

画像生成された面接志願者の第一印象の主観評価

我那覇・日置・西山

背景: 仮想空間で実施される面接において姿勢プロンプトとテキストプロンプトの組み合わせがデジタルヒューマン志願者の第一印象を良くすることができるかは評価されていない

目的: プロンプトの組合せにより生成された志願者の第一印象の良し悪しを主観評価

仮想空間



- 基本テキストプロンプトを用いてAIで生成



良い第一印象を感じない

デジタルヒューマン志願者

面接官

仮想空間



- 基本テキストプロンプト
 - 姿勢プロンプト
 - 追加テキストプロンプト
- を組み合わせるとAIで生成



良い第一印象を感じる

デジタルヒューマン志願者

面接官

結果: 姿勢プロンプトを与えていない時に追加テキストプロンプトを与えても第一印象は良くなるとはいえない。ただし、両者を組み合わせることにより第一印象が良くなる

姿勢プロンプトとテキストプロンプトとの 組み合わせにより画像生成された 面接志願者の第一印象の主観評価

発表番号: A-1-3

鳥取大学大学院 工学研究科 工学専攻

パターン認識工学研究室

我那覇 航 日置 翔太 西山 正志

2024年12月11日

研究背景

- 仮想空間の面接においてデジタルヒューマンを使う取り組みが注目されている

人間の外見や振る舞いを写実的に模倣した仮想空間上の存在

- 現実空間でも仮想空間でも、面接は**面接官**と**志願者**で構成



- 仮想面接でデジタルヒューマンを使用する既存研究

- デジタルヒューマン面接官と仮想空間における面接トレーニングし本番面接に備える

[Smith et al., ASD'20]

- 現実空間における本人自身の外見とは異なる外見をもつ

デジタルヒューマン志願者が仮想空間で面接官とインタラクションを行う

分析研究が最近登場 [Kim et al., CCSCW'23]

**本研究ではデジタルヒューマン志願者が現実の本人の外見とは異なり
架空の人物の見た目をもつ場合について考える**

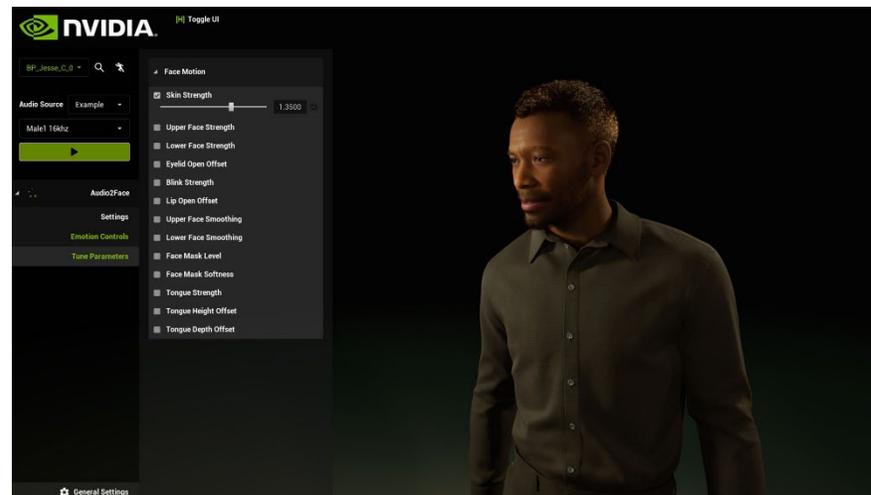
デジタルヒューマン志願者の生成時に考えること

面接官がデジタルヒューマン志願者の外見から感じる第一印象は面接において重要
デジタルヒューマンの外見の生成手法



MetaHuman

<https://www.unrealengine.com/ja/metahuman>



AvatarCloudEngine

<https://developer.nvidia.com/ace>

第一印象を良くするためには現実の人間の外見に近い方が好ましい場合が多い
ただし、上記サービスでは上半身に注力したデジタルヒューマンを生成

第一印象を良くするために
志願者の全身が面接官から見えるよう画像生成されることが望ましい

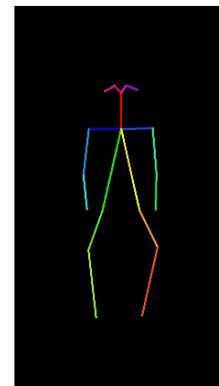
生成AIによる志願者デジタルヒューマン

全身のデジタルヒューマン画像を生成するために
Stable Diffusion 等の画像生成AIを使用することを考える
[Robin et al., CVPR'22]

