2024 전기 졸업과제 중간 보고서

주제	Dominance Factor 에 따른 사용자 인지 변화 분석
팀명	NotHuman
팀 번호	14
팀원	이영진, 조성환, 조주은

목차

1. 요구조건 및 제약사항 분석에 대한 수정사항

1.1. 요구조건

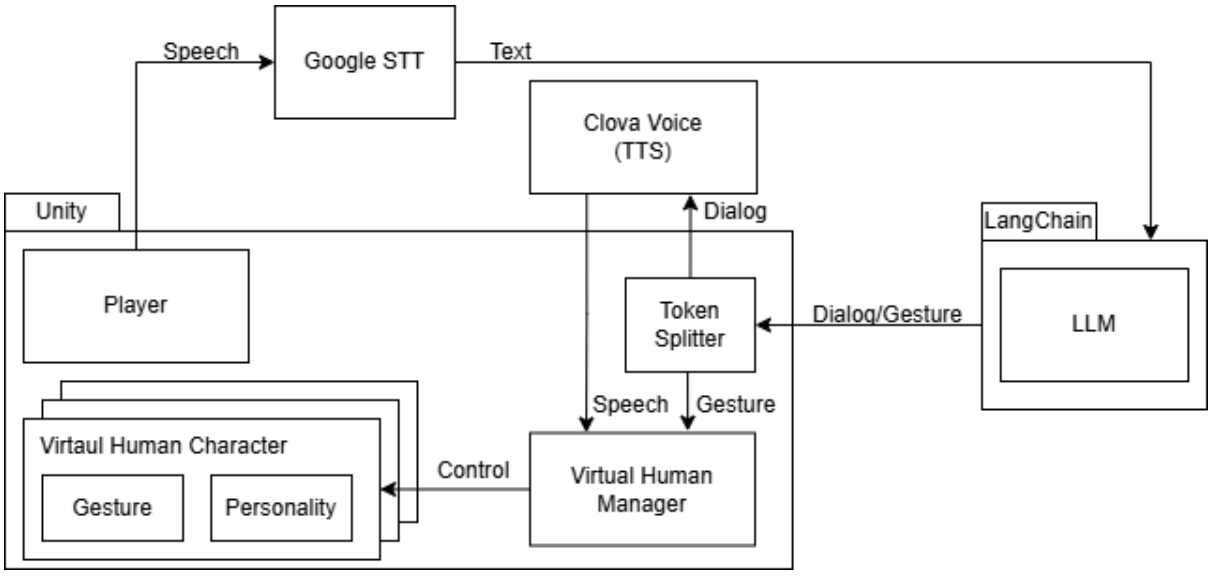
- Dominance Level 기반 대화 생성
  - 시스템은 높은(High), 중간(Middle), 낮은(Low) 수준의 Dominance 를 반영한 대화를 생성할 수 있어야 한다.
  - 이러한 대화 생성은 LLM(Large Language Model)을 통해 이루어지며, 연구에 따르면 간단한 언어적 표현만으로도 지배성 수치를 조절할 수 있다.[1]
- 비언어적 요소
  - 비언어적 요소의 영향을 분석하기 위해 행동, 백채널, 목소리, 시선들을 각각 구현하여야 한다.
- VR 기기 활용
  - 각각의 성격을 가진 Virtual Human 들을 사용자가 선택할 수 있어야한다.
- Virtual Human Manager
  - Meta Quest 를 이용한 VR 기기를 통해 Virtual Human 과 마주보며 대화할 수 있는 환경을 제공한다.

1.2. 제약사항 분석 및 수정사항

- 가상 캐릭터의 성격에 맞게 행동의 범위를 조절할 때 LLM 이 어느정도의 범위로 행동할지 텍스트로 알려주면 텍스트를 모션으로 생성해주는 모델을 이용하여 모션을 생성하기로 하였으나, 가상 캐릭터의 행동은 미리 정해둔 모션들 중 대답에 알맞은 모션 하나를 LLM 이 결정한다. 모션의 범위는 가상 캐릭터의 dominance level 에 따라 일정 비율로 조절된다.

2. 설계 상세화 및 변경 내역

2.1. 구조도



* Token Splitter : 입력 데이터를 처리하여 행동과 대화를 분리하고, 각각의 요소를 Gesture 배열, Dialog 문자열로 정리한다.

