The Impact of Metaverse and Virtual Idols Technologies to Teaching and Learning

Kenny Yu, Aaron Yuen and Tinson Ngai,

Kenny Yu[*,a]
[a] Vocational Training Council, Discipline of Information Technology, Senior Lecturer, Hong Kong,
Aaron Yuen[*,b]
[a] Vocational Training Council, Discipline of Information Technology, Lecturer, Hong Kong,
Tinson Ngai[*,c]
[b] Vocational Training Council, Discipline of Information Technology, Lecturer, Hong Kong.

* kennyyu@vtc.edu.hk, aaronyuen@vtc.edu.hk, and ^ tinson0428@vtc.edu.hk

Abstract

Virtual idols are not a new concept. The first generation of these artificial celebrities were developed for the ACG industry of Japan (animation, comic and games) in the 1980s. In recent years, the rapid advancements in social media platforms, Metaverse, cloud computing, big data and A.I. technologies, helps the development of photo-realistic digital characters to be applied in different areas to express themselves in more intimate, immediate ways and garner massive fan bases.

In China, the rise of Bilibili in 2020, a video platform that is favoured by ACG fans, stimulated the vast development and application of virtual idols on the platform. Based on the statistics provided by Bilibili released in 2020, there were around 32,400 virtual idols hosted livestreams on Bilibili in 2020. The people behind them are known as "Vtubers".

Our project team has been investigating in the relevant field starting in 2020, and developed the Phase I virtual idol MetaHuman, supporting three distinguished characters, namely, IT Sarah, IT Hana and IT Sophia with Unreal Engine. IT Sarah and Hana are dubbed behind the scenes and captured in real-time, while IT Sophia is an AI chatbot.

With the enhanced experiments provided to the students through virtual classroom, workshop collaboration and project cooperation through the virtual idols, positive results have been found. The virtual idol project team therefore starts the Phase II focusing on the adaptation of AI technology which empower the virtual idols with higher degree of interactivity, content filtering and the latest advancements in natural language programming (NLP).

.

Keywords: *metaverse, virtual idols, AR, VR, MR, XR NLP, artificial intelligence (AI), chatbot, google cloud service, Unreal Engine, OpenAI, Azure cloud service*

1. Introduction

1.1 The Evolution of Virtual Idols

The evolution of virtual idols has been a fascinating phenomenon. Virtual idols are computer-generated characters that are designed to appear as singers, dancers, and other performers. They are typically used in marketing campaigns and have become increasingly popular over the past decade. Virtual idols are created using a combination of 3D modeling, motion capture, and voice synthesis technology. As technology advances, the capabilities of virtual idols have become increasingly sophisticated, allowing them to be more interactive and engaging with their fans.

Vtubers often project visuals which are designed to be aesthetically appealing. The majority of virtual idols fans are the demographic cohort succeeding Millennials and preceding Generation Alpha (Generation Z). Over 70% of these followers are aged between 18 and 23 (iiMedia research, 2021). More international brands and organisations like Louis Vuitton, Tesla, KFC and Givenchy have commissioned virtual idols for promotional campaigns world-wide. It is for sure that in the future, metaverse environments will also offer new places for virtual idols to interact with followers and fans. They will be involved in various digital entertainment sectors and industries.

1.2 AI Voice Overs with Emotions

AI voice overs with emotion refer to the use of artificial intelligence (AI) technology to generate spoken content such as speeches, narrations, and audio recordings that exhibit a range of emotions. This technology uses machine learning algorithms and natural language processing (NLP) techniques to analyse and interpret text input, then generate human-like voices that convey various emotions such as happiness, sadness, anger, and excitement. The goal of AI voice overs with emotion is to create more engaging and personalised audio content that can connect with audiences on a deeper level. This technology has a wide range of applications in industries such as advertising, entertainment, and education.

Natural Language Processing (NLP) is a sub-discipline of artificial intelligence and linguistics. This field of NLP has been discussed in the previous paper released by the team in 2021, so that it will not be covered

again in this paper, instead the AI generated voice over will be researched and how virtual idols can be benefited from it.