□ 画像生成のために一般的に**テキストプロンプト**を駆使

- 基本テキストプロンプト: 人物の特徴や服装や姿勢やコンテキストを表す
- 追加テキストプロンプト: 生成された結果に補足したい要素を表す
テキストプロンプトのみでは, 志願者の姿勢を適切に制御することが難しい

→ ControlNet [Zhang et al., ICCV'23] を用いて**姿勢プロンプト**を与えることで
姿勢が制御された画像を生成することをねらう



課題

姿勢プロンプトとテキストプロンプトの**組み合わせ**が
デジタルヒューマン志願者の第一印象を良くすることができるかは評価されていない

本研究の目的と仮説

姿勢プロンプトと追加テキストプロンプトとの組合せにより生成された志願者について仮想空間における面接を想定した際の第一印象の良し悪しを主観評価

仮説: 姿勢制御の姿勢プロンプトと生成結果を補足する追加テキストプロンプトとの組合せにより生成されたデジタルヒューマン志願者から仮想空間の面接官は良い第一印象を感じる

仮想空間



デジタルヒューマン志願者

■ 基本テキストプロンプト
を用いてAIで生成



面接官

良い
第一印象を
感じない

仮想空間



デジタルヒューマン志願者

■ 基本テキストプロンプト
■ 姿勢プロンプト
■ 追加テキストプロンプト
を組み合わせるAIで生成



面接官

良い
第一印象を
感じる

実験条件: 刺激画像の設定

デジタルヒューマン志願者の刺激画像を設定し、実験協力者にその第一印象を回答させる主観評価を実施

□ 刺激画像の設定

- 以下の4条件でデジタルヒューマン志願者を生成



C1: 姿勢プロンプトなし
追加テキストプロンプトなし

C2: 姿勢プロンプトあり
追加テキストプロンプトなし

C3: 姿勢プロンプトなし
追加テキストプロンプトあり

C4: 姿勢プロンプトあり
追加テキストプロンプトあり

- 画像生成AI: Stable Diffusion [Robin et al., CVPR'22]
- AIモデル: BeautifulRealisticAsians <https://civitai.com/models/25494/beautiful-realistic-asians>
- 実験協力者1名あたりに提示する画像数: $6(\text{枚}) \times 2(\text{性別}) \times 4(\text{条件数}) = 48(\text{枚})$

実験条件: 使用したプロンプト

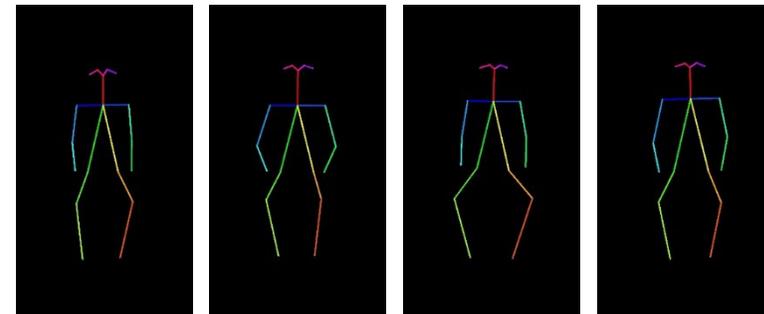
□ テキストプロンプト

基本テキストプロンプト (男性)	full-body, best-quality, realistic, a-20-year-old-Japanese-man, wearing-black-business-suit, derby-tie, seated-posture-on-a-folding-chair, no-background
基本テキストプロンプト (女性)	full-body, best-quality, realistic, a-20-year-old-Japanese-woman, wearing-black-business-suit, seated-posture-on-a-folding-chair, no-background
追加テキストプロンプト	smile, hands-on-lap, look-at-viewer