2.2. 설계 상세

① Unity 실험 환경 구성

Unity 에서 사용자가 가상 캐릭터에게 말한 내용이 Speech-to-text(STT) 기능을 통해 텍스트로 변환된다. 텍스트는 통신을 통해 LLM 에 전달되고 LLM 에서 가상 캐릭터의 dominance level 에 부합하는 대답 내용과 제스처가 Unity 에 전달된다. Unity 에서 Text-to-speech(TTS) 기능을 통해 대답 내용을 음성으로 변환한다. 변환할 때 목소리톤을 가상 캐릭터의 dominance level 에 맞게 변환한다. 동작은 .fbx 파일로 적용하여 가상 캐릭터가 대답중에 적절한 동작을 하게 한다. 이 과정 중에 가상 캐릭터는 설정된 dominance level 에 맞게 사용자를 일정 비율로 응시(Eye Gaze)하게 된다. 가상 캐릭터는 Q 와 E 를 통해 원하는 가상 캐릭터를 선택할 수 있으며, 스페이스 바를 통해 대화할 수 있도록 한다.

② LangChain 구성

Unity 로부터 사용자가 가상 캐릭터에게 말한 내용을 받아 LLM 이 가상 캐릭터의 dominance level 에 따라 대답하고 그에 알맞는 동작을 지정하게 한다. 이 기능을 수행하기 위해 프롬프트 엔지니어링을 사용한다. 프롬프트에 persona description 을 작성하고 이에 기반하여 대답을 하게 하면 인간과 유사한 성격 특성을 모방할 수 있다.[2]

이를 이용해 Unity 로부터 받은 가상 캐릭터의 dominance level 값을 프롬프트에 적용시켜 LLM 이 dominance level 에 알맞는 대답을 할 수 있게 한다. LLM 이 대답 내용에 알맞는 동작을 선택할 수 있게 하기 위해 Few-Shot 프롬프팅을 사용한다. Few-Shot 프롬프팅을 수행하기 위해 annotation 된 대화 데이터셋을 사용한다. 대화 데이터셋은 대사가 시작되는 시간, 종료되는 시간, 동작이 시작되는 시간, 종료되는 시간이 annotation 되어있고 화자의 Dominance 레벨도 라벨링되어있다.

③ 대화/행동 데이터셋

활용할 Genea 데이터셋은 대화 중 발생하는 제스처와 같은 비언어적 행동을 연구하기 위해 설계된 데이터셋이다. 이 데이터셋은 대화 상황에서 사람들의 제스처를 기록한 자료로, 특히 제스처 생성 모델을 학습하고 평가하는 데 중요한 역할을 한다. 이 팀은 Genea 데이터셋을 annotation 하여 각 인물의 대화/행동 쌍으로 구성된 새로운 데이터를 생성한다.

이를 위해 Genea 데이터셋에서 시간대별로 기록된 행동과, 단어 단위로 쪼개진 대화내용(각 단어의 발화 시점 포함)을 병합하여 대화/행동 데이터셋을 구성한다. 이 데이터 셋은 각 인물의 Dominance Level 에 따라 적절한 비언어적 행동을 생성하는 데 사용된다.
또한, 주변 지인 3 명 이상에게 부탁하여 약 20 개의 영상을 보고 각 인물의 Dominance Level 을 평가하게 하며, 평가를 위해 설문지를 준비한다.

해당 설문지는 '대인관계 원형모델에 따른 한국판 대인관계 형용사척도의 구성' 논문을 기반으로, 대인관계 원형모델에서의 PA(자기확신/주장), HI(비주장/소심) 형용사를 사용하여 구성되었다.[3]

'자신만만하다', '당당하다', '주장적이다', '추진력있다', '자기확신이 있다', '수동적이다', '비주장적이다', '자신없다', '소심하다', '유약하다' 총 10 개의 항목이 있으며, 각 항목에 점수를 매겨 Dominance Level 을 결정한다.

④ 비언어적 요소: 행동

LLM 의 프롬프트에 Dominance 레벨이 high 일 때, middle 일 때, low 일 때 대화 예시를 제공하고, 이를 기반으로 LLM 이 In-Context Learning 을 수행하여 가상 캐릭터의 Dominance 레벨 값에 적절한 대사를 생성한다. 행동은 가상 캐릭터가 각 단어를 말할 때 시작될 수 있고 어떤 종류의 행동인지, 행동의 시작 시간과 종료 시간, 그리고 행동의 dominance level 이 high 인지 middle 인지 low 인지가 parameter 로 주어진다. 아래는 행동과 대사를 LLM 이 생성한 예시이다. 행동과 관련한 요소들은 대괄호로 감싸져 있다.

ex) [head_nod; 0;0.3;param(~~~)] yes, [talking_one_hand;0.4;0.9;param(~~)] i have an [talking_two_hand;1;0.3;param(~~)] apple.

