

# Effectiveness of Facial Animated Avatar and Voice Transformer in eLearning Programming Course

Rex Hsieh  
Kanagawa Institute of Technology  
Atsugi, Kanagawa, Japan

Akihiko Shirai  
GREE VR Studio Lab, GREE, Inc.  
Tokyo, Japan

Hisashi Sato  
Kanagawa Institute of Technology  
Atsugi, Kanagawa, Japan

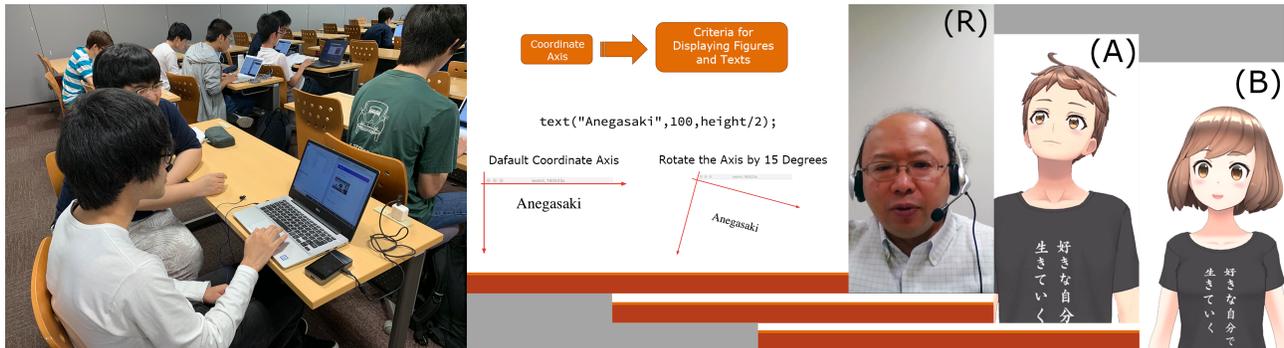


Figure 1: From left to right: Friday Processing Workshop, (R) Male Lecturer Video, (A) Avatar A Video, and (B) Avatar B Video

## ABSTRACT

The advancement in technology brought about the introduction of eLearning to educational institutes. By supplementing traditional courses with eLearning materials, instructors are able to introduce new learning methods without completely deviating from standard education programs [Basogain et al. 2017]. Some of the most popular forms of E-Learning include online courses [Aparicio and Bacao 2013], [Goyal 2012], video clips of lectures, and gamification of courses and materials [Plessis 2017]. This paper introduces and evaluates the performance of eLearning videos featuring anime styled avatars (a.k.a vTuber) speaking in vocoder transformed audios and how they compare with the traditional lecturer videos.

## CCS CONCEPTS

• Social and professional topics → Cultural characteristics.

## KEYWORDS

eLearning, VR, Avatar, User Experience, Voice Transform, vTuber

## ACM Reference Format:

Rex Hsieh, Akihiko Shirai, and Hisashi Sato. 2019. Effectiveness of Facial Animated Avatar and Voice Transformer in eLearning Programming Course. In *Proceedings of SIGGRAPH '19 Posters*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3306214.3338540>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
*SIGGRAPH '19 Posters*, July 28 - August 01, 2019, Los Angeles, CA, USA  
© 2019 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-6314-3/19/07...\$15.00  
<https://doi.org/10.1145/3306214.3338540>

## 1 RESEARCH METHOD

186 students from a Processing class at the Information Science Department of Kanagawa Institute of Technology is currently participating in this experiment which runs from April 2019 to July 2019. Every Friday, students are asked to watch videos containing one of the 6 videos labeled: RO, AO, BO, RT, AT, BT which stands for: Real Visual (Male Lecturer's visual) Original Audio, Avatar A (Male Style Avatar) Original Audio, Avatar B (Female Style Avatar) Original Audio, Real Visual Transformed Audio, Avatar A Transformed Audio, Avatar B Transformed Audio. Avatar A and B are created using "REALITY" application on iPhone, which projects a real time 3D avatar with fast facial capture runs on 30 to 75 fps. The transformed audios are generated using Roland VT-4 vocoder by adjusting the Pitch, Formant, Balance, or Reverb.

### 1.1 Student Division Method

The 186 students are divided into 15 groups labeled from A to O with each assigned to 1 or 2 videos as follows: [A: RO-RT, B: AO-AT, C: BO-BT, D: AO-BO, E: AT-BT, F: RO-AO, G: RO-BO, H: RT-AT, I: RT-BT, J: RO, K: AO, L: BO, M: RT, N: AT, O: BT]. Group A to I were asked to watch either of the 2 videos they were assigned to each week to measure their subjective impression differences for the 2 videos while groups J to O served as control group and watched only 1 of the 6 videos.

### 1.2 Subjective Data Gathering

Students attending the course are asked to fill out a survey before and after class as well as each week after having watched the eLearning videos in order to gather their subjective impression data. 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. The survey questions for the weekly videos are listed below. In order to make sure the students have finished each video and are viewing the correct video, we have asked each student to put down the start and end time as well as the keyword of the video they have watched.

