



Print ISSN: 1738-3110 / Online ISSN 2093-7717
 JDS website: <http://www.jds.or.kr/>
<http://dx.doi.org/10.15722/jds.20.11.202211.121>

Development and Distribution of Deep Fake e-Learning Contents Videos Using Open-Source Tools

Won HO¹, Ho-Sung WOO², Dae-Hyun LEE³, Yong KIM⁴

Received: August 04, 2022. Revised: September 01, 2022. Accepted: November 05, 2022.

Abstract

Purpose: Artificial intelligence is widely used, particularly in the popular neural network theory called Deep learning. The improvement of computing speed and capability expedited the progress of Deep learning applications. The application of Deep learning in education has various effects and possibilities in creating and managing educational content and services that can replace human cognitive activity. Among Deep learning, Deep fake technology is used to combine and synchronize human faces with voices. This paper will show how to develop e-Learning content videos using those technologies and open-source tools. **Research design, data, and methodology:** This paper proposes 4 step development process, which is presented step by step on the Google Collab environment with source codes. This technology can produce various video styles. The advantage of this technology is that the characters of the video can be extended to any historical figures, celebrities, or even movie heroes producing immersive videos. **Results:** Prototypes for each case are also designed, developed, presented, and shared on YouTube for each specific case development. **Conclusions:** The method and process of creating e-learning video contents from the image, video, and audio files using Deep fake open-source technology was successfully implemented.

Keywords: e-Learning, Deep learning, Deep fake, Open-source, Distribution

JEL Classification Code: A29, H52, I20, I29

1. Introduction

The Fourth Industrial Revolution has brought innovative changes in human life due to the convergence of technologies developed in various fields. The development and utilization of artificial intelligence occupy a significant proportion of the Fourth Industrial Revolution. The development of artificial intelligence has received much attention with the advent of AlphaGo. The rapid development of Deep learning technology is not only due to

the rapid development of neural network theory but also the improvement of the computational power of computers.

The influence of open-source or open-service also contributed significantly to the diffusion of Deep learning technology. The codes of related research papers are made public, and a service such as Google Collab (<https://colab.research.google.com>) is relatively low in the cost for cloud resource access. People can run the source code directly on the cloud for less expenditure. These open resources allow the general public to access Deep learning

1 First Author. Professor, Department of Electrical, Electronics, and Control Engineering, Kongju National University, South Korea. Email: wonyho@gmail.com, ORCID: 0000-0002-7673-7039

2 Second Author. Assistant Professor, Department of e-learning, Korea National Open University, South Korea.

Email: hugwhoo@kno.u.ac.kr, ORCID: 0000-0001-8059-9524

3 Third Author. CEO, Intube Co., Ltd, South Korea.

Email: eogus0709@gmail.com ORC ID: 0000-0001-7829-5150

4 Corresponding Author. Professor, Department of e-learning, Korea National Open University, South Korea.

Email: dragonknou@gmail.com, ORC ID: 0000-0002-5305-2213

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

technologies in various fields easily.

E-learning content is a field that is suitable for Deep learning technology. Improving educational outcomes through improving educational methods is a public value everyone anticipates.

Initially, the term 'e-learning content' was used broader than 'e-learning video', but this paper mainly considers a video as a primary content medium. Various efforts and studies have been conducted to develop good e-learning videos. Kizilcec et al. studied how adding the teacher's face character to the e-learning video will affect the educational effectiveness. This paper explained that the inclusion of the instructor's face might yield different results depending on the situation and do it carefully according to the instructional design to avoid the learner's cognitive burden and concentration dispersion; the production of a video that includes the instructor's face costly for development.

On the other hand, we can confirm in Mio et al's paper that good scenario-based e-learning videos can increase the learning effect by increasing the learner's immersion. Therefore, when developing an e-learning video, it is very desirable if we can include the instructor's character appropriately and without cost burden in a range that does not cause cognitive dispersion of learners depending on the situation. It is practical to produce a video following the story's plot; artificial intelligence's Deep learning technology can create characters for this. In this paper, we intend to present various methods of developing e-learning videos using Deep fake technology that can produce various characters from voice audio information.

