LS Scale of 4)
- The overall content of this week’s video is good (LS Scale of 4)
- Please give us any comments (SAQ)

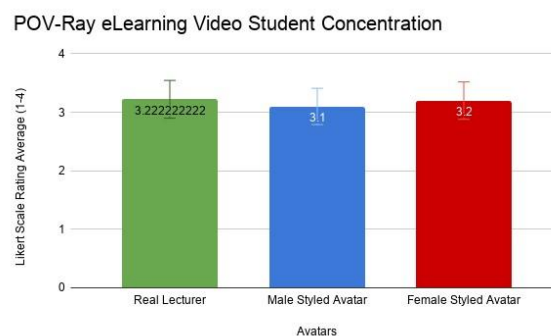
By the 5th week of the semester, 43 students have filled the surveys. Based on the data obtained, it was determined that when given the freedom to pick the avatar students want to view, the most popular avatar is the real lecturer (R) with 18 views (41.8%) followed by female styled avatar (B) with 15 views (34.9%) and lastly male styled avatar (A) with 10 views (23.3%).

Fig 21. POV-Ray Class Avatar Watched by Week 5



In regards to how concentrated students are when watching the video, all 3 avatars are very evenly matched with Real Lecturer (R) being slightly ahead at 3.22 followed by Female Styled Avatar (B) at 3.2 and lastly Male Styled Avatar at 3.1. The ratings given range from 3 to 4 for R and A but for B there are 2 instances where students marked 2 for their answer. The differences between the LS scale averages is also insignificant to come to the conclusion that any one avatar is responsible for the decrease in concentration.

Fig 22. POV-Ray eLearning video student concentration



The question: “This week’s visual is good” also yields very similar results between all videos with R and B at 3.33 and A at 3.3. The score range for the different avatars is once again from 3 – 4 for R and A but for B there is an instance of 2. “This week’s audio is good” yields more only slightly more differences between avatars with B having the at 3.26, A at 3.2, and R at 3.11. The range of answers for R is from 1 to 4, A from 3 to 4, and B from 2 to 4.

Fig 23. POV-Ray eLearning video visual impression

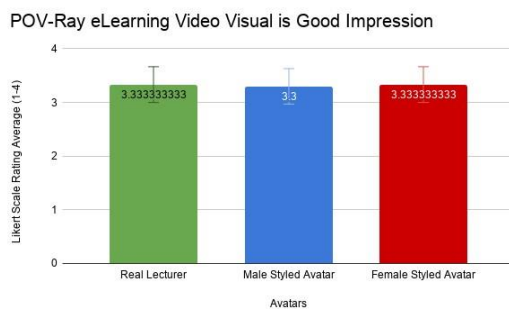
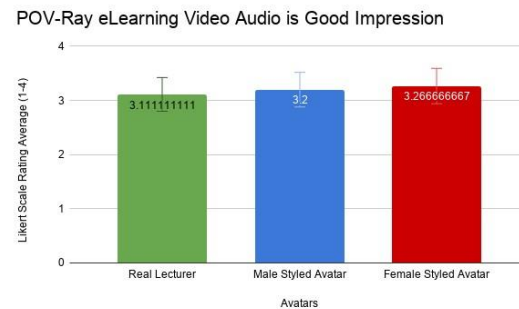
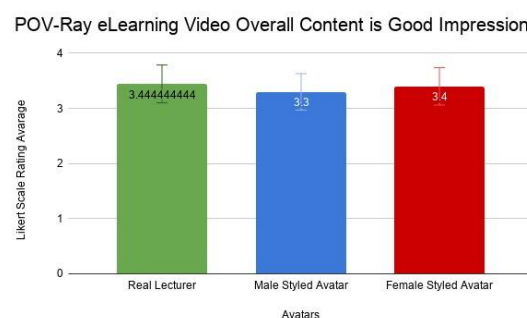
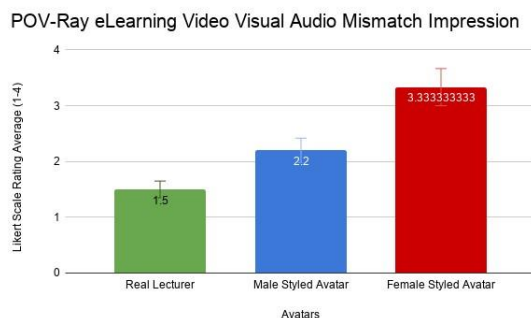


