Evaluation of Avatar and Voice Transform in Programming E-Learning Lectures

Rex Hsieh
rex@shirai.la
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

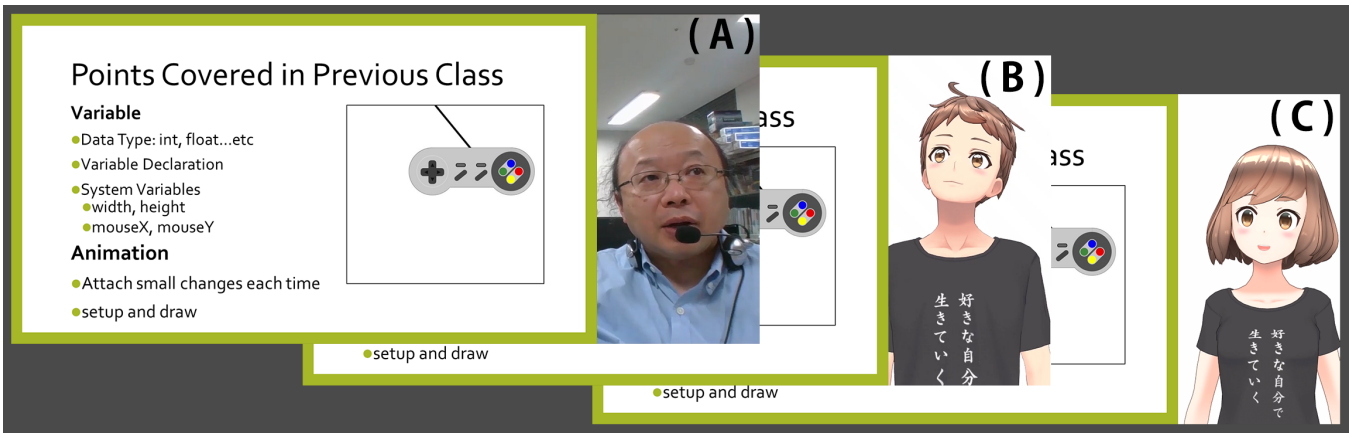

Figure 1: Human Lecturer Video (A), Avatar A Video (B), and Avatar B Video (C)

ABSTRACT

This article reports the effectiveness of high frame rate facial animated avatar and voice transformer in eLearning. Three avatars: (Real male professor, Male avatar, Female avatar) were combined with male professor's voice or VT-4 vocoder transformed voice to create 6 distinguished videos which were then viewed by university freshmen students. A total of 186 students divided into 15 groups participated in this experiment. Female avatar was the most appealing avatar visually, but its combination with voice transform severely hinders its overall score. This research can be extended to real time live with with preferences of students and draw more connections between student perception of avatar and actual lecturers.

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. Evaluation of Avatar and Voice Transform in Programming E-Learning Lectures. In *ACM International Conference on Intelligent Virtual Agents (IVA '19), July 2–5, 2019, PARIS, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3308532.3329430

IVA '19, July 2–5, 2019, PARIS, France

ACM ISBN 978-1-4503-6672-4/19/07…$15.00
https://doi.org/10.1145/3308532.3329430

1 MOTIVATION

The 21st Century has seem a rapid advancement and employment of new technologies such as E-Learning technology and Virtual Reality. In Japan especially the Virtual Reality Youtubers trend (VTubers for short) have gained popularity in recent years and its introduction enables ordinary citizens to transform themselves into anime celebrities online. 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 seeks to examine the possibility of using VTuber like avatars in a university setting where freshmen students were exposed to E-Learning online supplementary video materials with the image of instructor replaced by that of VTubers and the voice transformed into anime sounding voices.

2 RELATED WORKS

Numerous researches have been done regarding how to best utilize E-Learning materials and course structures. Zou [5] created and evaluated what was called the Mulsemedia (Multi-Sensational Media) aimed at providing supplementary video material to aid the education of STEM course works. Several other researchers such as Giannoukos [2], Kim [4] and He [3] 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. [1]. Despite of the richness in E-Learning 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 change the very visual and audio of instructors in an effort to increase the learning interest of students. Instead of encouraging more direct contacts between lecturers and students in E-Learning, this research aims at digitizing the human element by modifying the lecturer into an Avatar.

