FUNREADING: A GAME-BASED READING ANIMATION GENERATION FRAMEWORK TO ENGAGE KIDS READING USING AI AND COMPUTER GRAPHICS TECHNIQUES (FOR SPECIAL NEEDS)

Jiayi Zhang[1], Jiayu Zhang[1], Justin Wang[2] and Yu Sun[3]

[1]Northwood High School, 4515 Portola Pkwy, Irvine, CA 92620
[2]The Peddie School, 201 S Main St. Hightstown, NJ 08520
[3]California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

Children with ASD or ADHD are having a hard time learning and understanding, and there's no perfect education system [1][2]. However, audios and animations can improve their reading effectiveness. This paper designs an application to have animated characters talking with audio based on the text using Optical Character Recognition, text to speech, and Natural Language Processing.

KEYWORDS

AI, Computer Graphics Techniques, Machine Learning.

1. INTRODUCTION

Children with ASD/ADHD are having a hard time learning and understanding the text. In the educational industry, there are usually two traditional teaching methods specifically targeting kids with ADHD or ASD [3]. The first method is establishing specific schools for students with ADHD or ASD [4]. This method also has a flurry of disadvantages. First, it is costly. Establishing a school requires millions of dollars. Second, it cannot help kids with ASD or ADHD learn how to communicate with other people since those kids are confined inside these special schools and cannot interact with others [5]. In conclusion, the general existing offerings or treatments for ADHD or ASD are ineffective, inefficient, and financially burdensome. The second method is video games. These attract kids' attention and were thought to effectively combat ADHD. Teachers integrate knowledge inside games letting students learn through video games. However, this teaching method has drawbacks. First, children usually only focus on the game itself rather than the knowledge behind the game. Second, video games have a proven addictive nature in their development. Thus, video games as a treatment are low in efficiency and not long-term feasible in full-time education. This paper develops an application to automatically convert traditional storybooks to interactive videos to help educate and treat children with ADHD or ASD [6].

David C. Wyld et al. (Eds): SIPP, NLPCL, BIGML, SOEN, AISC, NCWMC, CCSIT - 2022
pp. 113-120, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121309

EndeavorRx is the first and only clinically proven prescription video game that participates as a treatment for ADHD children [7]. The game is designed to improve attention control by strengthening focus, interference processing, and multitasking through an immersive video game experience.

Studies show that video stories have vast potential for helping with the academic, social, and emotional education of children with ADHD and autism. Therefore, we developed a product that automatically converts traditional storybooks to interactive videos to help educate and treat children with ADHD or ASD.

Our treatment is more efficient and feasible than games or specialized schools. First, our solution is scalable. Video game programmers need lots of time to design the plot games [8]. Thus, the production of video games is generally slower, so it takes more time for customers to get benefits from the product. On the contrary, our treatment is converting stories into videos at the click of a button. Most crucially, the conversion of stories does not need time, because the plot was already designed by the story's author. Second, our solution is more specific and targeted because our product is designed to custom for each student. Every user has a unique background and story recommendation based on their interest. Third, our solution is more effective. We teach kids with ADHD or ASD knowledge in areas of math, science, and language arts while incorporating features engaging their communication functions through interaction functions. Kids can interact with AI or community members. Lastly, our proposal is a more accessible education method than traditional schools since we present less opportunity cost to parents, students, and government organizations. Our products are more scalable, effective, and affordable than the traditional methods of educating students with neurodivergence.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met while creating this application and the further development we think about and trying to achieve; Section 3 focuses on the details of our current solutions; Section 5 shows related work; Section 6 gives the conclusion remarks and the future development of the application.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Selecting Model (Apply SALSA)

Choosing a model is extremely important to the first impression and the willingness to use the application for users. These steps become challenging as not every model could cooperate with the SALSA Lipsync. Besides the animation and lip sync, the outfit of the character is also important for children to develop interest.

2.2. Using API flexibly

While building the app, it is important to have API apply and connect to the systems. In order to minimize the resources used in the app, we have to use the API flexibly and reduce space and resources.

2.3. Developing more interactions

We want to create a precise mental health system for the app and would like to cooperate with professional psychologists and psychiatric clinics to develop an algorithm that evaluates the user's condition to accomplish real-time tracking and optimization of content presentation, as well as the animation, and push media according to the user's preferences. For now, we are developing more interactions to allow users to receive more reactions. Making the app flexible to ask the right questions is also a risk, and the online video platform along with the group chat system we will build in the future is challenging.