Fig 24. POV-Ray eLearning video audio impression



“There is a feeling of mismatch between visual and audio” yields very significant different between avatars, with R performing the best at 1.5, A coming in second at 2.2, and B with the highest level of mismatch at 3.33. It is worth noting that there does not seem to be a correlation between audio quality and mismatch level as R received the lowest score for audio and yet has the least amount of mismatch while B received the highest score for audio but has the highest amount of mismatch. The overall content indicates R as the best at 3.44, followed by B at 3.4 and lastly A at 3.3. The differences in score is insignificant in the section of overall content as well. There is also no correlation between mismatch and overall content impression.

Fig 25. POV-Ray eLearning video visual audio mismatch Fig 26. POV-Ray eLearning video overall impression



8. Conclusion

The research team constructed the 2nd stage of anime avatar based eLearning course based on findings and shortcomings from the 1st stage of experiments, including: transformed audio (T) yields no positive impact in regards to student’s exam and project grade performance and when mixed with original audio (O) results in grade performance drop. It was also found that transformed audio results in about 10% drop in video view duration. Due to these issues, this new anime avatar based eLearning course, which does not feature different audio components has allowed students to pick the avatar and subject of their choosing. The result is real lecturer (R) being the most picked avatar followed by female styled avatar (B) and male styled avatar (A). While there is still significant difference between avatars in regards to mismatch between visual and audio, there are no significant differences when it comes to visual, audio, and overall content quality.

This research can be expanded with new avatars and different avatar styles that is not anime styles to observe differences in concentration and grade performance when using avatar styles students are not familiar with.

Acknowledgment

The authors of this research paper would like to thank all students from the POV-Ray Introduction class at the Kanagawa Institute of Technology for participating in this research.

References

- [1] Longhao Zou, Irina Tal, Alexandra Covaci, Eva Ibarrola, Gheorghita Ghinea, and GabrielMiro Muntean. 2017. Can Multisensorial Media Improve Learner Experience?. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 315-320. DOI: <https://doi.org/10.1145/3083187.3084014>
- [2] Ioannis Giannoukos, Ioanna Lykourantzou, Giorgos Mpardis, Vassilis Nikolopoulos, Vassili Loumos, and Eleftherios Kayafas. 2008. Collaborative e-learning environments enhanced by wiki technologies. In Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments (PETRA '08), Fillia Makedon, Lynne Baillie, Grammati Pantziou, and Ilias Maglogiannis (Eds.). ACM, New York, NY, USA, Article 59, 5 pages. DOI: <https://doi.org/10.1145/1389586.1389657>
- [3] Kim, Jihyun & Jung, Yujung & Lim, Yoonsun & Kim, Myung & Noh, Sunsook. (2009). An e-learning framework supporting personalization and collaboration. 635-638. 10.1145/1516241.1516352.
- [4] He, Zheng & Ueno, Haruki. (2012). Designing m-learning system based on WebELS system. 115-118. 10.1145/2382336.2382368.
- [5] Rex Hsieh, Akihiko Shirai, and Hisashi Sato. 2019. Effectiveness of facial animated avatar and voice transformer in elearning programming course. In ACM SIGGRAPH 2019 Posters (SIGGRAPH '19). ACM, New York, NY, USA, Article 82, 2 pages. DOI: <https://doi.org/10.1145/3306214.3338540>