鳥取大学 学位論文

実写アバタのインタラクション開始前動作に関する研究

令和3年7月

工学研究科 情報エレクトロニクス専攻

D18T2103H 宮内 翼

指導教員 岩井儀雄　教授

概 要

本研究では，実際の人物の見た目に近い実写アバタを用いた案内システムの開発を目指している．実写アバタの案内システムを用いることで，実際に案内を行う人の代替となり人同士のようなインタラクションが可能となり，労働力不足の解決や情報リテラシの低い人に対して今までの人と同等のような案内を行うことが期待できる．本研究で提案する案内システムは，等身大の大型ディスプレイに実写アバタを表示し，訪れたユーザに対してインタラクションを行って案内を行う．また，大型ディスプレイの案内が終了した後は，目的地までの道中でも案内を行えるようにモバイルデバイスの小型ディスプレイに実写アバタを表示して案内を行う．本研究では，実写アバタの案内システムの中でも，実写アバタがユーザとインタラクションを開始する前の動作に注目する．実写アバタのインタラクションを開始する前の動作は，インタラクションを開始するタイミングや誰とインタラクションを開始するのかをユーザに伝えることができる重要な要素である．本研究では，実写アバタがインタラクションを開始する前の３つの状況を取り扱う．

１つ目の状況は，ユーザが訪れるのを実写アバタが直立姿勢をとって待つ状況である．実写アバタの直立姿勢に人間に近い動きを加えるために，あたかも人間と同じような身体動揺をもつ実写アバタを生成する手法について考える．提案手法では，実際の人間から計測した身体動揺を実写アバタの待機状態で再現する．提案手法は，直立姿勢を対象としカメラ映像中の人物領域から身体の各部位の振動量を計測する．得られた振動量の時間変化から特徴を抽出し，映像中の時刻をランダムに遷移することで，実写アバタの身体動揺を任意の時間長で生成する．提案手法により身体動揺を再現した実写アバタは，インタラクション開始前において人間に近い動きをすることを主観評価で明らかにした．

２つ目の状況は，実写アバタを利用したいユーザが複数存在し，複数ユーザが実写アバタの周りに立って案内を待っている状況である．この状況では，実写アバタは次の案内を行う対象者として複数ユーザの中から一人を選択してインタラクションを開始する必要がある．提案手法では，インタラクションを開始する前に実写アバタが対象者の方を向く動きを加えることで，対象者は実写アバタが自身

の方を向いていると感じ，非対象者は実写アバタが自身の方を向いていないと感じさせることを狙う．これにより，実写アバタは対象者のみを選択してインタラクションを開始することができる．実写アバタに実装する動きは，実写アバタの体を対象者の方向に向ける動作効果，射影変換を用いて実写アバタが回転して見えるように対象者の方向に向ける回転効果，動作効果と回転効果の組み合わせの動きである．主観評価の結果，インタラクション前に動きを加えることで，対象者は実写アバタが自身の方向を向いていると感じ，実写アバタに選択されていると感じることを明らかにした．

3つ目の状況は，小型ディスプレイにおける実写アバタがユーザとインタラクションを行っていない時の直立姿勢をとっている状況である．この状況では，1つ目の状況の時と同様に，実写アバタに身体動揺の動きを実装することを考える．しかし，小型ディスプレイに実写アバタを表示する場合，身体動揺の動きは非常に小さくユーザが視認することができずに実写アバタが止まって見える可能性がある．そこで，実写アバタの身体動揺の動きを強調することで，実写アバタの動きが自然に見えるようにすることを狙う．強調する方法は，映像中の特定の周波数を持つ動きを強調する既存手法を用いる．主観評価の結果では，実写アバタの身体動揺の動きをノイズなく強調することができれば，ユーザは実写アバタの動きが自然であると感じる可能性があることを明らかにした．

本研究では，実写アバタのインタラクション前の動作に注目し，3つの状況において実写アバタにインタラクション前の動作を実装した．その結果，ユーザは実写アバタの動きは人間の動きに近づいて自然な動きであると感じ，実写アバタはユーザにインタラクションを開始するタイミングや誰にインタラクションを行っているのかを伝えることが可能であることを新たに示した．

目 次

第1章　はじめに

1.1　背景

現在，超少子高齢化社会により若者の人口が減っており労働力不足が社会的問題となっている [1]．これにより，社会に存在する様々なツールやサービスがデジタルに置き換わる DX(デジタルトランスフォーメーション) が進んでいる．その中でも，公共施設の案内や窓口などの実際に人が行っているような対人サービスをデジタル化することは，限界集落のようにサービスを行うはずである若者が少ない場所や人通りの多い公共施設で十分な人員を動員できないような場所で活躍が見込める．しかし，このような元々人が行っていたサービスをデジタル化したものは複雑な操作が必要となり，急激なデジタル化が進んでいる近年では高齢者などの情報リテラシが低い人々はサービスを受けるのに使用方法が分からないなどの障壁が存在する．

こういった問題を解決するために，情報リテラシが低い人々でも使用できるように，従来通りの人同士の会話のようなインタラクションで操作できることが求められている．人同士のインタラクションでは，一人が一方的にインタラクションを行っている訳ではなく，複数の人がインタラクションを交代しながら行っている [2]．そういったインタラクションの中では，インタラクションを行っている場面だけでなく，インタラクションを開始するための様々なアクションが必要となる．このような人同士の複雑なインタラクションを行うシステムはユーザビリティが高く，情報リテラシが低い人々でも簡単に使用できると考えられる．特に公共施設では案内や受付などのサービスを利用する人の往来が多く，人同士のインタラクションを行ってサービスを受けることができるインタラクションシステムの活躍が期待できる．

公共施設などのパブリックスペースで人に置き換わって対人サービスを行うイ

ンタラクションシステムについて考える．このようなインタラクションシステムには，ロボットを用いて人とインタラクションを行うシステムがある．例えば，公共施設やショッピングモールなどで案内を行うロボット [3, 4, 5] ，実際の人物に似せたアンドロイド型のロボット [6, 7]，実際に製品として数多くの施設で活躍しているロボット [8, 9] などが存在する．しかし，ロボットを用いたシステムは導入する場合に安全に十分考慮する必要があり，ユーザに対して危険がないように広い設置スペースが必要となる．小型ロボットを導入することも考えられるが，公共施設のような人通りの多い場所に設置した場合に人に気づかれずに利用されないおそれがある．また，ロボットは可動部が多く定期的にメンテナンスが必要となる．そのため，ロボットを多数の場所に設置したり，公共施設などの人通りの多い場所に設置するのは困難である．他のシステムとしては，ディスプレイにアバタを表示して人とインタラクションを行うシステムがある．例えば，エントランスで施設の案内をするアバタ [10]，看護実習における患者アバタ [11]，博物館で案内をするアバタ [12]，過去の体験を語るアバタ [13] などが実現されている．ディスプレイにアバタを表示するシステムでは，既存のディスプレイを使用することが可能でロボットと比べて少ないスペースに設置することができ，メンテナンスも少なくすむため容易に導入することが可能である．そのため，ディスプレイにアバタを表示して人とインタラクションを行うシステムについて考える．

アバタの種類としては，デフォルメされたキャラクターである CG アバタを用いる場合と実際の人の見た目に近づけた実写アバタを用いる場合がある．CG アバタ [10, 11, 12] は，現実ではありえないキャラクターを作成でき，システムを利用するユーザに対して仮想的な空間を提供することができる利点がある．一方，実写アバタ [13, 14, 15] は実際の人の見た目に近いことから，システムを利用するユーザに対して現実的な空間を提供することができる利点がある．ここでは，実際の人に置き換わって対人サービスを行うことを考えるため，実際の人の見た目に近い実写アバタを用いてユーザと人同士のようなインタラクションを行うシステムについて取り扱う．

実写アバタを用いたインタラクションシステムを取り扱い，その中でも公共施設におけるインフォメーションセンターの代替となるような案内システムについて検討を行う．インフォメーションセンターではチケットの案内や道案内などのサー

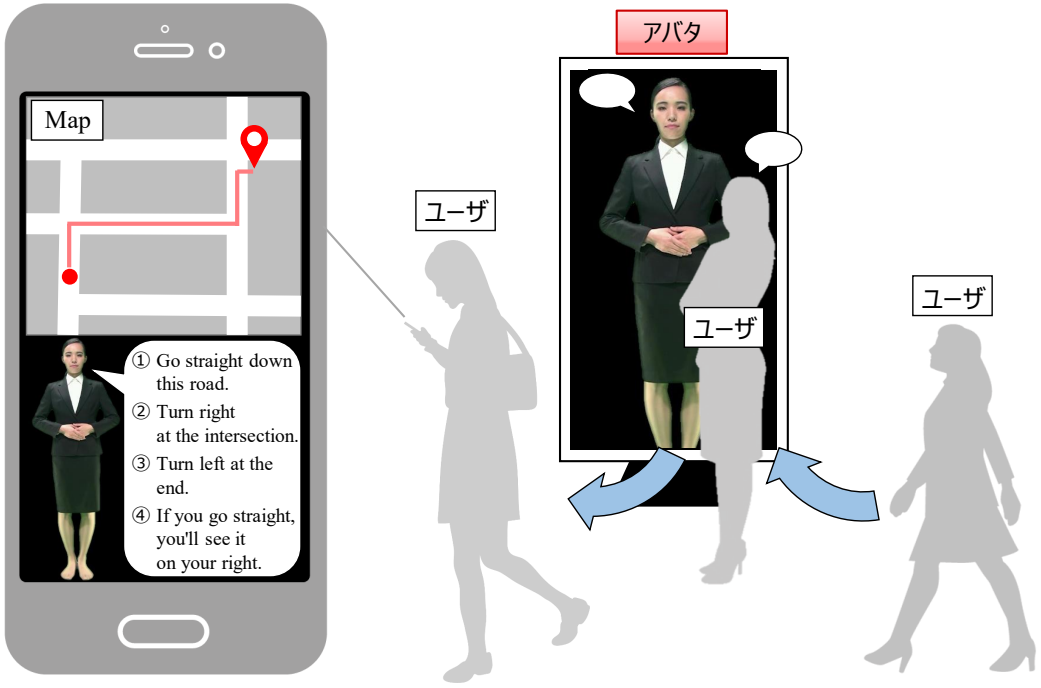


図 1.1: 大小のディスプレイに実写アバタを表示する案内システムの概要.

ビスを行っており，提案する案内システムにおいても同様のサービスをユーザに提供することを考える．実写アバタを用いた案内システムを用いることで，ユーザは実際の人同士のようなインタラクションで操作してサービスを受けることができるため，高齢者のような情報リテラシの低い人でも簡単にシステムを利用できると考えられる．

ここでは，提案する案内システムが提供するサービスの中でも，特に需要が多い道案内における案内の流れを説明する．図 1.1 は提案する案内システムにおける道案内の流れである．まず，実写アバタを等身大の大型ディスプレイに表示する．案内システムを利用したいユーザは大型ディスプレイに表示された実写アバタに近づき，実写アバタとインタラクションをすることで道案内を受けることができる．ここで，目的地までの道のりが非常に複雑な場合で，ユーザが道順を覚えるのが難しい状況を考える．このような状況においてユーザが道に迷わず目的地まで到着するために，提案システムでは大型ディスプレイ上の実写アバタの案内が終了した後に，携帯デバイスなどの小型ディスプレイ上で実写アバタが案内

を行う．既存の研究 [16] によると，地図を指し示すアバタはシステムの好感度を大幅に向上させることが分かっている．そこで，提案システムでは実写アバタと地図を表示してユーザに対して案内を行う．このように大小のディスプレイ上で実写アバタがユーザとインタラクションを行いながら案内を行うことで，多くのユーザが直感的で分かりやすい操作でシステムを利用でき目的地まで迷わずに案内を受け続けることができる．

実際の案内者のように実写アバタがインタラクションを行う案内システムを実現するためには，多くの技術が必要となる．実際の案内者に近いと感じさせる実写アバタを実現するためには，文献 [17] でも述べられているように，インタラクションシステムの全体構成として，ユーザの状況やユーザとの会話を理解する要素技術に加えて，アバタの画像と音声を自然に生成しユーザに伝達する要素技術が必要である．特に本研究では，実写アバタの見た目や動作に直結する重要な技術であるアバタの画像生成 [18, 19, 20, 21, 22] に注目する．アバタの画像生成には，質感や陰影などをつけてリアルなアバタの見た目を生成する手法と動きや表情などのアバタの動作を生成する手法が存在する．その中でも，アバタの動作生成はインタラクションをする上でユーザに視覚的に情報を提供できる重要な要素である．

ここでは，案内システムにおける実写アバタの動作生成について考える．実写アバタを用いた案内システムでは人同士のようなインタラクションが求められている．そのため，実写アバタにも実際の人同士のインタラクションで行われる動作を実装する必要がある．人同士のインタラクションでは，一人が一方的にインタラクションを行っている訳ではなく，複数の人がインタラクションを交代しながら行っている．この場合，インタラクションを行っている最中の動作だけでなく，インタラクションを開始する前の動作も考慮する必要がある．インタラクション開始前の動作は，実写アバタとユーザが円滑にインタラクションを開始するために非常に重要である．既存手法 [18, 19, 20, 21, 22] では，インタラクション中の動作を主に取り扱っており，インタラクションを開始する前の動作は十分に考慮されていなかった．そこで，本研究では実写アバタとユーザがインタラクションを円滑に開始するために，実写アバタにインタラクション開始前の動作を実装することを考える．

図 1.2: インタラクション開始前動作が必要な状況.

1.2 インタラクション開始前動作が必要な状況

本研究で提案する実写アバタを用いた案内システムでは，実写アバタはユーザとインタラクションを行って案内する．そのため，実写アバタとユーザがインタラクションを開始して案内を行うために，実写アバタはインタラクション開始前に動作を行ってユーザと円滑にインタラクションを開始する必要がある．提案する案内システムにおける案内の流れにおいて，インタラクション開始前動作が必要な状況を図 1.2 に示す．提案する案内システムにおける案内の流れは，ユーザが大型ディスプレイに表示された実写アバタの案内を受けるために，実写アバタに近づき声をかける．その後，ユーザは据置の大型ディスプレイに表示された実写アバタから案内を受けた後に，携帯モバイルなどの小型ディスプレイの案内に切り替わる．この流れにおいて，実写アバタとユーザがインタラクションを開始するためのインタラクション開始前動作が必要な状況について考える．

1.2.1 状況 1

図 1.3 のように実写アバタを用いた案内システムは常にユーザとインタラクションして案内を行っている訳ではなく，ユーザの案内を行っていない時はユーザが訪れるのを待っている．そのため，実写アバタはユーザとインタラクションして案内を行っている行動状態に加えて，ユーザが訪れるのを待つ待機状態を取り扱

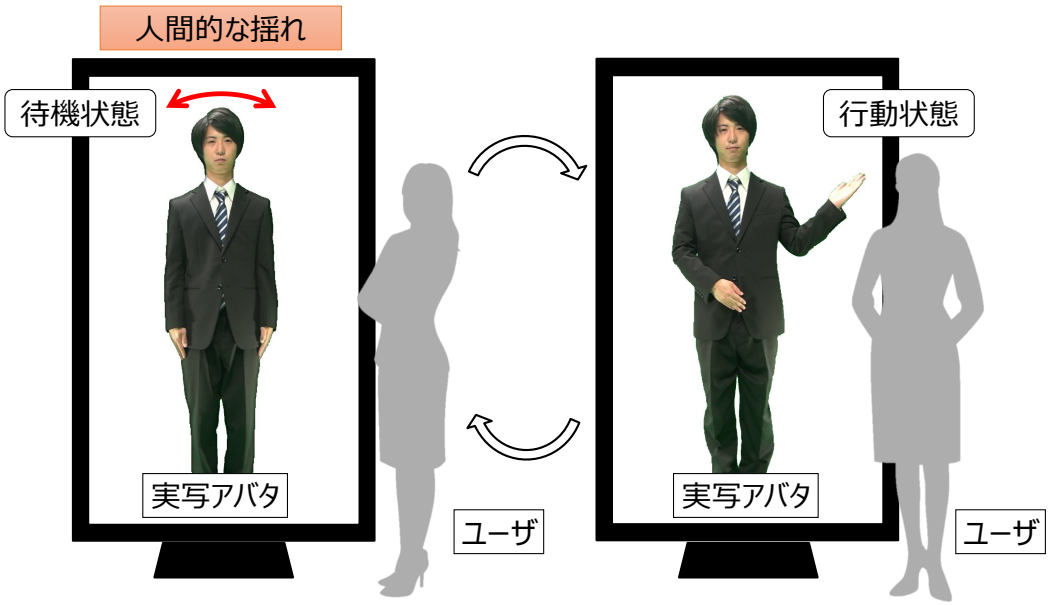


図 1.3: 実写アバタに必要な待機状態と行動状態.

う必要がある．インタラクション前の待機状態が適切に実写アバタに組み込まれていないと，ユーザは実写アバタに話しかけて良いかどうか判断できず，インタラクションが円滑に開始されない問題がある．例えば，待ち姿勢の静止画を表示し続けた場合や，人間の動きとは違う不自然な映像を表示した場合，ユーザはシステムがインタラクションを待っている状態とは判断できず，システムが異常動作していると判断する恐れがある．そのため，ユーザが訪れるのを実写アバタが待っている待機状態において，実写アバタがインタラクションを待っている状態だとユーザが判断できるようにインタラクション開始前動作を実装する必要がある．

1.2.2　状況 2

実写アバタが行動状態に遷移すると，実写アバタはユーザとインタラクションを開始して案内を行う．インフォメーションセンターでは，大人数のユーザが常に訪れることは少ないが，図 1.4(a) のように二人から三人のユーザが時折訪れた場合に，実写アバタを囲うように待つことは多い．また，人通りが多いところにインフォメーションセンターは設置されていることが多く，案内を待つユーザが列を作ると道を塞いでしまうため，ユーザが実写アバタを囲うように待つことが

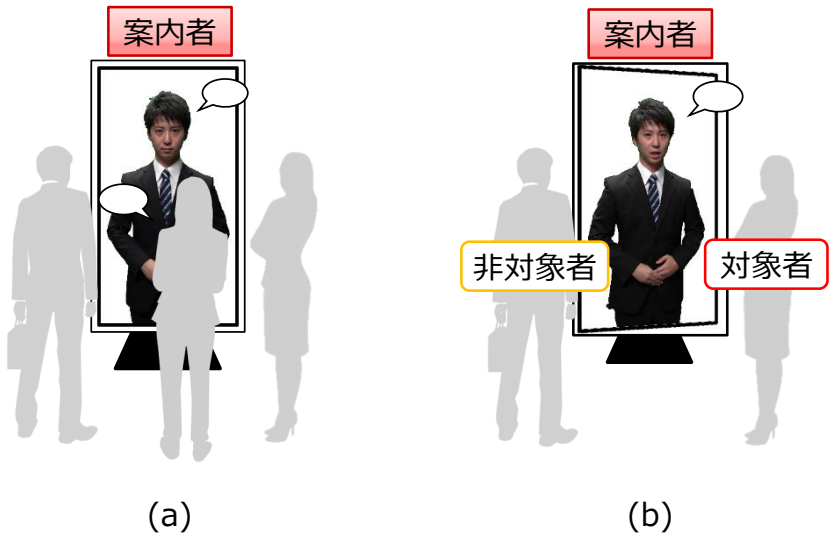


図 1.4: 実写アバタの周りに複数ユーザが待っている状況の例.

多い．インタラクション開始時にユーザが実写アバタを囲うように待つ状況において，ユーザは案内を受ける順番が次に来る対象者の役割と，案内を受ける順番が未だ来ない非対象者の役割に分けることができる．ユーザの役割を対象者と非対象者に分離した例を図 1.4(b) に挙げる．実写アバタが非対象者が存在する中で対象者のみを指定する時に，何の動作もなくインタラクションを開始してしまうと対象者と非対象者の両方が指定されていると感じる．そのため，実写アバタが非対象者が存在する中で対象者を指定する時のインタラクション開始前に動作を実装する必要がある．

1.2.3 状況 3

据置の大型ディスプレイで表示した実写アバタの案内が終了すると，ユーザは携帯デバイスなどの小型ディスプレイに表示された実写アバタの案内に切り替わる．小型ディスプレイに実写アバタを表示している時の案内では，ユーザが目的の場所までの道順が分からなくなった場合や曲がり角などのランドマークの案内を行う場合などで実写アバタとユーザが頻繁にインタラクションを行う．このようなインタラクションとインタラクションの間では，図 1.5 のように実写アバタはユーザとインタラクションを開始するために直立姿勢で待機している．この時，状

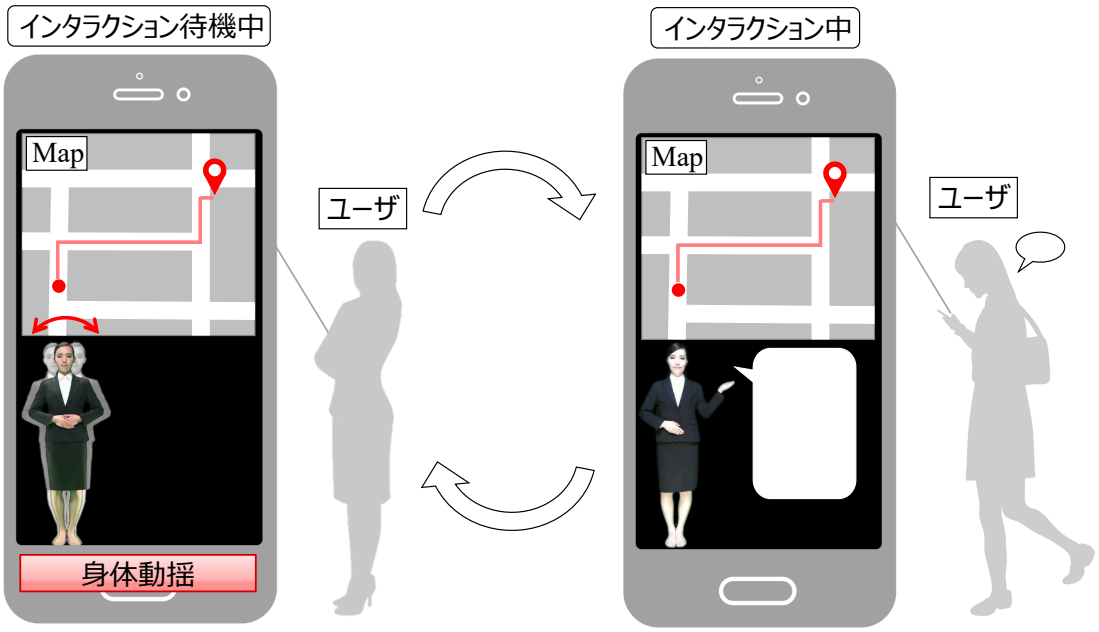


図 1.5: 小型ディスプレイにおける実写アバタの状態.