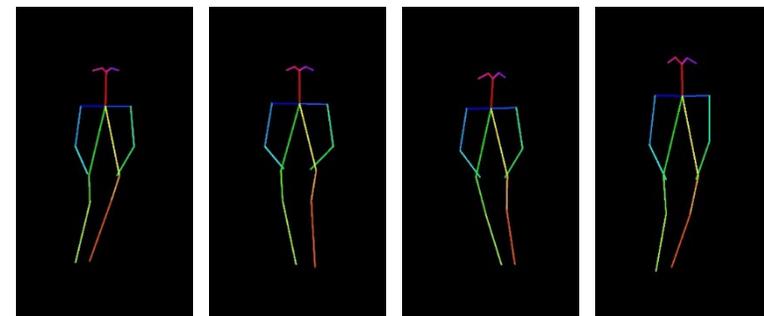
- 基本テキストプロンプト:
人物の特徴・服装・姿勢・コンテキストを表す単語
- 追加テキストプロンプト:
表情・姿勢・視線のばらつきを統制するよう
生成結果を補足する単語

□ 姿勢プロンプト

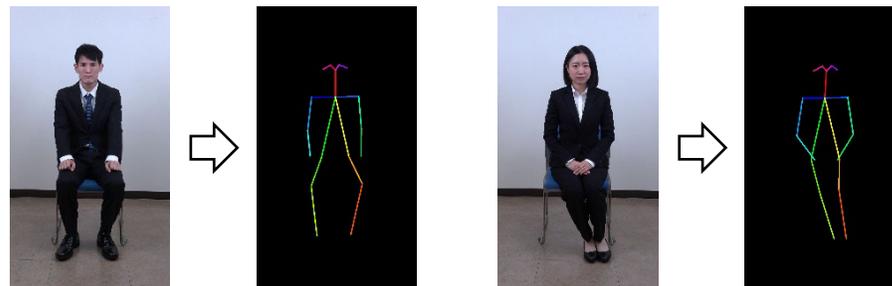
姿勢プロンプト
(男性)



姿勢プロンプト
(女性)



- 人物画像から姿勢推定し姿勢プロンプトを作成



実験条件: 主観評価

□ 前提条件

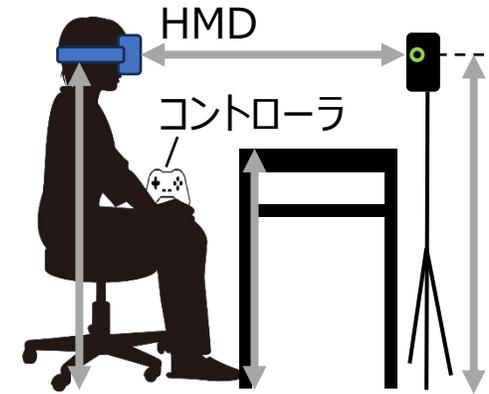
- 実験協力者: 15名 (男性14名, 女性1名, 平均年齢 23.9 ± 2.8 歳, 日本人)
- 仮想空間を想定しHMDを使用して刺激画像を提示

□ 質問と回答方法

- 質問Q1: あなたは志願者の画像から良い第一印象を感じるか
- 質問Q2: あなたは志願者の画像から悪い第一印象を感じるか
- 回答方法: 5段階で評価しコントローラで回答

□ 実験の流れ

- P1: 質問と回答方法を説明
- P2: 基準画像を2秒間表示
- P3: 刺激画像を9秒間表示
- P4: 実験協力者が質問Q1とQ2とを回答
- P5: 48枚の刺激画像について完了するまでP2~P4繰り返し