3 RESEARCH METHODOLOGY

This research is conducted in a Processing lecture class required for all freshmen belonging to the Information Science Department of Kanagawa Institute of Technology. The class consisted of 186 students and spans over a period of four months from April to September 2019. During this course, aside from the main lecture that took place every Tuesday, students will be supplied with a lecture video to enable them to review the course material after class. Instead of regular videos capturing the lecturer, however; this experiment substitutes the online lecture videos containing the professor with VTuber videos created using REALITY Avatar that contains the same speech as the original lecture video only this time the professor's imagery was substituted with that of the VTuber with the classroom changed to a virtual anime like background and the audio transformed to that of anime characters. The videos are uploaded onto a Youtube channel.

3.1 Student Pool Division Method

The 186 students in this class are divided into 9 groups of 12 and 6 groups of 13 with the 9 groups of 12 watching 2 out of 6 videos and asked to compare the 2 videos they were given. The 6 groups of 13 are made to only watch 1 video and they represent the control group of this experiment. The videos are labelled as follows: RO, AO, BO, RT, AT, BT. The R stands for Real and it implies that the imagery in the video is that of the original lecture footage. A stands for Avatar A and is an anime styled male avatar, B stands for Avatar B and is an anime styled female avatar. The O stands for original voice and as implied the audio of the video was not transformed by the Roland VT-4 Voice Transformer while T stands for voice transformed by the Roland VT-4 Voice Transformer. With this in mind the 15 groups are exposed to at least one of the following videos every Wednesday: Lecturer visual with original audio, male Avatar with original audio, female Avatar with original audio, lecturer visual with transformed audio, male Avatar with transformed audio, and female Avatar with transformed audio.

The 15 groups are labelled from A to O. Below are the video(s) each group is assigned to watch and the aim behind this division.

As indicated by the chart, groups from A to I are made to watch 2 videos with 1 video per week alternating between the two and asked about their thoughts on each video as as to compare the effectiveness of each. The objective of group A, B, and C is to access the effectiveness of original voice versus transformed voice and therefore the visual avatar each group watches stays the same in both weeks and it is only the voice that changes. The objective of Group D and E is to measure the effectiveness of Avatar A and B compare to each other and so the audio they are exposed to stays the same and it is only the visual that changes. Groups F, G, H, and I was assigned to determine the effectiveness of Real human versus digital avatar and therefore the audio stays the same while the visual changes between human and digital avatar.

Group	Video	Objective
A	RO, RT	Audio Comparison
B	AO, AT	
C	BO, BT	
D	AO, BO	Male/Female Avatar Comparision
E	AT, BT	
F	RO, AO	Real/Avatar Comparision
G	RO, BO	
F	RT, AT	
I	RT, BT	
J	RO	Control Group
K	AO	
L	BO	
M	RT	
N	AT	
O	BT	

Figure 2: 15 groups by 186 students

3.2 VTuber Video Creation

The lecturer videos containing the lecture will be accessible to students every Wednesday. These VTuber videos were created using the REALITY Avatar, a free online VTuber streaming application developed and published by GREE in 2018 while the audios were transformed using the Roland Voice Transformer VT-4. REALITY Avatar is able to track the users' eye and mouth movements and can easily create videos with the Anime character's lips matching that of the audio clips. The audio was created by the lecturer speaking in front of the VT-4 and adjust the PITCH, FORMANT, BALANCE, and REVERB until the voice resembles that of the anime boy or girl depending on whether the video was intended for Avatar A or B. Afterwards the videos and the audios are edited together using Adobe Premiere Pro and uploaded onto Youtube for students to view.All in all six videos were uploaded onto Youtube every week.

3.3 Quizzes and Surveys

Students attending the course are asked to fill out a survey before, after each weekly video viewing, and after the course. The surveys are structured in multiple choice (MCQ), semantic differential (SD), or short answer questions (SAQ) composed of a scale of 4 from Disagree to Agree. The research team purposefully made the scale an even number to eliminate the neutral option. The survey questions before the experiment are as follows:

- Which video do you want to watch? (MCQ of 6)
- I am looking forward to the class. (SD)
- I have experience with VTuber. (SD)
- I like Human Lecturer. (SD)
- I know a lot about Processing. (SD)
- I will take this class seriously. (SD)
- I have experience with online lectures. (MCQ of 2)

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 keyword of the video they have watched.

- How focused are you when watching the video? (MCQ of 6)
- This week's visual is good. (SD)
- This week's audio is good. (SD)
- This visual and audio fit well together. (SD)
- Overall the video is good. (SD)
- Regarding the Audio (MCQ of 2)
- Regarding the Learning Content (MCQ of 2)
- Regarding the Avatar (Checkbox of 13)

An after class survey aims at gathering the final thoughts of this experiment from students will be given at the end of this course. All students are also given weekly quizzes to measure their learning progress. These measures are conducted in hopes of both measuring the students' emotional response and academic performance.