Murf.AI Studio

Murf.AI Studio is an AI enabled, real people's voices platform that created studio-quality voice over in minutes. It converts from text to speech with a versatile AI voice generator, to be used by virtual idols.

Microsoft VALL-E

A Microsoft new text-to-speech AI model that can closely simulate a person's voice when given a three second audio sample. Once it learns a specific voice, VALL-E can synthesise audio of that person saying anything. The API also provides probability to simulate emotion.

So we started to research and use the

So we started to research and use the API to do it in a way that attempts to preserve the speaker's emotional tone to be applied in the virtual idols for course promotional and teaching purposes.

1.3 AI Powered Virtual Idols

The project team has been researching the Microsoft Azure OpenAI technology in the Virtual Idol project Phase III development. The project team's vision speculates that VALL-E could be used for high quality text-to-speech application, speech edition where a recording of a person could be edited and changed from a text transcript and audio content creation when combined with other generative AI models like GPT-3.

Based on the technology, the new virtual idols could generate discrete audio codec codes from text and acoustic prompts. It basically analyses how a person sounds, breaks that information into discrete components and uses training data to match how that voice would sound if it spoke other phrases outside of the three-second sample.

Figure 1 Virtual Idol - Ava & Max

2. Promotion & Education

2.1 AI Powered IT Hana & IT Sophia

Similar to IT Sarah, the difference is that IT Sophia and IT Hana are in a Metaverse. Students use MOTION CAPTURE to transform into IT Hana, interact with artificial intelligence IT Sophia, and promote our activities and information.

2.2 Observe learning progress

All Screens is designed to allow teachers to "close seven at a glance" and view all of their students' computer screens. Our team builds a cloud platform through Amazon Web Services (AWS) server architecture and uses AR and 3D animation to design "virtual assistant teacher Ma". Through a system where teachers ask students to share screens at all times, learn about students' computer activity, and through their pre-existing webcam setups, AI technology is used to detect the range of head movements to determine if students are distracted or suspected of cheating.

2.3 Education

When students turn on their webcams, artificial intelligence technology is used in the background to recognize students' faces and detect their eyes, head movements and computer screens. Once suspicious, it will issue a warning or notify the teacher, increasing the difficulty of cheating in disguised form and acting as a deterrent.

3. Methodology

3.1 IT Hana & IT Sophia

The team used Unreal Game Engine from Epic Inc. and MetaHuman to create unique realistic human models for the Virtual Idos in Phase I. The motion capture system vividly recreates the real-time facial expressions which enables IT Hana to show movements and expressions in virtual environments. In Phase I, the motions of the virtual idols are pre-recorded and captured with a motion capture system. The limitation with the previous method is that the virtual idol will have limited variations in the motions and facial expression based on the limited number of captured motions.

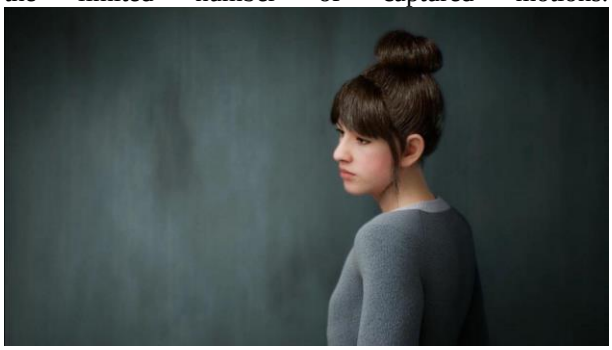

Figure 2 motion capture movement of Metahuman IT Hana

To improve the motion contents to IT Hana and IT Sophia, Procedural Animation will be tested and further implemented as the Live Animation System. Generally speaking, a procedural animation is a type of computer animation, used to automatically generate animation in real-time to allow for actions, which then could otherwise be created using pre-defined animations. Procedural animation is used to simulate particle systems, such as rain, fire and fog, cloth and clothing, rigid body dynamics, hair and fur dynamics, as well as character animation. In video games, it is often used for simple or complex actions such as turning a character's head whenever a player looks around. These techniques are programmed to have Newtonian physics acting upon them, therefore they are very realistic effects that can be generated that would be pretty hard to recreate with traditional animation. More complex examples of procedural animation are user-created creatures or 3D models, which will automatically be animated to all actions needed in the game from walking, to driving, to picking things up.