参考：最も第一印象の良い/悪い志願者画像(Q1)

Q1の
スコアが
高い画像



2.5



4.0



4.3



4.0



3.3



2.8



4.6



4.9

平均スコア

Q1の
スコアが
低い画像



1.7



1.3



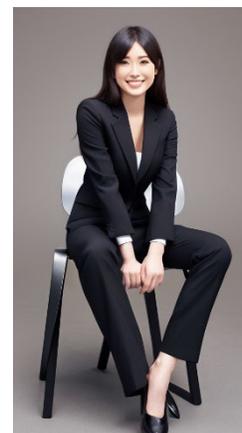
1.6



2.5



2.0



1.9



3.7



4.2

平均スコア

C1: 姿勢プロンプトなし
追加テキストプロンプトなし

C2: 姿勢プロンプトあり
追加テキストプロンプトなし

C3: 姿勢プロンプトなし
追加テキストプロンプトあり

C4: 姿勢プロンプトあり
追加テキストプロンプトあり

参考：最も第一印象の良い/悪い志願者画像(Q2)

Q2の
スコアが
高い画像



4.4



4.5



4.2



3.2



3.8



3.9



2.3



1.8

平均スコア

Q2の
スコアが
低い画像



3.0



1.9



1.3



1.4



2.7



2.9



1.3



1.1

平均スコア

C1: 姿勢プロンプトなし
追加テキストプロンプトなし

C2: 姿勢プロンプトあり
追加テキストプロンプトなし

C3: 姿勢プロンプトなし
追加テキストプロンプトあり

C4: 姿勢プロンプトあり
追加テキストプロンプトあり

実験結果：二元配置分散分析

Q1とQ2の主観スコアにAligned Rank Transformを適用し二元配置分散分析

[Wobbrock et al., HFCS'92]

質問	要因	F値	p値	主効果	交互作用
Q1 (第一印象が良い)	姿勢プロンプト	255.69	< .001	あり	-
	追加テキストプロンプト	43.38	< .001	あり	-
	姿勢プロンプト× 追加テキストプロンプト	35.48	< .001	-	あり
Q2 (第一印象が悪い)	姿勢プロンプト	247.25	< .001	あり	-
	追加テキストプロンプト	15.41	< .001	あり	-
	姿勢プロンプト× 追加テキストプロンプト	13.75	< .001	-	あり

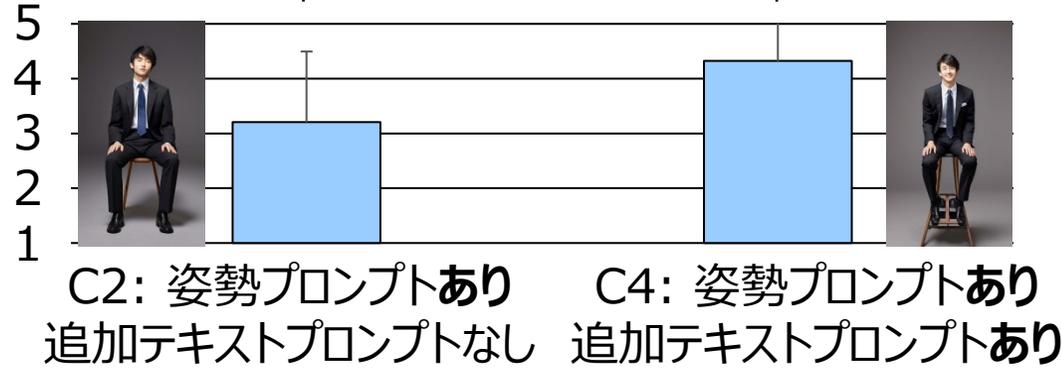
**主効果や交互作用があることから、姿勢プロンプトと追加テキストプロンプトを
組合せることで第一印象が変化するといえる**