- How focused are you when watching the video? (MCQ of 6)
- The video's visual is good. (LS)
- The video's audio is good. (LS)
- The visual and audio do not mix well. (LS)
- Overall the video is good. (LS)
- Regarding the Audio (MCQ of 2)
- Regarding the Learning Content (MCQ of 2)
- Regarding the Avatar (Checkbox of 13)

## 2 HYPOTHESIS

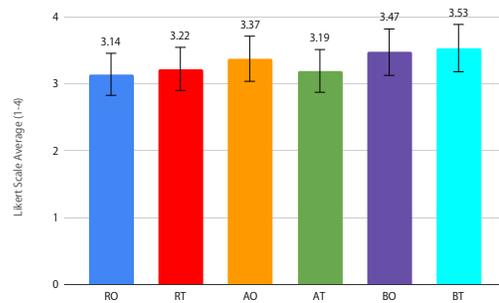
Due to the popularity of anime characters amongst Japanese youth, particularly female anime characters, the research team predicted that videos featuring Avatar B will be better received and will allow students to outperform the other groups grade-wise followed by Avatar A and finally Original Visual. The research team also predicted voice transformer will work best with avatar visuals but not with lecturer visuals. Therefore the predicted academic performance from best to worst is illustrated as follows: BT, BO, AT, AO, RO, RT.

## 3 SUBJECTIVE IMPRESSION RESULTS

According to the devised subjective evaluation using Google Form survey, the team has found Avatar B to be the most favorably ranked avatar amongst all three avatars with Avatar A and R ranked just slightly behind Avatar B. Audio wise; however, Avatar B performs considerably less well in comparison to Avatar A and B ranking number 1 in the Question: "The visual and audio do not match. (LS)" as well as having the lowest score in "The video's audio is good. (LS)" Original Audio also outperforms transformed audio by a significant margin and while O ranked 3 or above for Audio is Good Likert Scale Average, T only ranked around 2.5. The mismatch impression for visual and audio is also the least RO at 1.85 while RT scored the 2nd at 2.25. BT was ranked as having the most mismatch at 2.90. Despite the difference in audio quality, all videos are rated a 3 or above in terms of impression for overall content. The below graphs are data from week 1 - 4 with a total of 524 answers.

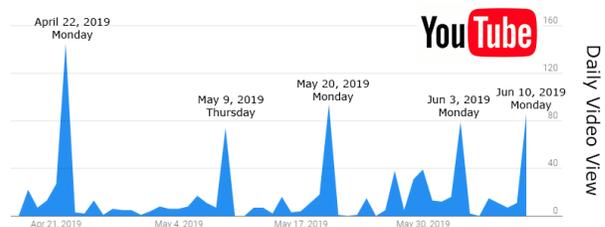
## 4 YOUTUBE DATA ANALYSIS

Aside from Google Form surveys, the research team has also been using YouTube Studio Analytics to conduct objective data analysis regarding the watch time and duration for the videos. The majority of the views took place on Monday and Thursday, which are the days right before the processing class with Tuesday being the Processing Lecture and Friday being the Processing Workshop. Furthermore by comparing the average view duration of each video with their length, the research team was able to obtain objective data regarding student's motivation towards watching each video. Overall videos featuring original audio received longer viewing time in comparison to videos featuring transformed audio by an average of 16.87 seconds. While the video duration varies



**Figure 2: Likert Scale Graph of Students' impression video visuals demonstrating Avatar B is the most preferred video.**

between each video, T videos with the exception of A series have longer duration therefore taking away the possibility O receives longer views because they are longer when in fact O is shorter than T and yet receives longer views. This demonstrates audio quality is very important and can play a crucial factor in motivating students to watch them.



**Figure 3: YouTube video view count showing students' most motivated to watch videos before lectures or after vacations**

## 5 CONCLUSION

The proposed method can analyze interests and behavior of students which includes preferred avatar, voice, and studying time. This method contributes to the development of next generation of eLearning Program utilizing motion tracking generated avatar visual and augmented audio. Due to how both the visual and audio are generated in real-time by tracking the visual and audio outputs from the lecturer, this research has the potential to become a live interactive education environment.

## REFERENCES

- Manuela Aparicio and Fernando Bacao. 2013. e-Learning Concept Trends. In *Proceedings of the 2013 International Conference on Information Systems and Design of Communication (ISDOC '13)*. ACM, New York, NY, USA, 81–86. <https://doi.org/10.1145/2503859.2503872>
- Xabier Basogain, Miguel Ángel Olabe, and Juan Carlos Olabe. 2017. Transition to a Modern Education System Through e-Learning. In *Proceedings of the 2017 International Conference on Education and E-Learning (ICEEL 2017)*. ACM, New York, NY, USA, 41–46. <https://doi.org/10.1145/3160908.3160924>
- Sumit Goyal. 2012. E-Learning: Future of Education. *Journal of Education and Learning (EduLearn)* 6 (09 2012), 239. <https://doi.org/10.11591/edulearn.v6i4.168>
- Ilona Du Plessis. 2017. The Effectiveness of Gamification as an E-Learning Strategy (*IJMAS*). *IRAJ*, 24–26.