Singing Video Generation with Music Separation by MCS08

Presented by :

Yeoh Ming Wei, Yew Yee Perng, Toh Xi Heng

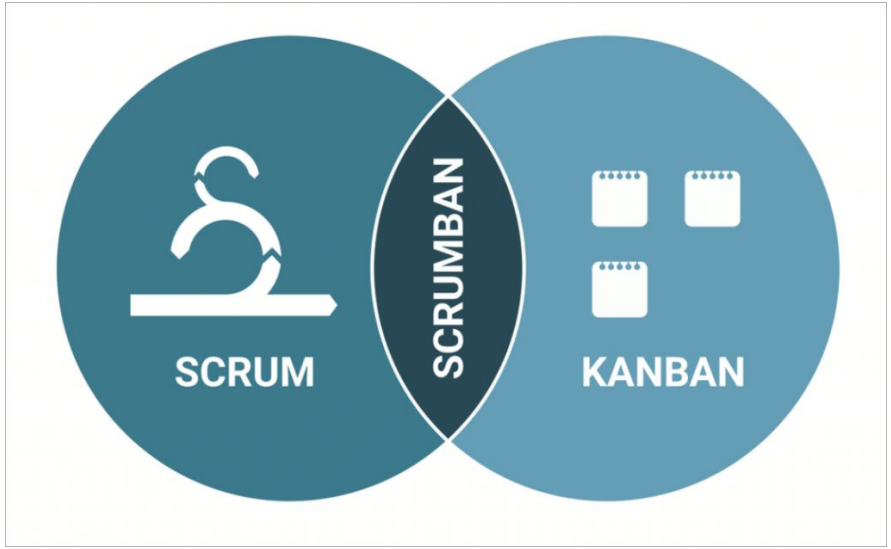
Introduction

- Create a cutting-edge system capable of generating lifelike singing face videos synchronised with music

- Works separated into two parts:
  1. Audio Separation
  2. Virtual Avatar Generation with Lip Synchronization

- Source code will be used : SadTalker

Referred Software Development Methodology



SCRUMBAN

The combination of Kanban and Scrum

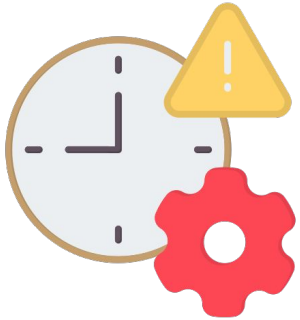
Scrum assists on team structure and principles

Kanban manages work through visual tasks

The Reason of Adapt Instead of Follow

There are many factors which our team decided to follow or modify some methods introduced by Scrum and Kanban:

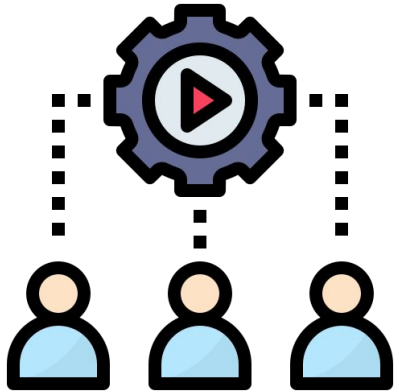
- Time Constraint
- Some methodology is unnecessary
- Modify methods until everyone in the team agrees



Our Software Development Methodology

Role and responsibilities

- Everyone has the same general role (Developer, Product Owner)
- Product Owner (Referenced by Scrum) is involved by everyone as communication is the key to focus on the project.
- In general, everyone is a developer but may be different in terms of sub-role (Quality Assurance, Backend Developer, Frontend Developer)



Our Software Development Methodology

Weekly meetings with supervisor (Once a week)

- Present our work to supervisor
- Gather feedback from supervisor
- Similar to sprint review from Scrum

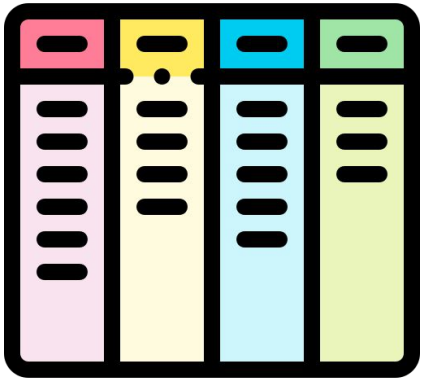
- Recap on the feedback given
- Brainstorming on what improvements can be made
- Similar to sprint retrospective from Scrum

Our Software Development Methodology

Kanban boards and principles

- Uses kanban board to track the progress of our task

- Apply continuous philosophy where our team dedicate some time to improve work quality every week



Work Completed

- In-depth on how generation virtual avatar model, SadTalker works
- Generate a sample video

Sample Video Observations:

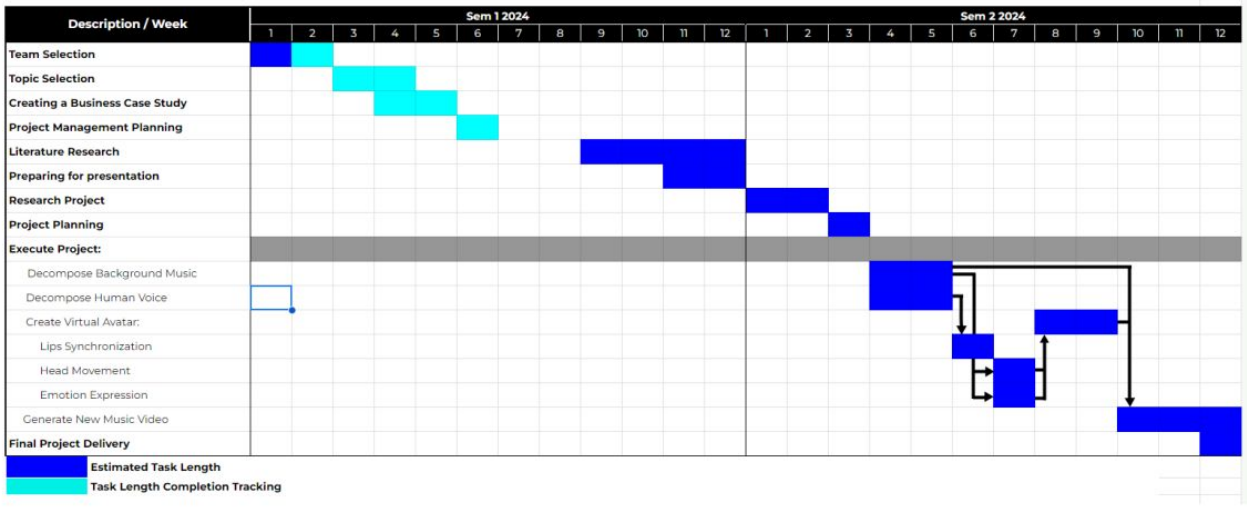
- Express emotion,
- Head movement
- Lips synchronisation
- Video quality is bad (Focus on improvement)

Sample Video generated (SadTalker)

Overview for Plan

Original Gantt Chart Created last semester

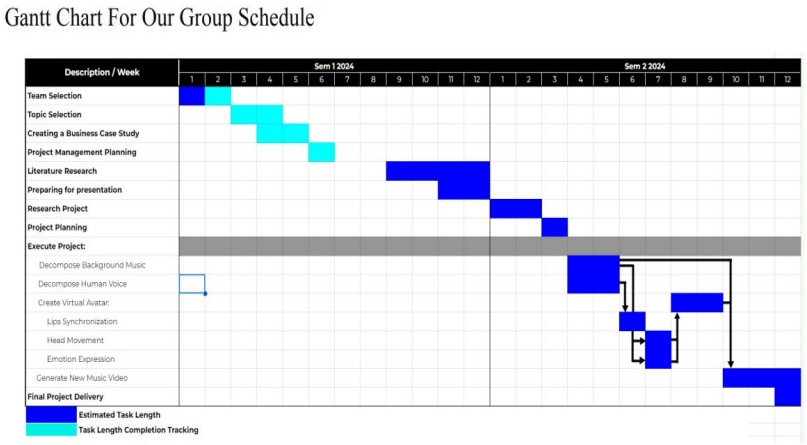
Gantt Chart For Our Group Schedule



Overview for Plan - continued

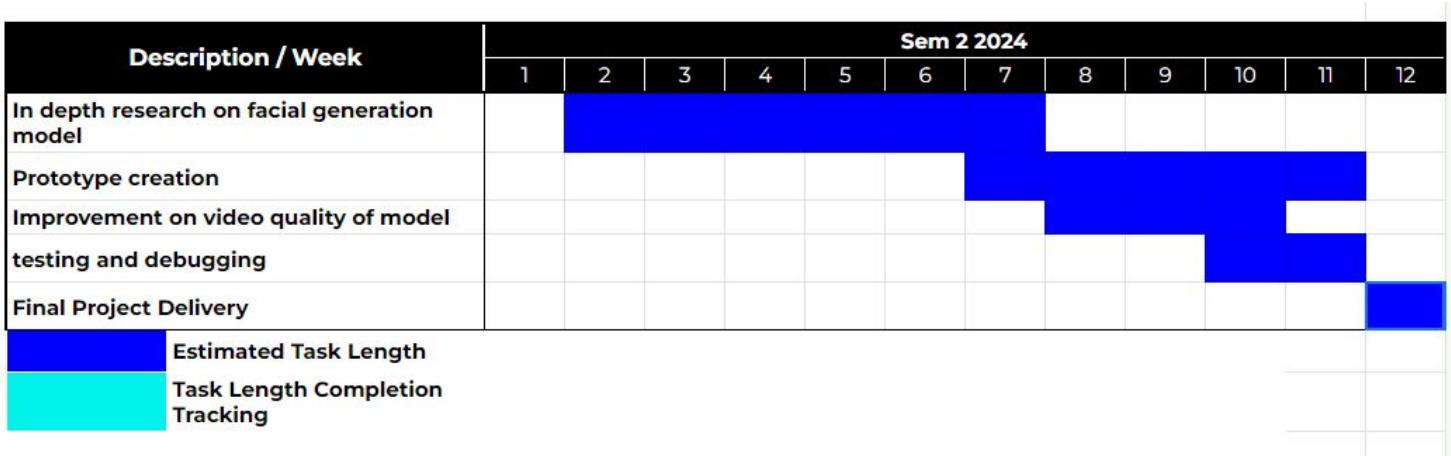
- Slight changes made on Original Chart
- Focusing on Facial generation instead of the whole
- Working on existing model and improving on desired parts(video quality)