3. Solution

Story Book is an application that provides users with an efficient, productive, and cost-effective product to automatically create animated videos from pictures of written material. The image is captured using the camera on our users' mobile devices. We implemented the google cloud vision service into the C# code on unity using API to complete our OCR phase [9]. To turn the text output from OCR into speech, we need to process a human voice simulation. We used IBM's online Watson text-to-speech service the same way we used the google cloud service to complete the text-to-speech function. We employed spaCy's pre-trained NLP model to extract keywords from the text such as places and organizations and organize them in a text file. Up to this point, all the information we need is ready, so the application starts to create the video. We need background images and a character for the videos. We used Ultimate Stylized Business Women to create a 3D figure. This 3D model allows us to control the muscles of the character and thus control their lip movements and facial expressions, which are both essential for the character to look natural during the video [10]. We also utilized the SALSA LipSync Suite asset to automatically match the lip movements of the character with the audio file. We have a library of pictures of different topics that can act as backgrounds. Using the keywords that we detected with NLP, we find pictures in that category from our library and set these pictures as interchangeable backgrounds for the video. Finally, after all these processes, a full video with the same context and meaning as in the written material is created.

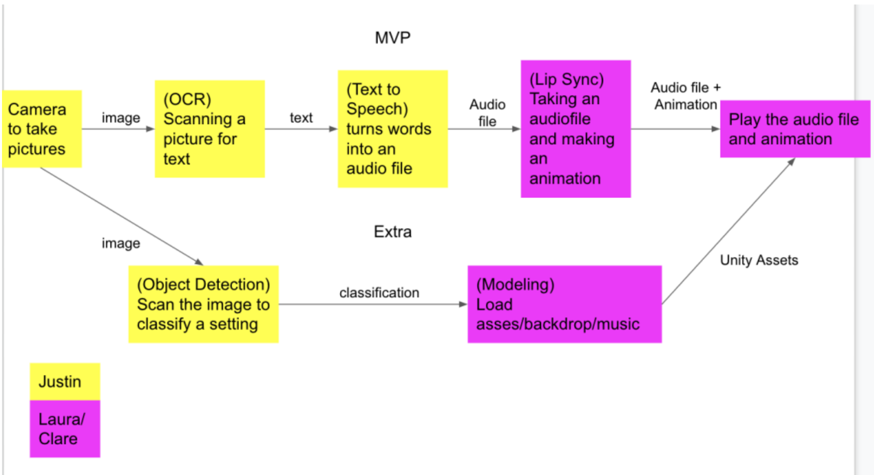

Figure 1. Overview of the solution

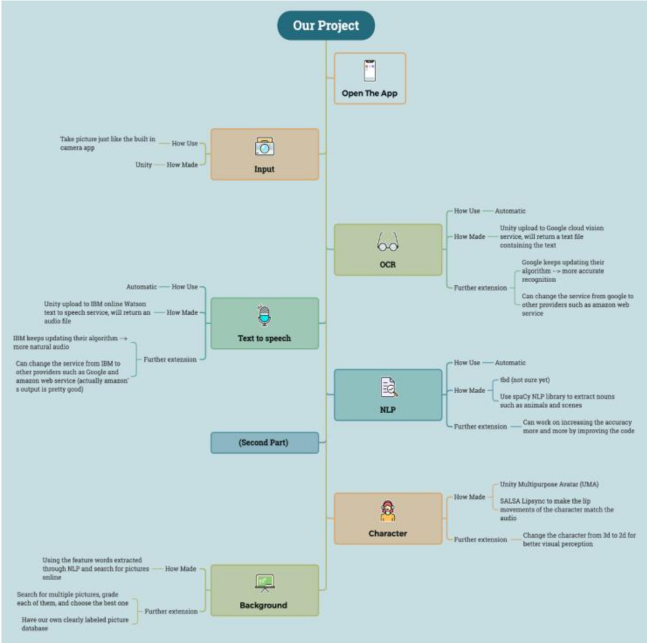

Figure 2. Project structure

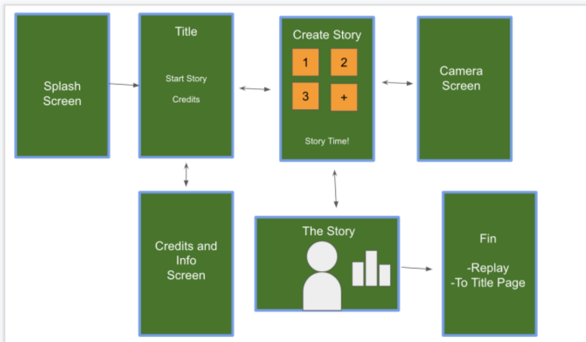

Figure 3. Overview of the APP

Figure 4. Screenshot of the APP

The application is developed using Unity 3D. Unity 3D is a cross-platform 3D engine for creating 3D games and applications for mobile, desktop, the web, and consoles. As shown in the picture above, the app has 3 main screens, title, create a story, and animation. These screens are different scenes in Unity and are connected using the LoadScene method. In the Create Story scene, the "+" sign allowed the users to take photos of the text materials. The camera is connected to the OCR, after the user takes the photo, it would automatically recognize the text and transform an audio file.

```
public IEnumerator ProcessImage()
{
    _status = OCR_Status.OCR_STARTED;
    string imagePath = Path.Combine(Application.persistentDataPath, ImageName);

    WWWForm form = new WWWForm();

    byte[] bytes = File.ReadAllBytes(imagePath);
    form.AddBinaryData("image", bytes, ImageName, "image/png");

    UnityWebRequest www = UnityWebRequest.Post(google_fucntion_url, form);