実験結果：単純主効果（Q1 良い第一印象を感じる）

Steel-Dwassによる多重比較

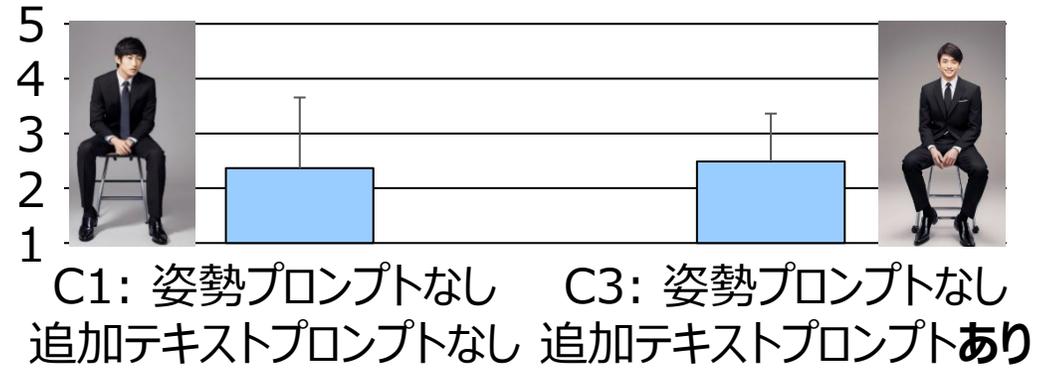
***: 単純主効果あり ($p < .001$)