Original Gantt Chart



Current Plan

Updated Gantt chart for the semester

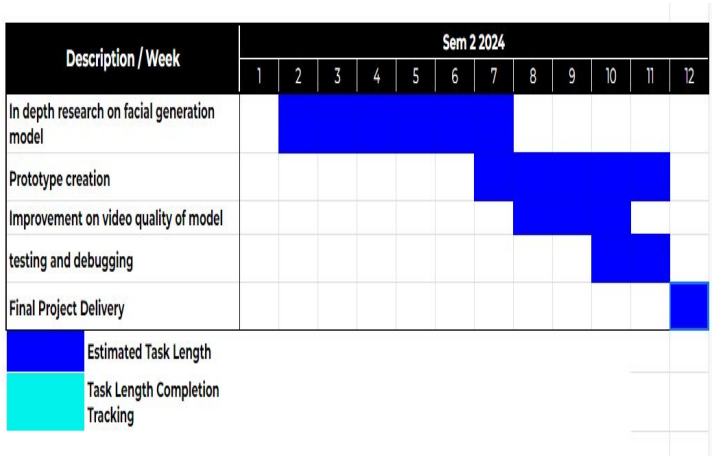


Current Plan - continued

- Changes made currently
- Perform in depth research on facial generation
- Begin to create the prototype
- Improve on video quality of model
- Testing and debugging to be planned at ending phase of the project

- Currently on track in progress

Updated Gantt Chart



Comparison with the state-of-the-art method on HDTF dataset

HDTF Dataset - High-definition Talking Face Dataset

FID - used to evaluate the quality of generated images

CPBD - measure the perceptual sharpness of an image

CSIM - evaluates the similarity between two images

Method	Eye Blink	Lip Synchronization		Learned Head Motion		Video Quality		
		LSE-C↑	LSE-D↓	Diversity↑	Beat Align↑	FID↓	CPBD↑	CSIM↑
Real Video	N./A.	8.211	6.982	0.259	0.271	0.000	0.428	1.000
Wav2Lip* [30]	N./A.	10.221	5.535	N./A.	N./A.	21.725	0.368	0.849
PC-AVS** [51]	from ref.	9.053	6.355	N./A.	N./A.	69.127	0.206	0.683
MakeItTalk [52]	automatic	5.110	10.059	0.257	0.268	28.243	0.283	0.838
Audio2Head [39]	automatic	7.357	7.535	0.181	0.267	24.392	0.281	0.823
Wang *et al.* [40]	automatic	4.932	10.055	0.226	0.268	22.432	0.295	0.811
SadTalker	controllable	7.290	7.772	**0.278**	**0.293**	**22.057**	**0.335**	**0.843**

Problems we may face in the future

- Lip Synchronization Degradation
  - lip movements don’t match the audio accurately
- Inconsistent Head Motion
  - head motion doesn't really match the speech rhythm or context
- Increased Computational Load
  - increase the computational requirements for real-time rendering or processing

Our Methodology Issue

Task breakdown not in-depth enough for Kanban Board

- Leads to confusion on how to start
- The process is not detailed enough

Does not have a member with Scrum Master role

- Members will have many conflicts during discussion

No Daily / Weekly standups for members meeting

- Lack of effective communication between members

Team Management

- Communication via Discord, WhatsApp among Team, Google Chat with supervisor

- Task allocation is done fairly by splitting the workload equally
  - Help out with each other’s tasks if assistance is needed

- Conflicts on ideas are solved by discussing and decisions are based on majority

Conclusion

In conclusion, we've made significant progress on performing thorough research on facial generation. While we've encountered changes in our plans, our approach moving forward includes the creation of our prototype and the improvement on the video quality for facial generation. We are confident that with continued effort and collaboration, we will successfully achieve our goals. I welcome any feedback or questions as we work towards the final phase of the project.

Thank you

Questions?