



2024 전기 졸업과제

Dominance Factor에 따른 사용자 인지 변화 분석

A
14번
NotHuman 팀

202055649 이영진
201725270 조성환
201924587 조주은
지도교수 이명호

목차

1

과제 개요

2

과제 과정

3

최종 결과

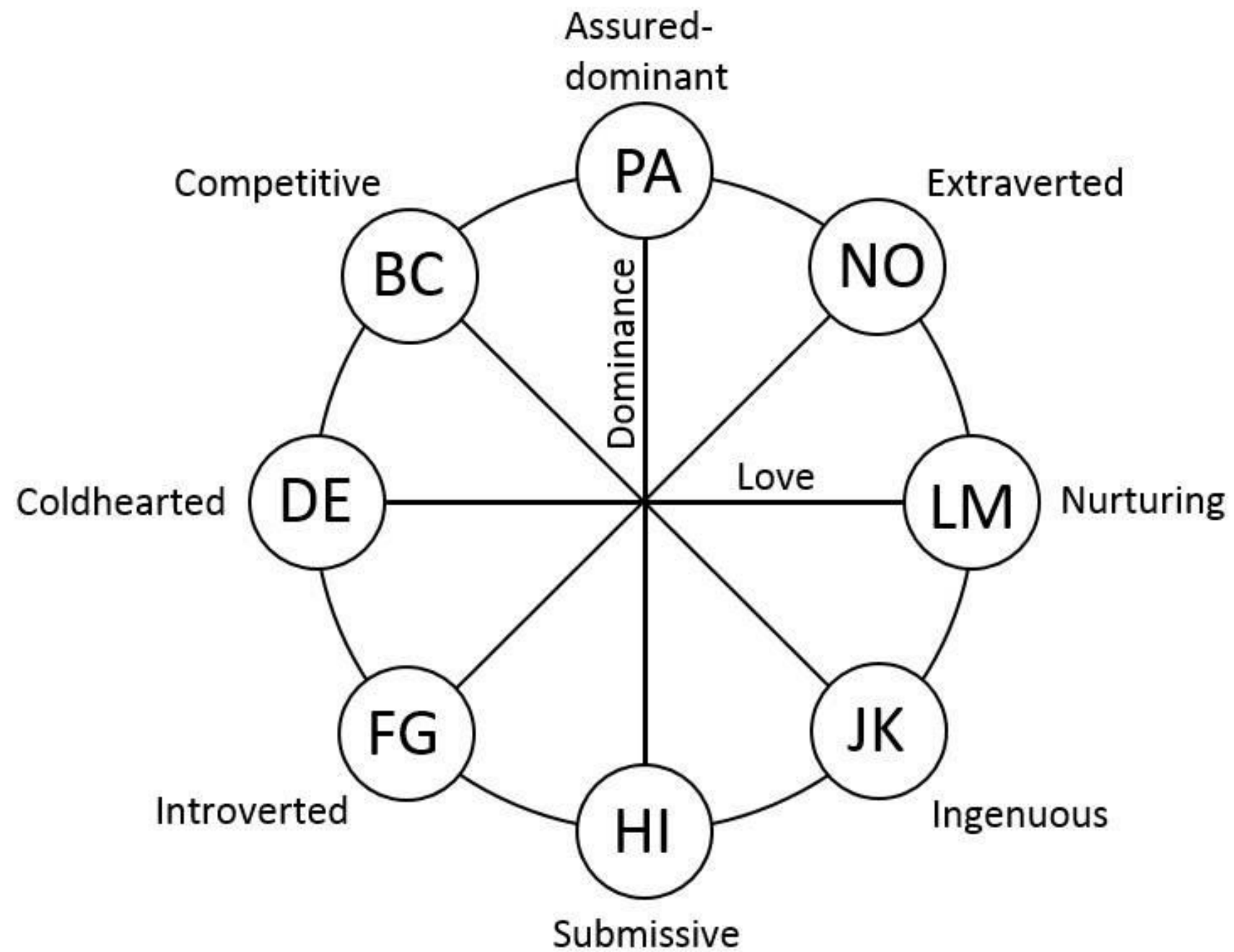
과제 개요

디지털 휴먼

- 디지털 휴먼은 사람과 유사하게 움직이고, 말하고 행동할 수 있도록 디지털 세상에서 구현한 존재
- 최근 ChatGPT와 같은 거대 언어모델(Large Language Model, LLM)이 발전함에 따라 같이 발전 중
- 이러한 디지털 휴먼은 사용자의 인지와 상호작용에 영향을 미칠 가능성을 보여주고 있음