    www.timeout = TimeOut;

    yield return www.SendWebRequest();

    if (www.result != UnityWebRequest.Result.Success)
    {
        Debug.Log(www.error);
        _status = OCR_Status.OCR_ERROR;

    }
    else
    {
        Debug.Log(www.downloadHandler.text);
        _status = OCR_Status.OCR_SUCCESS;
        _text = www.downloadHandler.text.Replace("\n", " ").Replace("\r", " ");
    }
}
```

Figure 5. Screenshot of code 1

```
// Watson Code section
private string Authenticate(string username, string password)
{
    string auth = username + ":" + password;
    auth = System.Convert.ToBase64String(System.Text.Encoding.GetEncoding("ISO-8859-1").GetBytes(auth));
    auth = "Basic " + auth;
    return auth;
}

public IEnumerator ProcessAudio()
{
    _status = OCR_Status.SPEECH_START;

    Debug.Log("Started");
    string method = @"/v1/synthesize";
    string urlmethodpath = watson_fucntion_url + method;

    string jsonText = "{\"text\":\"" + Text + "\"}";
    byte[] jsonBytes = System.Text.Encoding.UTF8.GetBytes(jsonText);

    Debug.Log(
    Debug.Log(


    string aut

    WWWForm fc
    //form.Adc

    UnityWebRequest www = UnityWebRequest.Post(urlmethodpath, form);

    www.SetRequestHeader("Authorization", authorization);
    www.SetRequestHeader("Content-Type", "application/json");
    www.SetRequestHeader("Accept", "audio/mp3");

    www.uploadHandler = (UploadHandler)new UploadHandlerRaw(jsonBytes);

    string path = Path.Combine(Application.persistentDataPath, $"{ImageName}.mp3");
    www.downloadHandler = new DownloadHandlerFile(path);

    Debug.Log("Sent Request!");

    yield return www.SendWebRequest();

    if (www.result != UnityWebRequest.Result.Success)
    {
        Debug.Log("Some error occurred in the Post Request");
        Debug.Log(www.result);
        Debug.Log(www.error);
        Debug.Log(www.responseCode);

        foreach (KeyValuePair<string, string> pair in www.GetResponseHeaders())
        {
            //print(pair.Key + " " + pair.Value);
        }

        _status = OCR_Status.SPEECH_ERROR;
    }
    else
    {
        Debug.Log("Success maybe in the Post Request");
        _status = OCR_Status.SPEECH_SUCCESS;

    }
}

}
```

Figure 6. Screenshot of code 2

As shown in the image above, we utilized the google cloud vision service to extract text from pictures with OCR and the IBM Watson text-to-speech service to generate audio of simulated human voice based on the text outputted from OCR. These two functions are closely connected so that the text will get delivered to the Watson online service immediately to generate audio files.

We based our NLP algorithm on the spaCy NLP library, which provides advanced natural language processing in Python [11]. With spaCy, we can filter out specific nouns, verbs, or both nouns and verbs in a specific orientation.