The open-source code is utilized and executed in Google's Collab environment. A video of the instructor's face is created by lip-syncing a picture or animation image of a character with audio information. Finally, it is added to the original e-learning video. The final video was created by editing and merging these videos. We used open-source tools such as Audacity and OpenShot Video editor in this series of processes. We also used Google's Collab Pro version. With these open-source tools and services, we can produce the final video at a low cost.

2. Literature Review

2.1. Deep Fake Technology

Among various fields of artificial intelligence technology, Deep learning has grown significantly. Deep learning is based on neural network theory. Among Deep learning technologies, a technology that synthesizes a human face or creates a fake face or voice is called a Deep fake. For example, creating a video of a celebrity giving a speech using Deep fake technology is possible. In this paper,

we will apply a Deep fake technology that creates a video of the instructor's face using their photo, an image obtained from the Internet, or a composite image. A different face can replace even a face in a video.

A technique for creating animation from audio information has been studied for a long time. Brand conducted a study in 1999 to extract changes in facial expression from a voice puppet show. On the one hand, a study to construct a dance motion animation from the rhythm and magnitude of audio information from music information was also studied by Shiratori et al. (2006). In addition, studies such as Ginosar et al. (2016). have been conducted to obtain information about the speaker's gestures and body movements from voice information.

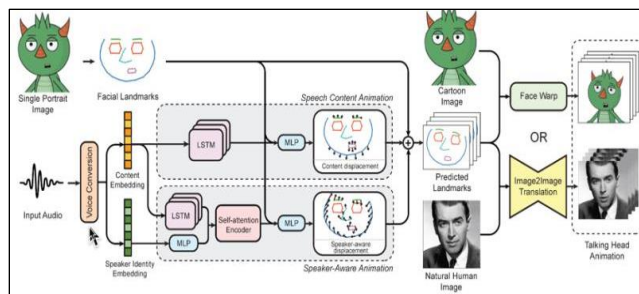


Figure 1: How MakeltTalk works (Zhou et al, 2020)

Regarding the method of extracting facial expression changes, many studies have been conducted by Suwajanakorn et al., Talyer et al., Karras et al., Zhou et al., Vougioukas et al., Chen et al., Xu et al., and Nießner et al. Most studies have been conducted using input information such as images, 3D face models, and structured facial skeletons. The scope of the research was extracting facial postures, recognizing voice characteristics personalized to the speaker, and restoring invisible face parts, depending on the source as either image or video information. This paper produced an e-learning lecture video using a Deep fake research paper called MakeltTalk as shown in <Figure 1> by Y. Zhou et al. and an open-source code.

In addition, various characters can be created using face morphing technology to obtain images of various speakers as shown in <Figure 2>.



Figure 2: Face-Morphing Example

2.2. Development of eLearning Video and Open-Source Tools

The development of e-learning videos generally proceeds in stages of analysis, design, development, execution, and evaluation according to the ADDIE model. In the video design stage, there are more detailed stages such as design outline preparation, content design, teaching/learning strategy design, learning flow chart preparation, and storyboard creation. Each development process is recommended to reflect this paper's technical video development method.

We used open-source tools such as OBS (Open Broadcast Software) shown in <Figure 6>, OpenShot Video Editor shown in <Figure 3>, Audacity, and Blender for the development. OpenShot Video Editor can be used on Windows, Mac, Linux, Chrome OS, etc. (<https://en.wikipedia.org/wiki/OpenShot>). It is easy to use, stable, and freely available. The project was started in 2008 and has been continuously improved since then. The development language was Python, PyQt5, C++, etc. Using the FFmpeg library supports various video formats supported by this library, and supports curve-based keyframe animation, track/channel-based video manipulation functions, and video clip manipulation functions. OpenShot Video editor can merge the original lecture video with a Deep fake generated video.



Figure 3: OpenShot Video Editor
<<https://www.openshot.org/>>

Screen recording is not supported in the OpenShot Video editor. Open Broadcast Software (OBS) can be used for screen recording, since OBS can record by overlaying various screens and cameras from the recording. Various screens can be configured and recorded in real-time, just like broadcasting equipment.

Audacity is also very useful as a voice editing open-

source software. Audacity can convert or extract different audio format files. Blender, an open-source 3D modeling software, can be used in conjunction with the OpenShot Video Editor. Blender is not only linked with OpenShot Video editor to obtain video special effects but can also be used to create high-quality animations independently. It is possible to develop highly immersive e-learning lecture videos using these various open-source tools.