4 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.

5 BEFORE CLASS SURVEY RESPONSE

Out of 186 students who are enrolled in the class, 182 students responded to the Before Class Survey. 174/182 (95.6%) of the students answered that they are looking forward to the class (rated 3 or 4 SD scale) while 172/182 (94.5%) said they will take the class seriously, pointing towards a positive attitude in the group. The majority of students also indicated they have no prior knowledge when it comes to E-Learning, Processing, or VTuber with 142 (78%) saying they have no experience with online learning, 169 (93.3%) saying they do not know Processing, and 168 (91.2%) saying they do not have experience with VTuber. There are no clear indication of which videos the students want to watch. BO received the highest vote at 49 (27.1%) followed closely by RO and BT each at 48 (26.5%). RT received 20 (11%) of votes while AO and AT each got 8 (4.4%). There are also no clear indication of opinion regarding if students like human lecturer with 123 students (67.6%) answering either 2 or 3 on the SD scale.

6 WEEKLY VIDEO SURVEY

The first video survey was conducted in April 18th for the week of April 16th and was participated by 160 students. For "The video's visual is good", BT received the highest SD score at 3.9 followed by AO at 3.44 then BO (3.37), RO (3.34), AT (3.22), and RT (3.15). It is important to note that despite of differences, all videos are rated above 3. For "The video's audio is good", RO ranked the highest at 3.39 followed by AO (3.29), BT (3.2), BO (2.68), RT (2.57), and finally AT (2.41). For "The visual and audio do not mix well", BT received highest score at 3.10 meaning students felt the visual and audio do not fit together, a result demonstrated in its positive visual reception and poor audio reception.

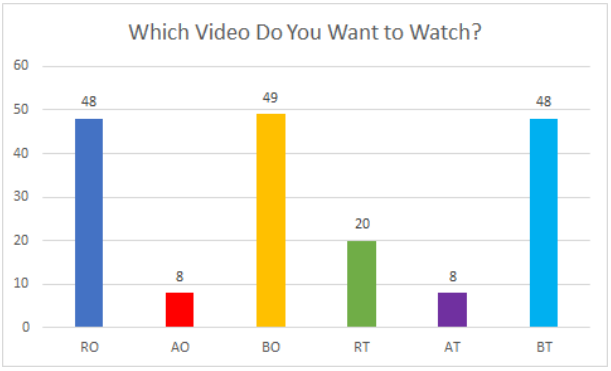

Figure 3: "Which Video Do You Want to Watch?" result. Answered by 181 Students

7 CONCLUSION

The initial video survey has found BT to be the most preferred video visual wise and while BO was also rated highly in terms of visual qualities, BT outshines BO by a large margin in terms of audio score. RO did fairly well in the visual field at 4th place and was ranked as having the best audio at 3.39. AO, despite of being rated as the video students wanted to watch the least in the "Before Class Survey" actually performed very well in both audio and visual coming out at 2nd place in both categories. The current results demonstrates that while female avatar is the most appealing avatar visually, its combination with voice transform severely hinders its overall score.

This research can be extended to real time live with preferences of students, and draw more connections between student perception of avatars and actual lecturers.

REFERENCES

[1] M. Samir Abou El-Seoud, Islam A.T.F. Taj-Eddin, Naglaa Seddiek, Mahmoud M. El-Khouly, and Ann Nosseir. 2014. E-Learning and Students' Motivation: A Research Study on the Effect of E-Learning on Higher Education. *International Journal of Emerging Technologies in Learning (iJET)* 9, 4 (2014), 20–26. https://online-journals.org/index.php/i-jet/article/view/3465

[2] Ioannis Giannoukos, Ioanna Lykourentzou, 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)*. ACM, New York, NY, USA, Article 59, 5 pages. https://doi.org/10.1145/1389586.1389657

[3] Zheng He and Haruki Ueno. 2012. Designing M-learning System Based on WebELS System. In *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service (ICIMCS '12)*. ACM, New York, NY, USA, 115–118. https://doi.org/10.1145/2382336.2382368

[4] Jihyun Kim, Yujung Jung, Yoonsun Lim, Myung Kim, and Sunsook Noh. 2009. An e-Learning Framework Supporting Personalization and Collaboration. In *Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication (ICUIMC '09)*. ACM, New York, NY, USA, 635–638. https://doi.org/10.1145/1516241.1516352

[5] Longhao Zou, Irina Tal, Alexandra Covaci, Eva Ibarrola, Gheorghita Ghinea, and Gabriel-Miro 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. https://doi.org/10.1145/3083187.3084014