The animation and background are built with assets downloaded from the Unity assets store, the items contain their own codes that are able to animate them in Unity. The background is built in different scenes using the assets downloaded, while the application processes the image, it would automatically select the background with the keywords extracted from the text using the NLP algorithm. The character is animated with the SALSA LipSync asset and the audio file generated from the Watson online service, allows the characters to have lip movements according to the text material. After the application finishes processing and clicking the "star" button, it would load the StoryScene, and the selected background and character would be in that scene.

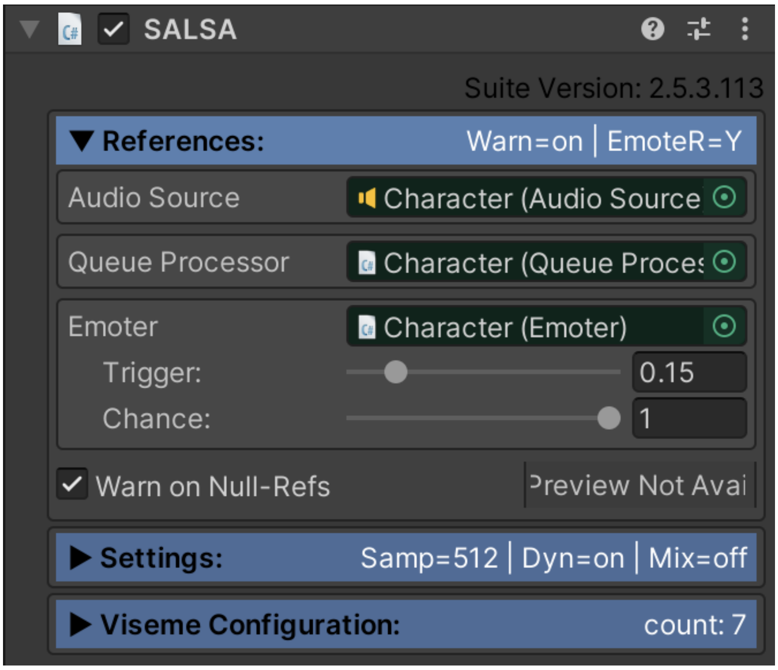

Figure 7. Screenshot of SALSA

4. EXPERIMENT

4.1. Experiment 1

In order to verify that our solution can effectively solve problems at different levels and have good user feedback, we decided to select multiple experimental groups and comparison groups for several experiments. For the first experiment, we want to prove that our solution works stable and continuously, so we choose a group size of 100 different NLP inputs text sentence different kind of target.The goal of the first experiment is to verify if IBM chat bot could analyze all the text sentence if the AI read the sentence right works good for all the target sentence. Experiments have shown that almost all targets in different types tested the right result. Moving enemy has the most correct rates, which means our user are works more better in aiming the moving enemy in the game. This experiment could explain that the data sets do have a obvious impact on the finding the targets and aim it, because the data we are using have a high rates of the moving enemy. The average correct rate of 100 different types of the aims shows below:

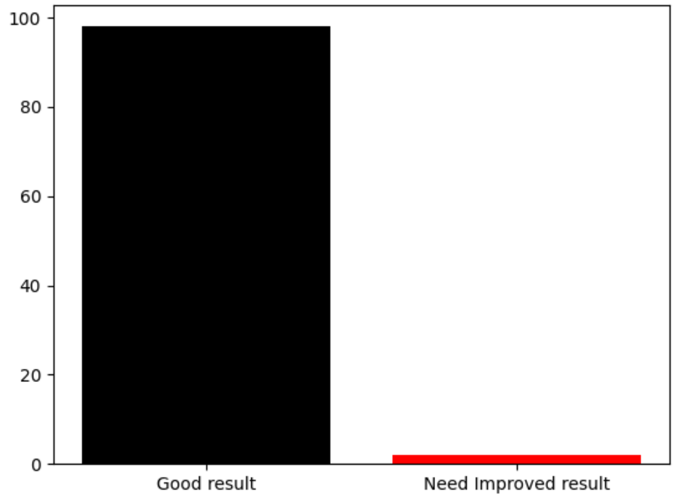

Figure 8. Number of good and nee improved result

5. RELATED WORK

Hayashi, Masaki, et al developed technology that generates CG animation generated automatically from a text-based script, it uses TVML, is a language with script structure used in actual TV production, to edit the text and FIL to complete text to speech [12]. Their project is similar to ours with the automatically generated animation from text. The technology of Hayashi, Masaki uses different language templates.

