

デジタルツインによる対話シミュレーション映像の事前視聴 を通じた関係構築支援 ～対話自体・対話相手の印象形成への影響～

石井 陽子^{a)} 石井 亮^{b)} Lidwina Andarini^{c)} 松尾 和哉^{d)} 大塚 淳史^{e)}

概要：本研究では、人デジタルツイン技術を初対面者間の関係構築支援に応用する可能性を検討した。参加者の内面（対話スタイル等）および外面（外見・音声・動作）を再現したアバターを用い、スピードデーティング場面において、対面前にアバター同士の対話映像を視聴することが関係形成に与える影響を検証した。実験の結果、映像視聴は対話前の心理的準備状態を高め、不安を軽減し、相手への親近感を促進する傾向が示された。さらに、映像を見た後に対話を行った相手の方が、映像を見ずに対話を行った相手よりも印象が良いという結果が得られた。

1. はじめに

科学技術の進歩により、我々は時間、距離、さらには身体的制約を超えたコミュニケーションを可能にしてきた。近年、仮想的なデジタル空間であるメタバース上では、アバターを操作することで多様な人とのコミュニケーションを取る環境が整いつつある。こうしたデジタル空間において、現実空間とほぼ同一のデジタルコピーを「デジタルツイン」と呼ぶ [6]。デジタルツインを活用することで、現実世界のシミュレーションや未来予測が可能となり、さらに現実世界ではコストがかかる分析や試験を繰り返し実施することができる [4]。現在、特に注目されているのは「人のデジタルツイン」の実現である [11][24]。これは人のデジタル表現であり、医療履歴や生理学的データ・行動パターンに基づいて、パーソナライズされた健康管理の提案などに活用されている [20]。また、心理的および物理的側面における行動や反応のシミュレーションが可能となり、最適な意思決定や行動パターンに関する貴重な洞察を提供することも可能である。

本研究では、人のデジタルツイン技術を人間関係の構築促進に応用した結果について報告する。良好な人間関係の構築はウェルビーイング (well-being) にとって重要であ

り、効果的な関係性の形成は孤立感の軽減や生活の質の向上に寄与すると考えられる [2]。人のデジタルツインは、言語的・非言語的なやりとりの双方を行うことが可能で、とりわけ初対面の同士の他者における関係性構築を促進する新たなアプローチとなる可能性がある。本研究では、初対面の人同士の対話の前にデジタルツインが対話を行ったシミュレーション結果を本人にフィードバックすることで、対話や相手への印象にどのような影響を与えるかを実験的に検証する。特に本稿では、SpeedDating といった構造化された環境下で初対面の人々が交流する場면을対象に検討を進める。

本研究におけるリサーチクエスションは以下である。
RQ：初対面の人同士が、自身および相手のデジタルツインによる対話映像を事前に視聴することが以下についてどのような影響を与えるのか。

1. 本人同士の対話への影響について
2. 対話後の対話相手の印象について

これらを通じて、デジタルツインを活用した社会的インタラクションモデルの有効性を評価し、現実世界でのコミュニケーションおよび対人関係の促進における応用可能性を検討する。

本研究は国際会議 INTERACT2025 において発表した論文を日本語に翻訳したものであり、内容は現論文と本質的に同一である [7]。

¹ NTT 株式会社 人間情報研究所

a) yoko.ishii@ntt.com

b) ryoct.ishii@ntt.com

c) lidwina.andarini@ntt.com

d) kazuya.matsuo@ntt.com

e) atsushi.otsuka@ntt.com

2. 関連研究

2.1 初対面同士のコミュニケーション支援

初対面同士のコミュニケーションを支援する方法として、さまざまなアプローチが提案されてきた。例えば、事前に収集した個人情報に基づきエージェントを活用し、対話の促進や関係性の構築を図る研究が行われている。

SpeedDatingのような二者間対話を対象とした研究ではないものの、複数人の対話にロボットを加えて話題を提供し、円滑な対話を促進する研究が報告されている [19]。この研究では、ロボットが参加者の経験を収集・整理し、意味的な類似性を発見することで、個人の写真と結びつけたグループ共有の物語を形成し、対話の活性化やグループの一体感を高めることを目的としている。また、Zhangの研究では、ヒューマノイドロボットがファシリテーターとして深い対話を引き出すことに成功したと報告されている [23]。これらの研究は、ユーザー同士の情報交換を円滑にする点で有効である一方、エージェントの介在によって本来の人間同士の自然な対話が妨げられ、直接的なコミュニケーションが抑制される可能性も指摘されている。

一方、本研究で提案する手法では、対話の前に参加者のデジタルツイン同士の対話を視聴するが、対話への自由度が担保される。そのため、人間同士の直接的な対話を抑制するような問題は生じにくい。初対面同士の対話においては、友人との対話と比較して一時的な情緒状態としての「状態不安」を感じやすいことが報告されている [16]。これは、初対面では相手についての情報が不足しているため、「話題や話し方に注意を払う必要がある」「話題探しに苦労する」といった認識を持ちやすく、対話の難易度が高く感じられるためである。もしこうした対話における困難さを軽減できれば、ロボットなどのエージェントの積極的な解材を必要とせず、円滑な対話を実現できる可能性があると考えられる。本研究では、外部からの直接的なファシリテーションを行うことなく、対話時の不安感を軽減し、自然な対話の流れを促進するアプローチを探索する。

2.2 人のデジタルツインの活用方法

デジタルツイン技術は、リアルタイムのモニタリングやシミュレーションを可能にする手法として、医療・スポーツ分野での活用が進んでいる。

医療分野では、個々の患者のデータを基にデジタルモデルを構築し、健康状態のモニタリングや治療効果のシミュレーションに活用されている [5][14]。一方、スポーツ分野においては、アスリートの動作解析やトレーニングプログラムの最適化に活用され、パフォーマンス向上を支援している [1]。

このように、現在のデジタルツイン技術は、個人の健康

や能力向上を目的とした個人単位での最適化に焦点を当てており、対人関係の構築を支援する応用例はこれまで報告されていない。本研究は、デジタルツインを人と人との相互作用の支援に応用するという新たな試みを提案するものであり、この分野における独自の貢献を示すものである。