01

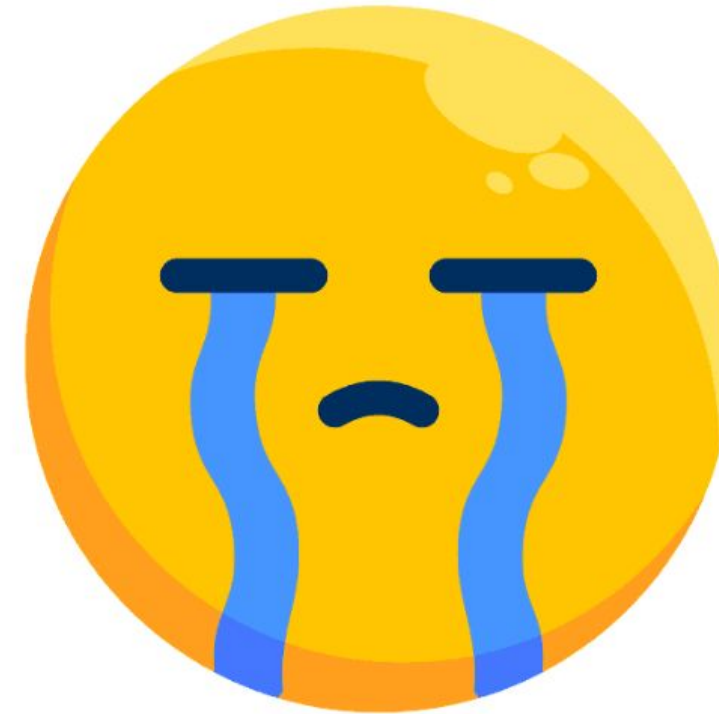
대인관계 원형모델



- 친애(Affiliation, Love)축과 통제(Control, Dominance)축으로 구성됨
- 대인관계에서 친애 축은 우호성을 가지는 행동을, 통제 축은 지배성을 가지는 행동을 하는 경향이 있음