2.3. Google Collab and Python

Google Collab is a cloud-based service provided by Google. Because we can use the functions of the Jupyter notebook in the cloud environment, it provides better computational power than an individual PC. Code that utilizes Deep learning technology that requiring heavy computation can be executed. It costs no fee for a general user. For the case of running computation-intensive code, the Google Collab pro version allows using more computing power on the cloud at a reasonable price.

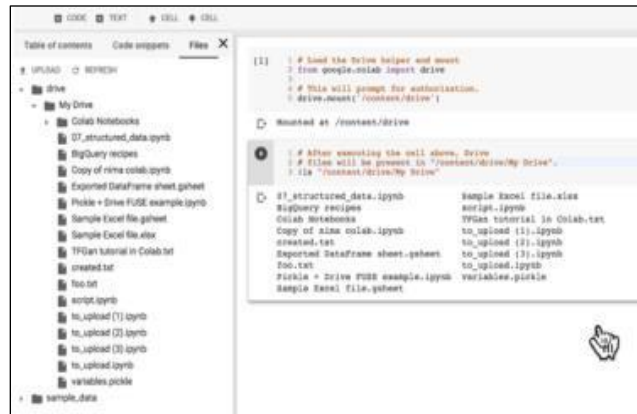


Figure 4: Google Collab
<<https://colab.research.google.com>>

Recently, researchers and programmers have published executable code using Google Collab as shown in <Figure 4>. The code can be directly copied to other Google Collab accounts and executed immediately. The source code of the papers used in this paper was also released as a Google Collab version by Y. Zhou et al.

Python provides Deep learning-related libraries in the form of a package and is one of the most widely used programming languages due to the fast program development and the ease of learning. The Jupyter notebook used by Google Collab also supports Python and is easy to use in a web-based cloud environment.

3. Methodology

3.1. Development Process

We developed an e-learning lecture video using Deep fake technology using various open-source tools and codes described in the previous section. The final e-learning lecture video is generated from the original lecture video, extracting the audio file from the original lecture video, creating a video of the instructor's face matched to the voice (we call it a talking head video), and merging it with the original lecture video. This procedure is shown in <Figure 5>, and each detailed procedure is explained as follows.

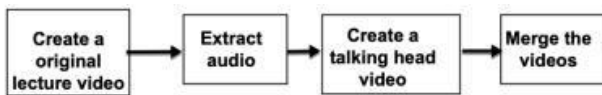


Figure 5: Development process

3.2. Creating Original Lecture Video

The original lecture video is recorded as the first step in development. Developers can develop the video in various ways, but they must decide whether to record the instructor's face. After completing the video, the developer can extract the voice and add the lip-synced instructor's face in the next step, so they do not have to record the instructor's face in this step. However, for the case of dubbing another face to the instructor's face in a video, the video needs to include the instructor's face. In general, non-face recording reduces the burden for the instructor and allows the instructor to focus more on the voice and lecture content.

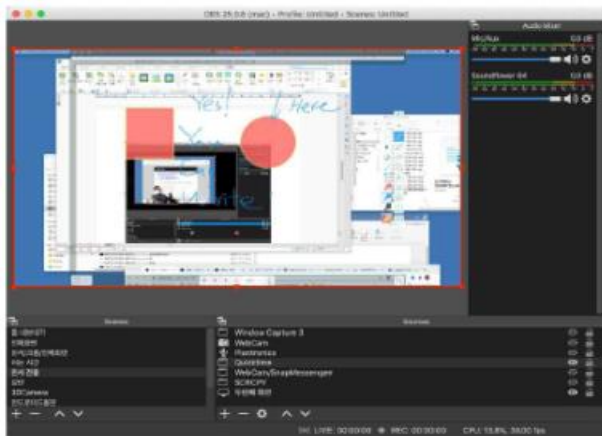


Figure 6: Broadcast Software

As in Ho et al.'s paper, using OBS or Blender to include a lot of various animations and audiovisual elements is encouraged. Producing videos with those various tools is desirable for good eLearning content creation. The sample output of Blender animation is shown in <Figure 7>.