2.3 人のデジタルツインの再現方法

本研究では、人の外的特徴と内的特徴の双方を再現するデジタルツインの実現を目指している。本節では、これまでの研究において提案されてきた、内面と外面の再現技術に関する関連研究を概観し、既存研究を踏まえて、関係性構築に必要なデジタルツインの要件を明らかにする。

2.3.1 内面の再現

人の内面とはその人の経験してきたことや趣味などの記憶をモデル化することであり、これにより一貫したキャラクター性を持たせることや自律的に発話を行うことができ、対話相手と自然な対話を行うことが可能となる [18]。内面は目に見えるものではないため、予測・理解しそれらを忠実に再現する、といったステップが必要になる。人の内面を形成するものは、その人の持つ価値観や思考、感性などが複雑に関係しており、そうした脳の機能を再現するために計測・再現の一部について成功しているものの、複雑な内面の完全な再現は困難な課題と言える。人の内面は複雑であり、対話といった複雑な感情を含んだコミュニケーションを行うため、大量の対話データを学習させることでその人自身らしい発話内容を出力する対話システムの実現手法が、現在主流の再現方法と言える [22]。本研究でもこのアプローチに従い、個人の対話データを用いて、その人物の発話特性を再現する対話システムを構築することで、内面の再現を目指す。

2.3.2 外面の再現

見た目や声を生成系 AI により再現する手法については、近年多種多様な手法が出てきており [13]、こうした技術を活用することで仮想空間内のアバターとして外面をフォトリアルに再現することもできるようになってきた [9]。そうした自己のアバター (Virtual Doppelgänger) を本人が操作した場合、アバターとの繋がり感といった主観的なエンゲージメントが向上することが確認されているとの報告もある [12]。しかしその一方で、フォトリアルなアバターを活用することは、プライバシー保護の観点などの懸念点が指摘されている [10]。特に本研究のように、初対面者間の関係構築支援を目的とする場合、実用的なプライバシー保護策を講じずに写実的なアバターを使用することは潜在的なリスクを孕む可能性がある。この問題に対応するため、本研究においては、視覚的外見についてフォトリアルでのアバターを再現するのではなく、本人に似たデフォルメさせたアニメ風のアバターでの外見の再現に留めることとする。

外面の再現は視覚的外観にとどまらず、ジェスチャーや身体動作、音声の再現も含まれる。たとえば、Yamahaは著名なピアニストであるグレン・グールドの演奏をAIにより学習し再現するシステムを開発しており、本人が演奏していない曲も演奏させることが可能である [21]。同様に本研究でも、デジタルツインのインタラクションにおけるリアリズムを高めるため、ジェスチャーや音声特性の再現精度を可能な限り高めることを目指す。

3. システム構成

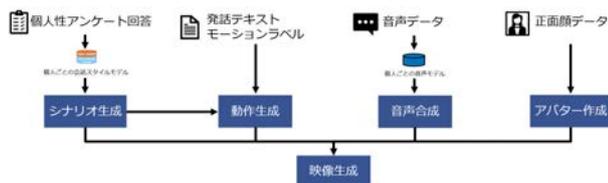


図 1 データフロー

本章では、デジタルツイン同士の対話をシミュレーションするために設計されたシステム構成について述べる。全体のデータフローは、図 1 に示す。

3.1 内面の再現：シナリオ生成

本節では、内面の再現として、参加者のデジタルツイン同士の対話シナリオを生成する手法について述べる。対話データを用いた事前学習においては、両者のプロフィール情報から、プロフィール情報に基づいて発話を生成できる発話生成モデル (Persona Pair Prompt Dialogue Model) を作成した [25]。このモデルは、X (旧 Twitter) の複数のリプライペアから成る対話ログとリプライ内容から推定したプロフィール情報とを用いて、パラメタ数 222M の Transformer Encoder-Decoder model を事前学習したモデルである。学習には、杉山らが取得した twitter (現:X) の返答のペア群と話者のプロフィールが含まれるデータセットを用いて [18]、Transformer の Encoder-decoder モデルの事前学習を行う。さらに、モデルを各参加者の個性に適合させるために、事前学習済みモデルを各参加者の過去の Speeddating 対話の書き起こしデータでファインチューニングした。これにより、各参加者固有の対話スタイルに合わせたモデルへの適応が可能となる。

モデルを用いた具体的なシナリオ生成方法は以下である。本モデルは、入力プロンプトに対し続く発話を生成するように学習させているため、まず話題となるはじめの発話の入力を準備する必要がある。このはじめの発話で何を話すかによって生成される対話の話題が異なるため、SpeedDating の場で事前に行われるプロフィール交換の場で提示され、かつ、対話の中でも良く話される話題を 7 つ選定した。選定した話題の具体的な内容は、後述する表 1

内、1: 個人性アンケートの話題生成カテゴリに記載した。この 7 つの話題について各参加者の回答を事前取得し、7 つの話題それぞれにはじめの発話を手動で作成する。はじめの発話を入力としてあたえ、話者 A のモデルを用い話者 A を模擬した発話を生成し、次にこの生成結果を入力とし話者 B のモデルを用い話者 B を模擬した発話を生成する。これら手順を繰り返すことで、一つの話題に対して 10 発話を生成し、7 つの話題全ての生成結果をつなげることで対話シナリオを生成した。生成された対話シナリオの中には一部意味の通らない箇所や支離滅裂な箇所があったため、それらについては、人手で確認し該当箇所を削除することで対応した。また、その他助詞が異なっている箇所については手動で正しい助詞へ修正を行っている。1 話題分のシナリオ出力例を以下に載せる。

- A: そういえば、ドライブやテーマパークデートがしたいんですね
 B: はい、そうですね。ドライブとかも好きです
 A: 私もドライブ好きです。どこ行きたいとかありますか?
 B: そうですね、最近では京都とか行きたいなと思ってます
 A: いいですね。京都はおしゃれな感じで
 B: そうですね。京都の、お寺とかも行きたいなあって
 A: お寺はいいですね。あと、京都の清水寺のあたりとかも
 B: 清水寺は、なんか雰囲気がいいですよ
 A: そうですね。あと、京都の八坂神社とかも
 B: 八坂神社、私も行ってみたいです
 A: はい。京都の八坂神社、めっちゃいいですよ

シナリオ中、A1 の出だしの発話は、B がデートで行きたいと事前に回答していた「ドライブやテーマパーク」について、「ドライブやテーマパークデートがしたいんですね」と A が確認する発話として手動で作成した発話となる。対話シナリオは参加者のペアごとに 3 種類出力し、出力されたシナリオの中から人手での修正後、最も発話文字数が多かった対話シナリオを映像生成用のシナリオとした。