01

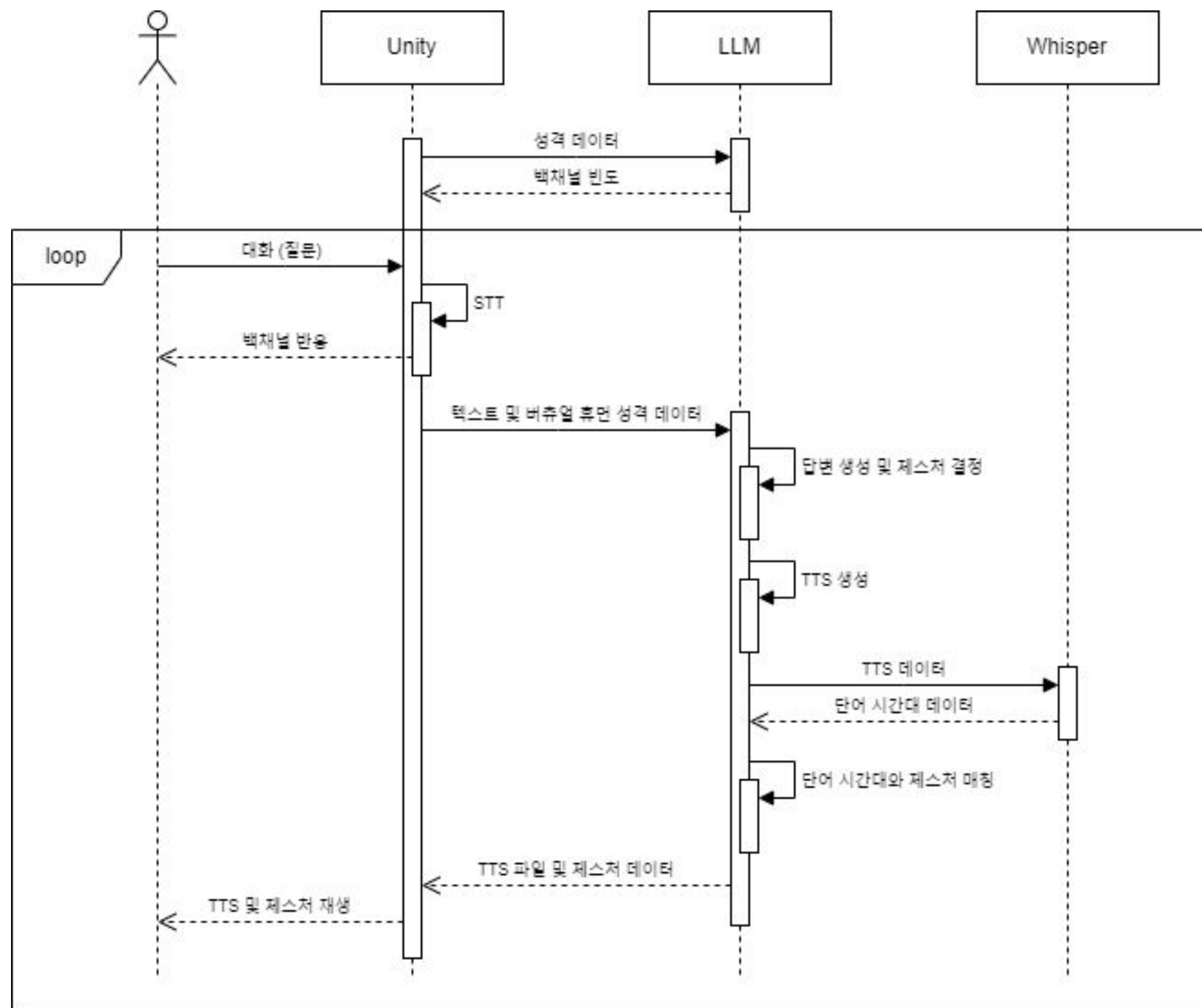
비언어적 행동



- 표정을 통해서 사람의 현재 감정을 추측할 수 있음
- 표정뿐만 아니라 말투, 행동, 시선, 목소리의 크기, 말의 빠르기 등에서도 정보를 얻을 수 있음
- 방금 언급한 것들은 비언어적 행동(Nonverbal behavior)의 일부

과제 목표

- LLM을 이용해 Dominance Factor에 따른 디지털 휴먼의 비언어적 행동 생성
- 디지털 휴먼의 비언어적 행동을 통해 사용자 인지 변화 분석



과제 과정

디지털 휴먼 성격 설계

지배적 성격(high dominance): 목소리가 크고 피치가 높으며, 상대방을 자주 응시하고, 큰 제스처를 사용합니다. 이러한 성격은 상대방에게 자신의 의도를 강하게 전달하고 대화의 주도권을 장악하려는 성격적 특성을 반영합니다.

중간 성격(mid dominance): 지배적이지도, 순종적이지도 않은 중간 정도의 성격으로, 적절한 목소리 톤과 중립적인 제스처를 사용하여 균형 잡힌 상호작용을 제공합니다.

순종적 성격(low dominance): 목소리가 작고 피치가 낮으며, 자주 상대방과의 시선을 피하며 작은 제스처를 보입니다. 이 성격은 상대방의 의견을 존중하고, 대화에서 소극적인 역할을 수행하려는 특성을 반영합니다.

프롬프트 설계

다음 작업에서, 제공된 설명에 맞춰 대화의 다음 단계를 생성하십시오. 응답은 제공된 인물의 설명을 반영하여야 합니다.

제공된 인물 : {character_description}. 나는 {personality}한 성격을 가진 인물이다.

대사는 자연스럽게 표현력 있게 작성되어야 하며, 제스처는 말과 일치하고 인물의 성격을 나타내야 합니다. 사용할 수 있는 제스처 목록은 다음과 같습니다: [{gesture_list_str}].

다음은 각 제스처에 대한 설명입니다.

[

Gesture1: 약간 아래를 보며 고개를 좌우로 꺾음. 여러 가능성을 제시하거나, 망설이며 결정을 미루거나, 불확실성을 표현하는 상황 등에 사용

Gesture2: 고개를 상하로 끄덕임

Gesture3: 고개를 좌우로 저음

Gesture4: 고민하거나 의아하다는 듯 고개를 옆으로 꺾음

Gesture5: 빈정대며/거만하게 고개를 상하로 끄덕임

Gesture6: 짜증내며 고개를 좌우로 저음

Gesture7: 고민하다 고개를 좌우로 저음

Gesture8: 장황하거나 지루한듯 고개를 상하로 끄덕임

Gesture9: 고민하다 고개를 상하로 끄덕임

]

예제의 word_or_phrase 는 제스처가 강조하는 대사(한국어)의 일부를 나타냅니다.