In the latest development phase, our team will design and implement simple procedural animation of body movement to the virtual idols, when the virtual idols are talking, greeting, laughing, etc. So these can improve the combination of body language of the virtual idols, as well as the realistic real-time response and interactions between the virtual idols and the users.

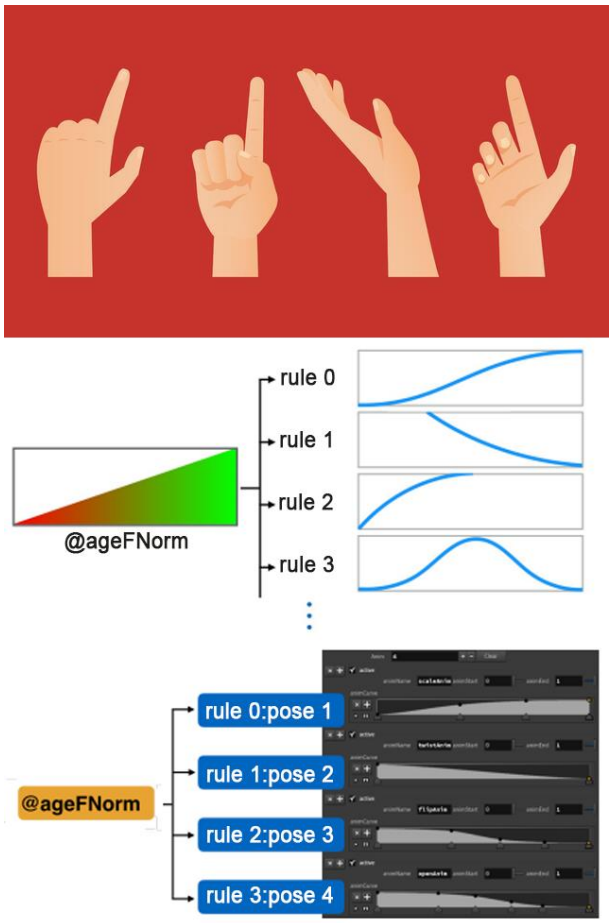

Figure 3 Norm sample of AI procedural animation learning of the AI model

3.2 Synthesise Personalized Speech

Synthesis is the voice content of Artificial Intelligence(AI) technology individuals' unique needs, preferences and characteristics. The technology uses algorithmic algorithms and natural language processing (NLP) technology to identify their age, gender, location, and even their tone of voice. Based on this analysis, a human-intelligence system can generate a personalised language that is more powerful and relevant to the individual.

Synthesised and Personalised voice and speech have been used in many commercial and industrial areas. For example, healthcare industry, digital entertainment industry and many more. It is used to create audio content that is tailored to customers' specific health conditions and provides tailored treatment plans. In education, personalised learning programs are designed and created for student to improve the student's proficiency level and learning style.

Collectively, the incorporation into a personalised language has the potential to revolutionise the way we create and consume audio content, providing individuals with greater appeal, interactivity, and relevance.

The project team will use Microsoft VALL-E to help improve the speech quality produced from text-to-speech

results of the virtual idol. So that it is expected to add "emotion" to the synthesised speech with ultra realistic AI generated human voices. The goal of implementing VALL-E to the virtual idol is aimed to create a more immersive and interactive language learning experience. By combining audio, visual, and language data, the next phase of virtual idol can be more engaging and effective.