況 1 と似た問題が発生する．さらに，小型ディスプレイの画面サイズは小さいため，直立姿勢で実写アバタの動きは小さくなってしまう．この場合，ユーザは実写アバタの動きを視認することができずに実写アバタが止まって見える可能性がある．そのため，ユーザが小型ディスプレイ上の実写アバタのインタラクション開始前に実写アバタがインタラクションを待機していると分かるように，実写アバタにインタラクション開始前動作を実装する必要がある．

1.3 課題解決手法

1.3.1 状況 1 に対する解決手法

実写アバタがインタラクションを待っている状態だとユーザが判断できるように，実写アバタにインタラクション開始前動作を実装することを考える．実写アバタは人間の代替を目的としているため，実写アバタに人間と同様の動きを実装することが理想的である．そのため，インタラクション前の待機状態として公共施設の受付などでよく見かける直立姿勢について考える．直立姿勢の人間が待機

状態においてどのような動きをするかを実際に観察すると，ある位置を中心とし絶えず揺れ動いていることが分かる．これは身体動揺 [23] と呼ばれており，一部の筋肉に負担が掛からないように上手く負担を分散するよう人間は無意識の内に体を制御している．そこで，実際の人間から身体動揺の振動量を計測し実写アバタで再現することで，待機状態でも人間的な動きができる実写アバタを生成する手法を提案する．

1.3.2 状況 2 に対する解決手法

非対象者が存在する中で対象者のみを実写アバタが指定するために，映像で動きを表現する手法 [24, 25, 26, 27] を適用することが考えられる．しかし，これらの既存手法は，ディスプレイの前にユーザが一人の状況を想定しており，非対象者が存在する中で対象者のみを指定する状況を十分に考慮していなかった．そこで本論文では，実写アバタが対象者の方を向く動きを加えることで，非対象者が存在する中から対象者のみを指定する手法を検証する．

1.3.3 状況 3 に対する解決手法

小型ディスプレイにおける実写アバタのインタラクション開始前に実写アバタがインタラクションを待機しているとユーザが分かるように，実写アバタの直立姿勢にインタラクション開始前動作を実装することを考える．実際の人物が直立姿勢を取っているときは，常に一定の位置で体が揺れている身体動揺という動きをしている．しかし，身体動揺の動きは非常に小さいため，小型ディスプレイではユーザが視認することが難しい可能性がある．そこで，小型ディスプレイ上における実写アバタの身体動揺の動きを強調して，ユーザが実写アバタの動きを自然であると視認できるかを検証する．

第2章　身体動揺の計測による待ち状態の実写アバタ生成

2.1　状況1の研究背景

実写アバタを用いた案内システムにおいて自然な実写アバタを再現するためには，ユーザとの対話などの行動状態に加えて待機状態を取り扱う必要がある．実写アバタは人間との間で何らかの行動を常にとっている訳ではなく，行動の前後に待機状態が存在する (図 1.3)．ここで述べる待機状態には，ユーザとの対話が開始される前で相手が来ることを待つフェーズと，ユーザとの対話中に相手の反応を待つフェーズに分けられる．本論文では，前者の待機状態を対象とし，アバタ1体とユーザ1名が存在する状況について議論する．対話前の待機状態が適切に実写アバタに組み込まれていないと，ユーザは実写アバタに話しかけて良いかどうか判断できず，インタラクションが円滑に開始されない問題がある．例えば，待ち姿勢の静止画を表示し続けた場合や，人間の動きとは違う不自然な映像を表示した場合，ユーザはシステムが対話を待っている状態とは判断できず，システムが異常動作していると判断する恐れがある．実写アバタの既存手法 [13, 14, 15] は行動状態を対象としていたものの，待機状態を十分に考慮していない課題があった．

ここで，対話前の待機状態として受付などでよく見かける直立姿勢について考える．直立姿勢の人間が待機状態においてどのような動きをするかを実際に観察すると，ある位置を中心とし絶えず揺れ動いていることが分かる．これは身体動揺 [23] と呼ばれており，一部の筋肉に負担が掛からないように上手く負担を分散するよう人間は無意識の内に体を制御している．

そこで本論文では，実際の人間から身体動揺の振動量を計測しアバタで再現することで，待機状態でも人間的な動きができる実写アバタを生成する手法を提案する．身体動揺を計測するために，カメラ映像から体の部位毎に振動量を求める．

計測された振動量の時間変化から特徴を抽出し映像をランダムに遷移させることで，実写アバタの身体動揺を任意の時間長で生成する．提案手法を用いることで，実写アバタが人間的な動きに近づくことを主観評価で確認した．

2.2 関連研究

情報端末における案内を想定したシステムにおいて，ユーザがアバタと自然にインタラクションするためには，アバタの立ち振る舞いに対し，あたかも本当の人間であるかのようにユーザが感じることが望ましい．実際の案内者に近いと感じさせるアバタを実現するためには，文献 [17] でも述べられているように，インタラクションシステムの全体構成として，ユーザの状況やユーザとの会話を理解する要素技術に加えて，アバタの画像と音声を自然に生成しユーザに伝達する要素技術が必要である．特に，アバタの画像生成は見た目に直結する重要な要素技術であり，身体や顔の動きについて近年盛んに研究が進められている．既存手法 [10, 11, 12] では，コンピュータグラフィクスのキャラクタにユーザの状況に応じた情報を与えることでアバタ画像を生成している．一方，実際の人物を撮影した映像を用いて実写アバタを生成する手法 [13, 14, 15] が提案されている．一般的にコンピュータグラフィクスに比べて実写の方がアバタの見た目は本物の人間に近づくことが多い．ただし，実写アバタはカメラで撮影した映像しか動きを再現できない課題が存在する．この課題を解くことを目指し，撮影した映像から新たな人物画像を生成する手法 [18, 19, 20, 21, 22] が提案されている．文献 [18, 19] では会話時の顔画像の生成について述べられており，文献 [20] では会話時の人物の全身画像の生成について述べられている．さらに文献 [21] では複数人による同時動作を編集する手法が提案されており，文献 [22] では複数の動作を時間方向に滑らかにつなぎ合わせる手法が提案されている．しかし，これらの既存手法は，行動状態の中で動きのある人物画像を取り扱っているが，待機状態の中で微動している人物画像については十分に考慮されていなかった．文献 [13] では，あるインタラクション映像と別のインタラクション映像をつなぐため，映像の終了フレームと開始フレームとの間をモーフィングで補完する手法が提案されている．この手法のように単純なモーフィングでは，人間の身体動揺を忠実に再現しているとは言えなかった．

また文献 [28] では，モーションキャプチャした被写体の関節位置の動きを用いることで，CG アバタの待機状態を制御する手法が提案されている．ただしこの手法は関節のみを対象としているため，全身の映像を用いる実写アバタにはそのまま適用できない問題があった．提案手法では，待機状態の人間で必ず発生する身体動揺を被写体の映像から解析し，実写アバタにおいて忠実に再現する課題に取り組む．

2.3 身体動揺の計測

2.3.1 実写アバタで要求される身体動揺

本論文では，エントランスでの受付など周囲からの目がある状況において，直立姿勢の人間に生じる身体の揺れを取り扱う．この身体動揺を実写アバタで再現するためには，カメラ映像中でどのように人間の身体が揺れているかを計測することが必要となる．さらに，身体の部位毎に揺れ方が異なるのかどうかも合わせて確認する必要がある．実写アバタの待機状態の映像は，直立姿勢の被写体が正面から撮影されることを想定している．ただし身体の揺れは微小であるため，カメラのフレームレートや画素数，カメラから被写体までの距離など撮影条件の影響，および，計測手法の誤差の影響で十分に観測されない可能性もあると考えられる．本論文では，実写アバタの被写体を撮影する環境において，提案手法で本当に身体動揺が計測されるかどうかを検証する．

まず身体動揺を計測する手法について議論する．一般的に重心動揺計 [29] を用いることが多いが，この装置は足元の圧力を計測するため，体の部位毎の揺れを扱うことはできない．一方，加速度センサや角速度センサを身体の各部位に装着して計測する手法 [30, 31] が提案されている．これらの手法はカメラとセンサの間で時間同期が難しく，体に取り付けたセンサが見え隠れする問題がある．近年では複数台カメラを用いて重心位置を計測する手法 [32] や距離センサを用いて人間の関節位置から重心位置を計測する手法 [33] も提案されている．ただし，これらの手法は重心位置のみを取り扱っており部位毎の変化までは十分に検討されていなかった．本論文ではカメラ映像のみを採用し，部位毎の揺れを身体と非接触で

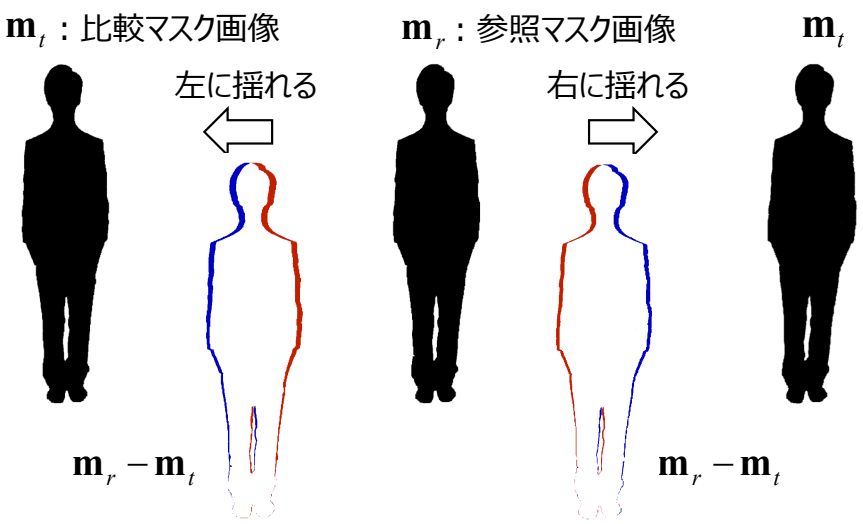


図 2.1: 各時刻におけるマスク画像の比較．赤色は人物から背景に変化した領域を表し青色はその逆を表す．

新たに計測する．以下では，類似する揺れ方が繰り返し表れると仮定した上で設計した計測手法を 2.3.2 で述べ，その性能を 2.3.3 で確認する．

2.3.2 計測手法

身体動揺を計測するために人物領域の画素値を 1 とし背景領域の画素値を 0 とするマスク画像を用いる．身体動揺の中心となる参照時刻 r が与えられると，カメラ映像から参照マスク画像 $\mathbf{m}_r$ を生成し，時刻 $t \in 1, ..., N$ における比較マスク画像 $\mathbf{m}_t$ との間で，図 2.1 のように時間方向の差分を求める．提案手法における身体動揺の振動量 [画素] は式 (2.1) で計算される．

$$d_i = \sum_{x \in \mathrm{parts}(i)} (\mathbf{m}_r(x) - \mathbf{m}_t(x)) \tag{2.1}$$

ただし，$\mathbf{m}_r(x), \mathbf{m}_t(x)$ は x で指定された位置の画素値を表し，$\mathrm{parts}(i)$ は身体部位 i で指定された領域を表す．これらの領域の大きさと位置は時間方向に変化させず固定とし，例えば図 2.2 のように設定する．指定された領域において，人物から背景に変化した画素数が多ければ d_i は正の値をとり，背景から人物に変化した画素数が多ければ d_i は負の値をとる．なお，参照マスク画像の参照時刻 r は，部位毎

の振動量の合計から Algorithm 1 で求まる．$\tilde{d}_i$ は，ある時刻 $\tilde{r}$ を仮の参照時刻とした際の振動量を表す．

Algorithm 1 参照マスク画像の参照時刻 r の決定

for $\tilde{r} = 1$ to N **do**
  $D_{\tilde{r}} \leftarrow 0$
  for $t = 1$ to N **do**
    compute $\tilde{d}_i$ using $\mathbf{m}_{\tilde{r}}, \mathbf{m}_t$
    $D_{\tilde{r}} \leftarrow D_{\tilde{r}} + \sum |\tilde{d}_i|$ for all body parts
  end for
end for
$r \leftarrow \arg\min D_{\tilde{r}}$

2.3.3 計測性能の評価

振動量の計測結果

提案手法の計測性能を評価するために実験を行った．被写体 5 名 (平均年齢 21.4±0.5 歳) に対し 180 秒間の直立姿勢を撮影した．被写体に撮影中はカメラへ視線を向け両足の踵をつけた姿勢を維持するよう指示した．カメラのフレームレートは 30 フレーム毎秒で解像度は 1920 × 1080 画素とした．カメラは床面から 90 センチメートルの高さに設置し，カメラと被写体の距離は 200 センチメートルとした．グリーンバックを用いた背景差分によりマスク画像を生成した．人物領域の外接矩形の大きさは約 300 × 950 画素であった．

被写体毎の振動量 d_i の標準偏差を表 2.1 に示す．身体部位は人手で与えた図 2.2 の頭，肩，手，足とし，各領域の大きさは 70 × 70 画素とした．実験結果より，被写体毎に振動量の標準偏差は異なるため，振動量には個人差が存在することが分かった．ただし，部位間の大小関係は被写体間で共通であったため，開眼時にカメラに視線を向けた直立姿勢において，振動量は身体の上部位ほど大きく身体の下部位ほど小さいことが確認された．なお，今回の実験では振動量 d_i が約 380 画

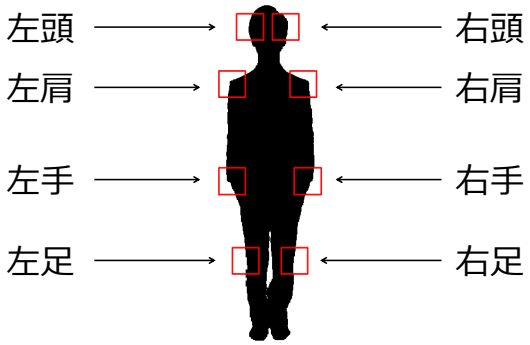


図 2.2: 身体動揺を計測する部位領域の例.

表 2.1: 身体動揺の振動量 d_i の標準偏差 [画素].

被写体	左頭	右頭	左肩	右肩	左手	右手	左足	右足
A	483	489	328	372	284	301	117	146
B	214	185	176	123	127	88	46	48
C	306	312	253	233	216	145	83	87
D	364	434	243	339	174	184	89	98
E	321	330	258	270	204	208	91	88
平均	338	350	252	267	201	185	85	93

素ほど変化すると，身体部位が実世界で約 1 センチメートル移動したことに相当した.

次に，身体の各部位における振動量 d_i の時間変化を図 2.3 に示す．図中の波形は被写体間で大きく違うため，振動量の時間変化にも個人差が存在することが分かった．ただし，部位毎の時間変化に着目すると，頭が左方向に動けば肩，手，足も共に同じ方向に動いており，左側の部位が動けば右側の部位が逆の方向に動いていた．他の被写体でも同様の結果がみられた．以上の結果より，振動量の時間変化の傾向は部位間でほぼ等しいことが確認された.

さらに振動量の時間変化の特徴について検証した．ここでは被写体 A の左頭の振動量について特徴を図 2.4 に示す．振動量がほぼ 0 の参照時刻が時間方向に繰り返し出現しており，それらの時刻の間に存在する波形は多峰性の弧を描いていた．参照時刻は 180 秒間の中で 25 個存在し，10 秒のような間隔が短い区間でも 4 個

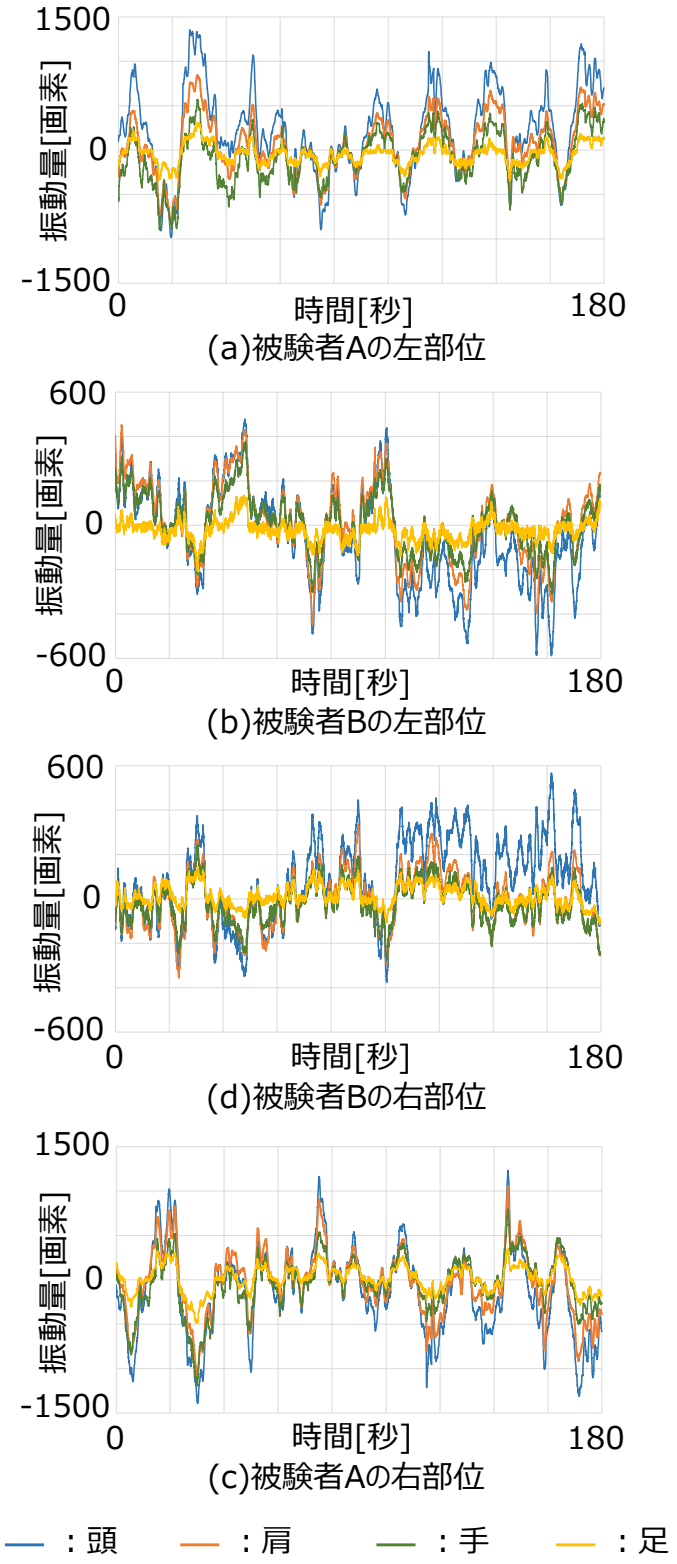


図 2.3: 身体の各部位の振動量 d_i の時間変化.

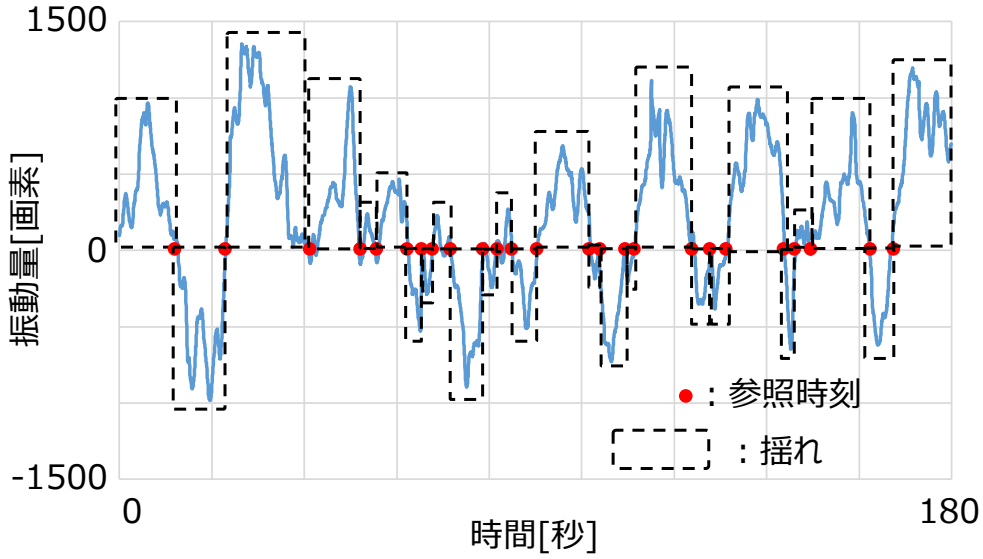


図 2.4: 振動量の時間変化の特徴.

存在していた．このように短時間の計測でも揺れの特徴が表れることが分かった．同じような特徴が他の被写体でも表れていた．以上より，振動量の時間変化の特徴として，参照時刻は映像中で複数回出現し，それら参照時刻の間における映像には身体の揺れが存在することを，実写アバタの被写体を撮影する環境において確認した．

計測誤差の検証

提案手法は背景差分による人物領域の抽出精度に依存しており，計測した振動量が抽出誤差に埋もれていないかを確認した．提案手法はグリーンバックを用いたが，相互反射の影響でグリーン成分が衣服領域に混じる場合や，人の影でグリーンバック領域の色が衣服の色に近づく場合が見られた．このように人物と背景の境目が曖昧な領域が存在するため以下の検証を行った．この実験では，背景差分から求めたマスク画像を用いた場合と人手で抽出したマスク画像を用いた場合とを比較した．人手による抽出は，グリーンバックと人物の境界を目視で確認しながら人物領域を設定した．被写体 5 名の参照マスク画像と，映像からランダムに選択した 15 枚の比較マスク画像を用いた．振動量の絶対値 $|d_i|$ について平均を比

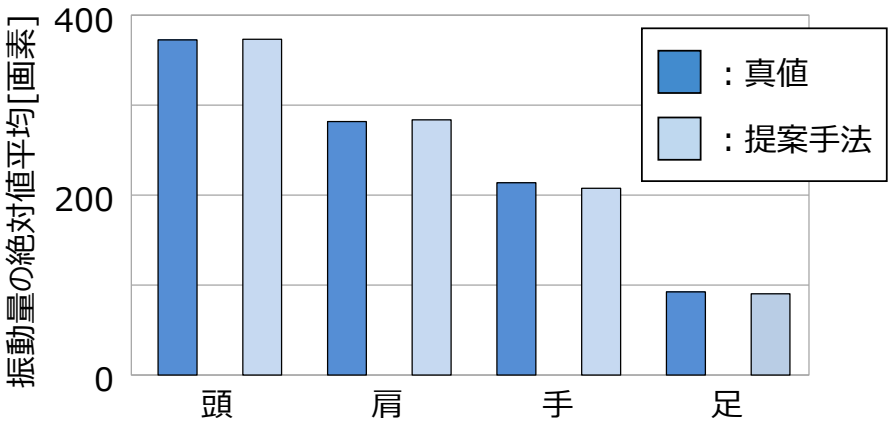


図 2.5: 背景差分による振動量の計測性能の比較.