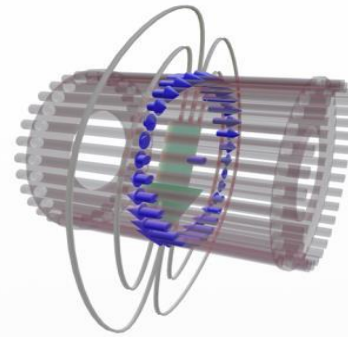


Figure 7: Development using Blender animation

3.3. Create Talking Head Video

Audio is extracted from the video created in the previous step to produce a video using Deep fake technology. Using Audacity, export the file in wav audio format from the video, upload this audio file, and run the MakeItTalk code in Collab to generate the Talking head video.

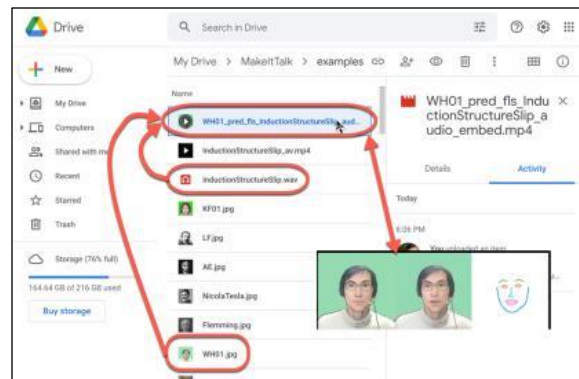


Figure 8: Run MakeItTalk

<Figure 8> show that a talking head video is obtained if you upload the voice file and the speaker's face image to the Google driver and execute the MakeItTalk code in Google Collab. Before executing the code, the speaker's face image is named as shown in <Figure 9>. Audio files do not need to be specified in the code because it automatically finds all wav files and creates videos for each wav file.

```

- Step 1: Basic setup for the animation

[ ] default_head_name = 'WH01' # the image name (with no .jpg) to animate
  ADD_NAIVE_EYE = True # whether add naive eye blink
  CLOSE_INPUT_FACE_MOUTH = False # if your image has an opened mouth, put this
  AMP_LIP_SHAPE_X = 1. # amplify the lip motion in horizontal direction original
  AMP_LIP_SHAPE_Y = 2. # amplify the lip motion in vertical direction original is
  AMP_HEAD_POSE_MOTION = 0.7 # amplify the head pose motion (usually smaller

Default hyper-parameters for the model.
    
```

Figure 9: Configuring MakeItTalk

3.4. Merging the Videos

In this stage, the instructor's video created with MakeItTalk needs to be merged with the original video. They can be merged better by making the talking head's image background transparent using the chromakey function. It is better to use the instructor's image with green background for this purpose. The merging can be executed using the chromakey function provided by the OpenShot Video editor. The final video is produced after cropping the talking head video and applying chromakey is shown in <Figure 10>.

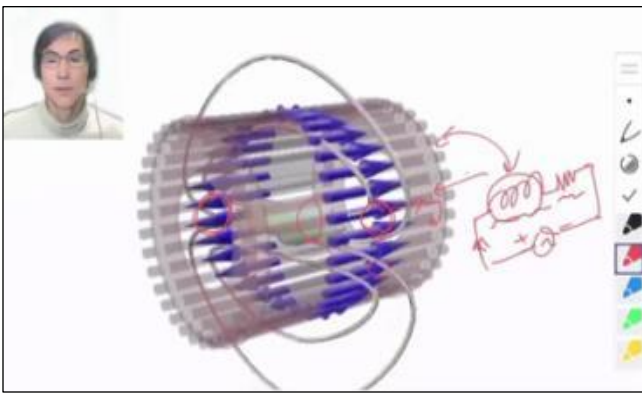


Figure 10: Final lecture Video Merged

4. Results and Discussion

According to the methodology presented in this paper, you can develop various types of e-learning lecture videos. It is also feasible to implement various scenario-based video development using Deep fake technology. The types of specific development cases are as follows.

4.1. Using Special Characters

Highly immersive videos can be created using famous experts or historical figures as characters in that fields. For example, a video explaining the principle of electricity or an electric motor featuring characters like Edison or Nikola Tesla can be created. The image can be searched with a Creative Commons license on Google. The output video for Nikola Tesla explaining the electrical machine is shown in <Figure 11>.