예제

1: {ex1}

2: {ex2}

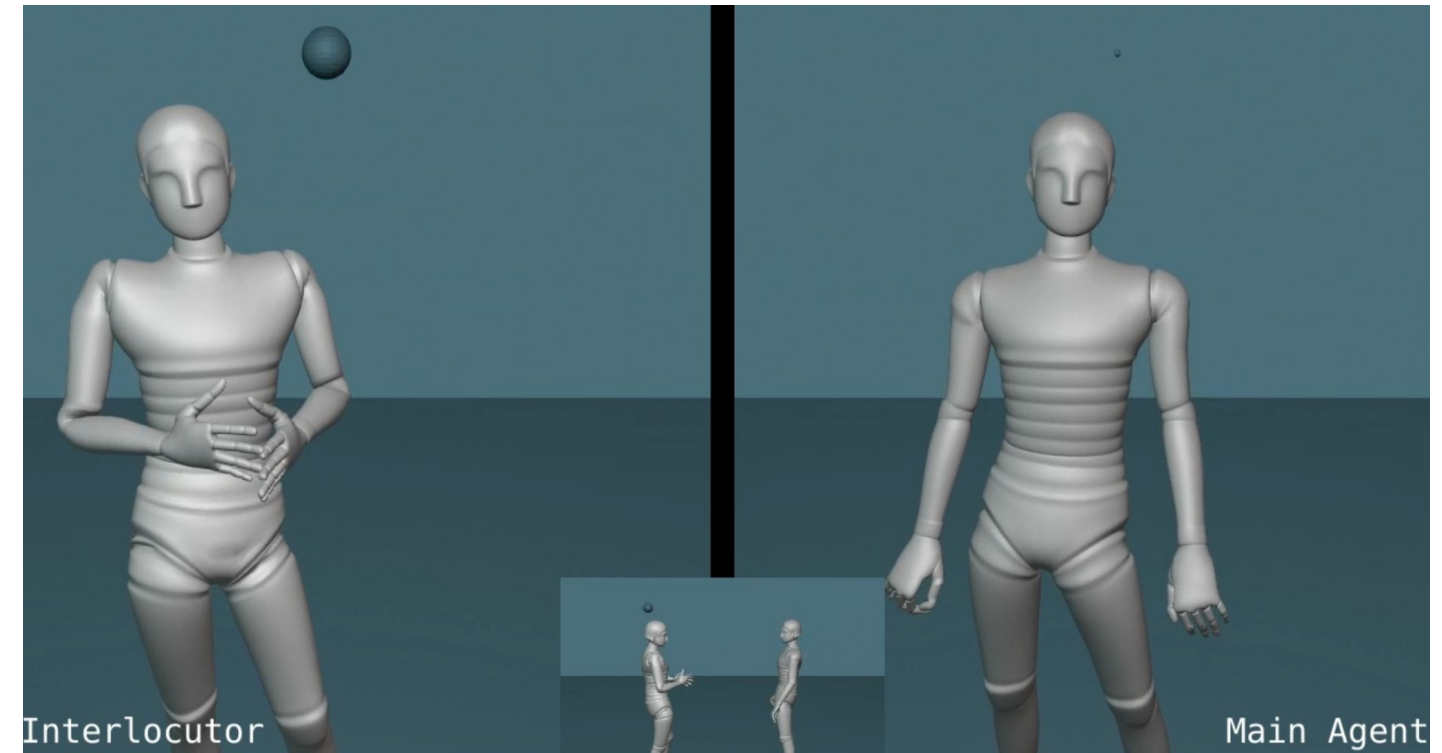
3: {ex3}

- 대화에 대한 답변 대사와 제스처를 생성하는 프롬프트

02

프롬프트 예제

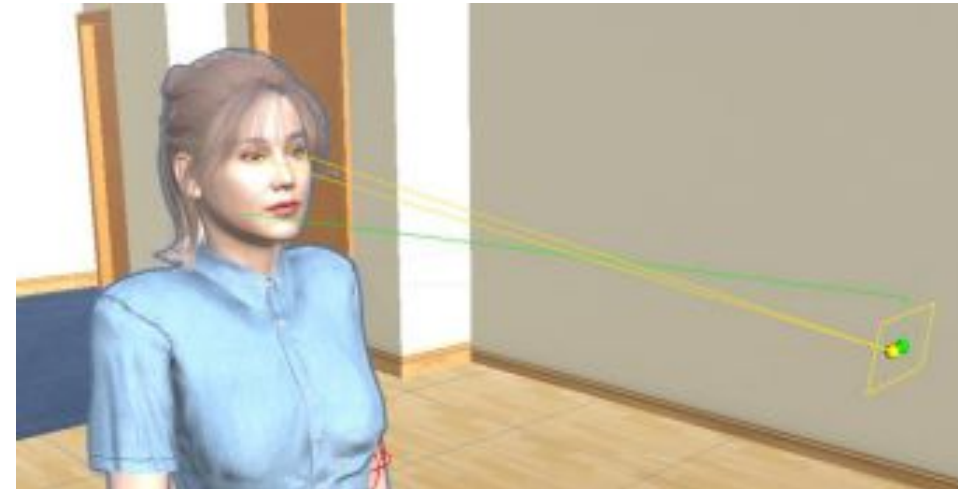
```
{
  "speech": "I liked watching that. I like Game of Thrones, but that's not
on there. I watched Vampire Diaries. Oh, my god. I loved Vampire Diaries.",
  "gesture": [
    {
      "word_or_phrase": "like Game of Thrones",
      "type": "Gesture7",
      "start": "1.30"
    },
    {
      "word_or_phrase": "not on there",
      "type": "Gesture2",
      "start": "2.10"
    },
    {
      "word_or_phrase": "god",
      "type": "Gesture8",
      "start": "6.87"
    }
  ]
}
```