較した結果を図 2.5 に示す．この結果より，人物領域の抽出誤差の影響は十分に小さいことが分かった．これにより，2.3.3 の実験結果は人物領域の抽出誤差にそれほど影響されないと言える．

次に，距離センサで獲得した関節位置を用いて揺れの大きさを計測する場合と比較した．ここでは既存手法 [33] でも用いられている Microsoft Kinect v2 で獲得した関節位置とカラー画像を利用した．Kinect から出力される関節位置と提案手法の身体部位の位置は異なるため，この実験では以下で述べる重心を評価に用いた．Kinect を用いる手法では，同時刻のカラー画像上に射影した二次元の関節位置から重心を求めた．なお Kinect が出力する全ての関節位置を重心計算に用いた．提案手法では，Kinect のカラー画像から抽出したマスク画像に含まれる人物領域の全身を用いて重心を求めた．さらに，Kinect のカラー画像から人手で抽出した人物領域の重心を求めた．評価指標として，一定時間が経過した後に重心がどれだけ動いたかを表す移動量を用いた．Kinect で撮影した映像において，第 1 時刻をランダムに設定し，その時刻から $(0, 80)$ 秒の区間で第 2 時刻をランダムに設定した．第 1 時刻と第 2 時刻のペアを 15 組準備した．第 1 と第 2 の時刻の間の平均移動量を計算した結果，人手で抽出した場合は 3.4 画素であったのに対し，提案手法は 3.5 画素，Kinect を用いた手法は 4.7 画素であった．提案手法は，Kinect を用いた手法と比べて人手で抽出した場合に移動量が近いことが分かった．この原因として，Kinect から出力される足や手の関節位置は誤差が大きかったことが考え

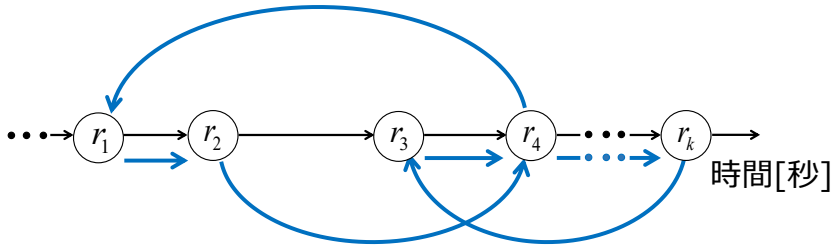


図 2.6: 参照時刻 r_k の集合を用いた身体動揺の再現．提案手法は参照時刻をランダムに遷移し映像を再生する．図中の黒矢印は撮影した映像の時間の流れを表し，青矢印は再現した映像の時間の流れを表す．

られる．以上より，身体の微小な動きを計測する場合，Kinect から出力される関節位置を単純に適用できるとは限らないことが分かった．

2.4 実写アバタにおける身体動揺の再現

2.4.1 再現手法の方針

実写アバタにおいて身体動揺をどのように再現するかについて述べる．ここでは，身体動揺を計測した被写体そのものを実写アバタとし，本人自身の揺れを任意の時間長で再現する場合を考える．この場合，図 2.4 で確認した身体動揺の特徴について，参照時刻が複数出現することを明示的に利用でき，参照時刻の間の揺れを暗に利用することができる．提案手法では，実写アバタのモデルを撮影した映像から身体動揺の参照時刻を複数抽出し，それらの時刻をランダムに遷移しながら映像を再生する．その流れを図 2.6 に示す．これにより遷移時の体の揺れを滑らかに繋げ，遷移時以外は揺れをそのまま再生することで，提案手法は身体動揺を任意の時間長で生成することができる．

ただし，ここまでの身体動揺の計測の議論では順方向の時系列しか取り扱っていないため，ランダム遷移により時刻が不連続に変化する場合は，再現時に新たな問題が発生する．ランダム遷移の参照時刻を抽出するためにマスク画像のみを用いた場合，人物領域の形状は似ているが，髪や衣服や手のずれなど身体前面のテクスチャは似ていない時刻が選ばれる可能性がある．このような参照時刻で遷

移すると，映像中に不連続な変化が含まれるためユーザが違和感を覚える．自然な身体動揺の映像を生成するためには，形状に加えて全身の見え方も考慮する必要がある．また，ユーザは顔の見え方の変化にも敏感である．例えば目や口など顔に含まれる部品が時間方向に不連続に変化し，急に閉じたり開いたりすると違和感を覚える．このため提案手法は，参照時刻を抽出する際に顔の見え方変化も考慮に加える．以下では再現手法の詳細について述べる．

2.4.2 形状と見え方の類似度の算出

身体動揺を再現するために，提案手法は人物領域の形状，人物領域の見え方および顔の見え方の類似度を各時刻で求める．時刻 t における人物領域の形状を表すマスク画像を $\mathbf{m}_t$，人物領域の見え方を表すカラー画像を $\mathbf{a}_t$，顔の見え方を表す顔画像を $\mathbf{f}_t$ とする．マスク画像 $\mathbf{m}_t$ は身体動揺の計測で述べた手法と同様に背景差分から算出する．カラー画像 $\mathbf{a}_t$ はカメラ映像の各フレームとする．顔画像 $\mathbf{f}_t$ は顔追跡手法 [34] で得た特徴点を利用し顔向きを正面に正規化した領域とする．初期参照時刻 r_0 が与えられたとすると，時刻 t における人物領域の形状の類似度 $s_{t,s}$ は式 (2.2) で求まる．

$$s_{t,s} \quad = \quad e^{-\lambda \sum |d_i|} \tag{2.2}$$

ここで，λ は定数とし，d_i は $\mathbf{m}_{r_0}, \mathbf{m}_t$ から算出した振動量とする．この式では身体部位毎に求めた振動量の大きさの合計値を用いる．この類似度 $s_{t,s}$ を，人物領域と背景領域との入れ替わりが少ない時刻を抽出するために用いる．次に，人物領域の見え方の類似度 $s_{t,a}$ は式 (2.3) で求まる．

$$s_{t,a} \quad = \quad \mathrm{SSIM}(\mathbf{a}_{r_0}, \mathbf{a}_t) \tag{2.3}$$

ここで，SSIM は人の主観に近い類似度を算出する Structural SIMilarity [35] を表す．この類似度 $s_{t,a}$ を，人物領域内のテクスチャ変化が少ない時刻を抽出するために用いる．次に，顔の見え方の類似度 $s_{t,f}$ は式 (2.4) で求まる．

$$s_{t,f} \quad = \quad \mathrm{FaceSimilarity}(\mathbf{f}_{r_0}, \mathbf{f}_t) \tag{2.4}$$

ここで，FaceSimilarity は顔画像のエッジから特徴量を求めコサイン類似度を算出する．この類似度 $s_{t,f}$ を，目の瞬きや口の開閉など顔の表情変化が少ない時刻を抽出するために用いる．これらの類似度を加算することで統合類似度 s_t を式 (2.5) で求める．

$$s_t \quad = \quad \alpha s_{t,s} + \beta s_{t,a} + \gamma s_{t,f} \tag{2.5}$$

ただし，$\alpha + \beta + \gamma = 1$ とする．各類似度の値域は $[0, 1]$ であるが，実際には偏りが存在するため α, β, γ でスケーリングを行う．また，統合する際にどの類似度を重視するかの制御も可能である．統合類似度を用いることで，映像を遷移する際に身体の変化が少なくなる時刻の集合を抽出することを狙う．なお，初期参照時刻 r_0 は，s_t を用いて Algorithm 2 を適用することで求まる．

Algorithm 2 初期参照時刻 r_0 の決定.

for $\tilde{r_0} = 1$ to N **do**
  $S_{\tilde{r_0}} \leftarrow 0$
  for $t = 1$ to N **do**
    compute $\tilde{s}_t$ using $\mathbf{m}_{\tilde{r_0}}, \mathbf{a}_{\tilde{r_0}}, \mathbf{f}_{\tilde{r_0}}, \mathbf{m}_t, \mathbf{a}_t, \mathbf{f}_t$
    $S_{\tilde{r_0}} \leftarrow S_{\tilde{r_0}} + \tilde{s}_t$
  end for
end for
$r_0 \leftarrow \arg\max S_{\tilde{r_0}}$

2.4.3 参照時刻の集合を用いた映像遷移

自然な身体動揺の映像を再現するために，提案手法は参照時刻 r_k の集合を利用する．この集合は，初期参照時刻の身体状態に近い状態を持つ時刻を Algorithm 3 で抽出することで求まる．提案手法は，参照時刻の条件として，s_t の時間変化において極大点となること，かつ，s_t が閾値 T_1 より大きいことを設ける．ただし，これらの条件だけでは非常に近い時間間隔で抽出されることもあるため，閾値 T_2 以上離れた時刻を選択する条件も設ける．

Algorithm 3 参照時刻 r_k の抽出.

for $t = 1$ to N **do**
  compute s_t using $\mathbf{m}_{r_0}, \mathbf{a}_{r_0}, \mathbf{f}_{r_0}, \mathbf{m}_t, \mathbf{a}_t, \mathbf{f}_t$
  if s_t is local maximum, $s_t > T_1$, interval $> T_2$ **then**
    add time t to set of reference times
  end if
end for

次に，参照時刻の集合を用いた映像遷移の手法について述べる．抽出された時刻 r_k の集合から，ランダムに 1 つの時刻を選択し，そこから次の参照時刻になるまでのカラー画像 $\mathbf{a}_t$ をディスプレイ上で再生する (Algorithm 4)．このランダム選択を繰り返すことで身体動揺を任意の時間長で再現する．これにより，身体動揺の振動量が異なる画像列をランダムにつなげ合わせることができ，自然な身体動揺が実写アバタで再現される．

Algorithm 4 身体動揺の生成.

while true **do**
  select r_k randomly from set of reference times
  for $t = r_k$ to r_{k+1} **do**
    display color image $\mathbf{a}_t$
  end for
end while

2.4.4 再現手法の評価

主観評価の結果

提案手法により身体動揺を再現した実写アバタについて評価した．図 2.7 の男女 2 体の実写アバタを用いた．どちらの実写アバタも直立姿勢としたが，女性の実写アバタは手を前で組むこととした．実写アバタのモデルとなる人物に直立姿勢を維持させ 30 秒間の映像を撮影した．この映像の先頭から 10 秒間を参照時刻の抽出対象とし，新たな 30 秒間のアバタ映像を生成した．抽出された参照時刻の個

図 2.7: 評価に用いる実写アバタ.

数は両アバタともに 4 個であった．提案手法のパラメータは $\alpha = 0.2, \beta = 0.7, \gamma = 0.1, \lambda = 1.0 \times 10^{-5}, T_1 = 0.97, T_2 = 1$ とした．類似度 $s_{t,s}$ を算出する際の部位領域は，男性アバタの場合は図 2.2 とし，女性アバタの場合は図 2.8 とした．比較のために次の 4 つの手法を用いて主観評価を行った．

V1 :(理想) 撮影した 30 秒間の映像をそのまま再生.

V2 :(提案手法) 生成した 30 秒の映像を再生.

V3 :(比較手法 1) 先頭 10 秒間の映像を 3 回連続再生.

V4 :(比較手法 2) 一枚の静止画像を 30 秒間表示.

主観評価にはサーストンの一対比較 [36] を用いた．被験者は 11 名 (男性 9 名，女性 2 名，平均年齢 21.6 ± 0.8 歳) とし，80 インチの縦置きディスプレイから 1.5 メートル離れた位置に立った．各手法の映像をペアで ${}_4C_2 = 6$ 回だけランダムな順番で提示した．被験者はペアの映像が順に再生された後に，どちらの映像がより人間の自然な動きに近いかを回答した．一対比較の各映像を 30 秒間とした理由は，順に映像を被験者が見ていくため，それ以上の長さとした場合は比較のために再生した前の映像の印象が薄れてしまい回答が曖昧になったためである.

主観評価の結果を図 2.9 に示す．図中のスコアが大きいほどアバタの動きが人間に近いとの同意が多かったことを表す．この実験結果より，提案手法 V2 は比較手法 V3, V4 と比べて人間の動きに近いことが分かった．図 2.10 に主観評価に用

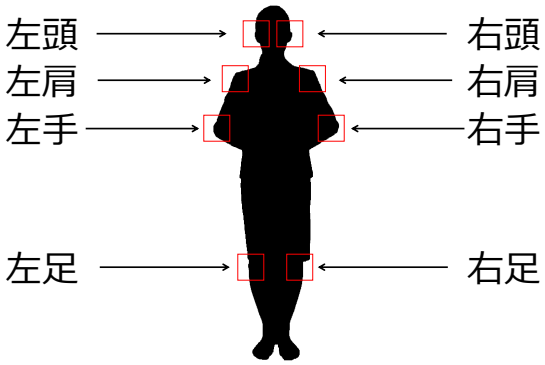


図 2.8: 人物領域の形状の類似度を算出するために用いた部位領域 (女性の実写アバタの場合).

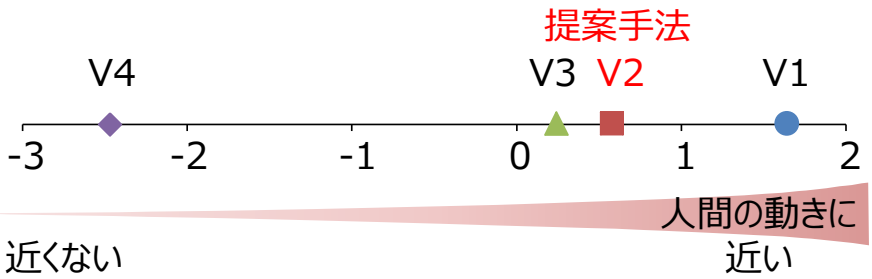


図 2.9: 実写アバタの身体動揺に対する主観評価の結果.

いた映像の例と振動量の可視化の例を示す．この図より，V3 は連続再生の遷移時に振動量 (図中の赤色と青色の領域) が大きいため揺れに不連続が生じていることが分かる．一対比較の後に自由記述のアンケートを取ったところ，V3 は映像の途中でアバタが急に動くことがあり不自然であったという意見がでた．なお，V2 と V3 を比較した際の得票数は V2 が 17 票で V3 が 5 票であった．一方，V2 は遷移時に振動量が残り，V1 と比べると不連続が発生していた．これは図 2.10 の上から 3 番目と 4 番目の振動量に注目すると，V1 と比べて V2 の方が僅かに多いことから分かる．このため V1 の主観評価の結果が V2 より高かったと考えられる．以上より，身体動揺を提案手法で再現することで，実写アバタの動きが人間に近づくことを確認した.

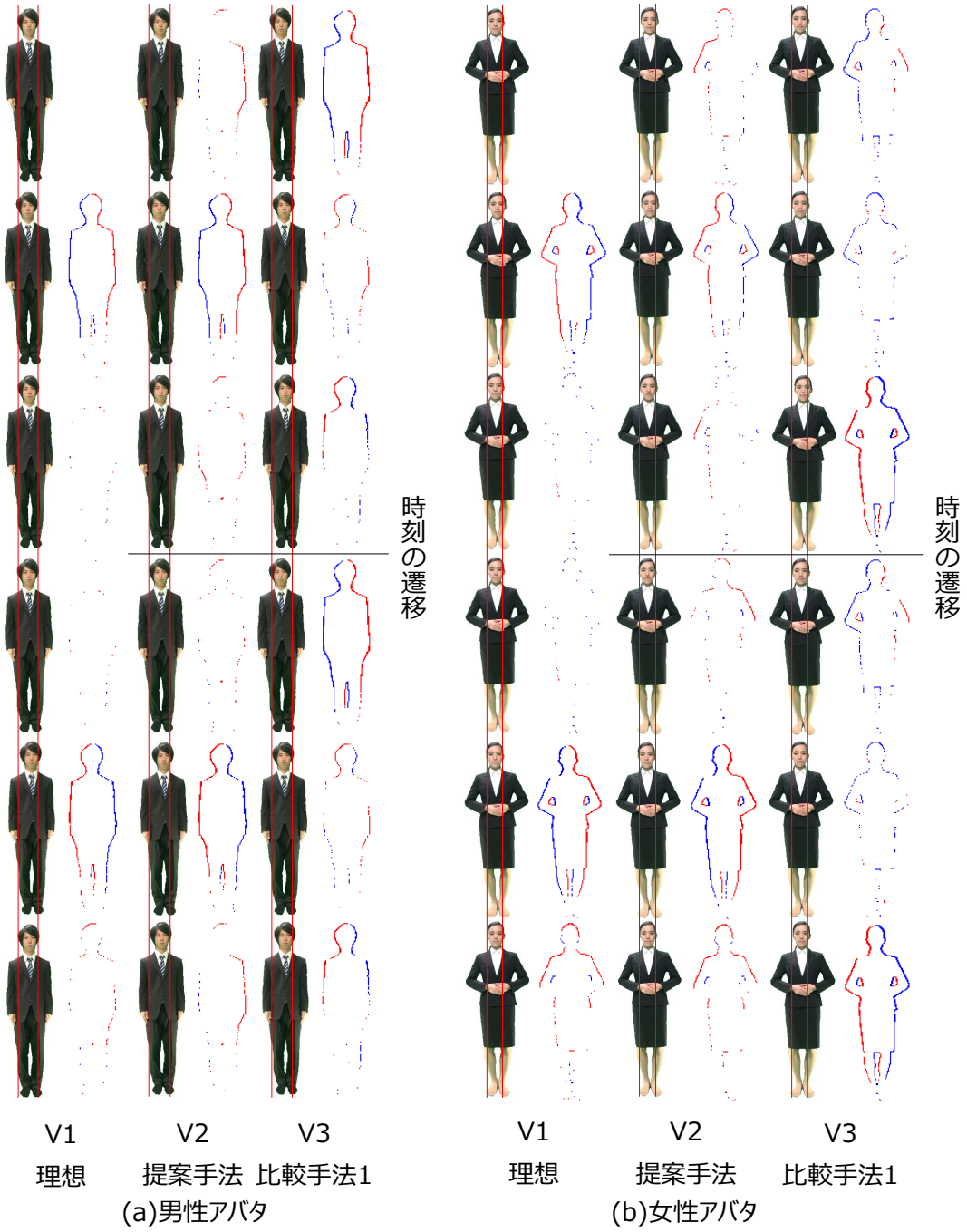


図 2.10: 主観評価に用いた映像と振動量の可視化の例．赤色は人物から背景に変化した領域を表し青色はその逆を表す．

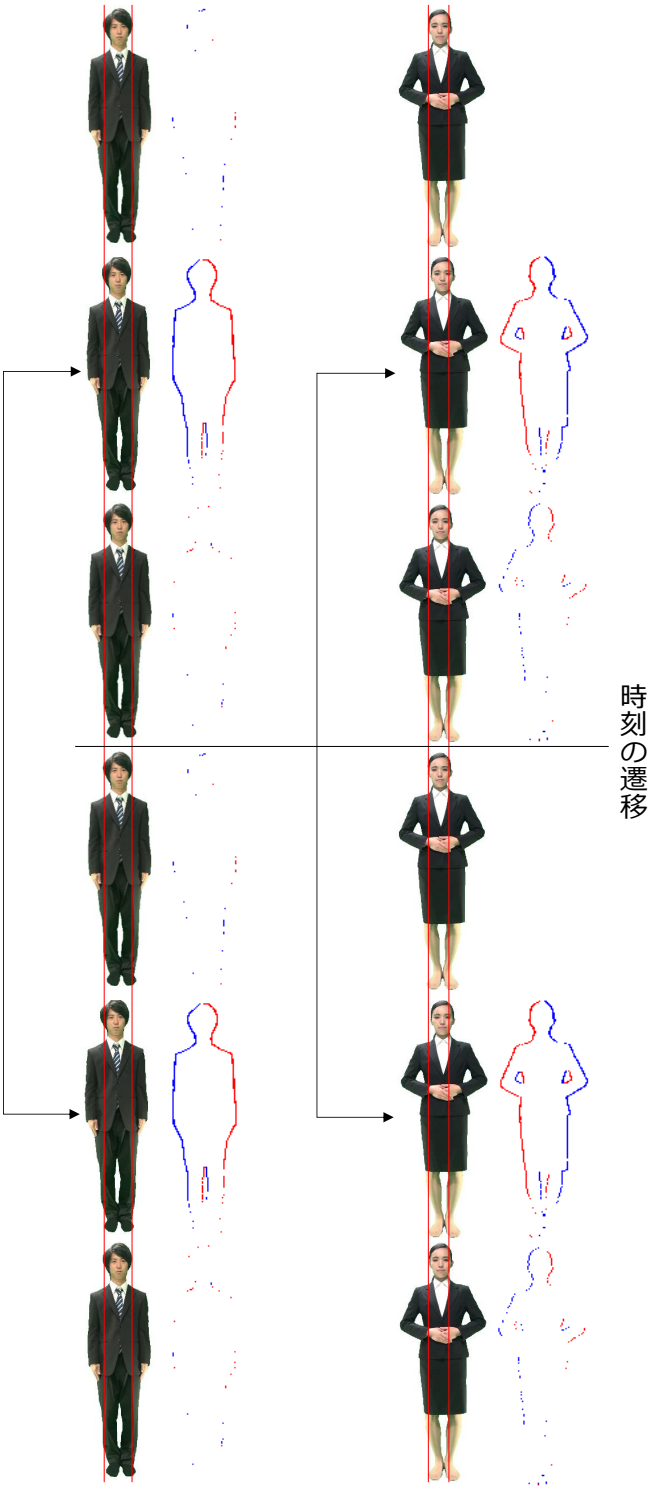


図 2.11: 単純遷移で生成した映像と振動量の可視化の例.

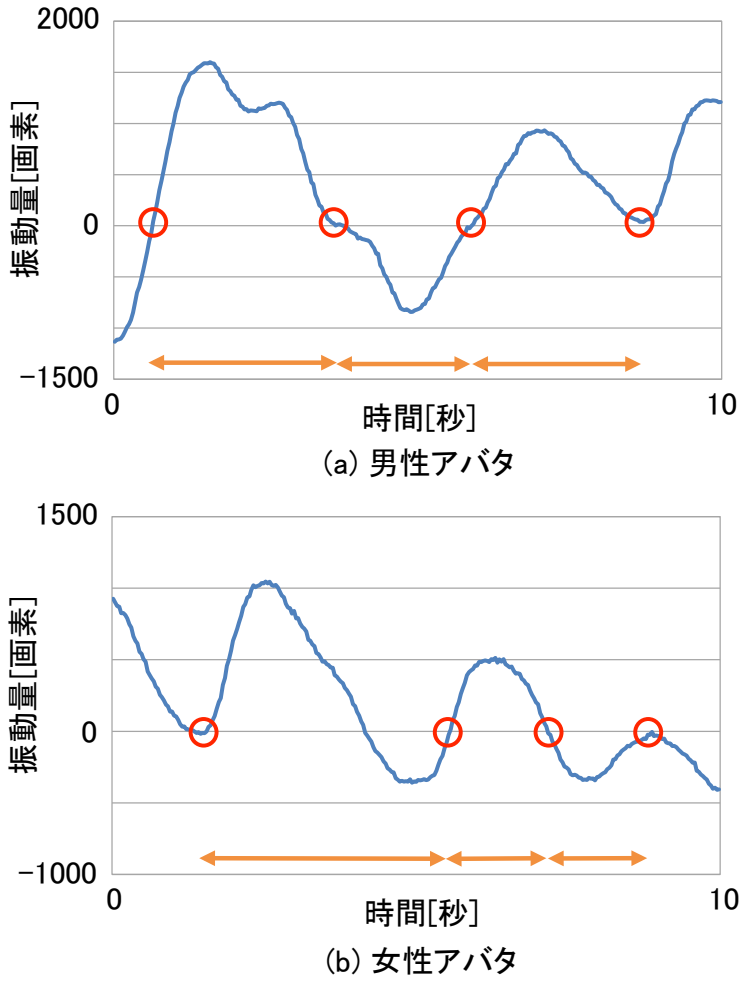


図 2.12: 各短区間における振動量の時間変化．丸印は参照時刻，矢印は短区間を表す．

ランダム遷移と単純遷移の比較

ランダム遷移を用いて身体動揺を生成する有効性について評価した．ここでは 2 つの参照時刻で挟まれる短区間を繰り返す単純遷移と比較した．2.4.4 で述べたように，10 秒間のアバタ映像には男女のアバタともに 4 個の参照時刻が含まれており，参照時刻で挟まれる短区間は 3 個存在した．それぞれの短区間の時間長は，男性アバタが 3.1 秒，2.0 秒，2.9 秒，女性アバタが 4.0 秒，1.6 秒，1.7 秒であった．図 2.12 に各短区間における振動量 d_i について，左部位の合計値の時間変化を示す．最も時間が長い短区間を繰り返す場合を単純遷移 1，その短区間と次に時間が長く

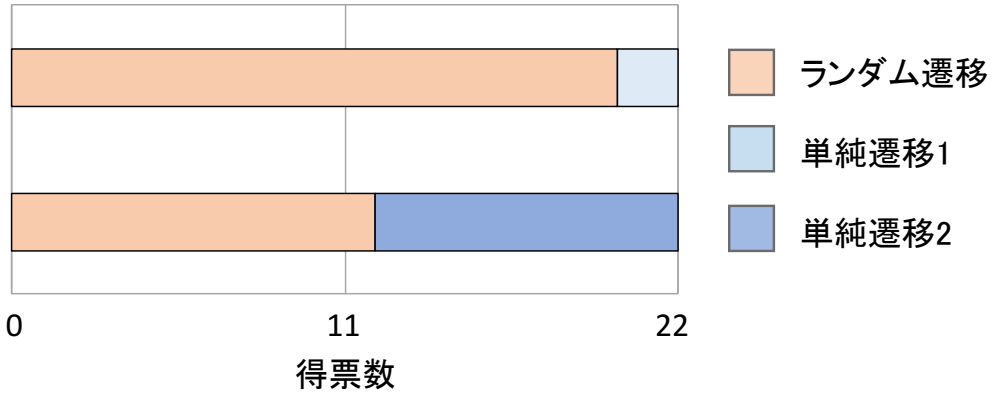


図 2.13: ランダム遷移と単純遷移の主観評価の得票数.