Images of famous actors can be used, but better be careful as Deep fake can cause legal issues. <Figure 11> shows the face morphing technique can synthesize the faces of several actresses and use the character as co-instructors.



Figure 11: Face-Morphed Actress

4.2. Conversational Video

Compared to a video in which one instructor conducts a lecture, a video with two or more characters in a conversational format can draw the learner's attention. <Figure 12> shows the type of video that a character asks a question to another character (primarily the instructor character) to introduce the lecture objectives in a question-and-answer manner at the beginning of the video.

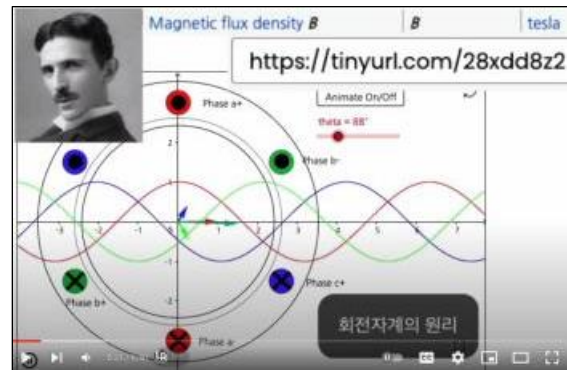


Figure 12: Nikola Tesla explaining Machine

4.3. Script-based Video Creation

Deep fake technology can imitate and generate voices, not just faces. There are already various services and open-source codes for generating voices from scripts. To utilize this, one can write a script, create a voice with TTS (Text To Speech) technology, create a talking head video with the instructor's image, and merge this video to produce a final output video. The script can be translated with Google Translator in various languages, and one can get the result with perfect dubbing of the fluent native-speaking voice.

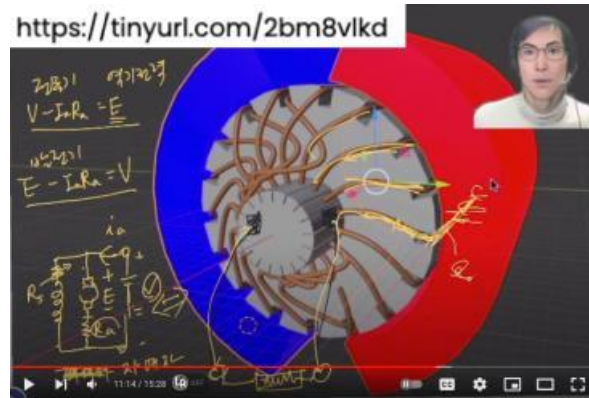


Figure 13: Various production tech.

4.4. Utilization of Various Production Techniques

What has been explained so far is how to combine the original lecture video with a talking head video. Since only the instructor's face is included, there is a limit to increasing the learner's attention, so it is desirable to provide an environment where the original lecture videos are developed using various technologies. <Figure 13> is the result of producing the original lecture video in a way that can be written on the screen while showing the structure of the 3D model using Blender. The use of Blender was effective because the electrical machine is a subject that explains the 3D structure. Still, it is necessary to use various service technologies related to each video for different subjects—for example, Google Earth for geography and history. Virtual lab tools for science are probably recommended. Recording original lecture videos in the metaverse is also a good idea.

5. Conclusions

5.1. Result Summary

In this paper, methods for producing various e-learning lecture videos using Deep fake technology and practical development guidelines are presented as follows.

- One can merge the instructor's talking head video to the original lecture video using audio information applying the Deep fake technology.
- One can include multiple instructors in the video in an interactive interview or situational play format.
- One can introduce celebrities or historical characters that connect with the video's subject.
- Use various open-source tools (Blender, OpenShot Video Editor, OBS, Audacity) or open-services (Google Earth, Google Maps, OER) as needed.

- One can create a video based on a script; at this time, it is possible to produce a video in multiple languages through script translation.

5.2. Contribution and Implication

We share the video results on YouTube and provide information through personal wikis and blog sites.

<<https://tinyurl.com/y9k6abl7>>

5.3. Limitation

This study focused on facial expression generation using Deep fake technology. Video development with body movements and gestures will provide better natural integration than only talking head video integration.