- 예제는 GENE Challenge 2023 데이터 셋을 이용해 생성하였음

02

디지털 휴먼



- 디지털 휴먼은 유니티를 이용하여 구현하였음
- Salsa LipSync Suite를 이용해 립싱크나 시선처리를 구현
- 네이버 Clova TTS를 이용해 목소리 톤을 구현

```
string url = "https://naveropenapi.apigw.ntruss.com/tts-premium/v1/tts"; // naver clova api url
HttpWebRequest request = (HttpWebRequest)WebRequest.Create(url);
request.Headers.Add("X-NCP-APIGW-API-KEY-ID", client_id);
request.Headers.Add("X-NCP-APIGW-API-KEY", client_secret);
request.Method = "POST";

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

byte[] byteDataParams = Encoding.UTF8.GetBytes($"speaker={avatar_name}&volume={volume}&speed=0&pitch={pitch}&alpha=-1&format=wav&text={sentence}");
```

설문지 제작

디지털 휴먼 1

동영상을 보고 동영상의 디지털 휴먼에 대해 평가해 주세요.
소리가 없는 동영상입니다.

디지털 휴먼 1

동영상이 재생되지 않으면 아래 링크를 통해 동영상을 시청해주세요.
[동영상 링크](#)



자신만만하다 *

1: 매우 그렇지 않다
5: 매우 그렇다



당당하다 *

1: 매우 그렇지 않다
5: 매우 그렇다



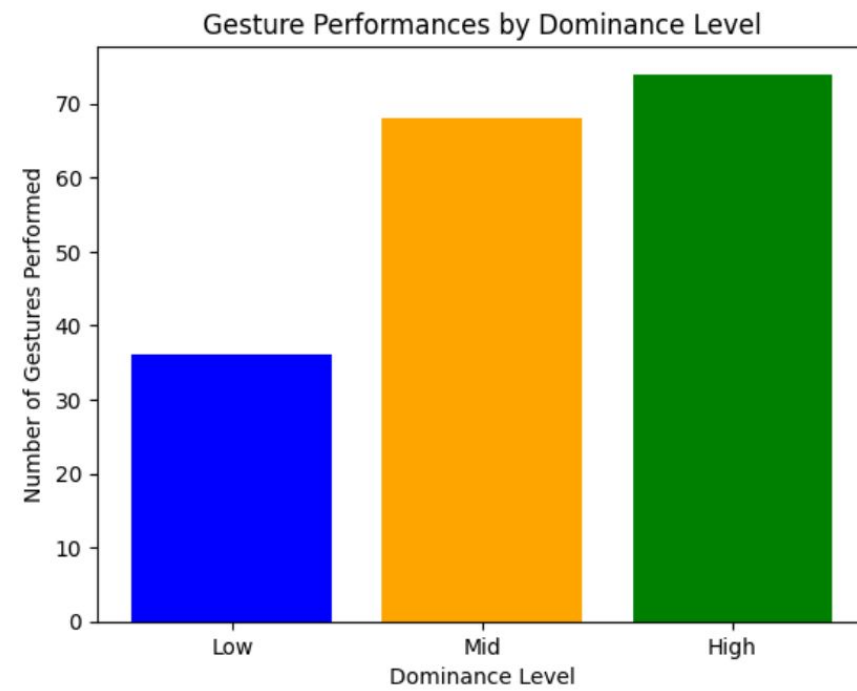
PA (자기확신/주장) 관련 질문: 자신만만하다, 당당하다, 주장적이다, 추진력 있다, 자기확신이 있다.

HI (비주장/소심) 관련 질문: 수동적이다, 비주장적이다, 자신없다, 소심하다, 유약하다.