かつ揺れの方向が異なる短区間を交互に繰り返す場合を単純遷移 2 とした．なお，女性アバタの 4.0 秒の短区間には瞬きの影響で左右の揺れが含まれていたため，単純遷移 1 および単純遷移 2 には用いなかった．それ以外の短区間は左右どちらかの振れが含まれていた．図 2.11 に単純遷移 1 で生成したアバタ映像の例と振動量の可視化の例を示す．この図より，遷移時の振動量が小さく映像は滑らかに繋がっているが，同じ動きが繰り返されていることが分かる．

ランダム遷移と単純遷移 1 と単純遷移 2 で生成した 30 秒間の映像を 11 名の被験者に見せ，人間的な動きに近い方を選択させた．図 2.13 にランダム遷移と比較した時の単純遷移 1 と単純遷移 2 の得票数を示す．その結果，ランダム遷移の得票数が，単純遷移 1 と比べて大幅に多かった．単純遷移 2 と比べるとランダム遷移が僅かではあるが得票数は多かった．以上のことから同じ揺れを単純に繰り返し再生するより，複数の参照時間をランダムで遷移させる方が人間的な動きに近づくことを確認した．

各類似度の有効性の検証

参照時刻を抽出するために用いた各類似度の有効性を評価した．統合類似度 s_t の場合，人物領域の形状の類似度 $s_{t,s}$ のみの場合，人物領域の見え方の類似度 $s_{t,a}$ のみの場合，顔の見え方の類似度 $s_{t,f}$ のみの場合で生成した映像についてサーストンの一対比較を用いて主観評価を行った．全ての場合において 2.4.4 で用いた 10 秒

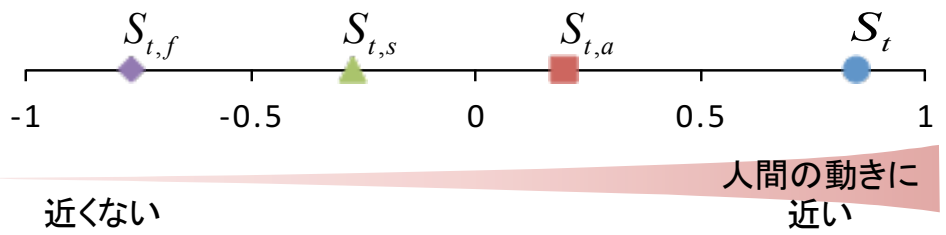


図 2.14: 各類似度を用いた場合の主観評価の結果.

間のアバタ映像から 30 秒間の身体動揺の映像を生成した．統合類似度を算出する際の α, β, γ は 2.4.4 と同じとした．生成前の 10 秒間の映像において各類似度の平均値と標準偏差を算出したところ s_t は 0.967 ± 0.010，$s_{t,s}$ は 0.987 ± 0.008，$s_{t,a}$ は 0.970 ± 0.008，$s_{t,f}$ は 0.912 ± 0.085 であった．

主観評価の結果を図 2.14 に示す．この結果より，$s_{t,a}$ を用いた場合は，$s_{t,s}$ や $s_{t,f}$ を用いた場合と比べてスコアが高くなることが分かった．また，s_t を用いた場合のスコアは，各類似度を単体で用いた場合と比べて高くなることが分かった．主観評価のスコアと α, β, γ の大小関係は同じであった．人物領域の見え方を重視し，かつ，類似度が取り得る値のスケーリング効果を考慮することが α, β, γ の設定に重要であると考えられる．以上より，各類似度を統合して参照時刻を選択することの有効性を確認した．

2.5 状況 1 まとめ

状況 1 では，待機状態の実写アバタで人間に近い動きを再現するために，身体動揺を計測し再現する手法について述べた．体の部位毎に振動量の時間変化を計測することで，身体動揺の中心となる参照時刻は映像中で複数回出現し，参照時刻の間には身体の揺れが存在することを確認した．これらの特徴を利用し参照時刻をランダムに遷移することで，任意の時間長の身体動揺を再現する手法について述べた．主観評価により，提案手法は比較手法と比べて人間の身体動揺に近いことを確認した．今後の課題として，身体動揺を計測した被写体とは別人の実写アバタでその揺れを再現する手法の検討，顔全体ではなく参考文献 [37] のように表情の基本単位に注目した類似度の設計などが挙げられる．

第3章　実写アバタ映像における動き表現を用いた対象者の指定

3.1　状況2の研究背景

実写アバタを用いた案内システムを設置するインフォメーションセンターにおける実際の人同士の案内を考察する．認知科学の分野における知見 [2] に基づいて，本論文では案内時の人物役割を三つに分類する．人物役割として，情報を提供する案内者，案内を受けている最中の参与者，案内を受けることを待つ傍参与者がある．傍参与者は案内を受ける順番になると案内者から指定され，参与者へ役割が遷移する．インフォメーションセンターでは，傍参与者が案内者を囲うように待つ状況が発生する．その状況の例を図 3.1(a) に示す．インフォメーションセンターでは，大人数の傍参与者が常に訪れることは少ないが，二人から三人の傍参与者が時折訪れた場合に，案内者を囲うように待つことは多い．また，人通りが多いところにインフォメーションセンターは設置されていることが多く，案内を待つ人が列を作ると道を塞いでしまうため，傍参与者が案内者を囲うように待つことが多い．

インタラクション開始時に傍参与者が案内者を囲うように待つ状況において，案内者が特定の傍参与者のみを，次の案内に向けてどのように指定するかについて考える．複数の傍参与者の役割を明確に分離するために，案内を受ける順番が次に来る対象者の役割と，案内を受ける順番が未だ来ない非対象者の役割を新たに設ける．傍参与者の役割を対象者と非対象者に分離した例を図 3.1(b) に挙げる．対象者は，案内者から指定されていると感じると，傍参与者から参与者に遷移する．非対象者は，自分が案内者から指定されていないと感じている間は，傍参与者の役割を維持する．案内者は，非対象者が存在する中で対象者のみを指定することが重要となる．案内者の代替である実写アバタでも同様のことが言える．

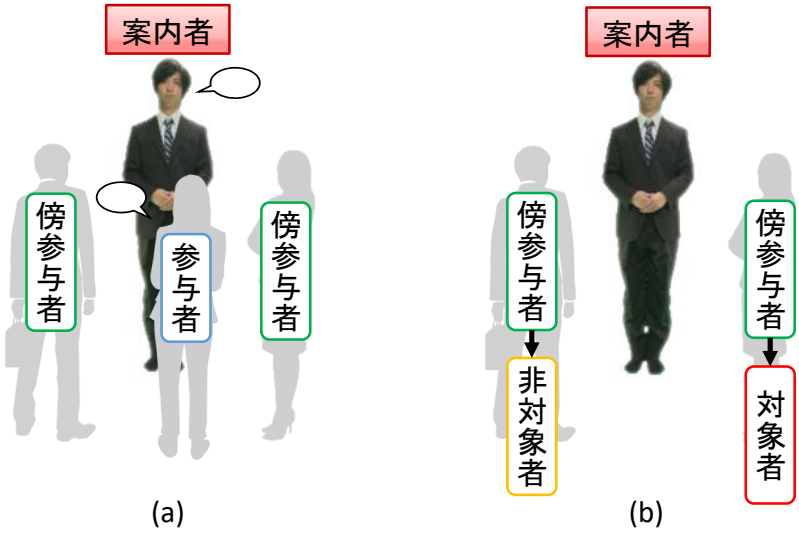


図 3.1: インフォメーションセンターにおける案内で発生する状況例.

非対象者が存在する中で対象者のみを実写アバタが指定するために，映像で動きを表現する手法 [24, 25, 26, 27] を適用することが考えられる．しかし，これらの既存手法は，ディスプレイの前にユーザが一人の状況を想定しており，非対象者が存在する中で対象者のみを指定する状況を十分に考慮していなかった．

そこで本論文では，実写アバタ映像に動きを加えることで，非対象者が存在する中から対象者のみを指定する手法を検証する．対象者のみを指定する上で，対象者と非対象者がどのように感じるかの主観評価が必要である．その中でも対象者の評価はインタラクションを開始するために特に重要となる．このため本論文では，対象者から良好な評価を得ることを主目的として検証を行う．対象者から良好な評価を得た上で，非対象者がどのように感じるかの評価も合わせて検証する．

3.2 関連研究

非対象者が存在する状況において実写アバタが対象者のみを指定するために，映像のみで動きを表現する手法 [24, 25, 26, 27] を適用することが考えられる．文献 [24, 25] では，アバタとユーザがアイコンタクトを行うために，映像上に表示されたアバタが視線や顔を動かす表現手法について述べられている．文献 [26] では，顔の表情に加えてジェスチャーの生成手法が述べられている．文献 [27] では，ユー

ザの視線から興味の有無を判定し，ユーザが興味をもつように誘導する動きを生成する手法について述べられている．映像のみで動きを表現する手法は，一般的な据置型ディスプレイへ簡単に適用することができる．ただし，既存手法ではディスプレイの前にユーザが一人の状況を想定しており，非対象者が存在する中で対象者を指定する状況を十分には考慮していなかった．また，インフォメーションセンターの状況に合った応用についても考慮していなかった．

非対象者が存在する中で対象者のみを指定するために，特殊な機構を組み込んだディスプレイを用いる手法 [38, 39, 40, 41, 42, 43] を適用することが考えられる．文献 [38] では，遠隔会議システムにおいてユーザが誰の方向を向いているかを明確に伝えるために，複数のカメラでユーザの視線を検知し，三次元空間内の会議室でユーザの向いている方向を表現する手法が述べられている．文献 [39] では，眼球型ディスプレイを提案し，注視方向をユーザへ伝えている．文献 [40, 41] では，ディスプレイに回転する機構を取り付けることで，ユーザの方向にディスプレイを向けることができる．文献 [42] では，ディスプレイにロボットアームをつけてポインティングを可能にする手法が述べられている．文献 [43] では，ディスプレイが付いているロボットを用いて，インタラクションを行っている．特殊な機構を組み込んだディスプレイを用いる手法は，直感的に分かりやすく対象者を指定できる利点がある．しかし，ディスプレイに特殊な装置を加える必要が有り，一般的な据置型ディスプレイと比べてメンテンナンス，安全性，設置スペースを配慮する必要がある．本論文では，一般的な据置型ディスプレイにおける映像のみで動きを表現する手法を用いる．非対象者が存在する状況において対象者のみを実写アバタが指定することを狙う．

3.3 状況 2 における仮説

実際のインフォメーションセンターで発生頻度が高く，対象者のみを指定する上で最もシンプルな状況において，実写アバタが案内することを考える．具体的には，参与者に対して左右の位置に傍参与者が立つ状況について検証する．実写アバタが対象者のみを指定するまでの流れを図 3.2 に示し，以下で時刻を順に追って説明する．ある時刻 t_1 で，ディスプレイ正面に立つ参与者が，実写アバタから

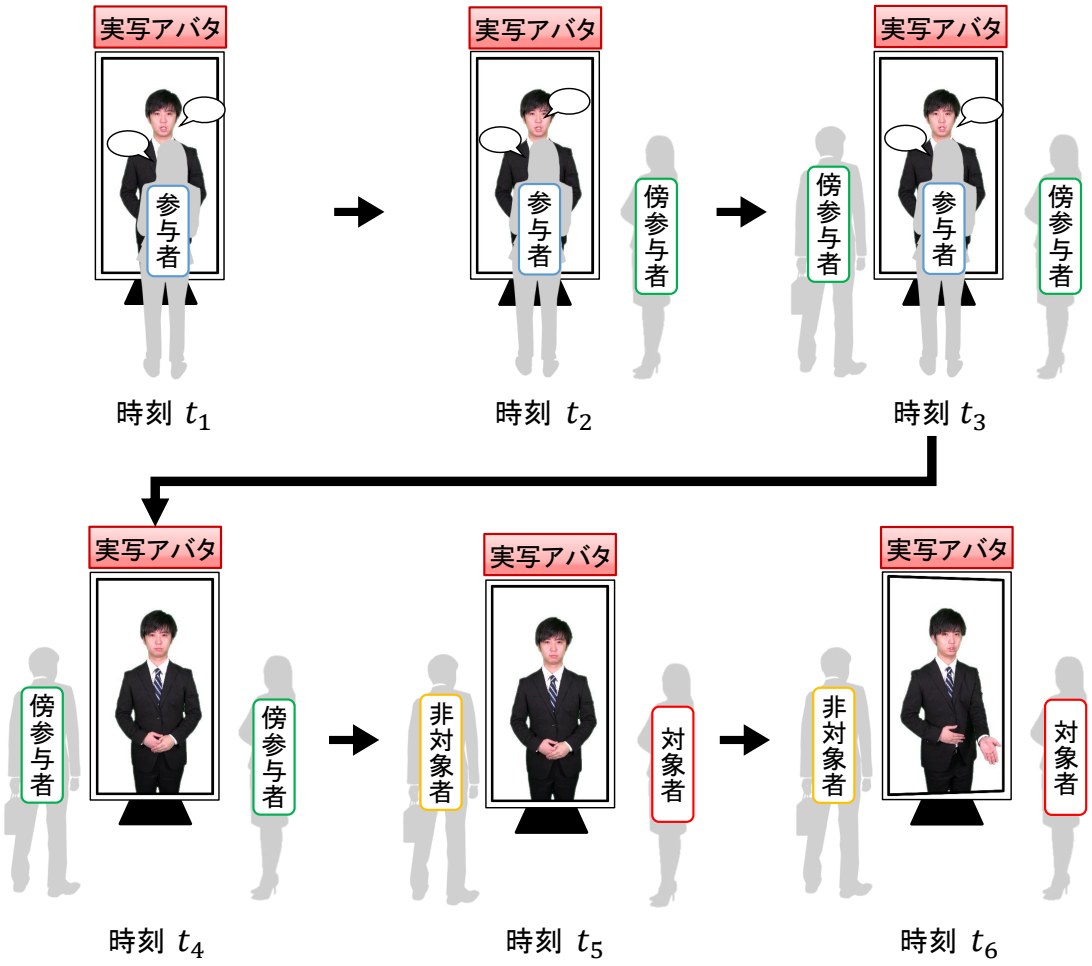


図 3.2: 本論文で想定する状況が発生する案内の流れ.

案内を受けているとする．次の時刻 t_2 で実写アバタを利用する傍参与者が新たに訪れると，参与者の右側の位置に立つ．さらに時刻 t_3 で傍参与者が来た場合，参与者のもう片方の位置に立つ．時刻 t_4 で実写アバタの案内が終了し参与者が立ち去る．時刻 t_5 で，参与者の右側に立っていた傍参与者が対象者となり，左側に立っていた傍参与者が非対象者となる．時刻 t_6 で，対象者のみを指定するために，実写アバタ映像に動きを加える．なお，時刻 t_2 で傍参与者が左に立った場合は，時刻 t_5 で対象者が左となり，非対象者が右となる．

本論文では，対象者と非対象者が左右に分かれて存在する状況において，実写アバタ映像に動きを加えて対象者のみを指定することを考える．対象者のみを指定する上で，対象者と非対象者がどのように感じるかの主観評価を実施する．そ

の中でも対象者は実写アバタから指定されたと感じない場合，そもそもインタラクションが開始されない問題が発生する．そこで，対象者から良好な評価を得ることを主目的として，実写アバタに動きを加える場合と加えない場合を比較することで，次の仮説が成立するかどうかを検証する．

H1: 対象者と非対象者が左右の位置に分かれて存在する状況において，実写アバタに動きを加えた方が，対象者 は自身の方を 実写アバタが向いた と感じ，さらに 実写アバタが指定した と感じる．

また，対象者から良好な評価を得た上で，非対象者の評価も考慮する必要がある．対象者の仮説 H1 が成立した上で，非対象者は実写アバタから指定されていないと感じることが望ましい．そこで，次の仮説が成立するかどうかも合わせて検証する．

H2: 対象者と非対象者が左右の位置に分かれて存在する状況において，実写アバタに動きを加えた方が，非対象者 は自身の方を 実写アバタが向いていない と感じ，さらに 実写アバタが指定していない と感じる．

3.4 映像表現

3.4.1 映像に加える動き

実写アバタが対象者のみを指定するために，実写アバタ映像にどのような動きを加えるべきかについて議論する．本論文では，人同士で発生するインタラクションにおける動きを映像に取り込むことを考える．また，人と機械の間で発生するインタラクションにおける動きも映像に取り込むことを考える．以下でそれぞれの動きについて詳細を述べる．

案内者の動き

実際の案内者がどのように動くことで対象者を指定するかについて述べる．案内者はまず対象者の方を向き，その後に声をかけて対象者を指定する．ここで重

要となる点は，案内者が対象者の方を向く動きである．この動きにより，対象者は自身の方を向いていると感じる．さらに，対象者は案内者から声をかけられることで，自身が指定されたと判断する．以下で，実写アバタ映像中の被写体にどのような動きを加えるかについて考える．

実写アバタ映像を使用する場合，モナリザ効果 [44, 45] を考慮する必要がある．視線をカメラに向けている実写アバタ映像を見ると，モナリザ効果により常に視線が合っていると感じる．逆に，視線を外している実写アバタ映像を，どの位置から見ても視線が合っていないように感じる．実写アバタが対象者とインタラクションを開始するためには，実写アバタが対象者と視線を合わせることが重要な要素の一つである．実写アバタが案内者のようにユーザの方へ視線を向ける映像を使うと，対象者は実写アバタと視線が全く合わない．そのため，実写アバタの視線はカメラ方向の正面に固定したままで，顔と体と手は対象者の方を向けることを考える．実写アバタが対象者の方に顔と体と手を向けた後に声をかけることで，対象者は自身の方を実写アバタから指定されたと感じることが期待される．

回転ディスプレイの動き

回転ディスプレイ [40, 41] がどのように動くことで対象者のみを指定するのかについて考える．回転ディスプレイは複数の傍参与者の中から対象者の方を向き，その後に音声を再生することで対象者を指定する．ここで重要となる点が，回転ディスプレイが対象者の方を向く動きである．本論文では，ディスプレイの物理的な枠を映像内で表現する．その枠と枠内の映像の見え方の変化を射影変換を用いて表現する．これにより，射影変換で映像を向けられた対象者は，自身の方へ実写アバタが向いたと感じることが期待される．この動きの後に映像中の被写体が声をかけることで，対象者は実写アバタから指定されたと感じることが期待される．

3.4.2 動作効果

実写アバタ映像に実際の案内者の動きを加える表現手法を動作効果と呼ぶ．3.4.1 で述べたように，視線を正面に固定したまま顔と体と手を向ける動きを映像に動作効果として組み込む．ただし，案内時に顔を動かす場合，自然と体は対象者に

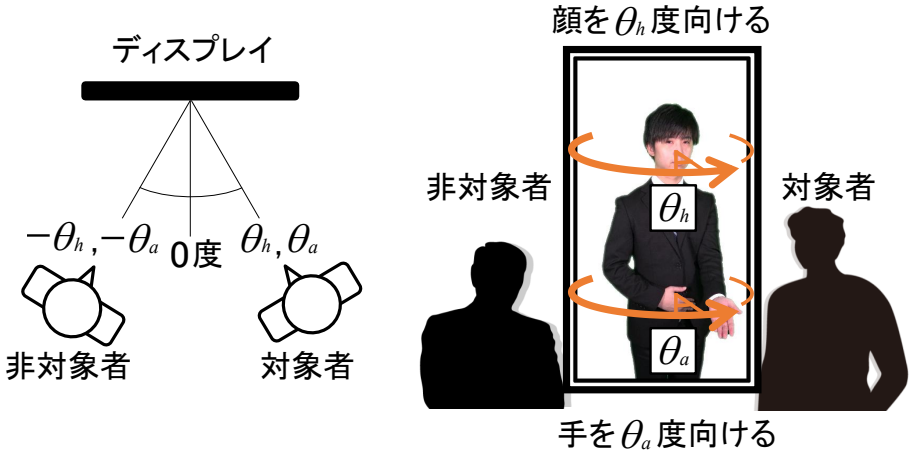


図 3.3: 動作効果のパラメータ.

向けることが多い．本論文では，体は顔の動きに連動するものとし，顔の角度のみを考えることにする．動作効果において，顔の向く方向を表す角度パラメータを θ_h，手を差し伸べる方向を表す角度パラメータを θ_a とする．それらの角度パラメータを図 3.3 に示す．θ_h と θ_a は正面方向を 0 度とする．

3.4.3 回転効果

実写アバタ映像に回転ディスプレイの動きを加える表現手法を回転効果と呼ぶ．3.4.1 で述べたように，実写アバタ映像に枠を加え，その映像自体が回転して見えるように射影変換を適用する．映像中の枠の回転する角度パラメータを θ_f とする．その角度パラメータを図 3.4 に示す．図中の y 軸とは，映像の中心を通る垂線である．θ_f は，映像上の y 軸の回転角度とし，正面方向を 0 度とする．

3.5 仮説の検証

3.5.1 セッティング

実写アバタを 80 インチの縦置きディスプレイに表示し，ディスプレイの正面から 1.2 メートルの距離に参与者が立つ環境を構築した．この参与者は，図 3.2 で説明した時刻 t_1 の人物とした．ペアとなる実験協力者らに，両者とも傍参与者とし

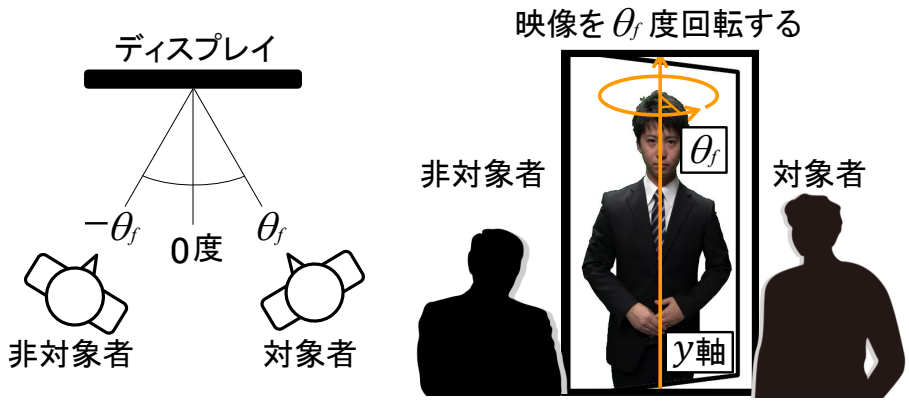


図 3.4: 回転効果のパラメータ.