동작은 Genea 데이터셋에서 화자가 발화할 때의 동작과 유사한 동작들을 Mixamo 홈페이지에서 구한다. 구한 동작들의 범위를 dominance level 이 high 인지, middle 인지, low 인지에 따라 설정한다. Dominance 특성에 따라 사지의 개방성과 관절의 움직이는 속도 등이 다르다는 것을 적용하여 동작들을 설정한다. 해당 설정은 Mixamo 의 'Character Arm-Space' parameter 를 활용하여 구현한다.

⑤ 비언어적 요소: 시선

캐릭터의 Dominance 수준에 따라 상대방을 바라보는 시선 처리 확률을 조정하여 자연스러운 시선 처리를 구현한다.

자연스러운 시선 처리를 위해 Crazy Minnow Studio 의 SALSA LipSync Suite 를 사용한다.

⑥ 비언어적 요소: 백채널

LLM 을 통해 얻은 백채널 빈도를 기반으로, Unity 에서 백채널 음성을 확률적으로 재생한다.

⑦ 비언어적 요소: 목소리

Domiance 수준에 따라 Pitch 와 Volume 이 다르게 설정된다.

Naver Clova TTS 를 활용하여 목소리 파라미터를 조정하고 음성을 생성한다..

2.3. 변경 내역

가상 캐릭터의 동작은 Text-to-motion 모델을 사용하여 동작을 생성하고 생성한 동작을 가상 캐릭터에 적용시키기로 하였으나, **생성된 동작이 Dominance 한지 알 수 없다, 원하는 결과물을 얻기 힘들다는 이유**로 인해 Text-to-motion 모델을 사용하는 대신 미리 지정된 동작들 중 대답 내용에 어울리는 동작을 LLM 이 선정하고 선정된 동작의 범위를 가상 캐릭터의 dominance level 에 맞게 설정하여 가상 캐릭터에 적용시키기로 한다.

3. 갱신된 계획 및 진척도

3.1. 갱신된 계획

	5월					6월				7월					8월				9월			
업무	1주	2주	3주	4주	5주	1주	2주	3주	4주	1주	2주	3주	4주	5주	1주	2주	3주	4주	1주	2주	3주	4주
자료조사																						
요구사항 분석 및 개발범위 산정																						
착수보고서 작성																						
유니티 환경 설정																						
랭체인 내부 및 API																						
유니티 내부 구현																						
중간 보고서 작성																						
유니티-랭체인 연결																						
최종 수정 및 테스트																						
IRB 연구계획서 작성 및 제출																						
데이터 분석																						
최종 보고서 작성 및 발표 준비																						
발표 포스터 제작																						

3.2. 구성원별 진척도

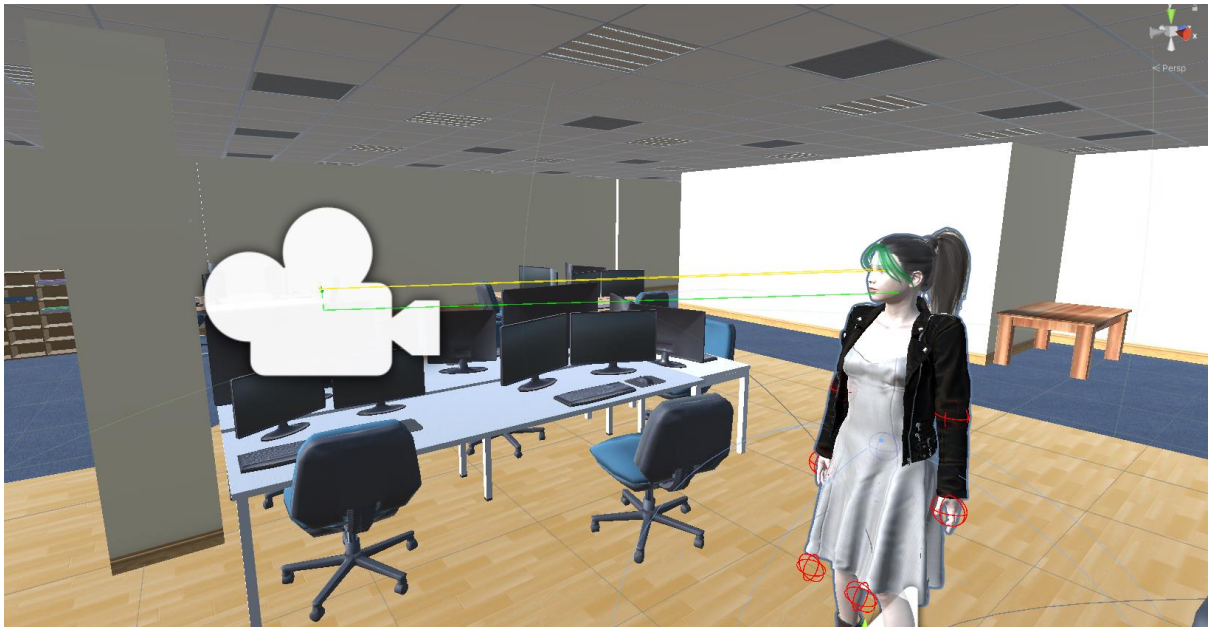
이름	역할	진척도
이영진	- 유니티 내부 구현 - 대화/행동 데이터셋	- 유니티 EyeGaze, TTS, STT, 컨트롤러 구현 - 데이터셋 대화내용 추출 진행중 - IRB 연구 계획서 작성 및 제출
조성환	- 랭체인 내부 구현 - 대화/행동 데이터셋	- 데이터셋 행동 추출 - 랭체인 내부 구현 진행중

