Reflections on the Ayatori project residency by Matthew Yee-King

In the Ayatori project, I was an AI specialist and musician. I became involved in the 'post-lockdown' phase of the project in its early planning stages, wherein we developed the idea of an interactive audiovisual space enhanced with AI technology. The initial idea was to work with the existing Ayatori audiovisual material and develop it into an audiovisual map using AI technology such as image and audio analysis. We would create a numerical latent space representing the space of audio and visual materials, and it would be possible to interactively move in that space, either by providing an image input, an audio input or by interacting directly with the map. These maps would allow for different forms of interaction, including audience and performer interaction.

Before the residency, I worked on the underlying technical infrastructure to make the maps possible. I developed the audio and image analysis layers using the Python language and various pre-trained neural networks, which could carry process images and audio fragments into map coordinates. I developed a prototypical visual presentation layer in the Javascript language, which meant it could be run on any machine and interact with the AI back end. Using web technology for the display layer would support the plans for multi-projector setup and provide for different types of interfaces.

I brought a raw prototype of this system along at the start of the residency. It was an exciting learning experience attempting to integrate this system in a creative workflow sense with what had been done before. Yuriko and Ed are used to a high level of control over their material, and the idea of an AI system analysing their work and presenting it in different ways was very new to them. Working with Simon, I integrated the AI system with an audiovisual performance system he worked on with Yuriko in the Touch Designer environment. This integrated system allowed a live audio feed to be fed to the AI, then for the AI to map the audio signal into the image 'latent space' and to select an image. It could then inform the Touch Designer system which image to select, and this would then be made available to Yuriko, who could use a MIDI control surface to manipulate the image. We could use a basic version of this system at the first performance at the end of the first week. I also developed a concatenative sound synthesis system that would reconstruct a live audio feed from pre-recorded audio using an AI-based audio distance metric to select audio fragments. I used this system at the first performance.

Reflecting on the first performance, I felt the map-based audiovisual AI system needed to be more tangible and controllable. It needed to be much more apparent what the AI system was doing. Also, during the first week, Yuriko and Ed had created a batch of new material and were keen to see what could be done. So, in the second week, I took a different approach. Rather than analysing and exploring existing material, I moved to a generative AI system. Generative AI is much more familiar to people as it can output material in the style of some material it has been trained on. I was initially not keen on working with generative AI as the Ayatori project had a particular audiovisual aesthetic. Constraining generative AI to work within that aesthetic and to output at an appropriate quality would be difficult to achieve in the time available. This is especially true if you wish to create an integrated audiovisual model. Also, the idea of generating 'in the style of' had been rejected during an earlier project meeting, with a stronger preference expressed for the concept of the map.

One problem with generative AI is that existing models, such as the well-known image generators, are trained on a large amount of data from various sources. You can then only really generate new images that are some sort of combination of the images it had already seen. Training an image generator from scratch normally requires a large amount of data (e.g. thousands of millions of images). I was not convinced that a pre-trained image generator such as stable diffusion would generate Ayatori-appropriate images. At the end of the first week, I did some experiments wherein I generated images using text prompts that described some of the Ayatori images. I could even automatically generate descriptions of the Ayatori images using my existing image analysis code, but they needed to be richer to generate images in the correct aesthetic, e.g., 'Mountains with clouds and lots of trees in front of them'. Next, I decided to experiment with training a network from scratch. I used the lightGAN model from LucidRain, designed to train quite fast. I fed it stills extracted from the new videos produced in the first week of the residency. As I expected, the trained model could not produce very high-quality images (not high resolution, not much detail), but the images were undoubtedly intriguing. I adapted the neural network Python code to generate a large set of videos showing what the neural network had learnt ('exploring latent space') and how the latent space changed as the neural network was learning.

When I presented the new videos to Yuriko, I found that these two concepts of learning and exploring were much more apparent than the mapping idea from the first week. We looked at all the videos I had produced - various animated grids of images and selected the most effective ones. We then loaded the selected videos into the Touch Designer system. Yuriko could interactively fade and resize the videos in different ways, fading between the AI-generated and original videos. Training

all these new models, generating the videos and then integrating them with the Touch Designer took up most of the second week. Towards the end of the second week, I switched my attention to the audio side of things. Unfortunately, there was not enough time to do the same process with audio-generating models. Still, I put together a machine listening system that would respond and re-process a live audio feed based on event triggers. We used this system to process the live drums and sax in the final presentation performance. I also revisited the original image maps, generating the maps for all the original Ayayori material. We also presented and discussed these maps after the live performance.

In conclusion, we created an exciting audiovisual space using AI techniques in combination with various other technical and creative methods. An interesting aspect for me was the `science communication' part, wherein I had to work with Ed and Yuriko towards an understanding of what the AI systems were doing and what they were capable of. I found that the generative AI systems had a much clearer, easier-to-understand capability than more subtle techniques such as mapping and analysis. I am looking forward to the next developments in the Ayatori project.