て参与者の横に立つように指示した．その後，時刻 t_4 の状況から主観評価を開始した．ペアとなる実験協力者らの中から一人を対象者，もう一人を非対象者としてランダムに役割を設定した．ただし実験協力者らには，自身がどちらの役割を与えられたかについて伝えなかった．ペアとなる実験協力者らに刺激映像を表示した後に，以下の設問に回答させた．

設問：実写アバタが

Q1：あなたの方を向いたと感じたか

Q2：あなたを指定したと感じたか

回答 (評価値)

- そう感じなかった (-1.5 点)
- ややそう感じなかった (-0.5 点)
- ややそう感じた (0.5 点)
- そう感じた (1.5 点)

設問の回答は上記 4 段階からの選択とした．なお，反転項目も設け，評価値を算出する際に得点を逆転した．実験協力者間のインタラクションについては特に指示を行わなかった．なお，実験中に実験協力者間のインタラクションは見られな

かった．実験者効果を排除するために，オペレーターを実験協力者から見えない位置に配置した．

刺激映像は，文献 [46] の撮影環境を用いて収集した．実写アバタに動作効果を加えるために，実写アバタの被写体が実際の動きを行った映像を撮影した．実写アバタに回転効果を加えるために，画像編集ソフトを用いて映像を三次元的に回転させた．

3.5.2 動作効果のパラメータ調査

角度パラメータ θ_h と θ_a の結果

動作効果のみを刺激映像に加えた場合について，実験協力者の評価値に角度パラメータが与える影響を調査した．実験協力者は男性 8 名，女性 2 名の合計 10 名とし，その平均年齢は 22.9 ± 0.8 歳であった．動作効果のみの 18 個の刺激映像を生成した．その際，実写アバタの顔の角度パラメータ θ_h を 0 度から 40 度の間で 5 度刻みに変化させ，手の角度パラメータ θ_a を 0 度に固定した．手の角度パラメータ θ_a も同様に角度を変化させ，顔の角度パラメータ θ_h を 0 度に固定した．実験協力者に対する表示の順序は，昇順と降順でランダムとした．ペアとなる実験協力者らに各刺激映像を表示した後に，設問 Q1 に回答させた．実験協力者は，刺激映像の 18 通りと役割の 2 通りの全 36 通りの評価を行った．

実験結果を図 3.5 に示す．図中では，対象者の評価値は大きくなる方が良く，非対象者の評価値は小さくなる方が良い．本論文では，対象者から良好な評価を得ることを主目的としているため，対象者の評価値が高い角度パラメータを採用することとした．また，対象者と非対象者の間で評価値の差が大きい角度パラメータを採用することとした．その結果，角度パラメータ θ_h，θ_a はそれぞれ 15 度であった．

動作効果として用いる動きの比較

動作効果に含める被写体の動きについて調査した．実験協力者は男性 10 名とし，その平均年齢は 22.3 ± 1.5 歳であった．刺激として表示する実写アバタ映像を以下

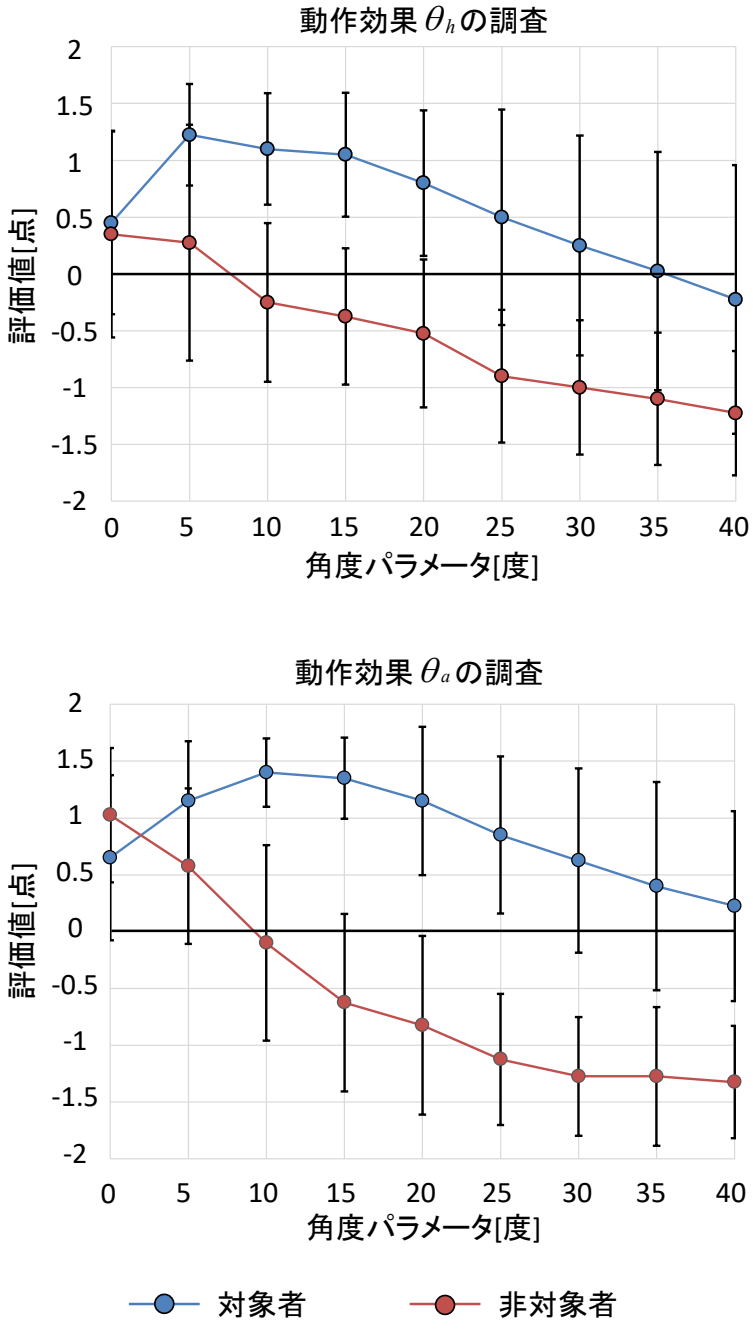


図 3.5: 動作効果の角度パラメータの調査結果.

の 4 つの手法で生成した．

A1 :動きなし

A2 :顔のみの動き

A3 :手のみの動き

A4 :顔と手の両方の動き

各動きの映像の例を図 3.6 に示す．A1 から A4 の全ての映像の最後で，実写アバタはセリフを発声した．セリフは，次の方どうぞとした．角度パラメータ θ_h，θ_a は，3.5.2 で採用した 15 度とした．ペアとなる実験協力者らに刺激として各実写アバタ映像を表示した後に，設問 Q1 と設問 Q2 に回答させた．各実験協力者は，手法 4 通りと役割 2 通りの全 8 通りの評価をランダムな順で行った．評価結果の解析方法として 2 要因の分散分析を行い，多重検定としてウィルコクソンの符号和順位検定と Bonferroni 補正を適用した．

対象者の結果を図 3.7(a) に示す．評価値は大きくなる方が，各設問に対して対象者は感じていることを表す．実験結果から Q1 に関して，顔の動きと手の動きの交互作用がないものの，顔の動きに主効果があることを確認した．手の動きには主効果がないことを確認した．対象者が向いていると感じる Q1 に関して，顔の動きが有効であると考えられる．Q2 に関して，顔の動きと手の動きの交互作用がないものの，手の動きに主効果があることを確認した．顔の動きには主効果がないことを確認した．多重検定の結果では，A1 と A3，A1 と A4，A2 と A4 の間で有意差が見られた．対象者が指定されたと感じる Q2 に関して，手の動きが有効であると考えられる．以上の対象者の結果より，各設問において顔または手の動きに効果があることが分かった．顔と手の両方の動きを用いた A4 に着目すると，各設問において評価値が A2 と A3 より低下する現象は見られなかった．本論文では，Q1 と Q2 の両方の評価値を高くすることを目的としているため，対象者には動作効果として手法 A4 を採用することとした．

非対象者の結果を図 3.7(b) に示す．評価値は小さくなる方が各設問に対して非対象者は感じていないことを表す．Q1 と Q2 で顔の動きと手の動きの交互作用がなく，各動きの主効果もないことを確認した．非対象者が向いていないと感じる

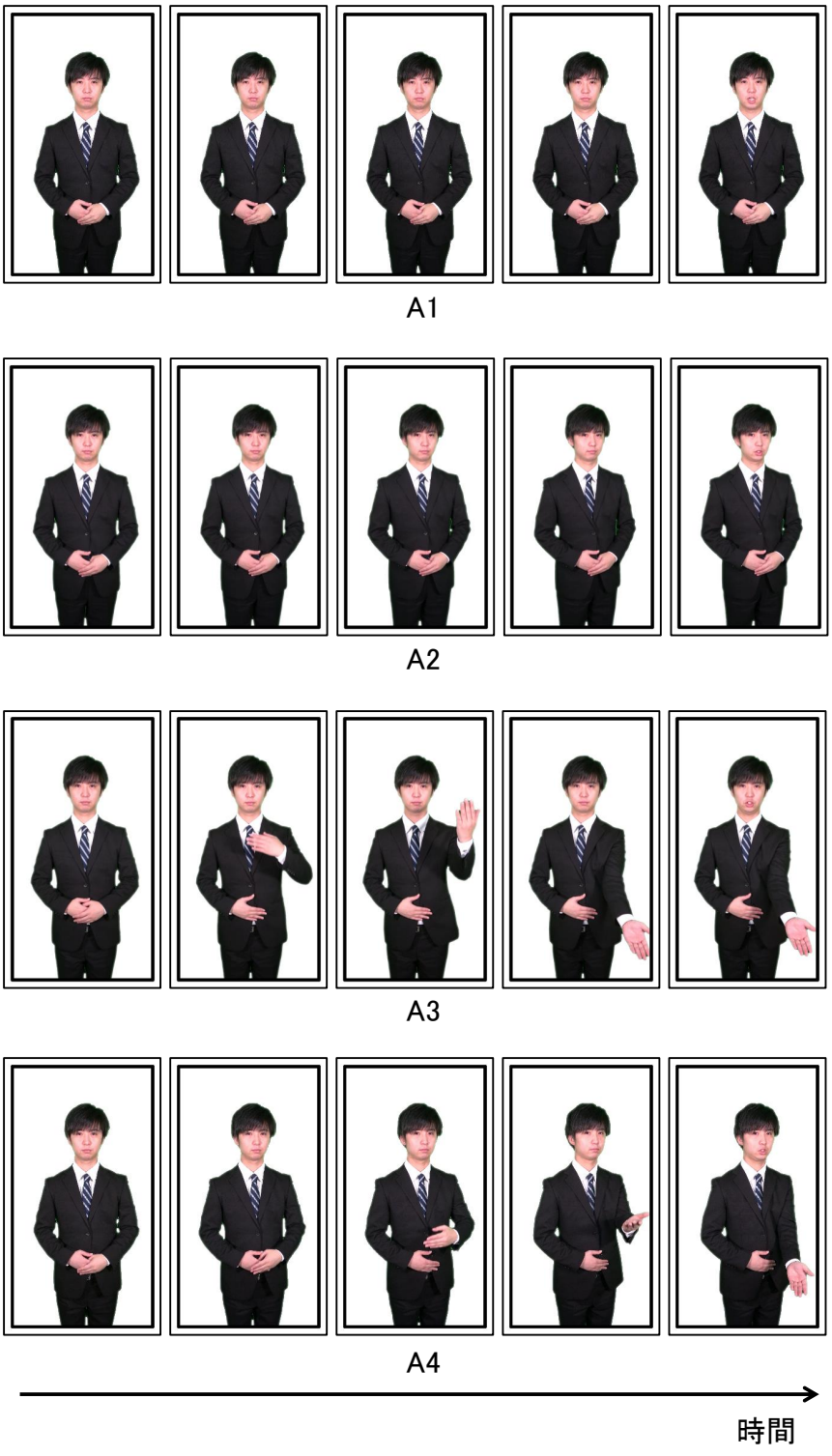


図 3.6: 動作効果に含める動きの比較に用いた表現手法の映像の例.

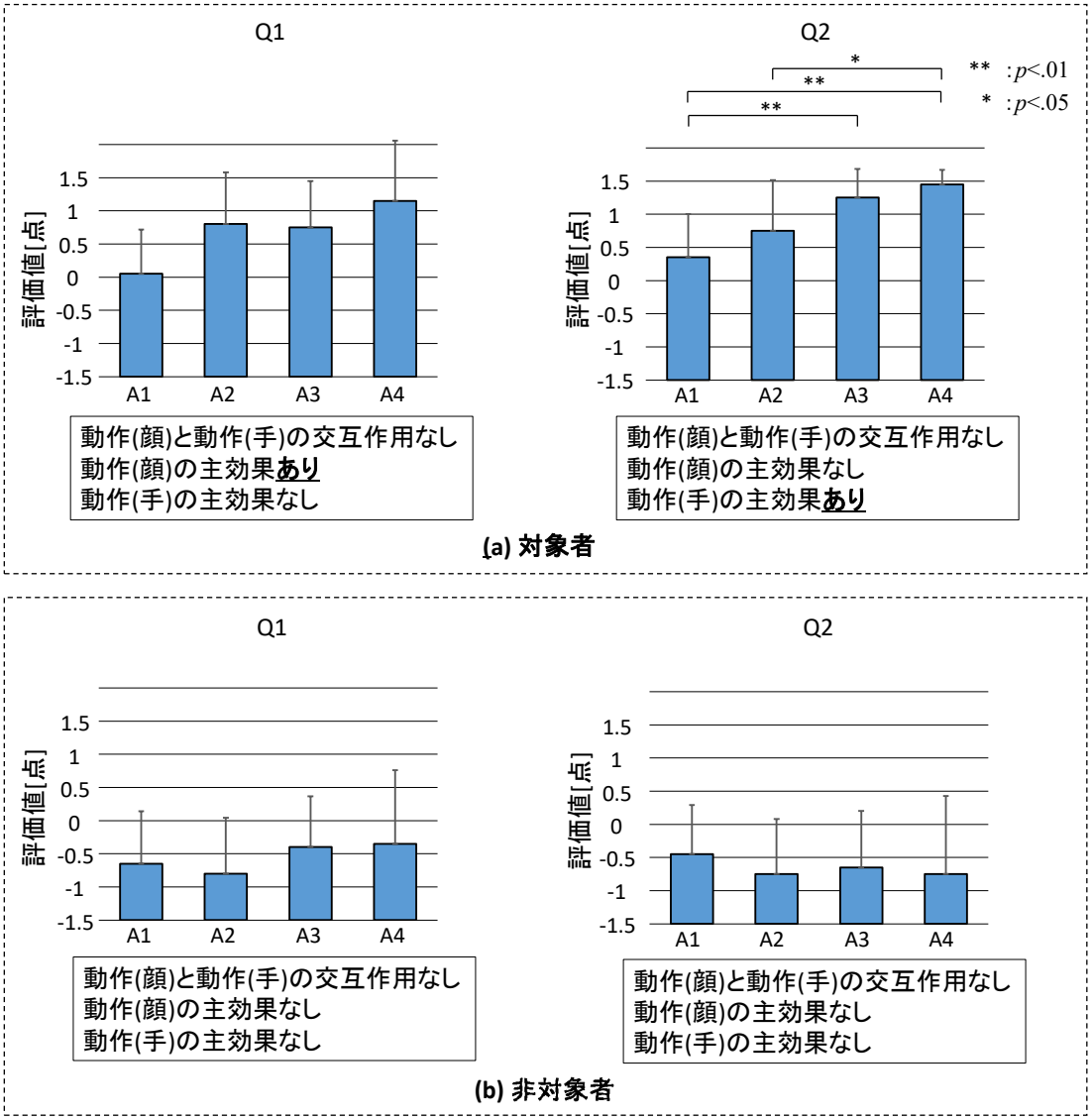


図 3.7: 動作効果の主観評価の結果.

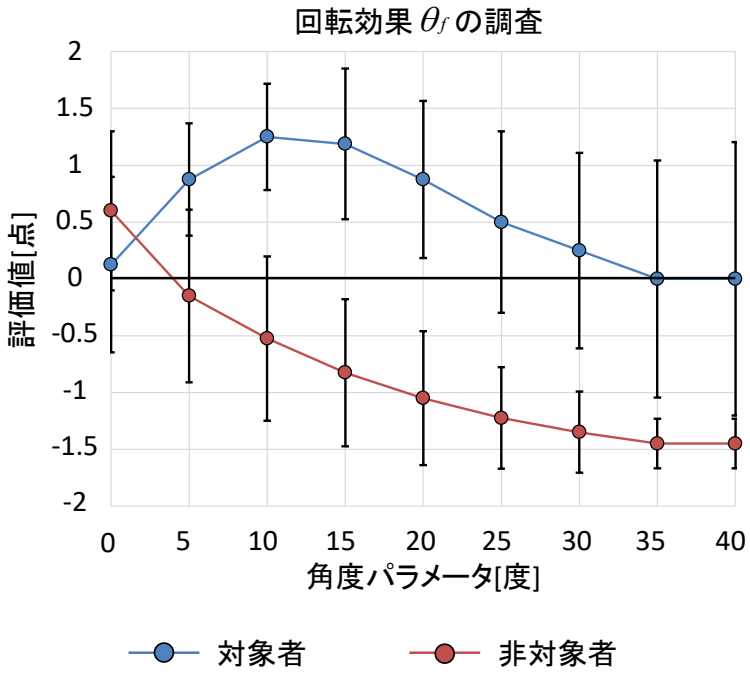


図 3.8: 回転効果の角度パラメータの調査結果.

Q1，非対象者が指定されていないと感じる Q2 に関して，A1 から A4 のどの手法を採用したとしても大きな差は見られないと考えられる．本論文では対象者から高い評価値を得ることを主目的としているため，非対象者には手法 A4 をそのまま用いることとした．

3.5.3 回転効果のパラメータ調査

枠の角度パラメータ θ_f のみを変化させた場合について調査した．実験協力者は男性 7 名，女性 3 名の合計 10 名とし，その平均年齢は 22.6 ± 0.9 歳であった．実験は，3.5.2 と同様の流れで行った．角度パラメータ θ_f を変化させた回転効果のみを含む 9 個の刺激映像を生成した．刺激映像に対してペアとなる実験協力者らが評価を行った．

実験結果を図 3.8 に示す．図中では，対象者の場合は評価値が大きくなる方が良く，非対象者の場合は評価値が小さくなる方が良い．3.5.2 と同様の方針で角度パラメータを採用することとした．その結果，角度パラメータ θ_f は 15 度であった．

3.5.4　動作効果と回転効果の組み合わせの検証

実験条件

本実験では，動作効果と回転効果を組み合わせた場合の検証を行った．実験協力者は男性 19 名，女性 1 名の合計 20 名とし，その平均年齢は 21.7 ± 2.1 歳であった．刺激映像を以下の 4 つの手法で生成した．

M1 :動作効果なし，かつ，回転効果なし

M2 :動作効果あり

M3 :回転効果あり

M4 :動作効果と回転効果の組み合わせ

各手法の映像の例を図 3.9 に示す．M1 から M4 の全ての映像の最後で，実写アバタはセリフを発声した．セリフは，次の方どうぞとした．ペアとなる実験協力者らに刺激として各実写アバタ映像を表示した後に，設問 Q1 と設問 Q2 に回答させた．各実験協力者は，手法 4 通りと役割 2 通りの全 8 通りの評価をランダムな順で行った．評価結果の解析方法として 2 要因の分散分析を行い，多重検定としてウィルコクソンの符号和順位検定と Bonferroni 補正を適用した．

M2 と M3 の角度パラメータ θ_h，θ_a，θ_f は，3.5.2 と 3.5.3 で採用した 15 度とした．M4 の角度パラメータは動作効果と回転効果を同程度ずつ適用するために，それぞれの角度パラメータを 7.5 度に設定した．動作効果の動きは，3.5.2 で採用した顔と手の両方の動きを適用した．

検証結果

対象者の結果を図 3.10(a) に示す．評価値は大きくなる方が，各設問に対して対象者は感じていることを表す．実験結果から Q1 に関して，動作効果と回転効果の交互作用がないものの，動作効果に主効果があることを確認した．回転効果にも主効果があることを確認した．対象者が向いていると感じる Q1 に関して，動作効果と回転効果が有効であることを確認した．さらに，多重検定の結果を見ると，動

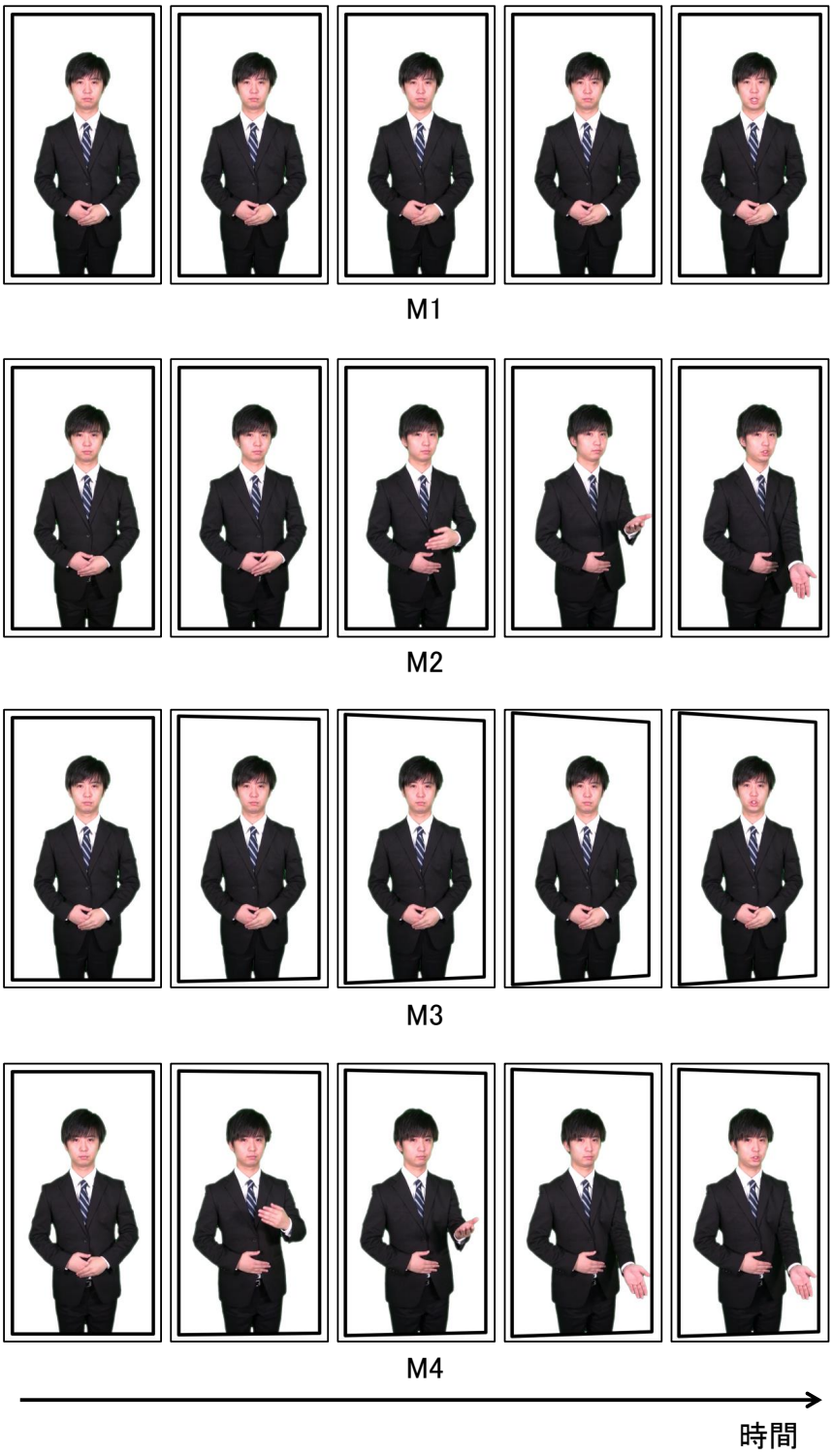


図 3.9: 動作効果と回転効果の組み合わせの主観評価に用いた表現手法の映像の例.