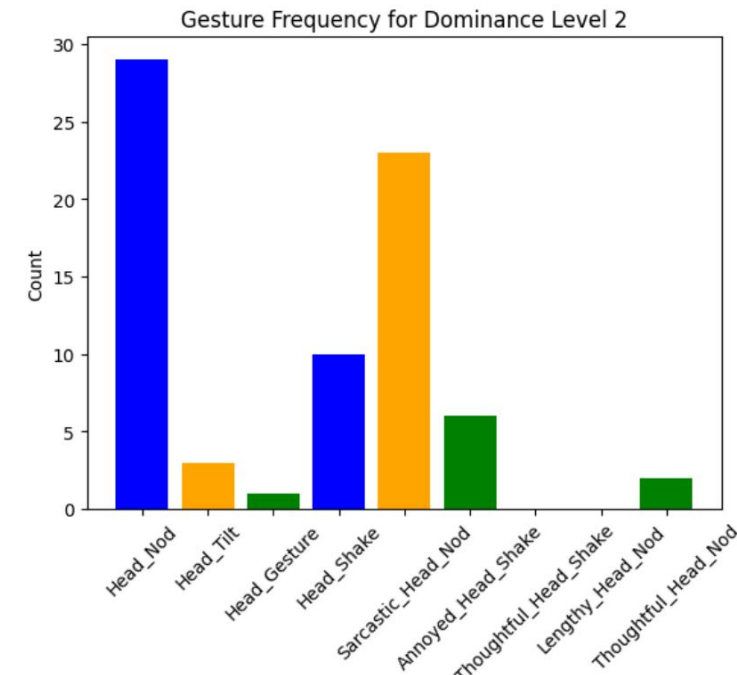
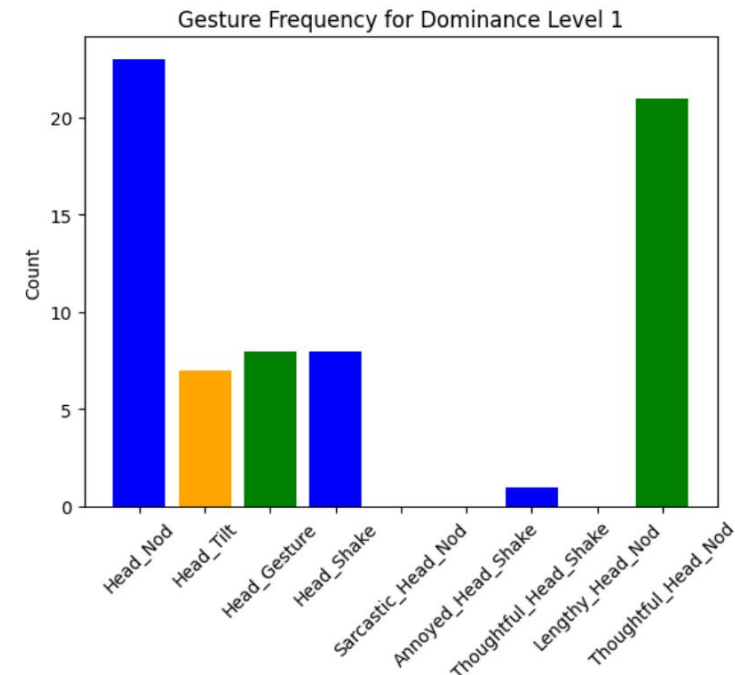
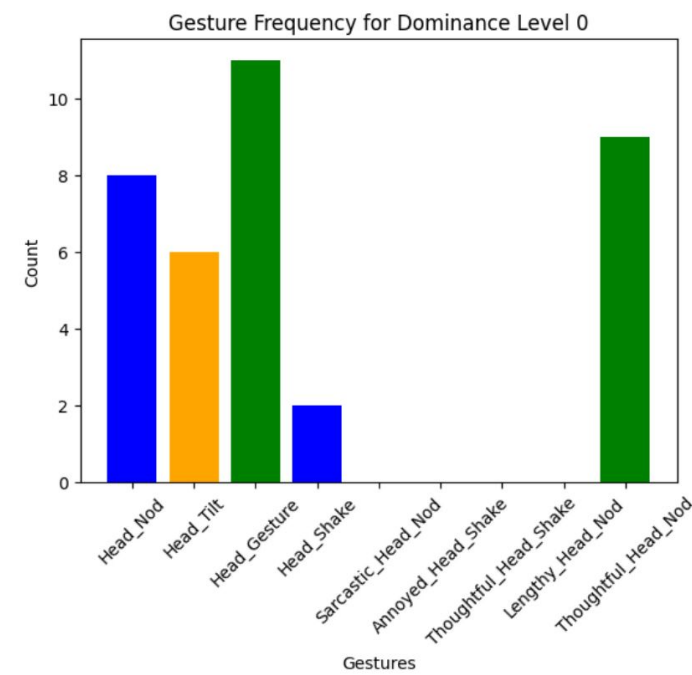
- Dominance 별로 원본 영상, 음소거 처리된 영상을 제작하여 설문지를 제작하였음
- 총 10개의 문항을 5점 리커트 척도로 측정하였음