		- IRB 연구 계획서 작성 및 제출
조주은	- 대화/행동 데이터셋	- Genea 데이터셋 영상 추출 - 대화/행동 데이터셋 구성 - IRB 연구 계획서 작성 및 제출

4. 과제 수행 내용 및 중간 결과

4.1. 가상 캐릭터 구현

① Eye Gaze



Salsa LipSync Suite 를 사용하여 사용자를 자연스럽게 쳐다보고, Dominance 에 따라 시선의 위치가 달라진다.

② 목소리톤

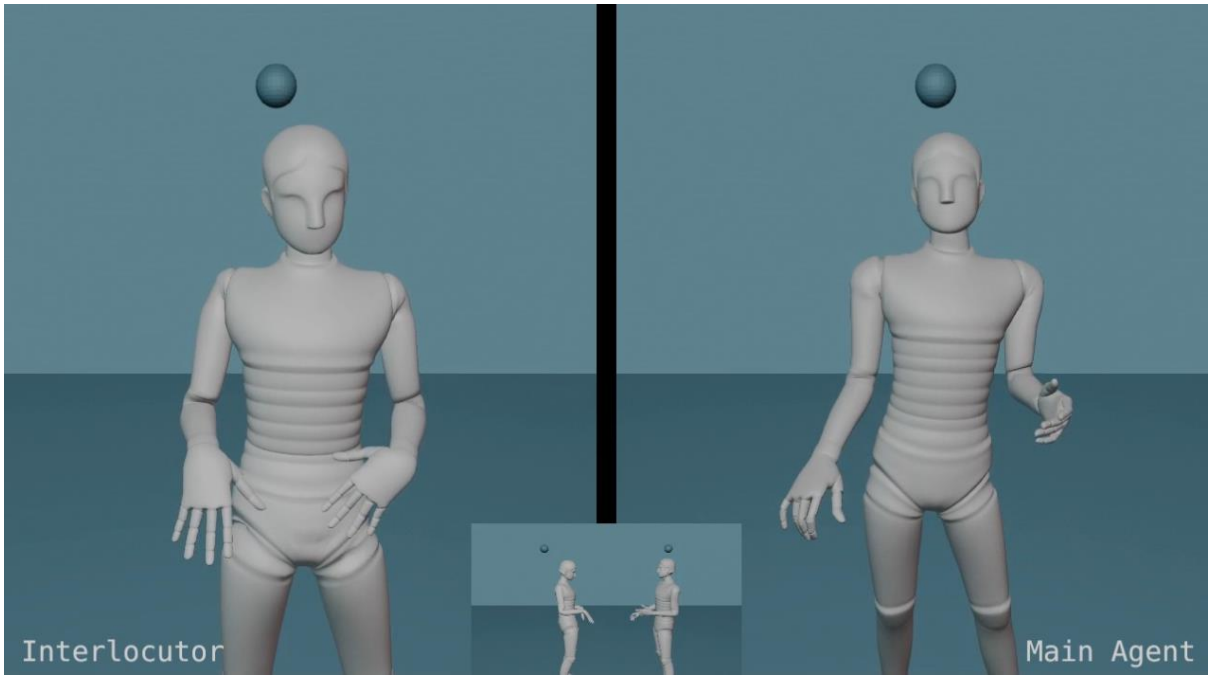
```
// volume과 pitch 출력
Debug.Log($"volume: {volume}, pitch: {pitch}");

byte[] byteDataParams = Encoding.UTF8.GetBytes($"speaker={avatar_name}&volume={volume}&speed=0&pitch={pitch}&format=wav&text={sentence}");

request.ContentType = "application/x-www-form-urlencoded";
request.ContentLength = byteDataParams.Length;

using (Stream st = await request.GetRequestStreamAsync())
{
    await st.WriteAsync(byteDataParams, 0, byteDataParams.Length);
}

HttpWebResponse response = (HttpWebResponse)await request.GetResponseAsync();
```