作効果と回転効果を組み合わせた手法 M4 が，他の全ての手法に対して有意差があることを確認した．よって，対象者が向いていると感じる Q1 に関して，動作効果と回転効果を組み合わせる M4 が対象者の評価値を高めると考えられる．Q2 に関して，動作効果と回転効果の交互作用があることを確認した．さらに，動作効果に単純主効果があることを確認した．回転効果には単純主効果はないことを確認した．対象者が指定されたと感じる Q2 に関して，動作効果が有効であると考えられる．動作効果と回転効果の交互作用があることから，動作効果に回転効果を組み合わせることで，対象者がさらに指定されたと感じると考えられる．よって Q2 では，組み合わせの M4 が，対象者の評価値を最も高めることを確認した．以上の結果より，対象者と非対象者が左右に分かれて存在する状況において，実写アバタ映像に動きを加えることで，対象者は自身の方を実写アバタが向いたと感じ，さらに実写アバタが指定したと感じる仮説 H1 は成立すると言える．

非対象者の結果を図 3.10(b) に示す．評価値は小さくなる方が各設問に対して非対象者は感じていないことを表す．Q1 に関して，動作効果と回転効果の交互作用がないものの，動作効果に主効果があることを確認した．回転効果には主効果がないことを確認した．また，多重検定の結果においても動作効果を適用した場合に有意差が見られた．このことから，対象者が向いていないと感じる Q1 に関して，動作効果を適用すると非対象者の評価値が逆に高くなることが分かった．Q2 に関して，動作効果と回転効果の交互作用がなく，主効果も確認されなかった．非対象者が指定されていないと感じる Q2 に関して，実写アバタ映像に動きを加えても非対象者の評価値に関して影響を与えないと考えられる．以上の結果より，実写アバタ映像に動きを加えても，非対象者に関する仮説 H2 は成立するとは言えなかった．

前節までででも述べたように，非対象者が存在する中で対象者とインタラクションを開始するためには，対象者が指定されたと感じることが最も重要であると考えられる．非対象者の仮説 H2 は成立しないものの対象者の仮説 H1 は成立するため，本論文では M4 を実写アバタ映像における動き表現として採用することとした．

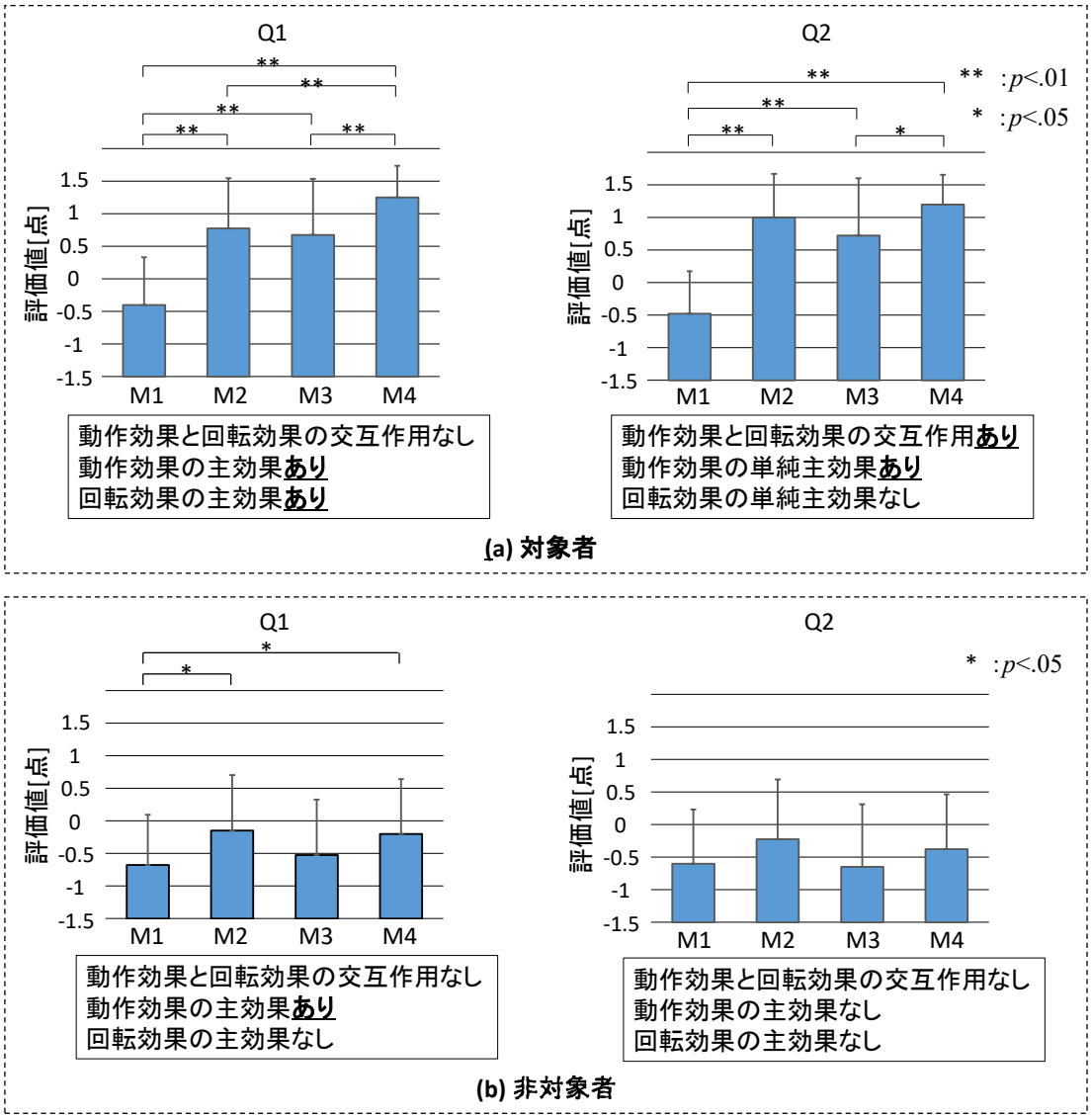


図 3.10: 動作効果と回転効果の組み合わせの主観評価の結果.

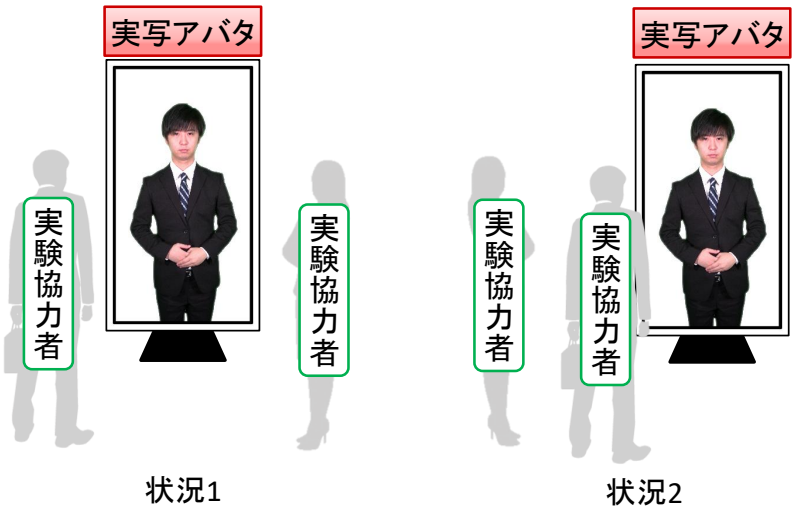


図 3.11: 立ち位置変化の状況.

3.6 傍参与者らの立ち位置変化の検証

3.6.1 実験条件

ここまでの実験では，対象者と非対象者が左右に分かれた状況を想定し調査を行った．以下では，対象者と非対象者が片側に集まった状況を想定し主観評価を行った．これにより，動きを加える映像表現の限界を調査した．実験協力者は男性 16 名とし，その平均年齢は 22.8 ± 1.0 歳であった．本実験では，以下の 2 つの条件で比較した．

状況 1 : 対象者と非対象者が左右のそれぞれに立つ状況 (3.3 で設定した状況)

状況 2 : 対象者と非対象者が共に左へ立つ状況

ペアとなる実験協力者らの立ち位置の例を図 3.11 に示す．ペアとなる実験協力者らの中から一人を対象者，もう一人を非対象者としてランダムに役割を設定した．ただし実験協力者には，自分がどちらの役割を与えられているかについて伝えなかった．ペアとなる実験協力者らには，3.5.4 で採用した手法 M4 の刺激映像を表示した．その後，設問 Q1 と設問 Q2 に回答させた．各実験協力者は，刺激映像の

1 通りと役割の 2 通りと状況の 2 通りの全 4 通りの評価を行った．評価値の解析方法として，対応のある t 検定を用いた．

対象者と非対象者が共に左へ立つ状況 2 において，実写アバタの角度パラメータ θ_h，θ_a，θ_f が両者の間で同じであると，対象者と非対象者で全く同じ評価値となる可能性が考えられる．状況 2 において，対象者が正面に近い位置に立つ場合は全ての角度パラメータを 5 度とし，対象者が正面から遠い位置に立つ場合は 10 度とした．一方の状況 1 において，全ての角度パラメータを 7.5 度とした．

3.6.2　調査結果

対象者の結果を図 3.12(a) に示す．Q1 では，状況 1 と状況 2 との間に有意差を確認した．対象者は，状況 2 より状況 1 の方が，実写アバタが向いていると感じることを確認した．また，Q2 でも同様の結果であった．次に，非対象者の結果を図 3.12(b) に示す．Q1 では，状況 1 と状況 2 との間に有意差を確認した．非対象者は，状況 2 より状況 1 の方が，実写アバタが向いていないと感じることを確認した．また，Q2 でも同様の結果であった．

対象者と非対象者との間で評価値を比較すると，状況 1 では差が見られた．一方の状況 2 では，ほとんど差が見られなかった．本論文の動作効果と回転効果のみでは，状況 2 において，実写アバタは対象者のみを指定することができないと考えられる．以上の結果より，動きを加える映像表現の限界を示した．

状況 2 でも実写アバタが対象者のみを指定するためには，本論文で行った簡単な音声だけでなく，詳細な音声での指定など映像の動き以外の伝達手法の導入も今後検討する必要がある．

3.7　状況 2 まとめ

状況 2 では，実写アバタ映像に動きを加える場合と加えない場合を比較することで，対象者と非対象者が実写アバタから向かれたと感じるか，また，指定されたと感じるかについて調査した．実写アバタ映像に含める動きとして，動作効果の有無，回転効果の有無の組み合わせを比較した．動作効果の角度パラメータと

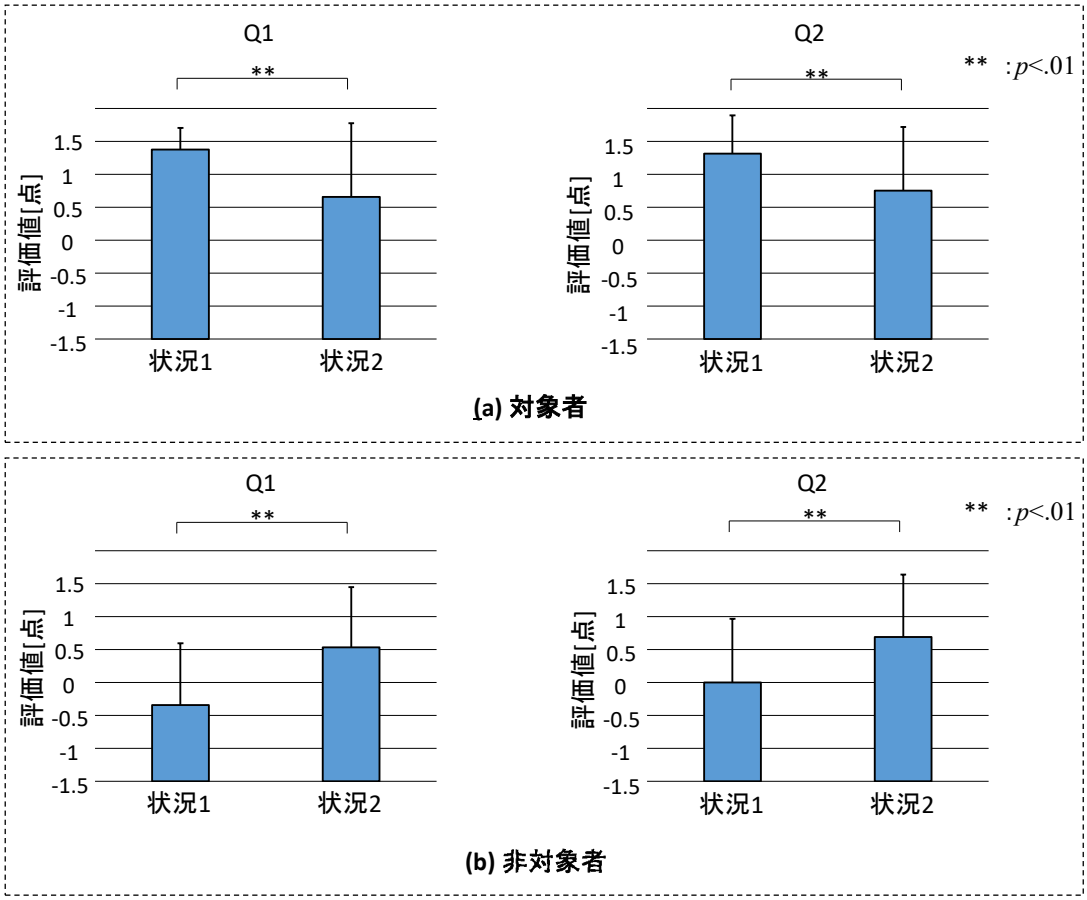


図 3.12: 立ち位置変化の主観評価結果.

実写アバタの動き，回転効果の角度パラメータを調査した．実写アバタが対象者のみを指定できるかどうかについて，対象者の仮説 H1 と非対象者の仮説 H2 を主観評価で調査した．さらに，動きを加える映像表現の限界も調査した．

今後の課題として，傍参与者の立ち位置が様々に変化した状態における伝達手法の検討，傍参与者の人数が増えた状態における伝達手法の検討，指向性スピーカーや回転ディスプレイなど実際の機構を用いた伝達手法との比較などが挙げられる．

第4章　小型ディスプレイにおける実写アバタの動き強調の検討

4.1　状況3の研究背景

提案する実写アバタを用いた案内システムでは，大型ディスプレイの案内と小型ディスプレイの案内に分けることができ，それぞれの大きさのディスプレイにおける動作生成について考える必要がある．大型ディスプレイの動作生成は，実物大の実写アバタを表示することができるため，実写アバタに実際の人物の動きを容易に再現することができる．一方，小型ディスプレイの動作生成は，モバイル端末の画面サイズが小さいため実写アバタの動きが小さくなってしまう．その結果，ユーザは実写アバタの動きを自然に認識することができない可能性がある．特に，直立姿勢の実写アバタがユーザとのインタラクションが始まるのを待っている状況では，実写アバタの動きが非常に小さく完全に静止しているように見える可能性がある．そしてこの場合，ユーザはアバタシステムが壊れていると誤って感じてしまい，インタラクションが円滑に開始できない問題が発生する．そこで，私たちは小型ディスプレイ上において自然な直立姿勢でアバタを表示するための動作生成について考える必要がある．

直立姿勢で立っている人の動作を観察してみると，常に一定の位置で体が揺れている．このような動作は身体動揺と呼ばれており，無意識のうちに筋肉に負担をかけないように体を動かしている．体の揺れを実写アバタで再現する方法は状況1で提案されている [47]．この手法は，短いビデオシーケンスから連続的で自然な体の揺れを生成するものである．しかし，この手法は大型ディスプレイ上の実写アバタで使用することを想定しており，小型ディスプレイ上で使用すると実写アバタの動きが非常に小さく不自然に感じられる可能性がある．そこで，直立姿勢の実写アバタの揺れを強調し，小型ディスプレイ上における直立姿勢の実写

アバタの動きをより自然に感じられるようにする必要があると考えられる．

本論文では，小型ディスプレイ上における直立姿勢の実写アバタの動きを強調する動作生成を検討する．具体的には，映像上の周期的な動きを強調する既存手法を用いて，揺れの大きさを制御するパラメータを小型ディスプレイ上における直立姿勢の実写アバタで評価する．そして，以下の仮説を検証する．

Hypothesis: 小型ディスプレイ上における実写アバタの身体動揺の揺れを強調することでユーザは実写アバタの動きをより自然であると感じる．

4.2 実写アバタの直立姿勢の重要性

実写アバタを用いたインタラクションシステムを実現するためには，実写アバタの状態を考慮する必要がある．実写アバタに必要な状態は，ユーザとインタラクションを行っている行動状態と，アバタが直立姿勢でインタラクションが開始されるのを待つ待機状態が存在する．実写アバタは常にユーザとインタラクションを行っている訳ではないため，実写アバタとユーザがインタラクションを開始するためには待機状態が重要な役割を果たす．例えば，実写アバタに待機状態が適切に適用されていない場合，ユーザはアバタとのインタラクションを開始して良いのか判断できない問題が発生する．

実写アバタの状態は図 1.5 に示すように，待機状態と行動状態が遷移しながらインタラクションが行われる．待機状態から行動状態へと円滑に遷移させるためには，待機状態において実写アバタに人間的な動きを実装する必要がある．ここで，空港やホテルのインフォメーションセンターなどにおける案内係がインタラクションを待っている時の直立姿勢を考える．人間が直立姿勢を維持する際には，立っていることへの筋肉の負担を分散させるために一定の位置を中心に絶えず揺れている．このような動きを身体動揺という．

ここで，実写アバタに身体動揺の揺れを再現することを考える．実写アバタを大型ディスプレイに表示する場合，ユーザが身体動揺の揺れの動きを視覚的に認識することができるため，実際の人間の身体動揺の揺れを実写アバタに再現することが可能である．しかし，身体動揺の揺れはとても小さな動きであり小型ディスプレイに実写アバタを表示する場合，実際の人間の身体動揺の揺れを視覚的に

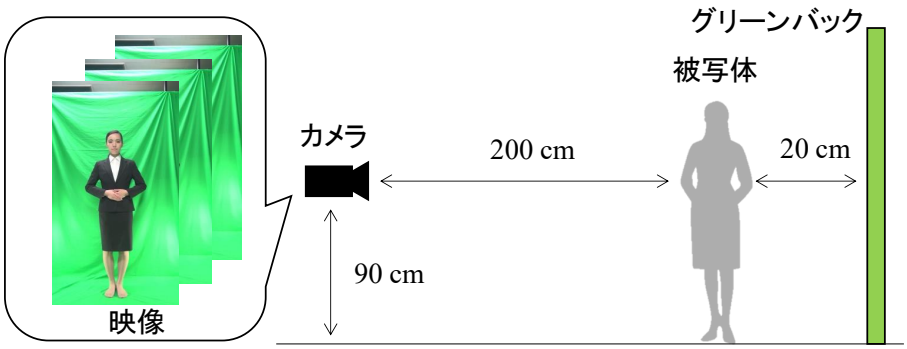


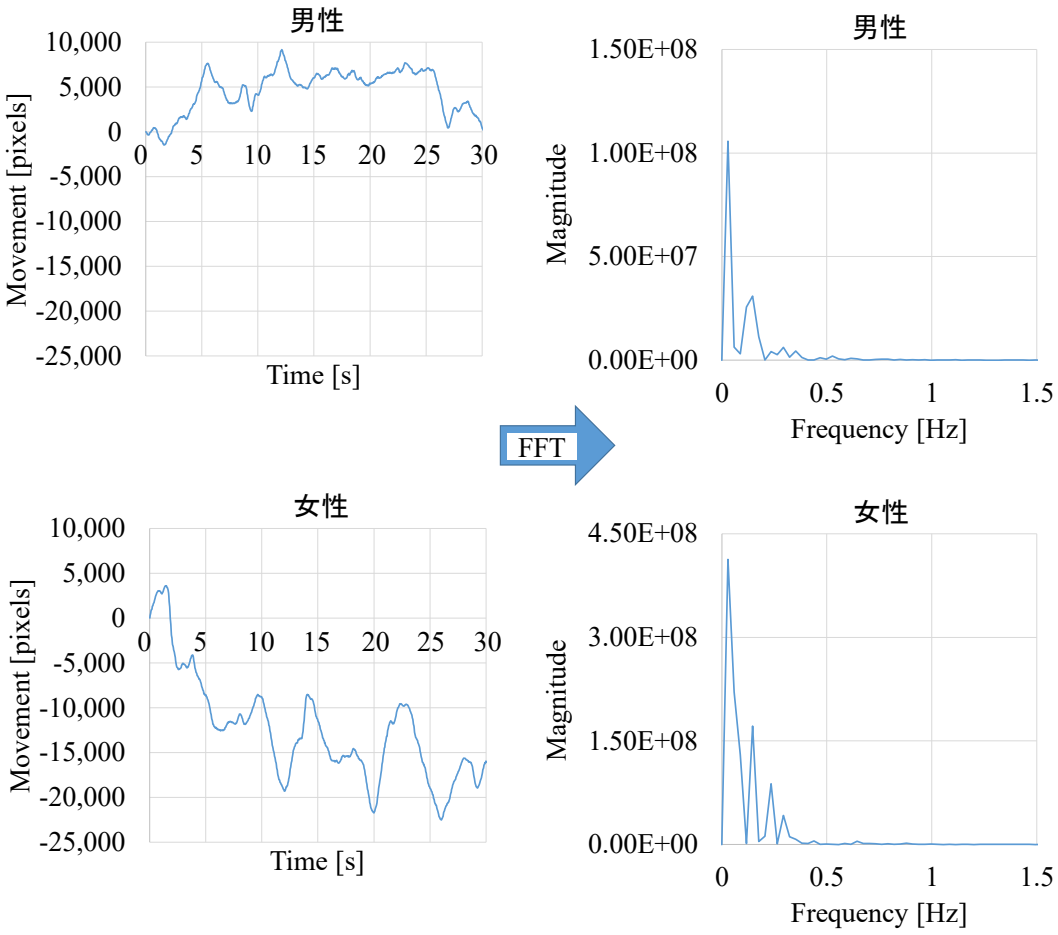
図 4.1: 実写アバタの撮影環境.