최종 결과

제스처 생성 분석

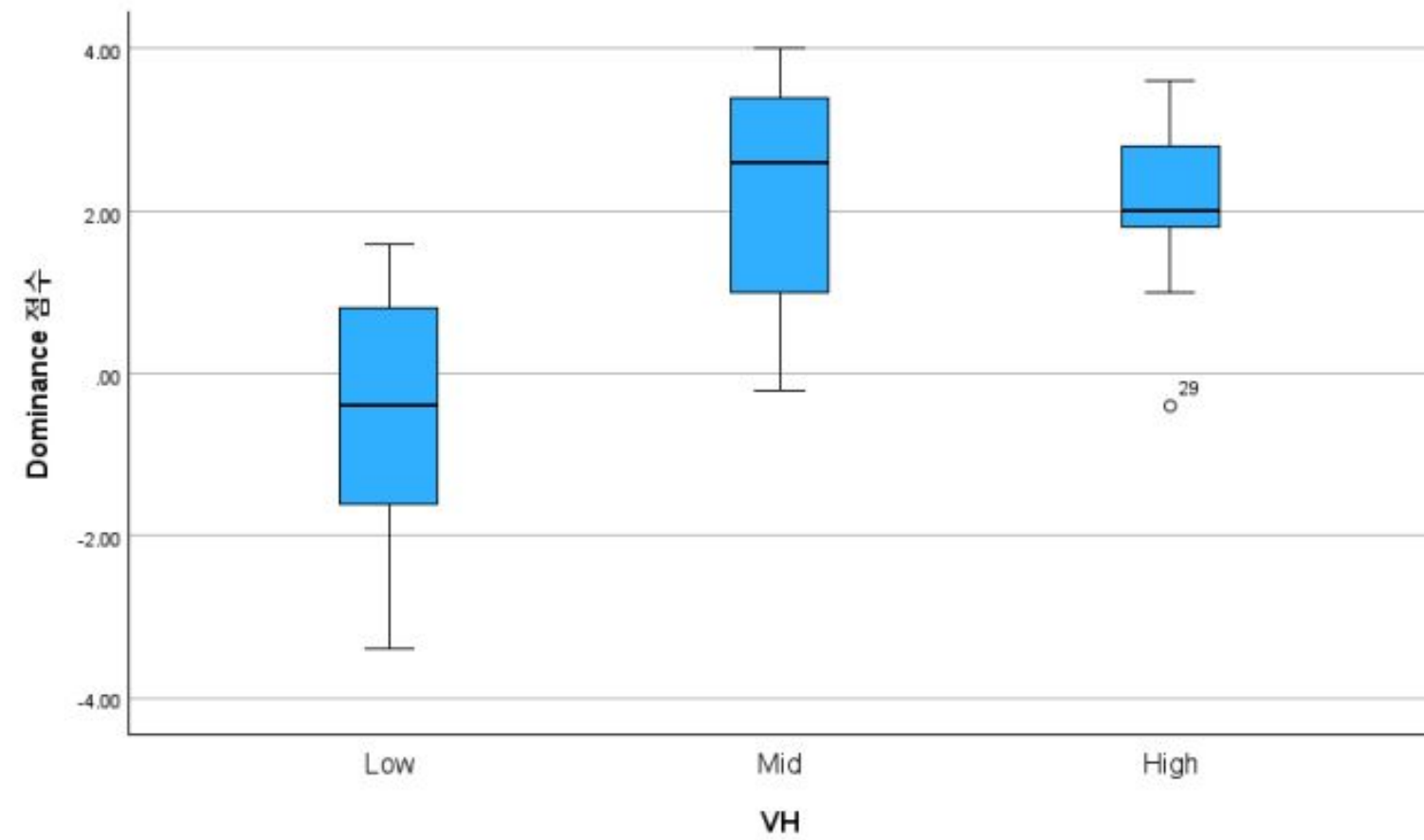


- 10개의 대본을 이용해 각 Dominance 별로 제스처를 생성하게 하였음
- Dominance가 높아질수록 제스처 생성 수도 증가함을 관찰하였음
- 또한 Dominance가 높아질수록 다양한 제스처를 사용하였음