References

- Brand, M. (1999). Voice puppetry. *In Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 21-28.
- Chen, L., Maddox, R. K., Duan, Z., & Xu, C. (2019). Hierarchical cross-modal talking face generation with dynamic pixel-wise loss. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7832-7841.
- Eri, W. I., & Susiana., (2019). Using the ADDIE model to develop learning material for actuarial mathematics. *Journal of Physics: Conference Series*, 1188. DOI:10.1088/1742-6596/1188/1/012052.
- Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., & Malik, J. (2019). Learning individual styles of conversational gesture. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3497-3506.
- Karras, T, Aila, T. Laine, S., Herva, A. & Lehtinen, J. (2017). Audiodriven facial animation by joint end-to-end learning of pose and emotion. *ACM Transactions on Graphics (ToG)*. 36(4), 94:1-94:12.
- Kim, Y. (2017). A Study on e-learning Contents Opening Information for Distribution Industry Labor Competence, *Journal of Distribution Science*, 15(8), 65-73. DOI:10.15722/jds.15.8. 201708.65.
- Kim, Y. (2018). A Design of Human Cloud Platform Framework for Human Resources Distribution of e-Learning Instructional Designer, *Journal of Distribution Science*, 16(7), 67-75. doi:10.15722/jds.16.7.201807. 67.
- Kizilcec, R., Bailenson, J., & Gomez, C. (2015). The Instructor's Face in Video Instruction: Evidence From Two Large-Scale Field Studies. *Journal of Educational Psychology*, 107(3), 724-739. <https://doi.org/10.1037/edu0000013>.
- Lee, D. H., Kim, Y., & You, Y. Y. (2018). Learning window design and implementation based on Moodle-Based interactive learning activities, *Indian Journal of Public Health Research and Development*, 9(8), 626-632. DOI:10.5958/0976-5506.2018.00803.3.
- Mio, C., Ventura-Medina, E. & João, E. (2019). Scenario-based

eLearning to promote active learning in large cohorts: Students' perspective. *Computer Applications in Engineering Education*. 27(4), 894-909. 10.1002/cae.22123.

Shiratori, T., Nakazawa, A., & Ikeuchi, K. (2006). Dancing-to-music character animation. *In Computer Graphics Forum*, 25(3), 449-458. doi:10.1111/j.1467-8659.2006.00964.x.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing obama: learning lip sync from audio. *ACM Transactions on Graphics (ToG)*, 36(4), 95:1-95:13.

Taylor, S., Kim, T., Yue, Y., Mahler, M., Krahe, J., Rodriguez, A. G., ... & Matthews, I. (2017). A deep learning approach for generalized speech animation. *ACM Transactions on Graphics (ToG)*, 36(4), 93:1-93:11.

Thies, J., Elgharib, M., Tewari, A., Theobalt, C., & Nießner, M. (2020). Neural voice puppetry: Audio-driven facial reenactment. *In European conference on computer vision*, 716-

731. Springer, Cham.

Vougioukas, K., Petridis, S., & Pantic, M. (2020). Realistic Speech-Driven Facial Animation with GANs, *International Journal of Computer Vision*, 128, 1398–1413.