認識できない可能性がある．そして，ユーザは実写アバタが停止していると勘違いし，案内システムが故障していると誤認してしまう．そこで，本研究では小型ディスプレイ上の実写アバタの身体動揺の揺れを強調することで，実写アバタの動きを自然に感じられるようにすることを考える．

4.3　身体動揺の強調方法

ここでは，実写アバタにおける身体動揺の揺れの動きを強調する方法について述べる．実写アバタの身体動揺の揺れを強調する方法では，実写アバタの被写体が大きな身体動揺の揺れを行って直立姿勢している映像を撮影して，データセットを作成する方法が考えられる．しかし，この方法では，ディスプレイの大きさによってデータセットを使い分ける必要があり，大量のデータセットを用意する必要がある．また，身体動揺の揺れは人間が無意識のうちに行っているものであり，身体動揺の揺れの振幅を意識的に大きくすることは困難である．そこで，実写アバタの身体動揺の揺れを映像内で強調する方法を考える．

身体動揺の揺れを強調する方法として，映像中の周期的な動きを強調する Phase-based video motion processing [48] という既存手法が応用できると考えられる．本論文では，この手法を用いて身体動揺の揺れを強調するために，まず映像中の実写アバタの身体動揺の揺れの周波数を調査する．既存研究 [49] によると，多くの身体動揺の揺れの周波数は 0〜1.5Hz であると述べられている．しかし，実際の映像



(a) 計測された体の揺れの時間変化　　(b) 身体動揺の周波数解析

図 4.2: 身体動揺の計測と周波数解析の結果.

中における実写アバタの身体動揺の揺れについては確認されていないため，映像中の身体動揺の揺れを計測し，身体動揺の揺れの周波数を調査した．まず，実写アバタのデータセットを作成するために，図 4.1 の環境で映像を撮影した．撮影された映像から実写アバタの人物領域を背景差分によって抽出し，映像中の人物領域を時間軸上で比較することで身体動揺の揺れを計測した．具体的には，映像中の基準となる時刻の人物領域と比較する時刻の人物領域の差を測定した．図 4.2（a）は画像中の頭の動きを測定した結果を示している（男性 1 名，女性 1 名）．グラフの縦軸は基準時刻と比較時刻の画像中の人物領域の差を表す．図 4.2 (b) は高速フーリエ変換 (FFT) を用いて得られた身体動揺の揺れの周波数の結果である．結果から映像中で確認できる身体動揺の揺れの周波数は 0〜1.5Hz であることが確認できた．また，頭部以外の部位でも同様の結果が得られた．そのため，実写アバタの身体動揺の揺れを強調するために，Phase-based video motion processing を用いて周波数 0〜1.5Hz の範囲の映像中の動きを強調させる．また，パラメータ α を用いて映像中の動きの大きさを制御する．α の値が大きいほど，映像中の動きが大きくなる．

4.4 主観評価

4.4.1 主観評価の実験条件

身体動揺の揺れを強調した実写アバタを表示して，主観評価を行った．強調した映像中の動きの周波数は，映像中で確認できた身体動揺の揺れの周波数 0〜1.5Hz とした．実写アバタは男性アバタと女性アバタの 2 種類を使用した．図 4.3 と図 4.4 は，男性と女性の実写アバタ映像である．映像の長さは 20 秒であり，以下の実写アバタ映像を比較した．

M1: 動きを強調していない実写アバタ.

M2: α 値が 0 でほとんど動きがない実写アバタ.

M3: α 値が 1 で少しだけ大きな動きに強調した実写アバタ.

M4: α 値が 2 で比較的大きな動きに強調した実写アバタ.

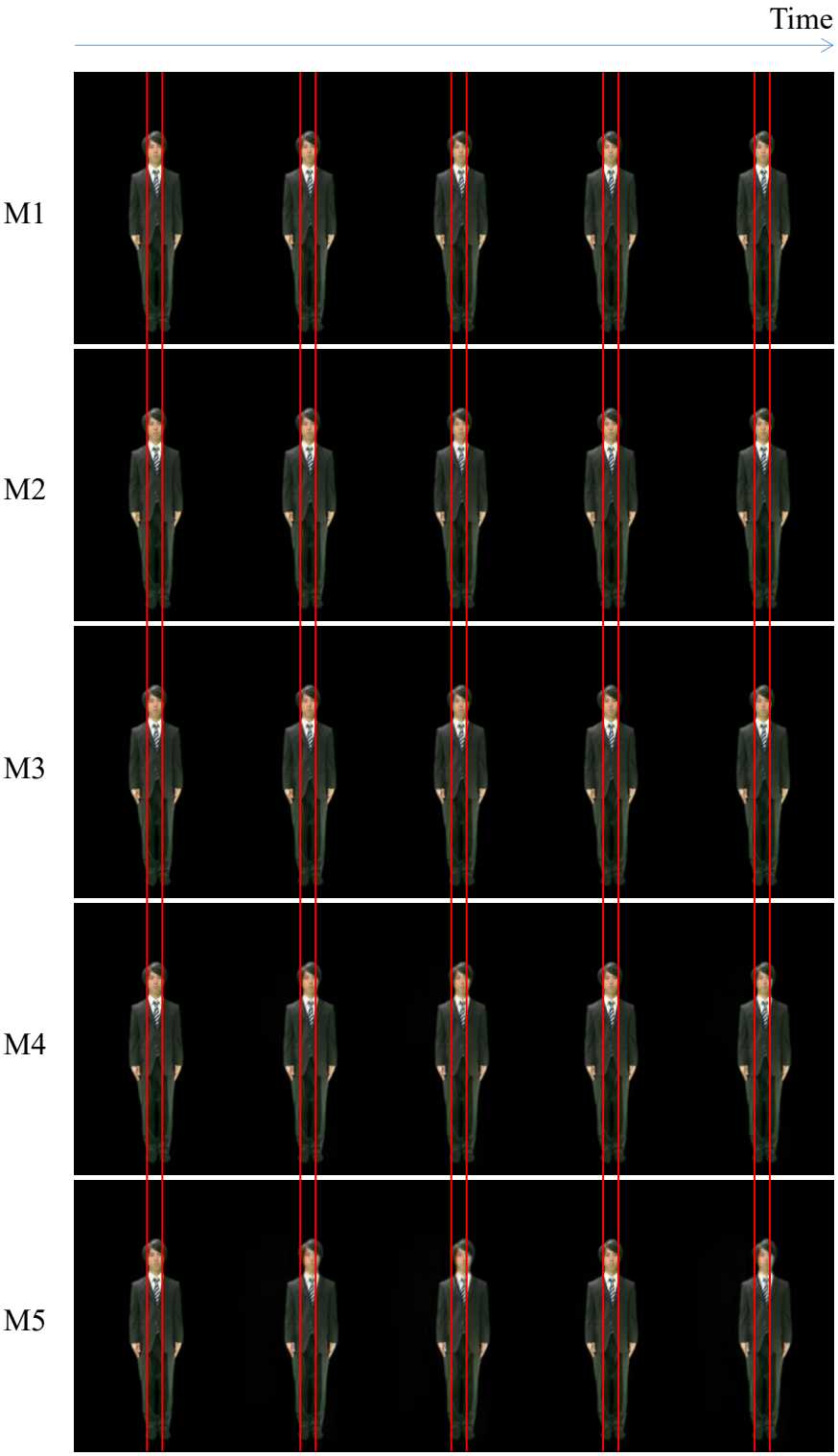


図 4.3: 主観評価に使用した男性アバタの映像の例.

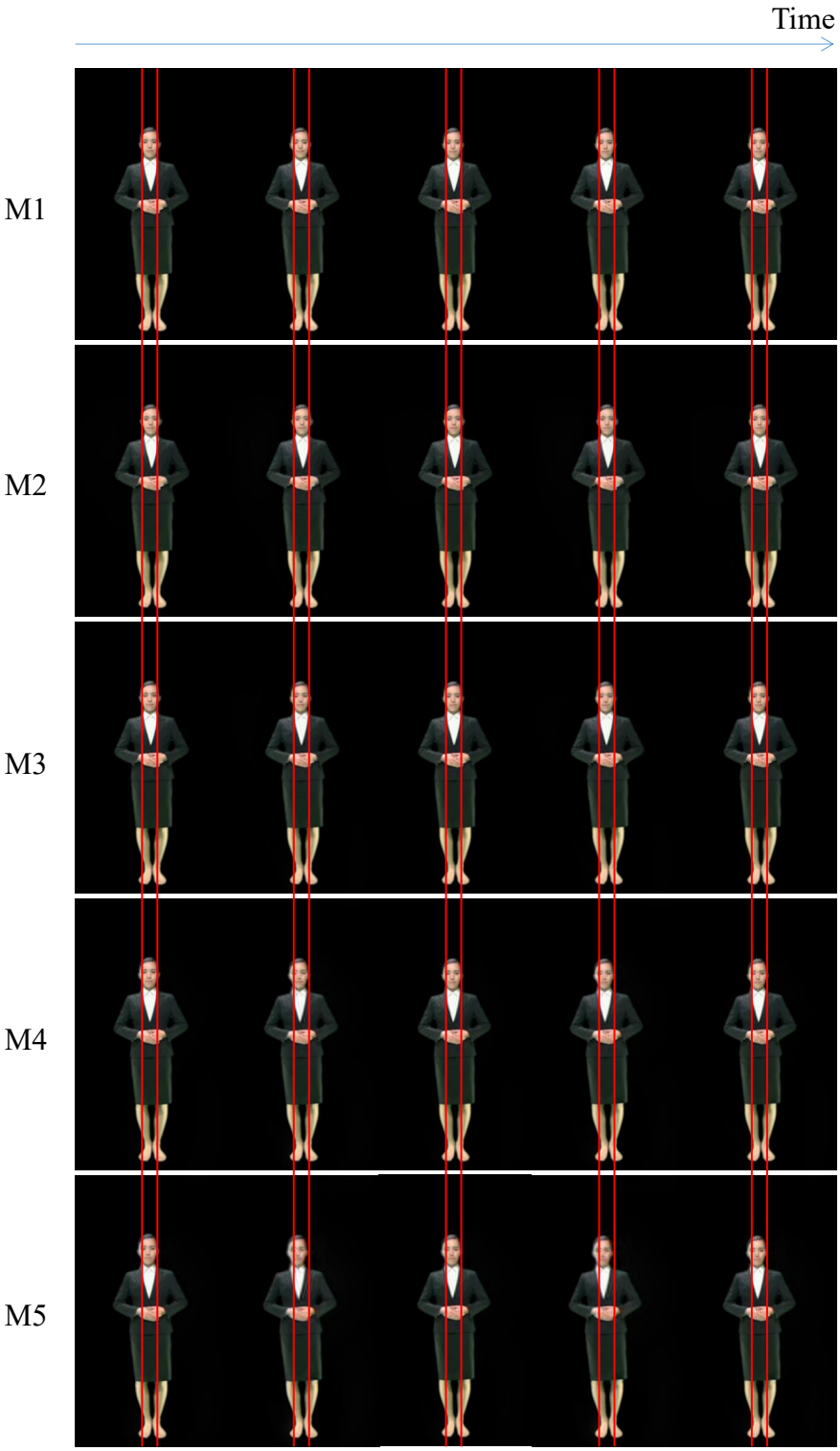


図 4.4: 主観評価に使用した女性アバタの映像の例.

図 4.5: 主観評価に使用した質問フォーム．

M5: α 値が 3 で大きな動きに強調した実写アバタ．

評価方法は，サーストンの一対比較法を用いた．22 名の実験協力者（男性 17 名，女性 5 名，平均年齢 22.2 歳）が，ディスプレイに表示された 5 つの実写アバタ（M1，M2，M3，M4，M5）を評価した．アバタは以下のサイズに調整した．

S1: 高さ 71.8mm (6.5 インチのディスプレイの半分のサイズ)

S2: 高さ 44.2mm (4 インチのディスプレイの半分のサイズ)

5 種類の実写アバタをペアにして表示し，実験協力者に以下の質問に回答させた．

Q: どちらの実写アバタの動きがより自然だと感じたか？

図 4.5 は実験協力者が質問の回答を行うフォームである．質問フォームの左側に表示されている 2 つの実写アバタを実験協力者が評価した．図 4.6 は主観的な評価を行った実験協力者の状況である．実験協力者は椅子に座り，机上の PC に表示された質問フォームを使用して評価を行った．実験協力者の目は平均して地面から 115cm，ディスプレイから 61cm のところにあった．ディスプレイの平均角度は 119 度であった．実験協力者は 2 つの表示サイズと 2 つの実写アバタの種類（男性

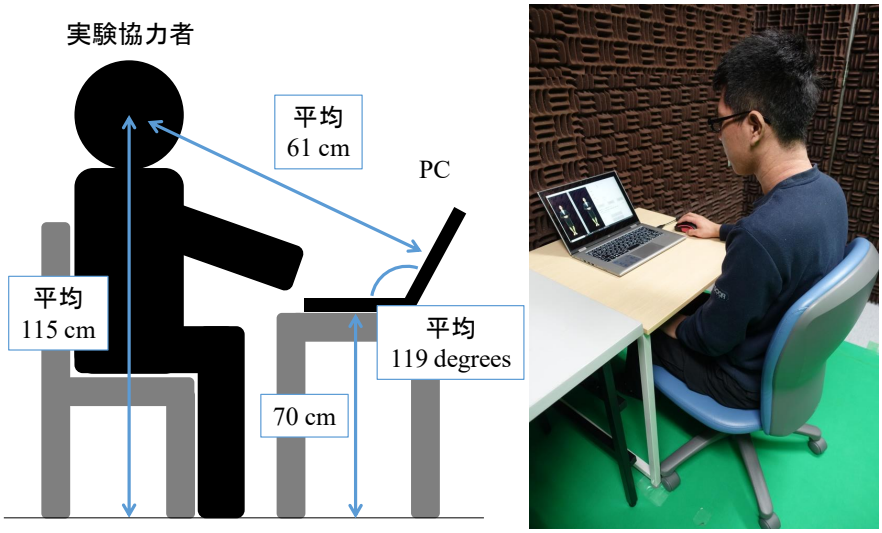


図 4.6: 主観評価の実験環境.

と女性）と実写アバタの揺れを変更した 5 種類の映像のペア（${}_5C_2 = 10$）をランダムに表示され，合計 40 回評価を行った．

4.4.2 主観評価の結果

図 4.7 は 5 種類の身体動揺の揺れを強調した実写アバタの主観評価結果を示したものである．グラフは，得られた票数から主観評価の結果を一軸にして表したものである．主観評価の結果は高い値であるほど良い評価が得られたことを示し，低いと悪い評価が得られたことをを示す．まず，男性アバタの結果に注目する．S1 では M1 の評価が最も高かった．これは，実写アバタの身体動揺の揺れを強調しなくても良いと実験協力者が感じたからだと考えられる．一方，S2 では M1 よりも M3 と M4 の評価が高かった．これは，ディスプレイサイズが小さい時には，身体動揺の揺れを強調した方が良いと感じられたからだと考えられる．M5 は身体動揺の揺れが大きくなりすぎて，不自然な動きであると感じられたために評価が低くなったと考えられる．また，M2 は動きが小さすぎて評価が低かったと考えられる．

次に，女性アバタの結果に注目する．S1 では M1 が最も評価が高かった．これは，男性アバタの時と同様に実写アバタの身体動揺の揺れを強調しなくても良いと実験協力者が感じたからだと考えられる．S2 では M2，M3，M4 の評価が M1 の評

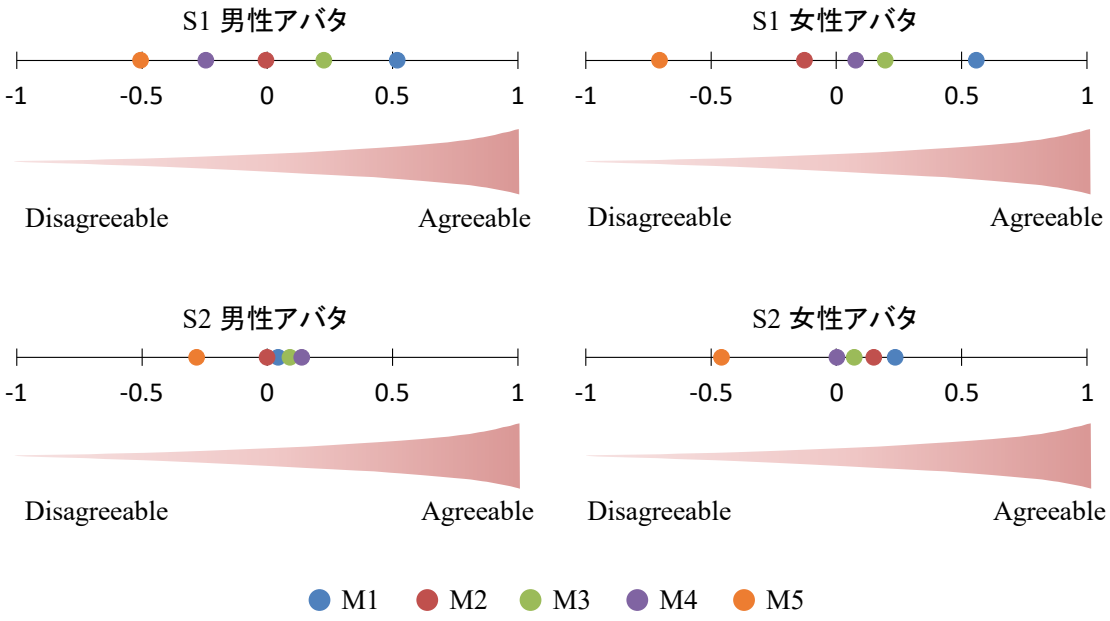


図 4.7: サーストンの一対比較を用いた主観評価の結果.

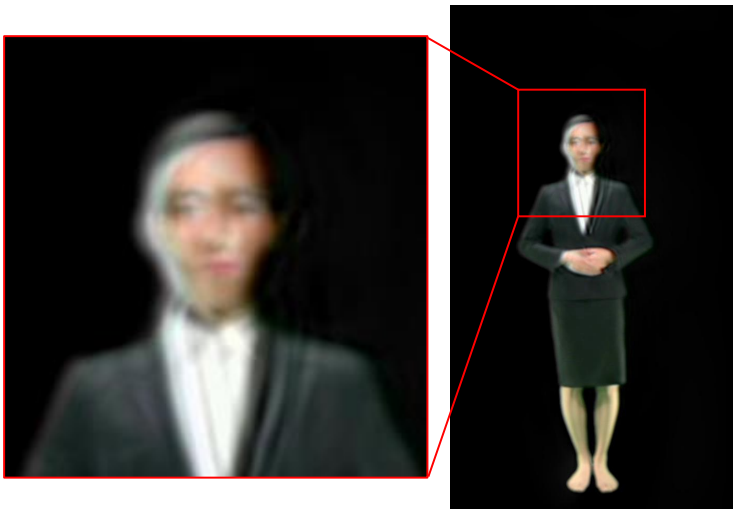


図 4.8: 女性アバタの動きを強調した際に発生したノイズ.

価に近づいてはいたものの，M1 の評価が最も高かった．この結果は，Phase-based video motion processing を用いて映像中の動きを強調した際に，実写アバタ映像にノイズが発生したからだと考えられる．図 4.8 は女性の実写アバタのノイズを示している．女性の実写アバタの動きを強調したところ M3，M4，M5 の映像にノイズが多く発生した．そのため，M3，M4，M5 の評価は M1 の評価よりも低くなったと考えられる．M2 はアバタの動きが小さすぎたため評価が低くなったと考えられる．

男性アバタの結果では，S2 に対して M3 と M4 の評価が M1 よりも高かった．女性アバタの結果では，S1 よりも S2 の方が M3 と M4 の評価が M1 に近かった．そのため，実写アバタの身体動揺の揺れをノイズを加えることなく強調することが可能であれば，ディスプレイサイズが小さい時には実写アバタが自然な動きをしていると認識すると考えられる．

4.5 状況 3 まとめ

直立姿勢の実写アバタの身体動揺の揺れを強調することで，小型ディスプレイ上でより自然に感じられるようにすることを狙った．主観評価実験では，実写アバタの体の揺れを強調してノイズが少なかった場合，実験協力者は小型ディスプレイ上で実写アバタの動きを自然に感じた．一方，ノイズが多い場合には，実写アバタの動きを自然であるとは感じなかった．今後の課題としては，ノイズを発生させずに実写アバタの体の揺れを強調する方法を開発することが挙げられる．

第5章　まとめ

5.1　概要

本研究では，超少子高齢化社会による労働力不足を解決するために，今まで人が行っていた公共施設における案内の代替となるような実写アバタを用いた案内システムを提案した．案内システムに実写アバタを用いて実際の人同士の会話のように操作できることで，高齢者などの情報リテラシが低い人々でも簡単に案内のサービスを受けることができる．提案する案内システムの案内の流れは，ユーザが案内を求めて等身大の大型ディスプレイに表示した実写アバタに近づき，実写アバタから案内を受ける．その後，携帯デバイスなどの小型ディスプレイに表示した実写アバタの案内に切り替えることで，目的地までの道中も案内し続けることができユーザは迷わずに目的地に到着することができる．本研究では，このような実写アバタを用いた案内システムの中で，実写アバタのインタラクション開始前動作に注目した．そして，提案した案内システムにおける案内の流れから，実写アバタのインタラクション開始前動作が必要となる 3 つの状況を上げ，各状況においてインタラクション開始前動作が実装されていない時の課題解決に取り組んだ．

5.2　インタラクション開始前動作が必要となる状況の取り組み

5.2.1　身体動揺の計測による待機状態の実写アバタ生成

本研究では，大型ディスプレイ上における待機状態の実写アバタがユーザの訪れを直立姿勢で待っている状況を状況 1 として取り扱った．状況 1 において，待ち

姿勢の静止画を表示し続けた場合や，人間の動きとは違う不自然な映像を表示した場合，ユーザはシステムが対話を待っている状態とは判断できず，システムが異常動作していると判断する恐れがある．そこで，待機状態の実写アバタで人間に近い動きを再現するために，身体動揺を計測し再現する手法を提案した．体の部位毎に振動量の時間変化を計測することで，身体動揺の中心となる参照時刻は映像中で複数回出現し，参照時刻の間には身体の揺れが存在することを確認した．これらの特徴を利用し参照時刻をランダムに遷移することで，任意の時間長の身体動揺を再現する手法について述べた．主観評価により，提案手法は比較手法と比べて人間の身体動揺に近いことを確認した．

5.2.2　実写アバタ映像における動き表現を用いた対象者の指定

本論文では，大型ディスプレイ上における実写アバタの周りに複数ユーザが案内を受けるのを待っている状況を状況 2 として取り扱った．状況 2 において，何の動きもなく実写アバタがユーザに対してインタラクションを開始した場合，ユーザは複数ユーザの中で誰を指定しているのかが分からず，実写アバタとインタラクションを行って良いのか混乱する可能性がある．そこで，実写アバタに次の案内を受ける対象者の方を向く動きを加える手法を提案した．そして，実写アバタ映像に動きを加える場合と加えない場合を比較することで，対象者と非対象者が実写アバタから向かれたと感じるか，また，指定されたと感じるかについて調査した．実写アバタ映像に含める動きとして，動作効果の有無，回転効果の有無の組み合わせを比較した．動作効果の角度パラメータと実写アバタの動き，回転効果の角度パラメータは実験にて有効なパラメータを明らかにした．実写アバタが対象者のみを指定できるかどうかについて，対象者の仮説 H1 と非対象者の仮説 H2 を主観評価で調査した．調査した結果，非対象者の仮説 H2 は成立するとは言えなかったが，対象者の仮説 H1 は成立することを確認した．また，対象者と非対象者の立ち位置を変化させて評価し，動きを加える映像表現の限界を明らかにした．

5.2.3　小型ディスプレイにおける実写アバタの動き強調の検討

本研究では，小型ディスプレイ上における待機状態の実写アバタがユーザとインタラクションを開始することを待っている状況を状況 3 として取り扱った．状況 3 において，大型ディスプレイ上と同様の身体動揺の動きを再現した場合，小型ディスプレイ上では画面サイズが小さく実写アバタの動きが小さくなってしまい，ユーザが実写アバタの動きを視認できずシステムが止まっているように判断する可能性がある．そこで，直立姿勢の実写アバタの身体動揺の揺れを強調することで，小型ディスプレイ上でより自然に感じられるようにすることを狙った．主観評価実験では，実写アバタの体の揺れを強調しノイズが少なかった場合，実験協力者は小型ディスプレイ上で実写アバタの動きを自然に感じた．一方，ノイズが多い場合には，実写アバタの動きを自然であるとは感じなかった．

5.3　今後の展望

今後の展望としては，実写アバタのインタラクション開始前動作以外の画像生成について検討を行う．また，実写アバタを用いた案内システムの構築に必要な技術として，画像生成以外のユーザの状況を理解する技術，ユーザの音声を理解する技術，アバタの音声を生成する技術などの検討を行い，実写アバタを用いた案内システムの構築を目指す．最終的には，実写アバタを用いた案内システムを空港などの公共施設に設置して評価し，超少子高齢化社会による労働力不足の解決を目指す．

謝辞

本研究は，鳥取大学大学院工学研究科情報エレクトロニクス専攻メディア理解研究室で行われたものです．本研究を進めるにあたり，ご指導，ご鞭撻を受け賜りました岩井儀雄教授，西山正志准教授，吉村宏紀助教に深く感謝致します．特に岩井儀雄教授と西山正志准教授には，ご多忙のところ個別で研究の打合せを実施して頂きましたこと心より感謝致します．打合せでは親身になって助言をして下さり，多くのことを学ぶことができました．また，論文投稿や学会参加の際には，論文の査読や発表資料の確認など多くの時間を割いて頂き誠にありがとうございます．論文投稿や学会参加など多くの機会を頂き，人生において貴重な経験を得ることができたと思っております．鳥取大学へ在籍していた長い間，非常にお世話になったこと感謝致します．メディア理解研究室に在籍の我那覇航氏には，本研究の実験を実施するにあたり，ご協力頂きましたこと深く感謝致します．また，実験の被験者として研究に協力してくださったメディア理解研究室の皆様には心から感謝致します．大日本印刷株式会社アドバンストデバイス研究開発本部の皆様には，業務を行いながら研究を進めるにあたり，業務スケジュールの調整や学会参加の支援などをして頂き深く感謝致します．最後になりましたが，本研究および学生生活でご協力，ご指導頂きました皆様へ厚く御礼申し上げます．

参考文献

[1] 日本経済 2016-2017 -好循環の拡大に向けた展望-. 内閣府政策統括官, 2017.