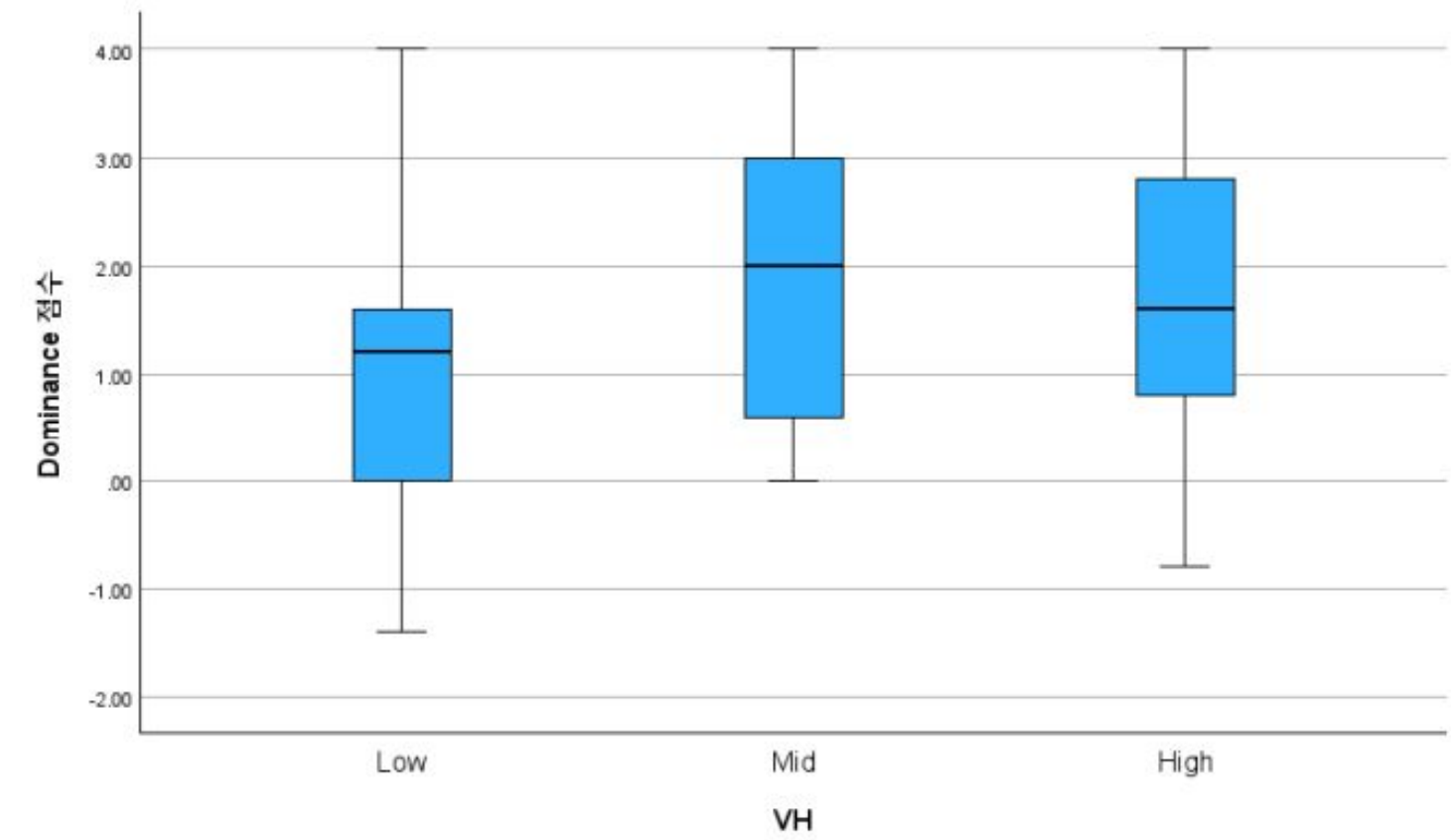


03

설문 결과



음소거 영상의 결과



원본 영상의 결과

설문 분석 결과

다중비교

종속변수: DominanceScore
Games-Howell

(I) VH	(J) VH	평균차이(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한	상한
Low	Mid	-2.62857 [*]	.43427	<.001	-3.6855	-1.5716
	High	-2.55238 [*]	.36751	<.001	-3.4522	-1.6525
Mid	Low	2.62857 [*]	.43427	<.001	1.5716	3.6855
	High	.07619	.36781	.977	-.8244	.9768
High	Low	2.55238 [*]	.36751	<.001	1.6525	3.4522
	Mid	-.07619	.36781	.977	-.9768	.8244

*. 평균차이는 0.05 수준에서 유의합니다.

음소거 영상

다중비교

종속변수: DominanceScore

	(I) VH	(J) VH	평균차이(I-J)	표준오차	유의확률	95% 신뢰구간	
						하한	상한
Tukey HSD	Low	Mid	-.96190	.41822	.063	-1.9670	.0432
		High	-.74286	.41822	.186	-1.7479	.2622
	Mid	Low	.96190	.41822	.063	-.0432	1.9670
		High	.21905	.41822	.860	-.7860	1.2241
	High	Low	.74286	.41822	.186	-.2622	1.7479
		Mid	-.21905	.41822	.860	-1.2241	.7860

원본 영상

- 유의확률이 0.05보다 작을 경우 유의하다고 판단함
- 음소거 영상에서는 Low-Mid, Low-High 쌍에서 유의함
- 원본 영상에서는 유의한 결과가 없음
- 미리 정해진 대본에 따라 생성된 영상이기에 말투와 같은 청각적 요소의 지배성이 중화되었을 가능성이 있음

향후 연구 방향

- 향후 연구에서는 지배적, 중간, 순종적 이외에 성격을 더욱 세밀하게 구분하여 비언어적 요소를 생성하는 방향으로 발전 가능
- 디지털 휴먼이 사용자와 상호작용할 때, 사용자의 감정을 인식하고 적합한 반응을 제공하는 연구
- 디지털 휴먼 기술이 보편화됨에 따라 상호작용 과정에서 발생할 수 있는 심리적 영향이나 사회적 변화 연구