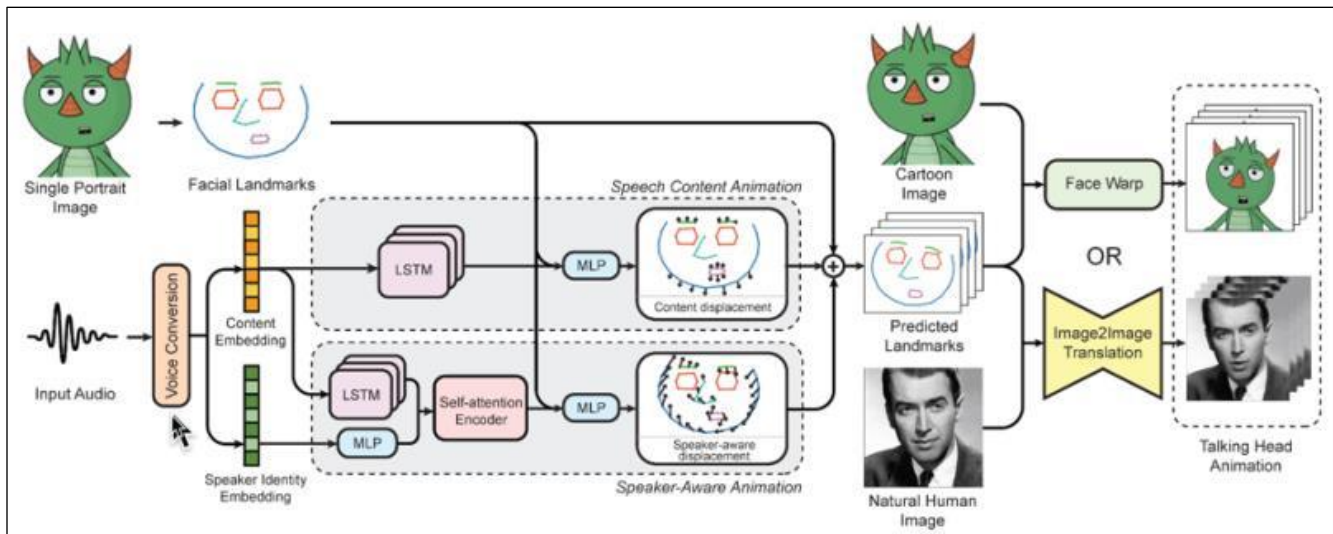
W, H., Lee, D. H., & Kim, Y. (2021). Implementation of an Integrated Online Class Model using Open-Source Technology and SNS, *International Journal on Informatics Visualization*, 5(3), 218-223. 10.30630/joiv.5.3.668.

Zhou, Y., Han, X., Shechtman, E., Echevarria, J., Kalogerakis, E., & Li, D. (2020). Makelttalk: speaker-aware talking-head animation. *ACM Transactions on Graphics (TOG)*, 39(6), 1-15.

Zhou, Y., Xu, Z., Landreth, C., Kalogerakis, E., Maji, S., & Singh, K. (2018). Visemenet: Audio-driven animator-centric speech animation. *ACM Transactions on Graphics (TOG)*, 37(4), 1:1-1:10.