また、生成された対話シナリオについては、話題のうちランダムに 2 つを選択し、関係性が一歩進められるよう、次の約束をする発話を話題の最後に挿入している。上記で示したシナリオ出力例では、最後の A の対話の後に、「B: 今度よかったら八坂神社一緒に行きませんか? (提案)」 「A: 良いですね。週末の予定を確認してみます (受け入れ)」といった 2 つの発話を挿入している。関係性が一歩進められる発話は、A (提案) + B (受け入れ)、B (提案) + A (受け入れ) の両方のパターンを対話の中で 1 つずつ再現した。

また、シナリオには各発話ごとに感情を付与する。付与する感情は Happy, Surprised, Sad, Angry, Neutral とした。この感情ラベリングには ChatGPT-3.5 を用い、各発話に対して最も適切な感情カテゴリを自動的に割り当てた。

3.2 外面の再現：アバター生成・音声合成・モーション生成

本研究において、人のデジタルツインはアバターとして仮想空間内に表現する。アバターは、正面顔の画像を与えることで、そこから顔の特徴、髪型、髪、肌の色といったアバター用の特徴を抽出し、本人に似たアバターをVRM形式で生成する。アバターの声については参加者の音声から声色の特徴を抽出し、その特徴を再現した音声を生成する形で実現し、生成されたシナリオの通り発話させる。アバターの動きは、発話テキストごとに表情およびジェスチャーを学習したモデルを構築し、テキスト入力からそれに対応する動きを生成できる、テキスト情報から動作を生成するアプローチ [8] を採用している。数十時間分の実際の間人同士の対話データから収集した「発話テキスト」と「全身モーションラベル」のペアデータを用いて機械学習モデルを訓練した訓練済みモデルは、入力された発話テキスト内の各単語に対応する全身モーションラベルを生成し、その結果に基づいて自然で人間らしいアバターの動作を実現する。また、表情については、各発話ごとに付けられた感情に合わせ変化させた。本来であれば、参加者ごとの実際の動作データや表情に基づいて個別にモーション生成を行うことが望ましい。しかし、こうしたデータを収集するには大規模なデータセットが必要となるため、本研究では代表的なサンプル個人のモーションデータを使用し、参加者ごとのモデルカスタマイズは行わなかった。

3.3 映像生成

映像生成における流れを述べる。参加者の声を用い参加者毎にアバターの音声モデルを生成する。次に感情ラベルが付与された対話シナリオを用いアバターの動作を生成する。次に、UnrealEngineを用い、対話を行うアバター2体を空間上に配置し、各発話単位で音声・モーションに合わせアバターの挙動を生成する。生成した映像上では、発話内容を吹き出しにも提示させた。参加者のペアごとにこれら映像を生成した。映像は、平均で3~4分程度の映像を生成した。出力映像例を図2に示す。



図2 出力映像例

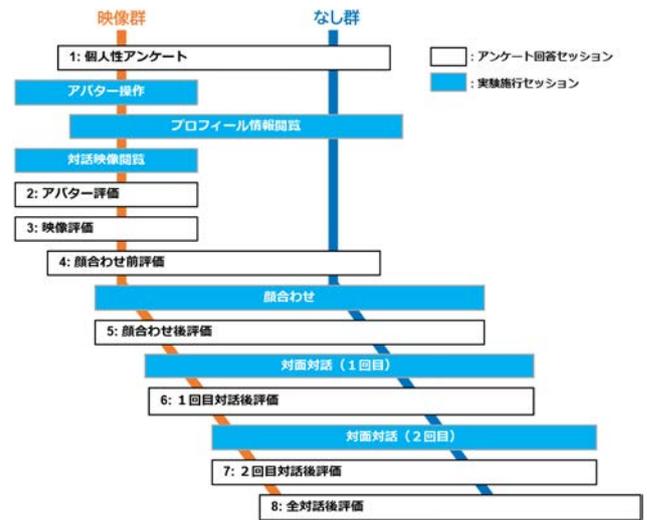


図3 実験の流れ

4. 実験

4.1 概要

本研究では、初対面の男女による SpeedDating 形式の対話を対象とした。参加者はいずれも SpeedDating を通じた出会いを希望する一般公募者であり、こうしたイベントの典型的な年齢層である 20~40 代を主な対象とした。SpeedDating では、引き続き連絡を取りたい相手同士で連絡先を交換することが行われており、本実験においても各対話の終了後、参加者には相手との連絡先交換の希望有無を尋ね、双方が交換を希望した場合のみ「マッチ成立」とみなした。

実験では、以下の2つの群を設定した：

- 映像群：参加者は、実際の対面対話の前に、自身と相手のアバター同士の対話映像を視聴した。
- なし群：参加者は、自身と相手のアバター同士の対話映像を視聴せずに対面対話を行った。

参加者は、本研究で使用する個人化対話モデルの学習データとして事前に対話データを提供した人物の中から選出された。

実験には、男性 16 名・女性 16 名（いずれも日本語母語話者）が参加した。女性参加者の平均年齢は 35.4 歳 (SD = 6.5)、男性参加者の平均年齢は 35.2 歳 (SD = 5.9) であった。各参加者は、それまで一度も対話したことがない新しい相手とペアを組み、実験に臨んだ。

実験では、各参加者が異なる 4 人の相手と対話を行い、そのうち 2 回を映像群、2 回を映像なし群で実施し、各群につき合計 32 試行が行われた。順序効果を最小限に抑えるため、各参加者においてどちらの群にて試行するかは可能な限りランダム化した。

