Natural Language-Driven AI Avatar Motion Detailed Project Plan

1 October 2023

Project Supervisor:
KOMURA, Taku

Project Team:
3035767724 ASJAD, Marzukh Akib
3035778175 LAU, Ho Yu
3035777975 OLANO, James Daniel Cubillas

Background

In order to create realistic human-like animations for games, 3D animations, etc, from a simple natural language prompt input. With our project, the user can just input a list of commands to generate their desired animated 3D motion. At our current stage, the final end point will be a webapp showcasing the 3D motion rendered as a video output.

Currently, 3D avatar movements are created through rotoscoping or motion capturing. Moreover, there are softwares/tools available that can extract motion data from 2D videos. EasyMocap and others alike are able to do so but not generate custom motions through natural language inputs.

The motivation behind the project is to make 3D avatar animations accessible and cost-efficient for every independent artist. By providing a ChatGPT-like experience to creating animations for 3D avatars, this eliminates the need for expensive and time consuming methods.

This project will benefit the virtual reality, gaming and digital entertainment sectors, as users will no longer need to learn complex codes or learn character animation to let 3D avatars perform a variety of movements. Instead, they will be able to intuitively command 3D avatars with natural language, even without knowledge in programming languages or animating 3D characters. Thus, people who are creative with no technical or artistic knowledge are able to generate animations for their desired medium.

Objective

The project aims to implement a solution where natural language processing can be used to generate custom motions of varying length for 3D avatars with the help of a natural language input.

Methodology

The project will be implemented through integrating motion data into a 3D avatar. The skeletal structure and the kinematics of the avatar will be fine-tuned, in order to ensure realistic motion.

After the motion data has been successfully integrated with the 3D avatar, a Natural Language Processing (NLP) system will be implemented and trained with movement-related commands through supervised and reinforcement learning methods, so that the system will be able respond to user input and perform a wide range of movements. We will also attempt to enable character-object interaction as well during the rendering of the video.

In order for the user to interact with the system, the front-end will be built on React for the web application and React Native for mobile application. Moreover, MongoDB will be picked as the database for both the web app and mobile app. As for the rendering of the 3D avatars/animations, personal laptops will be used for hosting the server. The back-end logic will be created using Python as the scripting for Blender is in Python as well.

Project Schedule and Milestones

The project will be divided into a total of five stages.

Initially, the **first** stage of the project will involve the collection of datasets for motion data, as well as filtering and breaking down long animations taken from databases. All tasks related to pre-processing the data, testing the data and preparing the data in a format that can be integrated with the 3D character will be done in this phase.

Subsequently, in the **second** stage of the project, we will create the database and create the logic for querying the database to retrieve the motion data. In order to store the motion datasets, MongoDB will be incorporated into the back-end so as to ensure fast read and write operations allowing for a seamless user experience at the end of this phase. It is to be noted that the

motion data will be stored with their corresponding annotated information which will be the crucial factor during the query part by the GPT API in the fourth stage.

For the **third** stage of the project, the final endpoint of the system, which is our software, will be created. For this stage, we plan to develop a web app with React which will enable the user to input the prompt and see the final rendered video in a user-friendly manner. The initial UI/UX designs will be first done on Figma then recreated over to React and React Native, which will be used to build the frontend of the web application and the mobile application respectively (if time does not permit, then we will only develop the web app). At this stage, we also plan to integrate the database with the front end of the software.

For the **fourth** stage of the project, we plan to integrate the ChatGPT API into our database querying system. The prompt from the front end by the user will be delivered to our local host into the computer. Afterwards, the API will then scan the database in order to retrieve the relevant motion data from the database. This stage also includes the training of the API in order to successfully retrieve the correct and relevant motion data corresponding to the user prompt.

Ultimately, for the **fifth** and final stage of the project, the rendering of 3D animations will be done on the personal laptop/computer which will have the blender software necessary for the rendering of the video. This phase will be broken down into two sub parts, subpart A and subpart B. Subpart A will cover the scripting and creating the algorithms for which we will incorporate the logic for proper character-object interaction. This will be done in Python as mentioned since Python is the main programming language used in scripting for Blender. Subpart B will incorporate the automation of the final video rendering and then uploading the video to a cloud storage from where the video will be downloaded into the front end software and then displayed to the user.

Allocated Learning Hours

In total, roughly 240 learning hours will be spent from 2 October to the end of April.

In the first semester, the first and second stage will take up to 104 hours in total. 48 hours will be allocated to the first phase, while 24 hours will be allocated to the second phase. 32 hours will be allocated to the third phase before the end of 2023.

For the second semester, the total learning hours for the remaining stages will be 136 hours, of which 24 hours will be allocated to the remaining parts of the third phase. 36 hours will be allocated to the fourth phase, and 72 hours will be allocated to the final phase.