Naver Clova TTS 를 사용하여 Pitch 와 Volume 을 조절한다.

③ Virtual Human Manager

Q 와 E 를 통해 원하는 Virtual Human 을 선택할 수 있으며, 스페이스 바를 통해 대화할 수 있다.

4.2. 대화 데이터셋



Genea 데이터를 가공하여, 행동과 대화내용을 Annotation 하였다.

```
# Exported using VGG Image Annotator (http://www.robots.ox.ac.uk/~vgg/software/via)
# CSV_HEADER = metadata_id,file_list,temporal_segment_start,temporal_segment_end,metadata
"1_XgtMXNuU","[""001.mp4""]",0.81167,2.01537,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Head_Nod_Yes""}"
"1_tSuzIUQj","[""001.mp4""]",19.99685,26.47833,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Hands_Forward""}"
"2_veni9LAn","[""001.mp4""]",0.849,1.849,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Hard_Head_Nod""}"
"2_XG7AX0rn","[""001.mp4""]",19.978,26.21908,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Hands_Forward""}"
"2_EiMjIWdz","[""001.mp4""]",2.81167,3.73759,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Hands_Forward""}"
"2_ezcFBTkO","[""001.mp4""]",8.645,10.20056,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Sarcastic_Head""}"
"2_wGuXl6J1","[""001.mp4""]",55.562,57.83862,"{""TEMPORAL-SEGMENTS"":""Interlocutor_Cocky_Head_Turn""}"
"2_3xuo3i7R","[""001.mp4""]",0.785,3.46628,"{""TEMPORAL-SEGMENTS"":""MainAgent_Head_Nod_Yes""}"
"2_MXydoLkP","[""001.mp4""]",35.58833,37.7247,"{""TEMPORAL-SEGMENTS"":""MainAgent_Shaking_Head_No""}"
"2_LHr5MrBQ","[""001.mp4""]",0.966,2.2853,"{""TEMPORAL-SEGMENTS"":""MainAgent_Happy_Hand""}"
"2_fDydLR41","[""001.mp4""]",6.528,8.82466,"{""TEMPORAL-SEGMENTS"":""MainAgent_Hands_Forward""}"
"2_8mZROXg9","[""001.mp4""]",43.669,47.95472,"{""TEMPORAL-SEGMENTS"":""MainAgent_Arm_Gesture""}"
```

<Genea 데이터셋, 행동 Annotation>

```
WEBVTT

00.00.000 --> 00:03.178
[speaker_01]: my base knowledge though of like medical stuff

00:03.233 --> 00:03.966
[speaker_01]: **laugh**

00:03.982 --> 00:04.388
[speaker_01]: So,

00:04.434 --> 00:05.296
[speaker_01]: **laugh**

00:02.411 --> 00:04.612
[speaker_00]: yeah yeah yeah **laugh**

00:04.263 --> 00:05.028
```

<Genea 데이터셋, 대화 내용>

데이터셋에서 시간대별로 행동 추출하는 것은 진행되었으며, 대화 내용을 추출은 진행중이다. 이후, 행동과 대화를 병합하여 하나의 데이터셋으로 만들 예정이다.

Ref

[1] Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995, May). Can computer personalities be human personalities?. In Conference companion on Human factors in computing systems (pp. 228-229).

[2] Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., ... & Matarić, M. (2023). Personality traits in large language models. arXiv preprint arXiv:2307.00184.

[3] 정남운. (2004). 대인관계 원형모델에 따른 한국판 대인관계 형용사척도의 구성. 한국심리학회지: 상담 및 심리치료, 16(1), 37-51.

[4] Randhavane, T., Bera, A., Kubin, E., Gray, K., & Manocha, D. (2019). Modeling data-driven dominance traits for virtual characters using gait analysis. IEEE transactions on visualization and computer graphics, 27(6), 2967-2979.