



プログラミング教育向け eLearning における 顔アニメーションアバターとボイスチェンジャーの影響

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

謝レックス²⁾, 白井暁彦¹⁾, 佐藤尚²⁾
HSIEH Rex, SHIRAI Akihiko, and SATO Hisashi

1) GREE VR Studio Lab, グリー株式会社

2) 神奈川工科大学大学院工学研究科 (〒 243-0292 神奈川県厚木市下荻野 1030)

概要: This paper reports the effectiveness of eLearning Videos composed of virtual avatar visuals and transformed audios generated from the lecturer's original footage. Measured elements include grades, subjective opinions, and video watch duration. From data drawn from 11 weeks of experiment, it can be concluded that although female styled avatars are rated as the most visually appealing in subjective impression surveys, students assigned to them performed the worst grade-wise while male styled avatar has the most positive academic impact on students.

キーワード: Avatar, E-Learning, vTuber, voice changer

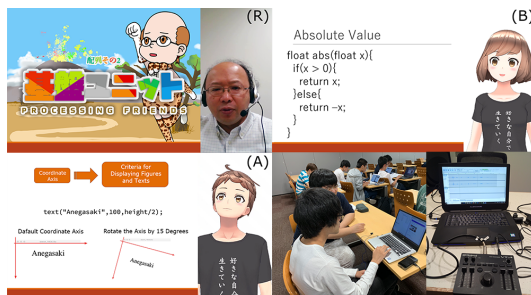


図 1: Real Avatar Visual (R), Male Styled Avatar Visual (A), Female Styled Avatar Visual (B), Processing Class, Video Equipment

1. Introduction

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 [3], [5]. Some of the most popular forms of E-Learning include online courses, video clips of lectures, and gamification of courses and materials [2]. The motivation behind developing eLearning platforms includes enabling students to learn anywhere anytime [1]. This paper evaluates the performance of eLearning videos featuring anime styled avatars (a.k.a VTuber) speaking in vocoder transformed audios and serves as the next step to previous published results [4].

2. Research Method

186 students from a Processing class at the Information Science Department of Kanagawa Institute of Technology participated in this experiment that lasts for 11 weeks running 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 Lecture'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. The lecturer's visual and original audio are captured using Open Broadcasting Software Studio (OBS Studio). Avatar A and B are created using "REALITY" application on iPhone, which projects a real time 3D avatar with fast facial capture that runs on 30 to 75 fps. The transformed audios are generated using Roland VT-4 vocoder by adjusting the Pitch, Formant, Balance, or Reverb.

2.1 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

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	
J	RO	RO	
K	AO	AO	Control Group: Only Watched 1 Video
L	BO	BO	
M	RT	RT	
N	AT	AT	
O	BT	BT	

図 2: 186 Students are divided into 15 Groups Labeled from A to O. Each group is assigned to 1 or 2 videos and watches 1 per week

videos. The experimental groups A to I can be further classified into 3 main sections each designed to test a specific feature of the videos. Group A, B, and C, although watches different videos in odd and even week, did not have their visual's changed and the only difference between the videos is their audio. Group D and E are designed to measure the difference between male and female styled avatar and thus the audio does not change in between weeks while the avatar switches between A and B. Group F, G, H, and I watch videos of the same audio but have their visual avatars switch between "REALITY" generated anime styled avatar and the real lecturer, therefore the visual change follows 1 of the 2 tracks: A -> R -> A, B -> R -> B. A detailed group division chart can be seen in Chart 2.

2.2 Google Form Survey

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 before class impressions are listed below.

- どのビデオを見たいですか? (MCQ of 6)
- 授業を楽しみにしている (LS)
- VTuber の経験がある (LS)
- 普通の人間の講師が良い (LS)
- Processing をよく知っている (LS)
- 真面目に授業に取り組みたい (LS)
- オンライン学習に馴染みがある (MCQ of 2)
- なんかコメントがあればよろしくお願ひします (SAQ)

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.

- 開始時刻 (Time Input)
- Keyword (SAQ)
- ご自身の集中度について (MCQ of 6)
- 視聴終了時刻 (Time Input)
- 今週のビデオの映像 (visual) は良かった (LS)
- 今週のビデオの音声 (audio) は良かった (LS)
- 映像と音声の違いに違和感を感じた (LS)
- 今回のビデオ (総合的内容) は良かった (LS)
- 音声品質について (MCQ of 2)
- 学習内容について (MCQ of 2)
- アバターについて (Checkbox of 13)
- この調査について何かコメントがありましたらお寄せください (SAQ)

3. Hypotheses

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.