Appendixes

Appendix 1: How MakeltTalk works



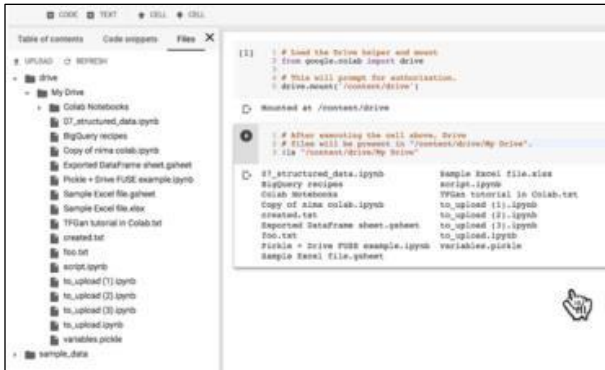
Appendix 2: Face-Morphing example



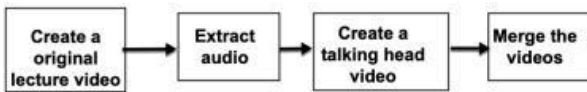
Appendix 3: OpenShot Video Editor



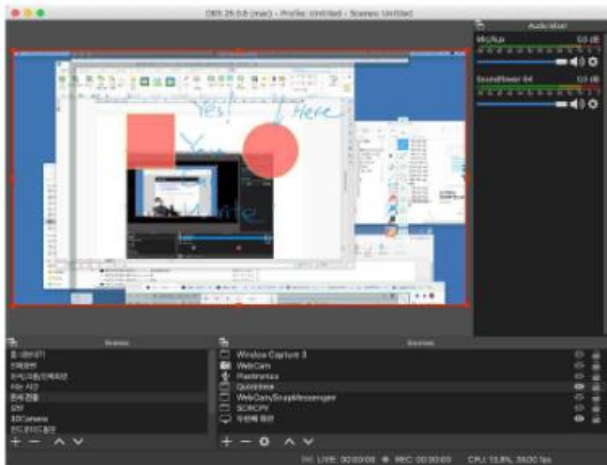
Appendix 4: Google Collab



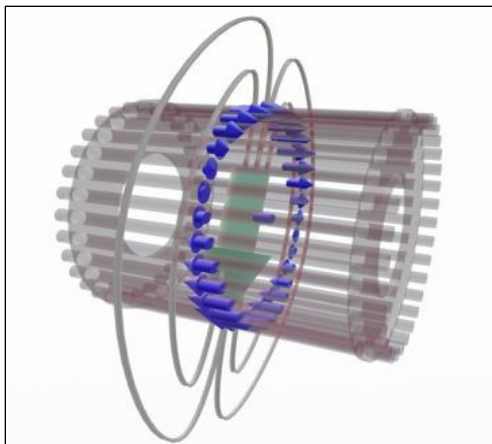
Appendix 5: Development process



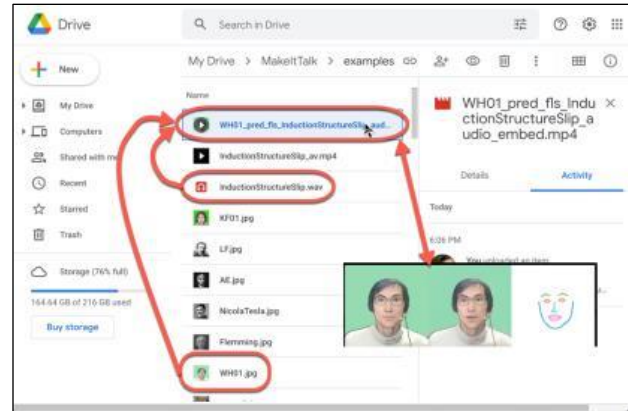
Appendix 6: Open Broadcast Software



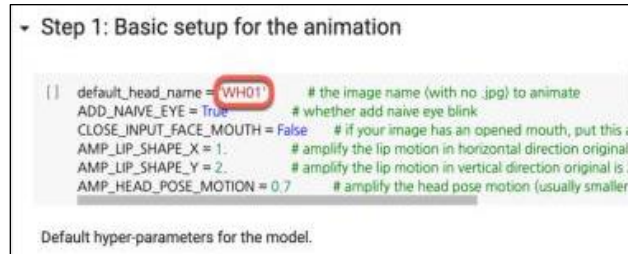
Appendix 7: Development using Blender animation



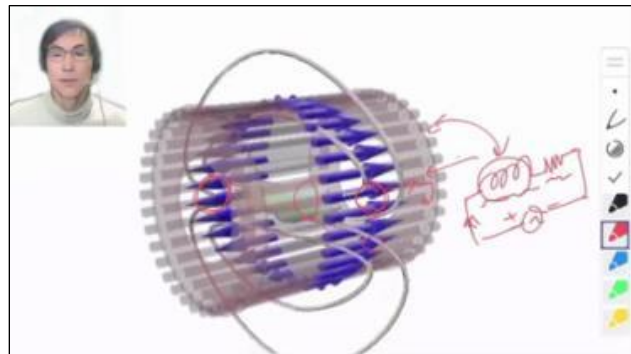
Appendix 8: Run Makeltalk



Appendix 9: Configuring Makeltalk



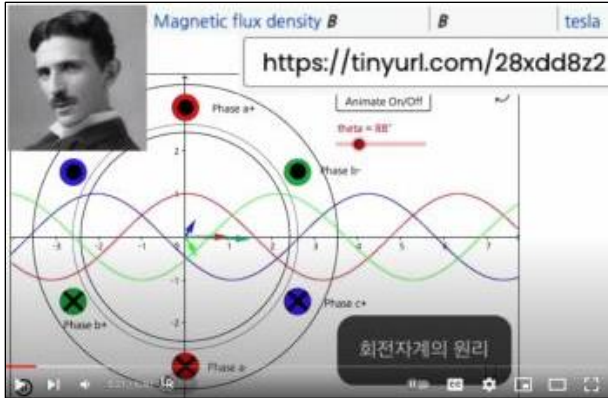
Appendix 10: Final lecture video merged



Appendix 11: Face-Morphed actress



Appendix 12: Nicola Tesla explaining machine



Appendix 13: Various production tech.