主観スコアの平均値



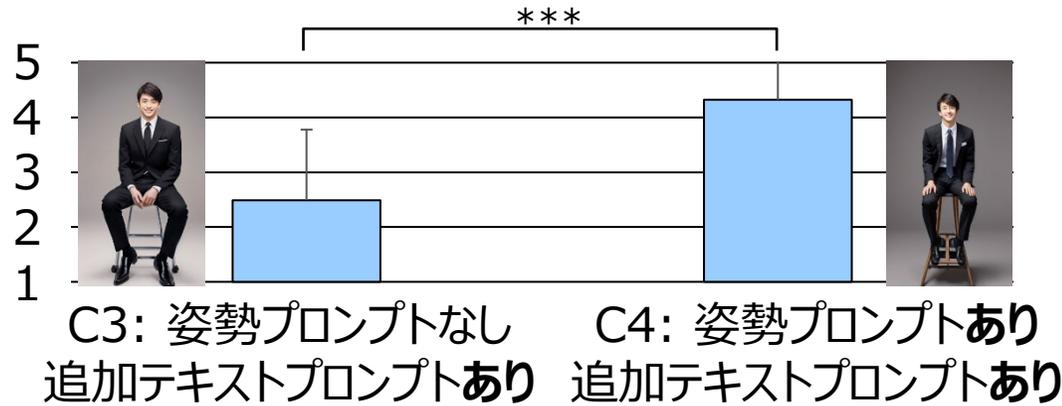
主観スコアが**向上**

主観スコアの平均値



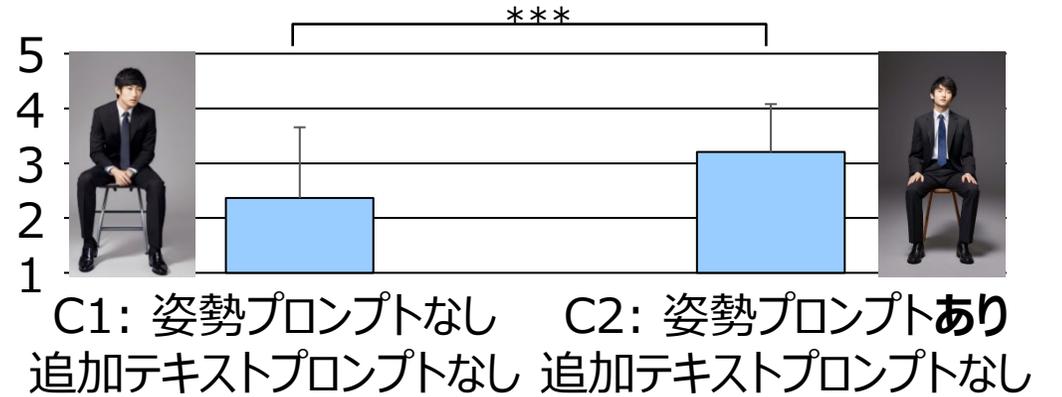
主観スコアに**有意差があるとはいえない**

主観スコアの平均値



主観スコアが**向上**

主観スコアの平均値



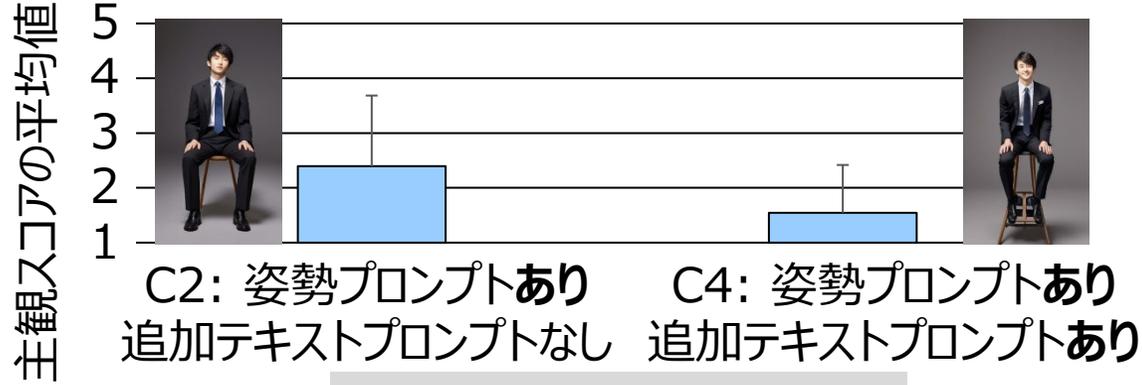
主観スコアが**向上**

姿勢プロンプトを与えていない時に追加テキストプロンプトを与えても主観スコアに有意差があるとはいえない。ただし、両者を組み合わせることにより主観スコアは向上

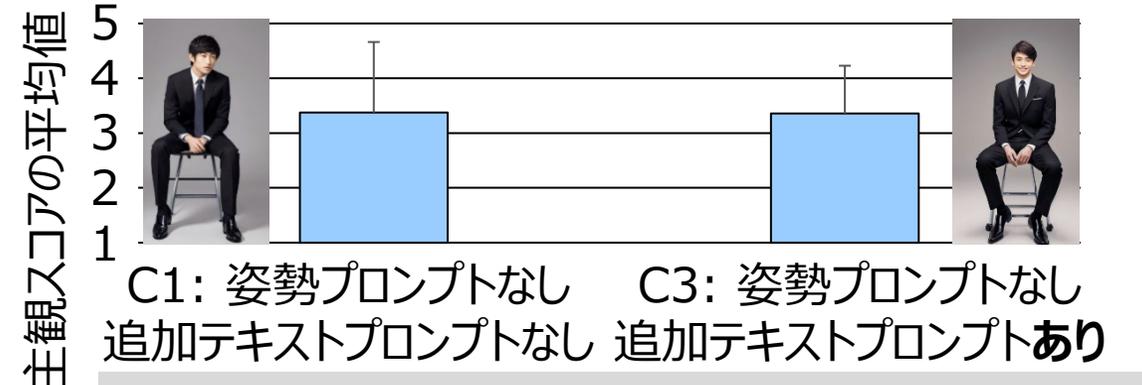
実験結果：単純主効果（Q2 悪い第一印象を感じる）

Steel-Dwassによる多重比較

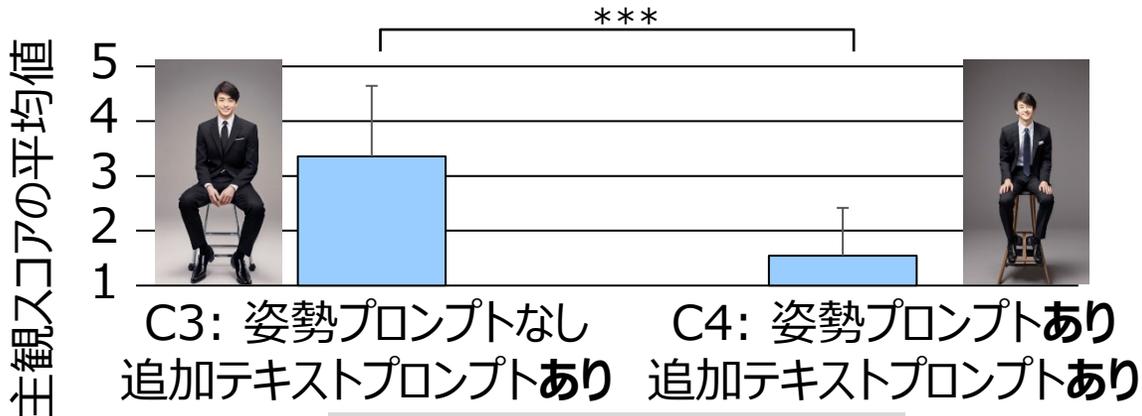
***: 単純主効果あり ($p < .001$)



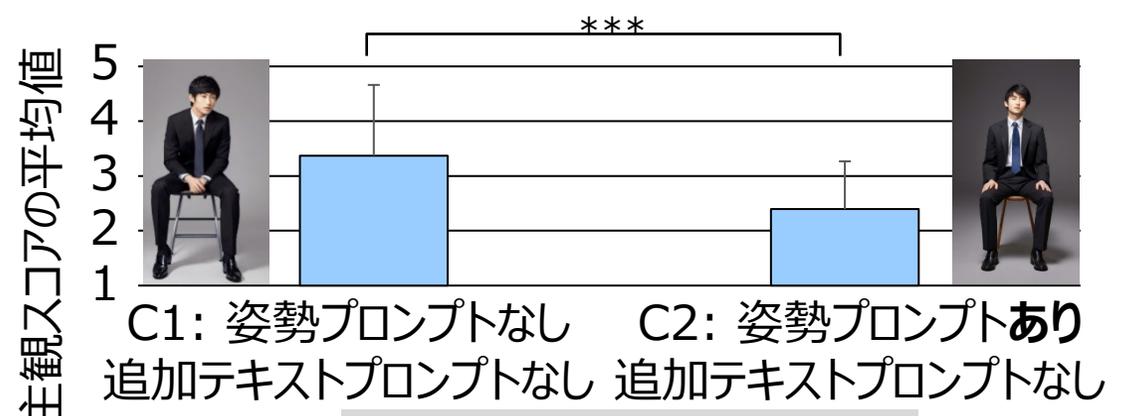
主観スコアが**減少**



主観スコアに**有意差があるとはいえない**



主観スコアが**減少**



主観スコアが**減少**

姿勢プロンプトを与えていない時に追加テキストプロンプトを与えても主観スコアに有意差があるとはいえない。ただし、両者を組み合わせることにより主観スコアは減少

まとめ

目的: プロンプトの組合せにより生成された志願者の第一印象の良し悪しを主観評価

仮想空間



- 基本テキストプロンプトを用いてAIで生成



面接官

良い第一印象を感じない

デジタルヒューマン志願者

仮想空間



- 基本テキストプロンプト
 - 姿勢プロンプト
 - 追加テキストプロンプト
- を組み合わせるとAIで生成



面接官

良い第一印象を感じる

デジタルヒューマン志願者

結果: 姿勢プロンプトを与えていない時に追加テキストプロンプトを与えても第一印象は良くなるとはいえない。ただし、両者を組み合わせることにより第一印象が良くなる

今後の課題: どのような姿勢や単語がデジタルヒューマン志願者の第一印象に強く影響するか調査