4. Before Class Subjective Impression Results

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 LS 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

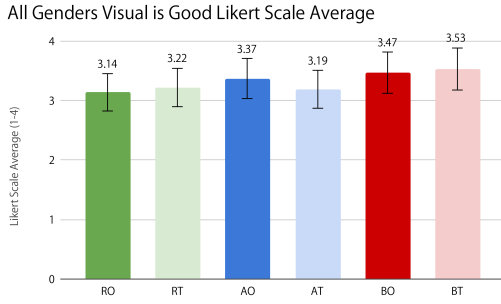


図 3: Video’s Visuals LS Score

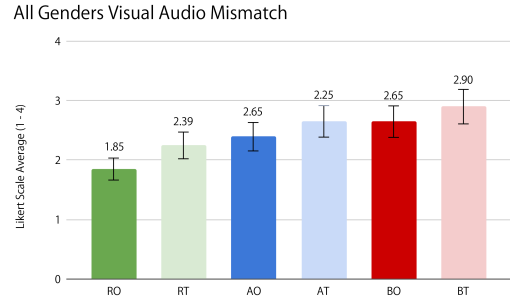


図 5: Audio Visual Mismatch LS Score

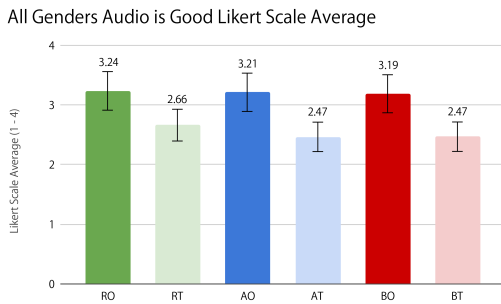


図 4: Video’s Audio LS Score

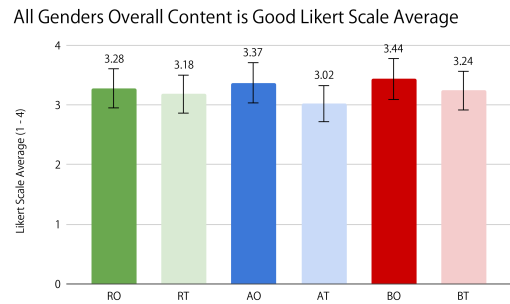


図 6: Overall Content LS Score

also no clear indication of opinion regarding if students like human lecturer with 123 students (67.6%) answering either 2 or 3 on the LS scale.

5. Weekly Video Subjective Impression

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.

6. 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 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.

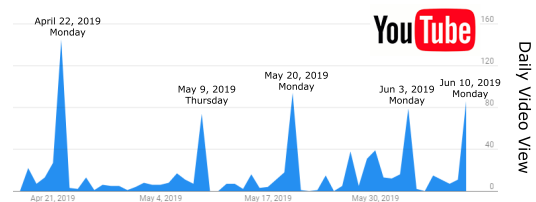


図 7: YouTube Video View Counts

Watched rate (Obtained by YouTube)								
Average / Duration (rate)	Week 1	Week 2	Week 3	Week 4	Week 5	Total	(StdEv)	Diff (RO)
RO	87/163 (0.53)	178/271 (0.66)	101/179 (0.56)	121/197 (0.61)	546/894 (0.61)	1033/1704 (0.61)	±0.043	0.00
RT	89/163 (0.55)	139/271 (0.51)	100/180 (0.56)	106/171 (0.62)	44/83 (0.53)	478/899 (0.53)	±0.038	-0.07
AO	130/181 (0.72)	183/272 (0.67)	102/181 (0.56)	101/183 (0.55)	56/84 (0.67)	572/901 (0.63)	±0.065	0.03
AT	122/181 (0.67)	157/300 (0.52)	141/180 (0.78)	39/142 (0.27)	54/76 (0.71)	513/879 (0.58)	±0.180	-0.02
BO	128/161 (0.80)	198/283 (0.70)	104/171 (0.61)	99/197 (0.50)	53/76 (0.70)	582/888 (0.66)	±0.099	0.05
BT	62/170 (0.36)	197/362 (0.54)	72/198 (0.36)	73/171 (0.43)	60/79 (0.76)	464/980 (0.47)	±0.149	-0.13