表 1 アンケートリスト

アンケート種類	内容 (太字はカテゴリ, () 内は設問数)	回答形式
1: 個人性アンケート	<p>話題生成用 (7): 7: 趣味, 特技, 好きな映像作品, 好きな音楽について, 休日の過ごし方について, マイブームについて, デートで行きたい場所について</p> <p>プロフィール情報 (21): 年齢, 居住地区, 出身地, 身長, 血液型, 最終学歴, 職業, 勤務地, 休日, 年取, 家族, ペット有無, 同居有無, 結婚歴, 子の有無, 喫煙有無, 飲酒有無, 希望結婚時期, 好きなタイプ, 好きなランキング 1-3 位, 話してみたい話題</p> <p>モチベーション (2): 知り合いになることが楽しみだ, 連絡先を交換したい</p>	自由記述 主に自由記述 (有無は選択肢) 6 件法
2: アバター評価	<p>自身のアバターについて (15): [A] は人間らしいか, [A] に違和感があったか, [A] は自身に似ていたか</p> <p>相手のアバターについて (10): [A] は人間らしいか, [A] に違和感があったか</p> <p>相手への期待 (5): 相手へ [A] が近いことを期待するか [A]: 見た目, 動き, 発話内容, 口調, 声</p>	6 件法 6 件法 6 件法
3: 映像評価	内容理解 (3): 対話の雰囲気は良かった, 対話は盛り上がっていた, 内容を理解できた	6 件法
4: 顔合わせ前評価	<p>対話について (14): 楽しそう, 真剣に取り組みそう, リラックスできそう, 満足できそう, 雰囲気良くできそう, やる気を持ってそう, 集中できそう, 関心を持ってそう, 親しみをもちそう, 盛り上がりそう, 緊張しそう, 不安感がある, 期待している, うまく対話ができそう</p> <p>相手について (6): 相手と仲良くなれそう, もっと知りたい, 会うことに緊張している, 会うことに不安感がある, 期待している, 会うのが楽しみだ</p> <p>共通話題について (4): 話題が見つかった, 話題がイメージでき安心感が出た/不安感が減った/緊張感が減った</p> <p>自分のことを知ってもらえる安心感 (4): 知ってもらえた, 知ってもらえて安心感が出た/不安感が減った/緊張感が減った</p> <p>相手の情報について (6): よく理解できた, 見た目/動き/発話内容/口調/声が予想できる</p> <p>対話への期待 (3): 話題をうまく広げてくれそう, 話の展開がうまそう, こちらに合わせて自己開示してくれそう</p>	6 件法 6 件法 6 件法 6 件法 6 件法 6 件法
5: 顔合わせ後評価	4: 顔合わせ前評価と同じ	6 件法
6:1 回目対話後評価	4: 顔合わせ前評価と同じ Love-Liking 尺度 (26)	6 件法 6 件法
7:2 回目対話後評価	6:1 回目対話後評価と同じ	6 件法
8: 全対話後評価	<p>対話相手に対するのランク付け (1): 対話相手へのランク付け</p> <p>連絡先交換 (1): 連絡先を交換しても良いか (各対話相手ごとに回答)</p>	1-4 yes/no

4.2 実験手順

実験の流れを図 3 に示す。図 3 中, 1-8 で記載されているアンケートについては次節にて説明する。映像群, なし群ともに, アバター生成・確認セッションにおいて, 参加者はまず自分自身のアバターを作成・確認した。アバターは, 参加者の正面顔画像から抽出した顔の特徴, 髪型, 髪色, 肌の色をもとに生成した。また, 参加者が希望する場合には, これらの特徴を手動で選択または修正することも可能とした。アバターの服装については 13 種類を用意し, 好みのものを自由に選択できるようにした。さらに服装の色についても変更を希望する参加者には対応した。

アバター作成後, 参加者は 10 分間のアバター操作セッションを行った。このセッションでは, PC カメラを通じてアバターが参加者本人の動作や表情をリアルタイムに反映するようにし, アバターの挙動や外観を確認できるようにした。この過程には WebcamMotionCapture^{*1}を用いた。

次に, プロフィール情報閲覧セッションにて, 対話相手に関する情報を互いに閲覧した。この際に閲覧する情報は, 表 1 の 1: 個人性アンケートのプロフィール情報カテゴリの 21 項目とした。

映像群においては, 対話映像視聴セッションにより, これから対話を行う相手のアバターと自身のアバターが対話を行う映像を視聴した。映像は, ペアの参加者同士は同じ映像を用いた。

次に映像群/なし群ともに顔合わせセッションを行った。このセッションでは「よろしくお願いします」程度の挨拶のみとし, まだ対話を行わないよう指示した。

その後, 対面対話セッションを実施した。本セッションは 5 分間と 10 分の 2 回行い, 対話の話題に制限は設けず自由対話形式とした。

各セッションの前, もしくは後にはアンケート回答を行っている。本実験で取得したアンケートを表 1 に載せる。各アンケートは図 3 の流れで取得している。2: アバター評価, 3: 映像評価は映像群のみ実施している。8: 全対話後評価は, 4 名の異性との対話が全て終わった段階で実施しており, 対話を行った 4 名について, 印象が良かった順に 1-4 でランクをつけさせた。さらに, 連絡先を交換したいかどうかを 4 名の対話相手ごとに回答させた。

ほとんどの質問項目は 6 件法のリッカート尺度で評価された。回答選択肢は以下のとおりである:

とてもそう思う (6), そう思う (5), ややそう思う (4), ややそう思わない (3), そう思わない (2), 全くそう思わ

*1 <https://webcammotioncapture.info/>

ない (1)

また、参加者が対話相手に対して抱く情動的反応を評価するために、本研究では Love-Liking Scale [15] を用いた。この尺度は、対人印象を評価する際に「愛情 (love)」と「好意 (liking)」を区別して測定するものである。

5. 結果

第一章にて記載した RQ を以下に再掲する。これらについて本章では分析結果を示す。

RQ: 初対面の人同士が、自身および相手のデジタルツインによる対話映像を事前に視聴することが以下についてどのような影響を与えるのか。

1. 本人同士の対話への影響について
2. 対話後の対話相手の印象について

3つ目の項目に対しては、対話前後での違いやすべての対話後における違いについて、対話や相手についての思いをアンケート結果から明らかにする。

分析対象とするアンケート回答今回、本人を再現したアバター同士の対話を生成するため、過去に我々の行った SpeedDating 実験へ参加した参加者を再度呼び実験を行っている。しかし、これら参加者の中には、過去の実験時には連絡先交換への意欲が高かったものの、今回の実験ではモチベーションが低下している可能性があるものも含まれる。この対話に対してのモチベーション低い参加者は対話や相手に対しての興味が低い可能性を懸念し、これら参加者の回答は本研究における分析対象からは外すこととした。低モチベーションの参加者は、1: 個性アンケートのモチベーションカテゴリの「連絡先を交換したい」において、3 以下 (ややそう思わない, 以下) と回答した参加者とし、これらを分析対象から除外した。結果、男性 14 名、女性 15 名の有効回答が得られ、合計 116 試行 (映像群: 58 試行/なし群: 58) を分析対象とする。有効回答のうち、47 試行では少なくとも一方の参加者が 8: 全対話後評価の連絡先交換カテゴリ「連絡先を交換しても良いか」に対し yes と回答し、そのうち 6 ペアが相互に連絡先交換を希望し、マッチ成立となった。

