

Animator Request for Speaking Portal B

This is a document pertaining to the specific folders of images we will require to make a seamless, professional lip-synced animation for Kukarella. Enclosed are 2 sample folders of images, the mouth folder, and the poses folder. We will need each folder recreated in the preferred animation style of Kukarella. The mouth folder consists of 22 separate mouth styles, each corresponding to one of the viseme's for the English language mentioned here:

<https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-speech-synthesis-viseme?tabs=visemeid&pivots=programming-language-csharp>

By creating an image that corresponds to each viseme, we can compile these to simulate lip syncing. The most important feature here is that each viseme frame looks like the corresponding sound is being spoken, as this is what accounts for all the realism. In the case that the provided images result in a 'choppy' transition, we have the option to have a transition frame. For example, the viseme for the sound 'ahhhh' has an open mouth. But for the sound 'mmmmm' the mouth would be completely closed. A frame in between showing a mouth halfway closed or appearing to be on its way to being closed would result in a smoother animation if need be. As of current, our animation only uses 22 frames for visemes, and this appears to function well. If additional smoothness was requested in the animation, we would need transition frames to add more fluid movement. For now, we will stick with the request of just one image for each of the 22 visemes. (One mouth .png for each viseme.) this concludes the requirement for the mouth folder of images.

As for the poses folder, there are 15 separate poses in which the stickman will cycle through when the script detects a pause or break in sentences. The pose consists of every part of the avatar, minus the mouth. 2 of the poses are simply one blinking model, and one not. This is solely to add realism and movement to the animation. These are not required to be specific poses, but the animator can use their best judgement on what to create. Additionally, there are another 65 poses that are used to convey emotion if the script requests it. This is an additional feature, and at the time we will only request the 15 standard poses

(pose0001.png - pose0015.png) until we find a way or decide to implement an emotion selection option on our API.

Additionally, should Kukarella want to implement more than one avatar, we thought it might be a good idea to have an additional folder of mouths and poses with a slight change made, such as the mouths having different colored lips, or the poses having different styled/colored hair or different shaped/colored eyes. This saves the animator effort as only small changes need to be made but creates a large library of avatar options for users.

Finally, the way the mouth is overlayed on the posed image is through a .csv file of coordinates for the mouth. It looks at both the pose and mouth image, and determines the center of the mouth used, and the center of where the mouth should be on the pose. We are hoping not to make too many changes to how the .csv coordinates are created, so we are hoping to keep the mouths and poses images, generally the same size, and appendages in generally the same location. If the place for the mouth is generally in the same area on each pose, and the mouths are relatively the same size as the sample frames provided, there should not be too much we need to change coordinate-wise to simply swap out our test images for Kukarella's final copy and generate a video!

Please reach out directly to us through the capstone discord if you have any questions or concerns with the request! Thank you.