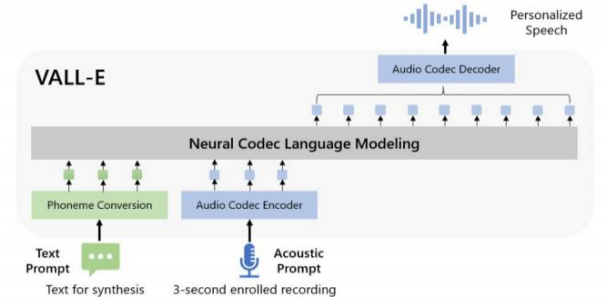

Figure 4 Model Overview of Microsoft VALL-E

3.3 AI Powered IT Hana & IT Sophia

The next generation IT Hana and IT Sophia will be an AI-powered virtual idol with a more realistic appearance, personality, and backstory. AI models will be developed for IT Hana and IT Sophia that can generate speech, facial expressions, and body movements for them. Further study on machine learning algorithms to analyse and interpret data, such as visual cues will be carried out within the project team. To collect data for training the new AI model. Audio recordings, videos, and images of mankind performers, as well as literature and text data will be collected to be used for generating speech and dialogue. Finally, the trained AI model will be integrated into the virtual environment, allowing the virtual idols to respond to user input in real-time.

Figure 5 AI Powered Smiling Facial Expression Testing of Metahuman IT Sophia

4. Result

The AI-powered Virtual Idols, namely, IT Hana and IT Sophia, have been enhanced in terms of appearance, facial expression, efficiency and body movement. They have been used in the departmental information day to help promote the courses provided by the institute to the public, fun-interaction with the guests, virtual tour-guide to the visitors, and animated greetings to welcome the honourable VIPs.

Besides, the researched AI algorithms and latest technology have been included in the course teaching materials for the I.T. in an attempt to equip the students with the most up-to-date professional knowledge and skills of AI and virtual idol. Synthesised voice technology has been applied in other digital game and VR App development projects.

5. Conclusion

Creating an AI-powered virtual idol is a complex process that requires a team of experts in AI, graphic and animation design, and programmers. However, with the right resources and expertise, it is possible to create a virtual idol that can engage with users in new and exciting ways. In the coming future, there will be more interactive ways between the teachers, students as well as the audience's virtual idols are also worth looking forward to.

Our team will continue to bring more learning opportunities to students by optimising the virtual idols, including tailoring various costumes to make the appearance of virtual idols more varied; and allowing more interaction between virtual idol and human users, and communication between AI-powered virtual idols to explore if more impacts can be created through these developments.. In addition, the information technology discipline will also encourage students to create other virtual idols to broaden their application and enrich their learning experience.

Acknowledgements

This work was supported in part by the Higher Diploma in Games and Animation and Higher Diploma in Multimedia, VR and Interactive Technology programmes.

References

Google. (n.d.). Dialogflow documentation | google cloud. Google. Retrieved June 29, 2022, from https://cloud.google.com/dialogflow/docs

Collins, E. (2021, May 18). LAMDA: Our breakthrough conversation technology. Google. Retrieved June 29, 2022, from https://blog.google/technology/ai/lamda/

Artist. MUSIC NATION SMTOWN. (n.d.). Retrieved June 29, 2022, from https://www.smtown.com/artist/musician/10766

What is a vtuber? What is a VTuber - And How Can You Become One? (n.d.). Retrieved June 29, 2022, from https://sensoriumxr.com/articles/what-is-a-vtuber

Bokyung Kye1, Nara Han1, Eunji Kim1, Yeonjeong Park2 and Soyoung Jo1. (2021). Educational applications of metaverse: possibilities and limitations, *Journal of Educational Evaluation for Health Professions.*

Mauro Conti, Jenil Gathani and Pier Paolo Tricomi. (2022). Virtual Influencers in Online Social Media, *IEEE Communications Magazine*.

The Morning After: Microsoft's VALL-E AI can replicate a voice from a three-second sample. (n.d.). Retrieved March 9, 2023, from https://www.engadget.com/the-morning-after-microsofts-vall-e-ai-can-replicate-a-voice-from-a-three-second-sample-121605576.html

Byun, D. J. (2018). A conceptual framework for procedural animation (CFPA) (pp. 1–16). *SIGGRAPH*.