以降、本章の表に示す分析結果は、グラフ中に特記が無い限り縦軸は平均評定値を表し、エラーバーは標準誤差を表している。

5.1 自身および相手のデジタルツインをどのように知覚するか (2: アバター評価)

今回、我々は人のデジタルツインをアバターとして仮想空間内に再現した。参加者が自分自身、もしくは相手のアバターをどのように感じたのか確認するため、2: アバター評価の分析結果を述べる。図表 2 の通り、アバターの人間らしさについては、自身のアバターも相手のアバターもそれぞれの要素 (見た目、動き、発話内容、声のトーン、音

声) においてニュートラルな結果となっている。違和感については有意さは無いものの、自身のアバターの評価においてやや平均値が高く、違和感を感じる値がやや高かったことがわかる。自身のアバターが自分に似ているかについても有意差は無く、動きの評価が他と比べてやや低くなっているが、ほぼニュートラルな評価であった (表 3)。

表 2 2: アバター評価 (人らしさ, 違和感) 結果

設問	自身のアバター M(SD)	相手のアバター M(SD)
見た目は人間らしい	3.89(1.11)	3.86(1.02)
動きは人間らしい	3.45(1.21)	3.51(1.11)
発話内容は人間らしい	3.85(1.25)	3.91(1.12)
口調は人間らしい	3.74(1.23)	3.81(1.10)
声は人間らしい	3.65(1.25)	3.80(1.10)
見た目に違和感があった	3.73(1.18)	3.25(1.16)
動きに違和感があった	3.86(1.22)	3.45(1.05)
発話内容に違和感があった	3.90(1.45)	3.09(1.23)
口調に違和感があった	3.48(1.35)	3.08(1.14)
声に違和感があった	3.23(1.34)	2.95(1.15)

表 3 2: アバター評価 (自身に似ていた) 結果

設問	M(SD)
見た目は自身に似ていた	3.16(1.25)
動きは自身に似ていた	2.70(1.04)
発話内容は自身に似ていた	3.24(1.35)
口調は自身に似ていた	3.25(1.43)
声は自身に似ていた	3.46(1.63)

5.2 デジタルツイン間のやり取りを、全体的印象の観点からどのように評価するか (3: 映像評価)

本節では、映像群がどのように映像を受け止めたのか確認するため、3: 映像評価の内容理解カテゴリの分析結果を述べる。表 4 に結果を示す。結果から、アバター同士の対話の雰囲気や盛り上がりはややポジティブな印象を示したことが分かった。また、「内容を理解できた」については高い評価が得られており、参加者が映像内の対話内容を比較的容易に理解できたことが示唆された。

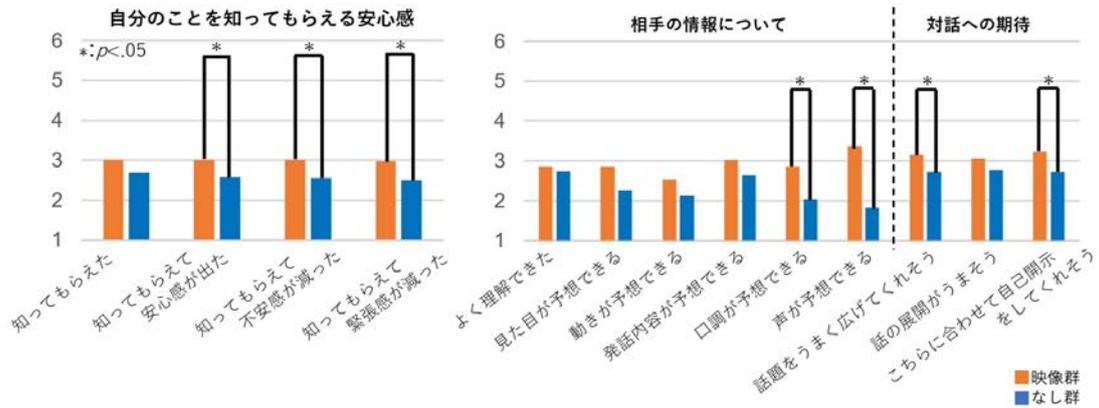
表 4 3: 映像評価 結果

設問	M(SD)
対話の雰囲気は良かった	3.83(1.26)
対話は盛り上がっていた	3.58(1.31)
内容を理解できた	4.75(1.08)

5.3 対話への影響: 対話前での比較 (4: 顔合わせ前評価, 5: 顔合わせ後評価)

映像群となし群の間で、対話前にどのような違いがみられるかを確認する。映像群/なし群間の比較を行うために t 検定を実施した結果、有意な差が確認された。

4:顔合わせ前評価



5:顔合わせ後評価

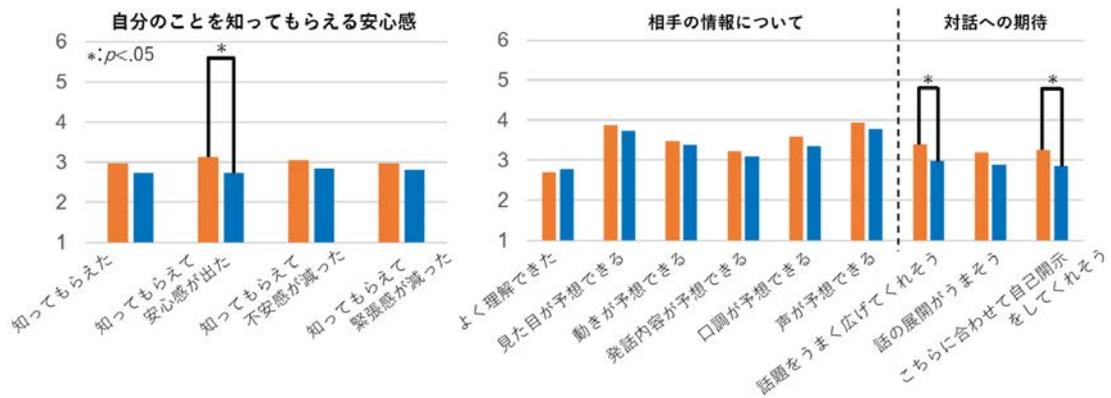


図 4 上 4: 顔合わせ前評価 結果 / 下 5: 顔合わせ後評価 結果

4: 顔合わせ前評価 (図 4 上) :