☒ 8: YouTube Video Watched Duration

7. Mid-Term Grades

On July 2nd, 2019, a week between week 9 and 10 of the experiments where students are not assigned new videos or course materials, the students took a mid-term exam. Of the 186 students originally registered for the class, 173 students completed the exam. The average mid-term exam grades for each group are as follows: A: 82.36, B: 82, C: 75, D: 82.36, E: 86, F: 75.6, G: 83.23, H: 78.73, I: 73.5, J: 85.17, K: 80.14, L: 78.31, M: 80.83, N: 89.2, O: 75.33. Here the group performing the best is N at an average of 89.2 assigning to AT while the group performing the worst is I at 73.5 assigning to RT/BT.

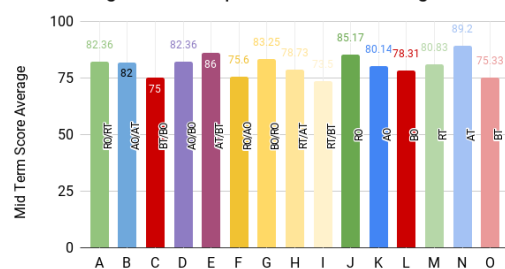
Within these groups, J to O served as control group and only watched 1 of the 6 videos during the class. From these groups, the grade ranking are as follows: AT (89.2) > RO (85.17) > RT (80.83) > AO (80.14) > BO (78.31) > BT (75.33). According to the results of these control groups, visually students assigned to Male Styled Avatar (A) performed the best followed by Real Lecturer (R) and finally Female Styled Avatar (B).

Groups A to C, while having watched 2 videos, only watched 1 of the 3 videos and it is the audio that changes between the weeks. For these 3 groups, the grade rankings are as follows: RO/RT: 82.35, AO/AT: 82, BO/BT: 75. Based on the current finding we concluded that female styled lectures, in contrary to what our hypothesis predicted, has an apparent negative impact on the student's grades while male styled avatar apparently aided the students to get the best grade with real avatar (lecturer's original visual) following not far behind. As to regarding whether original or transformed audio has more of a positive impact on the student's grades, the result is much less apparent with groups representing original audio: D (AO/BO), F (RO/AO), G (BO/RO) at 82.36, 75.6, and 83.23 respectively while groups representing transformed audio: E (AT/BT), H (RT/AT), I (RT/BT) at 86, 78.73, 73.5 respectively.

8. Conclusion

The proposed method can analyze interests and behavior of students which includes preferred avatar, voice,

Processing Class Groups A~O Grade Average



☒ 9: Average MidTerm Grades of 15 Groups

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. Based on the findings of this experiment, we concluded that while female styled avatars are rated as the most visually appealing avatar, students assigned them performed the worst grade-wise while those assigned to male styled avatar performed the best.

参考文献

- [1] Longhao Zou, Irina Tal, Alexandra Covaci, Eva Ibarrola, Gheorghita Ghinea, Gabriel-Miro Muntean : Can Multisensorial Media Improve Learner Experience?, Proceedings of the 8th ACM on Multimedia Systems Conference, pp. 315–320, 2017.
- [2] Ilona Du Plessis : The Effectiveness of Gamification as an E-Learning Strategy, IJMAS, pp. 24–26, 2017.
- [3] M. Samir Abou El-Seoud, Islam A.T.F. Taj-Eddin, Naglaa Seddiek, Mahmoud M. El-Khouly, Ann Nosseir : E-Learning and Students' Motivation: A Research Study on the Effect of E-Learning on Higher Education, International Journal of Emerging Technologies in Learning, 2014.
- [4] Rex Hsieh, Akihiko Shirai, Hisashi Sato : Evaluation of Avatar and Voice Transform in Programming E-Learning Lectures, IVA '19 Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, pp. 197–199, 2019.
- [5] Xabier Basogain, Miguel Olabe, Angel, Juan Carlos Olabe : Transition to a Modern Education System Through e-Learning, Proceedings of the 2017 International Conference on Education and E-Learning, pp. 41–46, 2017.