Large, Andrew, et al conclude that animation helps children to identify the major plots more successfully than reading the text through his experiment [13]. Their project also uses animation to create solutions for social problems, but is designed for special science experiments.

Shaw, Rebecca, and Vicky Lewis reported ADHD children to focus more and give more accurate responses to the animated computerized tasks [14]. Different from our project, this work is an investigation and study on the positive impact on education of computers to ADHD children, and our project is to implement the help.

6. CONCLUSIONS

The project provides users with an efficient, productive, and cost-effective product to automatically create animated videos from pictures of written material; we design it to help educate and treat children with ADHD or ASD. The current prototype can automatically create videos from images of written sources. The user merely needs to click on the “start” button on our homepage and start taking pictures of books, our application handles everything behind the stage. After a few seconds, the users would be able to watch the animations. The most important characteristic of this product is its ability to incorporate multiple functions and design them to work together at the fastest speed possible. All parts of our project are already implemented into different commercialized applications and services. A large portion of our application’s functions uses pre-existing services in our databases. These include Google’s online vision service to achieve optical character recognition (OCR), IBM’s Watson text-to-speech service to generate audio files, Unity’s Multipurpose Avatar with SALSA LipSync Suite asset to match the characters' lip movements with audio data, and the spaCy library to accomplish natural language processing (NLP) [15].

We still have a lot more proposed functions in development, including the interactive feature in videos, an online video sharing platform, and a group chats feature. The interactive feature in videos is aimed to help children with ADHD/ASD concentrate and learn communication skills. The online video sharing platform is akin to YouTube with the difference that our users are not video creators. All videos on this sharing platform are automatically generated through our application. When users use our application to generate a video for themselves, they are able to choose whether to upload this video to our sharing platform so that others without the original written material can also access it. The group chat feature is implemented so that users can join specific learning group chats and share related videos and support each other. This feature also helps people with ADHD/ASD set up their own virtual community.

REFERENCES

[1] Jensen, Peter S., David Martin, and Dennis P. Cantwell. "Comorbidity in ADHD: Implications for research, practice, and DSM-V." Journal of the American Academy of Child & Adolescent Psychiatry 36.8 (1997): 1065-1079.
[2] Kasari, Connie, et al. "Social networks and friendships at school: Comparing children with and without ASD." Journal of autism and developmental disorders 41.5 (2011): 533-544.
[3] Sonuga-Barke, Edmund JS, et al. "Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments." American Journal of Psychiatry 170.3 (2013): 275-289.
[4] Reichow, Brian, et al. "Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD)." Cochrane database of systematic reviews 10 (2012).
[5] McAlister, Alfred L., Cheryl L. Perry, and Guy S. Parcel. "How individuals, environments, and health behaviors interact." Health Behavior 169 (2008).
[6] Moody, Amelia K., Laura M. Justice, and Sonia Q. Cabell. "Electronic versus traditional storybooks: Relative influence on preschool children's engagement and communication." Journal of Early Childhood Literacy 10.3 (2010): 294-313.
[7] Dondlinger, Mary Jo. "Educational video game design: A review of the literature." Journal of applied educational technology 4.1 (2007): 21-31.
[8] Rahimi, Farzan Baradaran, et al. "A Game Design Plot: Exploring the Educational Potential of History-Based Video Games." IEEE Transactions on Games 12.3 (2019): 312-322.
[9] Nguyen, Trong Duc, et al. "Exploring API embedding for API usages and applications." 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE). IEEE, 2017.
[10] Jain, Vishal. "3D model of attitude." International Journal of Advanced Research in Management and Social Sciences 3.3 (2014): 1-12.
[11] Vasiliev, Yuli. Natural Language Processing with Python and SpaCy: A Practical Introduction. No Starch Press, 2020.
[12] Hayashi, Masaki, et al. "T2V: New Technology of Converting Text to CG Animation." ITE Transactions on Media Technology and Applications 2.1 (2014): 74-81.
[13] Large, Andrew, et al. "Multimedia and comprehension: The relationship among text, animation, and captions." Journal of the American society for information science 46.5 (1995): 340-347.
[14] Shaw, Rebecca, and Vicky Lewis. "The impact of computer-mediated and traditional academic task presentation on the performance and behavior of children with ADHD." Journal of Research in Special Educational Needs 5.2 (2005): 47-54.
[15] Joorabchi, Mona Erfani, Ali Mesbah, and Philippe Kruchten. "Real challenges in mobile app development." 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. IEEE, 2013.