- “自分のことを知ってもらえる安心感 カテゴリ”
“知ってもらえて安心感が出た, $t(114)=-2.36, p=.02$ ”, “知ってもらえて不安感が減った, $t(114)=-2.45, p=.01$ ”, “知ってもらえて緊張感が減少した, $t(114)=-2.46, p=.016$ ”, の項目において、映像群の方が有意に高い評価を示した。
- “相手の情報について カテゴリ”
“見た目が予想できる, $t(114)=-2.56, p=.012$ ”, “口調が予想できる, $t(114)=-3.66, p=.00$ ”, “声が予想できる, $t(114)=-7.69, p=.00$ ”, の各項目において映像群が有意に高い評価を示した。
- “対話への期待 カテゴリ”
“話題をうまく広げてくれそう, $t(114)=-2.42, p=.017$ ”, “こちらに合わせて自己開示してくれそう, $t(114)=-3.2, p=.00$ ”, において映像群が有意に高い評価を示した。

5: 顔合わせ後評価 (図 4 下) :

- “自分のことを知ってもらえる安心感 カテゴリ” と “対話への期待 カテゴリ”:
“知ってもらえて安心感が出た, $t(114)=-2.02, p=.046$ ”, “話題をうまく広げてくれそう, $t(114)=-$

2.1, $p=.038$ ” and “こちらに合わせて自己開示してくれそう, $t(114)=-2.06, p=.041$ ” の各項目において、映像群の方が有意に高い評価を示した。

5.4 対話への影響：対話相手への Love-Liking 尺度での比較 (6:1 回目対話後評価, 7:2 回目対話後評価)

対話相手に対する印象の違いを検討するため、Love-Liking Scale での 1 回目と 2 回目の対話後の結果で両群を比較した。

6 : 1 回目対話後評価

- Love 尺度 (Love Scale) : 有意差は認められなかった ($t(114)=-0.46, p > .05$).
- Liking 尺度 (Liking Scale) : 映像群の方が有意に高い評価を示した ($t(114)=-2.00, p=.04$).

7 : 2 回目対話後評価

- Love 尺度 : 有意差は認められなかった ($t(114)=-0.36, p > .05$).
- Liking 尺度 : 有意差は認められなかった ($t(114)=-1.46, p > .05$).

これらの結果から、1 回目対話後においては映像群の方が「好意 (Liking)」得点が有意に高かったものの、2 回目対話後ではその差は消失し、群間での有意差は見られなかった。

5.5 対話相手の印象への影響 (8: 全対話後評価)

参加者は合計で4名の対話相手と交流し、うち2名は映像群とし、残り2名との対話はなし群として対話を行った。4名の対話相手との対話が全て終わった後、参加者はそれぞれの対話相手に対する総合的な印象に基づき、4名の対話相手の印象が良かった順に1-4でランクをつけさせた結果を図5に載せる。ランクの1,2位については、映像群が多く、ランクの3,4についてはなし群が多かった様子が図から読み取れる。

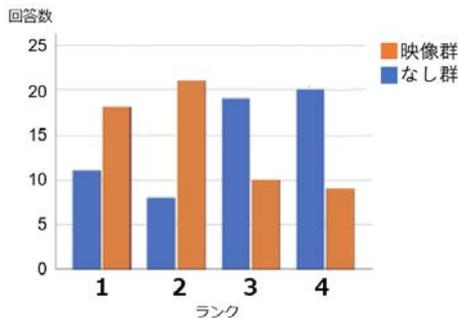


図5 対話した相手のランク付け結果

Wilcoxon の符号付き順位検定を用いて平均順位と比較を行った結果、映像群 (平均=2.2, SD=0.7) となし群 (平均=2.8, SD=0.7) の間に有意な差が認められた ($Z = -2.4$, $p = .016$)。このことから、映像群における対話相手へのランクが、なし群よりも有意に高かったことを示している。

6. 考察

6.1 RQ: 初対面の人同士が、自身および相手のデジタルツインによる対話映像を事前に視聴することの影響

本研究では、参加者の内的と外的の両方を再現するアバターを構築した。本節では、これらのアバターがどの程度実際の人物を反映していたか、また参加者がその相互作用をどのように知覚したかについて考察する。

6.1.1 アバターや対話の受け止め方、本人同士の対話への影響について

2: アバター評価の分析結果によると、参加者は自分自身のアバターに対して強い不自然さや違和感を感じてはいなかったが、同時に自分に非常に似ているとも感じていなかった。すなわち、全体としてニュートラルな評価が示された。また、アバター間の対話映像における雰囲気や盛り上がり等の評価もおおむねニュートラルであった。一方で、「内容を理解できた」に関しては高い評価が得られており、生成された対話が理解しやすい内容であったことが示唆される。さらに、アバター同士の対話映像を視聴した参加者からは、否定的な意見はほとんど見られず、全体として肯定的で受容的な反応が確認された。これらの結果から、参加者はアバター同士の対話映像を本人同士の対話前に視聴することに対して抵抗感なく受け入れたことが示唆される。

映像群では、「口調が予想できる」「声が予想できる」「話題をうまく広げてくれそう」「こちらに合わせて自己開示してくれそう」といった対話に関する期待が高かった。さらに、「知ってもらえて安心感が出た」「不安が減った」「緊張感が減った」といった情動的安定に関する項目でも、映像群が有意に高い評価だった。このことから、アバター同士の対話映像を事前に見ることで相手の情報を知り、その結果として相手と対話をする前の安心感が高まり、緊張感や不安感が減り、なし群よりも状態不安を感じにくくなる効果があることを示唆している。さらには、話題をうまく広げてくれそう、こちらに合わせて自己開示してくれそう、といった対話への期待も上がり、対話前の心理的な準備が促進されている可能性が考えられる。

さらに、Love-Liking Scale のうち「Liking 尺度」においては、1回目の対面対話セッション後に映像群が有意に高い評価を示した。このことから、アバター同士の対話映像の視聴が、互いに親しみを感じる基盤を作ることに有効である可能性がある。一方で、この有意差は2回目の対面対話セッションでは減少、もしくは消失した。これは、対話を重ねるにつれて、映像群の参加者も相手への理解を自然に深めたためと考えられる。したがって、これは映像群の効果が失われたことを意味するのではなく、なし群の参加者が対話を通じて追いついた結果、映像群/なし群の差が縮まったと解釈できる。

6.1.2 対話後の相手の印象への影響について

対話を行った相手の印象についてランクづけを行ってもらった結果について、映像群の平均順位が有意に高かったことから、なし群と比較して、映像群の参加者同士は相手に対して肯定的な印象を形成していたことが推察される。つまり、対話映像を事前に視聴することで、対話自体への影響のみならず、対話後の対話相手の印象自体にも影響を及ぼす可能性が示唆された。この結果には、いくつかの要因が関係していると考えられる。

- 要因 1: 心理的負担の軽減と期待一致効果

前述の結果からも示唆されるように、映像群の参加者は状態不安を感じにくくなる効果が示唆されており、対話に対する心理的負担が軽減されていたことが考えられる。

社会心理学の分野では、気分と判断の一貫性 (mood-judgment consistency) に関する研究が広く行われている [3]。これらの知見に基づけば、ポジティブな気分や低い不安状態にある参加者ほど、対話相手に対しても好意的な印象を形成しやすいことが示唆されている。また、映像視聴によって対話に対する期待が高かった状態が確認された。こうした期待が実際の対話体験と一致することで、ポジティブな印象が強化された可能性が考えられる。今後その効果の確認、および大きさについても確認していくことが求められる。

● 要因 2: 対話シナリオと非言語的要素の影響

もう一つ考えられる要因として、対話シナリオ内容による影響が考えられる。今回の映像群で提示されたシナリオには、対話の最後に「次の約束をする」という発話が意図的に挿入されていた。このような発話は、相手との関係を次のステップへ進める意図を明示し、互いに受け入れられているという感覚を生じさせる効果があった可能性がある。さらに今回のシナリオでは 1: 個人性アンケートから話題を生成したことから、互いに自己開示を行う形でシナリオが進んでいる。この様子を見ることで、相手への印象にポジティブな影響を及ぼした可能性も考えられる。初対面の対話において、互いに自己開示を行う対話の方が相手への好意は親密度が高くなることが研究で報告されており [17]、本研究においても同様のメカニズムが働いた可能性がある。加えて、アバターの表情、口調、動作といった非言語的要素も印象形成に寄与したと考えられる。これら非言語的要素については、対話シナリオから生成されるため、映像を通じて言語的、非言語的情報ともに相手がオープンで協調的な態度を示す様子を見たことで、実際の対面対話において、より好ましい印象形成につながった可能性がある。

本研究では、デジタルツインを仮想空間におけるアバターとして再現し、そのアバター同士の対話の視聴が対話相手に対する印象形成に影響を与えることが明らかになった。しかしながら、どの要因がポジティブな印象形成に寄与したのかについては依然として明確ではない。今後の研究ではこれらの効果を構成する具体的要因を個別に抽出・分析し、そのメカニズムをより詳細に検討する必要がある。

6.2 倫理的影響

音声クローン技術の倫理的影響や、AI によって再現された「自己」による違和感など、潜在的な負の影響については十分に考慮する必要がある。近年の研究では、音声生成技術がアクセントによって性能差を示すことが報告されており、その結果として言語的特権の強化やデジタル排除のリスクが生じる可能性が指摘されている [29]。また、Hutiri らはこの種の技術に関連する倫理および安全性上の懸念を体系的に整理し、包括的な分析を行っている [28]。さらに、音声クローン技術の悪用防止を目的とした技術的セーフガードに関する研究も進められており、Bird らおよび Sakka らによる取り組みが報告されている [30][31]。

今回の実験では、自分のアバターに対しての結果は 5.2 節に記載したとおりネガティブな評価ではなく、ニュートラルな評価として出ており、不快感や負の影響のもととなりそうなネガティブな反応は見られなかった。しかし、こうした反応は状況依存적であり、発話内容や提示環境の文脈によって影響を受ける可能性がある。したがって、今後

の応用に際しては慎重なデザインと文脈への配慮が不可欠である。特に、これら手法を社会実装あるいはサービスの提供として行う場合、本人の同意・説明可能性など、複製された人格表現に関して慎重に検討を行う必要がある。したがって、今後の発展に向けて、倫理的リスク評価の強化と、インフォームド・コンセントを確実に担保するための枠組み構築の必要性を強調する。

6.3 本研究の限界

本研究にはいくつかの限界が存在する。

まず第一に、実験参加者が日本人男性 16 名・女性 16 名と比較的小規模であり、かつ特定の文化的背景に限定されている点が挙げられる。したがって、本研究の結果を他の文化圏に一般化するには注意が必要である。本研究は日本文化圏を前提としており、先行研究では文化的要因が認知や対人評価に影響を与えることが指摘されている [27][26]。Markus らの研究によれば、東アジア文化圏では「相互依存的自己観 (interdependent self-construal)」が特徴的である一方、西洋文化圏では「独立的自己観 (independent self-construal)」がより強調される傾向にある [26]。このような文化差を踏まえると、本研究の知見が西洋社会にそのまま適用できるとは限らない。したがって、今後の研究では、異なる文化的文脈における再現性の検証や、その背後にある心理的メカニズムの比較分析が求められる。

第二に、本研究で観察された効果が、初対面後の実際の関係構築過程や長期的な対人評価においても持続するかどうかは明らかではない。この点を解明するためには、より大規模なサンプルを用いた実験的検証や、長期的な追跡評価を行う必要がある。

7. まとめ

初対面の人どうしの対話において、デジタルツインによる事前シミュレーション映像の視聴が関係構築に寄与するかを明らかにするため、本稿では、SpeedDating といった男女の出会い場面を対象に、互いのデジタルツインを仮想空間上のアバターとして再現し、そのアバター同士の対話映像の視聴が関係形成に与える影響を検証した。参加者の内的側面 (人格・対話スタイル) および外的側面 (外見・音声・動作) を再現したアバターを構築し、それが対話前の情動・相手への印象形成・対人関係構築に及ぼす影響を分析した。

その結果、アバター同士の対話映像を事前に視聴することで、対話前の不安が軽減され、心理的な準備状態が促進される可能性が示唆された。さらに、参加者は実際に対面する前から相手に対してよりポジティブな第一印象を形成できることが示された。一方で、映像群における効果は初期段階で有意に認められたものの、時間の経過とともにその差は減少または消失した。これは、実際の人間同士の自

然な相互作用によって、最終的に映像群となし群の差が埋められることを示唆している。今後の研究では、どの要因が最もポジティブな印象形成に寄与するのかを明らかにするとともに、デジタルツインを活用した対人交流が長期的にどのような影響をもたらすかを、多様な社会的文脈において検証する必要がある。

参考文献

- [1] Barricelli, B.R., Casiraghi, et.al.: Human Digital Twin for Fitness Management, in IEEE Access, vol.8, pp.26637-26664(2020)
- [2] Diener, E., and Seligman, M. E. P.: Very Happy People. *Psychological Science*, 13(1), 81–84(2002).
- [3] Forgas, J. P.: Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117(1), pp.39-66(1995).
- [4] Fuller, A., Fan, Z., et.al.: Digital Twin: Enabling Technologies, Challenges and Open Research, in IEEE Access, vol. 8, pp.108952-108971(2020).
- [5] Gabrielli, S., Piras, E. M., et.al.: Digital Twins in the Future Design of Digital Therapeutics. In Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC2023 Adjunct), pp.602–605(2023)
- [6] Grieves, M., Vickers, J.: Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems; In: Kahlen, J., Flumerfelt, S., Alves, A. (eds) *Transdisciplinary Perspectives on Complex Systems*. Springer, Cham, pp.85-113(2017).
- [7] Ishii, Y., Ishii, R., et.al.: Support for Building Relationships in Speed Dating Through Observation of Pre-dialogue Simulations Using Digital Twins, In *Human-Computer Interaction(INTERACT2025)*, pp.25–44(2025).
- [8] Ishii, R., Higashinaka, R., et.al.: Methods of Efficiently Constructing Text-dialogue-agent System using Existing Anime Character, *Journal of Information Processing*, Vol.29, pp.30-44(2021).
- [9] Li, Z., Zheng, Z., et.al.: Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2024)*, pp.19711-19722(2024).
- [10] Lin, J., Latoschik, M. E.: Digital body, identity and privacy in social virtual reality: a systematic review, *Virtual Reality*, vol. 3(2022).
- [11] Lin, Y., Chen, L., Ali, A. et.al.: Human digital twin: a survey, *J Cloud Comp* 13, 131(2024).
- [12] Lucas, G., Szablowski, E., et.al.: The effect of operating a virtual doppelganger in a 3D simulation, *MIG2016*, pp.167–174(2016).
- [13] Mirsky, Y., and Lee, W.: The Creation and Detection of Deepfakes: A Survey; *ACM Comput. Surv.*54, 1, Article7(2022).
- [14] Papachristou, K., Katsakiori, P. F., et.al.: Digital Twins' Advancements and Applications in Healthcare, Towards Precision Medicine. *Journal of Personalized Medicine*, 14(11), 1101(2024).
- [15] Rubin, Z.: Measurement of romantic love. *Journal of Personality and Social Psychology*, 16, pp.265-273(1970).
- [16] Sasaki, Y.: Study on the Relationship between State-Anxiety and Agari of Japanese College Students and Their Living-Abroad Experiences and Self-Construals in Dyadic Conversations with Strangers and Friends, *Japanese Journal of Communication Studies* Vol.52 No.1, pp.41-64(2023).
- [17] Sprecher, S., and TregerThe, S.: benefits of turn-taking reciprocal self-disclosure in get-acquainted interactions, *Personal Relationships* 22(3), pp.460-475(2015).
- [18] Sugiyama, H., Mizukami, M., et al.: Empirical Analysis of Training Strategies of Transformer-Based Japanese Chat-Chat Systems, *SLT2022*, pp. 685-691(2023).
- [19] Uchida, T., Ishiguro, H., et.al.: Improving Quality of Life with a Narrative Robot Companion: II – Creating Group Cohesion via Shared Narrative Experience, *29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp.906-913(2020).
- [20] Vallée, A.: Digital twin for healthcare systems. *Front. Digit. Health* 5:1253050(2023).
- [21] Yamaha.: Yamaha Dear Glenn Project AI System Gives Concert in Style of Legendary Pianist Glenn Gould at Ars Electronica Festival, Retrieved from(2019).
- [22] Saizheng Zhang, S., Dinan, E., et.al.: Personalizing Dialogue Agents: I have a dog, do you have pets too?, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp.2204–2213(2018).
- [23] Zhang, A. W., Lin, T.-H., et.al.: IceBreaking Technology: Robots and Computers Can Foster Meaningful Connections between Strangers through In-Person Conversations, in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp.1–14(2023).
- [24] Lauer-Schmaltz, M., Cash, P., et.al.: Towards the Human Digital Twin: Definition and Design – A survey, *Ithaca, NY*, pp.1-30(2024).
- [25] Otsuka, A., Matsuo, K., et.al.: User-Specific Dialogue Generation with User Profile-Aware Pre-Training Model and Parameter-Efficient Fine-Tuning, *arXiv:2409.00887(2024)*.
- [26] Markus, H. R., and Kitayama, S.: Culture and the self: implications for cognition, emotion, and motivation, *Psychological Review*, 98, pp.224-253(1991).
- [27] Varnum, M E., Grossmann, I., et.al.: The Origin of Cultural Differences in Cognition: Evidence for the Social Orientation Hypothesis, *Curr Dir Psychol Sci.*2010 19(1), pp.9-13(2019).
- [28] Hutiri, W., Papakyriakopoulos, O., et.al.: Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators, In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, pp.359–376(2024).
- [29] Michel, S., Kaur, S., et.al.: It's not a representation of me: Examining Accent Bias and Digital Exclusion in Synthetic AI Voice Services, *arXiv:2504.09346(2025)*.
- [30] Bird, J. J., Lotfi, A.: Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion, *arXiv:2308.12734(2025)*.
- [31] Sakka, S., Liagkou, V., et.al.: Preserving Voice Message Integrity: Strategies Against AI Cloning, Reliability and Statistics in Transportation and Communication, Human Sustainability and Resilience in the Digital Age. *RelStat* 2024, vol.1337, pp.443–452(2025).