[2] H. H. Clark and T. B. Carlson. Hearers and speech acts. Language, Vol. 58, No. 2, pp. 332-373, 1982.

[3] 塩見昌裕, 神田崇行, Dylan F. Glas, 佐竹聡, 石黒浩, 萩田紀博. 複数の案内ロボットが連携してサービス提供するネットワークロボットシステムの実現. 日本ロボット学会誌, Vol. 29, No. 6, pp. 544-553, 2011.

[4] 塩見昌裕, 神田崇行, 石黒浩, 萩田紀博. 人々の興味を引きつける案内ロボット-後ろ向きに移動する案内の効果. 情報処理学会論文誌, Vol. 51, No. 2, pp. 301-313, 2010.

[5] 宮下善太, 神田崇行, 塩見昌裕, 石黒浩, 萩田紀博. 顧客と顔見知りになるショッピングモール案内ロボット. 日本ロボット学会誌, Vol. 26, No. 7, pp. 821-832, 2008.

[6] 井上昂治, ララディベッシュ, 山本賢太, 中村静, 高梨克也, 河原達也. 自律型アンドロイド erica による傾聴対話システムの評価. 言語・音声理解と対話処理研究会, Vol. 87, pp. 19-24, 2019.

[7] 渡辺美紀, 小川浩平, 石黒浩. ミナミちゃん : 販売を通じたアンドロイドの実社会への応用と検証. 情報処理学会論文誌, Vol. 57, No. 4, pp. 1251-1261, 2016.

[8] Softbank. Pepper. https://www.softbank.jp/robot/pepper/.

[9] Aruze Gaming Technologies. Arisa. https://www.aruzegamingtech.com/pickup/arisa.html.

[10] 大浦圭一郎, 山本大輔, 内匠逸, 李晃伸, 徳田恵一. キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム. 人工知能学会誌, Vol. 28, No. 1, pp. 60-67, 2013.

[11] 高林範子, 小野光貴, 渡辺富夫, 石井裕. 看護実習生-患者役アバタを介した看護コミュニケーション教育システム. 人間工学, Vol. 50, No. 2, pp. 84-91, 2014.

[12] S. Robinson, D. Traum, M. Ittycheriah, and J. Henderer. What would you ask a conversational agent? observations of human-agent dialogues in a museum setting. In Proceedings of LREC, 2008.

[13] R. Artstein, D. Traum, O. Alexander, A. Leuski, A. Jones, K. Georgila, P. Debevec, W. Swartout, H. Maio, and S. Smith. Time-offset interaction with a holocaust survivor. In Proceedings of the 19th International Conference on Intelligent User Interfaces, pp. 163-168, 2014.

[14] A. Jones, J. Unger, K. Nagano, J. Busch, X. Yu, H. Y. Peng, O. Alexander, M. Bolas, and P. Debevec. An automultiscopic projector array for interactive digital humans. In Proceedings of ACM SIGGRAPH 2015 Emerging Technologies, No. 6, p. 1, 2015.

[15] 原健太, 堀磨伊也, 武村紀子, 岩井儀雄, 佐藤宏介. 実画像アバタを用いた対人インタラクションシステムの構築. 電気学会論文誌 C, Vol. 134, No. 1, pp. 102-111, 2014.

[16] M. Inoue, A. Shiraiwa, H. Yoshimura, M. Nishiyama, and Y. Iwai. Evaluating effects of hand pointing by an image-based avatar of a navigation system. In 20th International Conference on Human-Computer Interaction (HCII), pp. 370-380, 2018.

[17] J. Gratch, J. Rickel, E. Andre, J. Cassell, E. Petajan, and N. Badler. Creating interactive virtual humans: some assembly required. IEEE Intelligent Systems, Vol. 17, No. 4, pp. 54-63, 2002.

[18] W. Mattheyses and W. Verhelst. Audiovisual speech synthesis: An overview of the state-of-the-art. Speech Communication, Vol. 66, pp. 182-217, 2015.

[19] R. Anderson, B. Stenger, V. Wan, and R. Cipolla. Expressive visual text-to-speech using active appearance models. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3382-3389, 2013.

[20] D. Okwechime, E. J. Ong, A. Gilbert, and R. Bowden. Social interactive human video synthesis. In Proceedings of the ACCV 2010, Vol. 6492. pp. 256-270, 2010.

[21] Q. Shi, S. Nobuhara, and T. Matsuyama. Augmented motion history volume for spatiotemporal editing of 3-d video in multiparty interaction scenes. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 25, No. 1, pp. 63-76, 2015.

[22] P. Huang, M. Tejera, J. Collomosse, and A. Hilton. Hybrid skeletal-surface motion graphs for character animation from 4d performance capture. ACM Transactions on Graphics, Vol. 34, No. 17, pp. 1-14, 2015.

[23] 産業技術総合研究所人間福祉医工学研究部門. 人間計測ハンドブック. 朝倉書店, 2003.

[24] 森井精啓, 岸野文郎, 鉄谷信二. 眼の CG アニメーションと視線の知覚に関する検討. 電子情報通信学会論文誌 A, Vol. J78-A, No. 4, pp. 512-522, 1995.

[25] 吉田直人, 米澤朋子. ユーザ視点位置に応じた描画エージェントを用いた実空間内注視コミュニケーションの検証. 電子情報通信学会論文誌 D, Vol. J99-D, No. 9, pp. 915-925, 2016.

[26] G. Caridakis, A. Raouzaiou, E. Bevacqua, M. Mancini, K. Karpouzis, L. Malatesta, and C. Pelachaud. Virtual agent multimodal mimicry of humans. Language Resources and Evaluation, Vol. 41, No. 3, pp. 367-388, 2007.

[27] 石井亮, 中野有紀子. ユーザの注視行動に基づく会話参加態度の推定-会話エージェントにおける適応的会話制御に向けて. 情報処理学会論文誌, Vol. 49, No. 12, pp. 3835-3846, 2008.

[28] A. Egges, T. Molet, and N. Magnenat-Thalmann. Personalised real-time idle motion synthesis. Proceedings of 12th Pacific Conference on Computer Graphics and Applications, pp. 121-130, 2004.

[29] 山本昌彦, 吉田友英. 重心動揺計を用いた体平衡機能検査: 重心動揺検査・電気性身体動揺検査. Equilibrium research, Vol. 70, No. 3, pp. 135-144, 2011.

[30] 竹内弥彦, 下村義弘, 岩永光一, 勝浦哲夫. 小型三軸加速度計による高齢者の動的バランス評価の有用性. 理学療法科学, Vol. 22, No. 4, pp. 461-465, 2007.

[31] 大西智也, 橘浩久, 武田功. 安静な立位における足位の違いが下肢帯および下肢の動揺に及ぼす影響. 理学療法科学, Vol. 30, No. 2, pp. 313-316, 2015.

[32] F. Wang, M. Skubic, C. Abbott, and J. M. Keller. Body sway measurement for fall risk assessment using inexpensive webcams. In Proceedings of International Conference of Engineering in Medicine and Biology Society, pp. 2225-2229, 2010.

[33] L. F. Yeung, K. C. Cheng, C. H. Fong, W. C. C. Lee, and K. Y Tong. Evaluation of the microsoft kinect as a clinical assessment tool of body sway. Gait & Posture, Vol. 40, No. 4, pp. 532-538, 2014.

[34] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. IEEE international Journal of Computer Vision, Vol. 91, No. 2, pp. 200-215, 2011.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600-612, 2004.

[36] L. L. Thurstone. A law of comparative judgment. Psychological Review, Vol. 34, No. 4, pp. 273-286, 1927.

[37] 坂口竜己, 山田寛, 森島繁生. 顔画像を基にした 3 次元感情モデルの構築とその評価. 電子情報通信学会論文誌 A, Vol. J80-A, No. 8, pp. 1279-1284, 1997.

[38] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 521-528, 2003.

[39] M. Otsuki, T. Kawano, K. Maruyama, H. Kuzuoka, and Y. Suzuki. Thirdeye: Simple add-on display to represent remote participant ' s gaze direction in video communication. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5307-5312, 2017.

[40] N. Yankelovich, N. Simpson, J. Kaplan, and J. Provino. Porta-person: Telepresence for the connected conference room. In Proceedings of CHI '07 Extended Abstracts on Human Factors in Computing Systems, pp. 2789-2794, 2007.

[41] I. Kawaguchi, H. Kuzuoka, and Y. Suzuki. Study on gaze direction perception of face image displayed on rotatable flat display. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1729-1737, 2015.

[42] Y. Onishi, K. Tanaka, and H. Nakanishi. Embodiment of video-mediated communication enhances social telepresence. In Proceedings of the Fourth International Conference on Human Agent Interaction, pp. 171-178, 2016.

[43] S. O. Adalgeirsson and C. Breazeal. Mebot: Arobotic platform for socially embodied presence. In Proceedings of 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 15-22, 2010.

[44] K. Masame. Perception of where a person is looking: Overestimation and understimation of gaze direction. Tohoku Psychologica Folia, Vol. 49, pp. 33-41, 1990.

[45] A. Kendon. Some functions of gaze-direction in social interaction. Acta Psychologica, Vol. 26, pp. 22-63, 1967.

[46] 宮内翼, 吉村宏紀, 西山正志, 岩井儀雄. 身体動揺の計測による待ち状態の実写アバタ生成. 電子情報通信学会論文誌 D, Vol. 100-D, No. 3, pp. 365-375, 2017.

[47] M. Nishiyama, T. Miyauchi, H. Yoshimura, and Y. Iwai. Synthesizing realistic image-based avatars by body sway analysis. In Proceedings of the Fourth International Conference on Human Agent Interaction, pp. 155-162, 2016.

[48] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Phase-based video motion processing. ACM Transactions on Graphics, Vol. 32, No. 80, pp. 1-10, 2013.

[49] J. W. Kim, G. M. Eom, C. S. Kim, D. H. Kim, J. H. Lee, B. K. Park, and J. Hong. Sex differences in the postural sway characteristics of young and elderly subjects during quiet natural standing. Geriatrics & Gerontology International, Vol. 10, pp. 191-198, 2010.

[50] V. Bruce and A. Young. In the eye of the beholder: The science of face perception. Oxford University Press, 1998.

付 録A　自然なインタラクションのための実写アバタにおける認識状態と反応状態の実装

A.1　付録の研究背景

実写アバタがユーザとのインタラクションを円滑に開始するためには，待機，認識，反応，行動の各状態が必要となる．各状態を図 A.1 に示す．待機状態では，近くにユーザがいないために実写アバタがユーザの接近を待機する．認識状態では，ユーザが接近していることに実写アバタが気づき，いつでもユーザに対応できるように準備する．反応状態では，実写アバタがユーザとのインタラクションを開始する準備ができていることを示すために反応を返す．行動状態では，実写アバタがユーザとインタラクションを行う．特に，実写アバタが待機状態から行動状態へと遷移してインタラクションを円滑に開始するためには，認識状態と反応状態が重要である．反応状態が実写アバタに埋め込まれていない場合，ユーザが近づいても実写アバタが反応しないため，ユーザがインタラクションを開始するタイミングを判断することが難しい．さらに，実写アバタに認識状態が埋め込まれていない場合，実写アバタが反応するタイミングが遅れてしまうため，近づいてきたユーザが混乱してしまう可能性がある．既存の実写アバタでは，実写アバタの行動状態 [13, 14] や待機状態 [47] に注目している．しかし，これらの既存の実写アバタでは，実写アバタの認識状態と反応状態について十分に考慮されていなかった．

本論文では，ユーザとのインタラクションを円滑に開始するために，実写アバタに認識状態と反応状態を実装することを考える．提案手法では，インフォメーションセンターにおける案内者とユーザとの会話の開始を観察し，観察結果に基

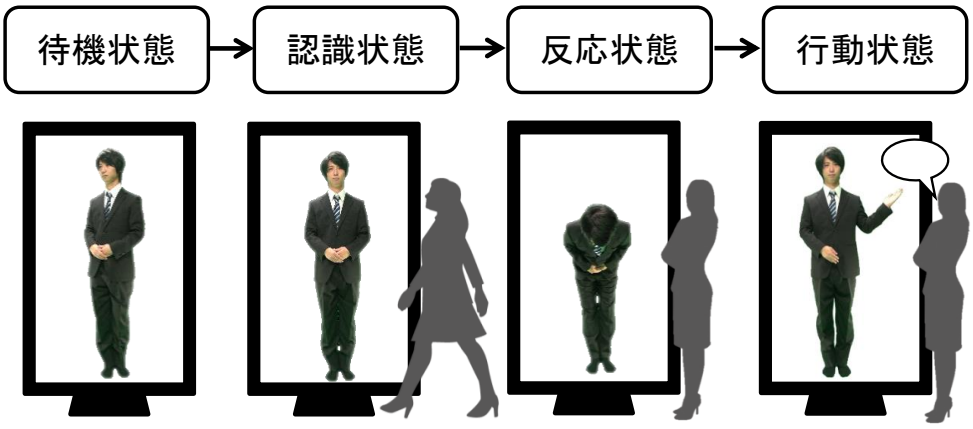


図 A.1: 実写アバタが円滑にインタラクションを開始するために必要な 4 つの状態.

づいて案内者の行動モデルを設計する．そして，行動モデルを実写アバタに実装することで，インタラクションを開始する際の案内者の行動を再現する．主観的な評価を行った結果，認識状態と反応状態を実装した実写アバタは，認識状態と反応状態を実装していない実写アバタよりも良い評価が得られたことを確認した.

A.2　インタラクション開始のための行動モデル設計

A.2.1　案内者の観察

インフォメーションセンターでユーザと案内者の会話を観察した．観察結果では多くの場合，ユーザは案内者にチケットカウンターや搭乗ゲートなどの場所を尋ねていた．案内者はユーザから話しかけられるまで受動的な行動をとっていた．観察後，案内者にインタビューを行い，案内者の行動モデルを設計した.

A.2.2　案内者の行動モデル

図 A.2 はユーザとのインタラクションを開始する際の案内者の行動モデルを示している．S1 では案内者の近くにユーザがいなかったため，案内者はインタラクションの発生を待っている．案内者がインタラクションの発生を待っている時は,

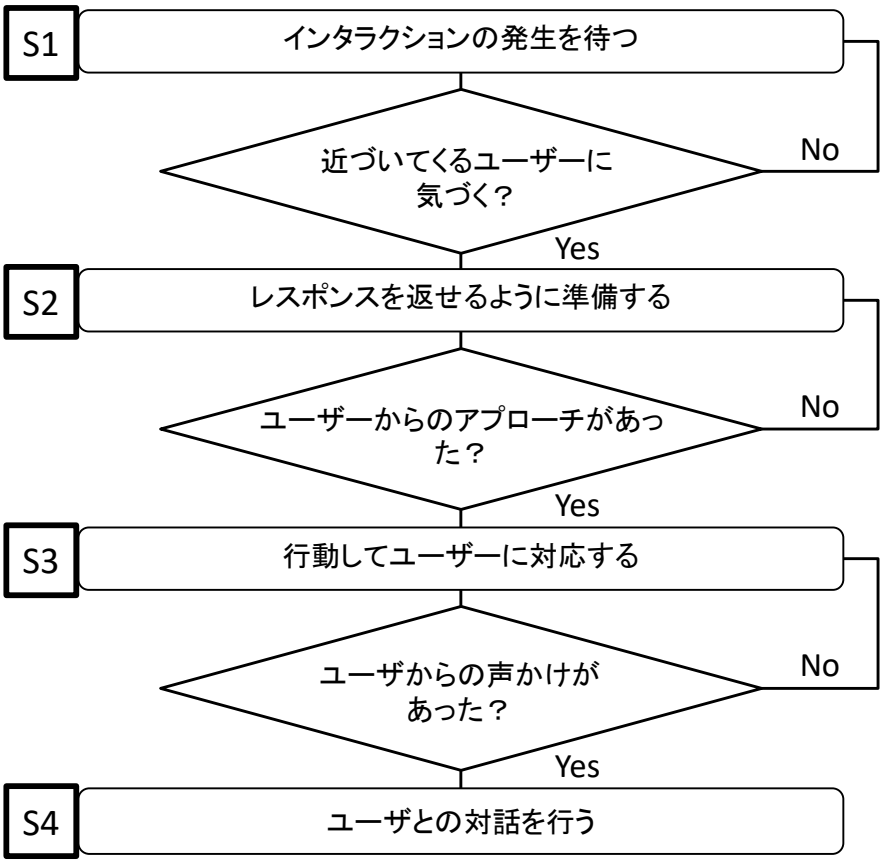


図 A.2: 案内者がユーザとインタラクションを開始する際の行動モデル.

案内者は前を見たり，周りを見たりする行動をとる．S2 では案内者がユーザの接近に気づき，ユーザに反応を返す準備をする．案内者の反応を返す準備としては，案内者がユーザに対して顔を向けるという行動を行う．S3 では案内者がユーザから話しかけられやすいように，案内者がユーザに反応する．案内者の反応としては，案内者がユーザの方を見たり，お辞儀をしたりする行動をとる．S4 では案内者がユーザとインタラクションを行う．具体的には，実写アバタがユーザに道案内やチケットの案内などのインタラクションを行う．案内者が上記のような行動をとる理由を聞いたところ，案内者は公共施設において遠距離から会話するのは難しく，ユーザが近づくまで待ってから会話を始める必要があるためであった．本研究の行動モデルでは，S1 から S2，S2 から S3 への遷移はユーザと案内者の距離で判断する．また，案内者とユーザのインタラクションは，ユーザが「すみません」と言ってから開始した．本研究の行動モデルでは，ユーザが案内者に向かって発話した内容をもとに，S3 から S4 への遷移を判断する．

A.2.3　実写アバタへの行動モデル適用

ここでは，実写アバタを使って案内者の行動モデルを再現する方法を設計する．図 A.2 に示した案内者の行動モデルを実写アバタの各状態に当てはめる．S1 を待機状態，S2 を認識状態，S3 を反応状態，S4 を行動状態とする．実写アバタは，S1 では「正面を向く」「周囲を見る」という行動をとり，S2 では「正面を向く」という行動をとり，S3 では「正面を向く」「お辞儀する」という行動をとる．S4 では「話す・聞く」という行動をとる．実写アバタの各状態の遷移方法を設定する．ユーザから実写アバタまでの距離 d_t が閾値 D_1 以下の場合，S1 から S2 へと状態が遷移する．d_t が D_2 以下の場合，S2 から S3 へと状態が遷移する．ユーザの立っている方向から取得した音の大きさ a_t の短期的な平均値が閾値 A 以上の場合，S3 から S4 へと状態が遷移する．閾値 D_1，D_2，A は，実験によってあらかじめ測定する．また，距離センサとマイクを使用してリアルタイムに d_t と a_t を測定する．

A.3　実写アバタへの行動モデル再現

A.3.1　実写アバタの行動の映像

実写アバタが行動している映像は，実際の人物を撮影した映像で表現される．実写アバタに行動モデルを再現する際に，行動モデルの一連の流れにおける案内者の行動の全ての組み合わせを事前に撮影するのは手間がかかる．そこで，提案手法では行動モデルの S1，S2，S3 において，単体の行動を撮影した映像を組み合わせて一つの映像とすることを考える．また，S4 の行動については，本研究ではインタラクションの開始に注目しているため考えないものとする．提案手法では，以下の映像を使用する．

- **C1**: 初期姿勢で正面を見続ける．
- **C2**: 初期姿勢から左を向いて初期姿勢に戻る．
- **C3**: 初期姿勢から右を向いて初期姿勢に戻る．
- **C4**: 初期姿勢からお辞儀をして初期姿勢に戻る．

正面方向の顔画像を平面ディスプレイに表示する場合，画像中の被写体が常に自分の方を見ているような感覚をユーザに与えるモナリザ効果 [50] が生じる．そこで，S2 や S3 でユーザを見るという行動を表現するために，正面を見る行動に C1 を使う．C2 と C3 は，S1 で周囲を見る行動に使う．C2 または C3 は，S2 でユーザの方に顔を向ける動作にも使う．C4 は S3 でユーザにお辞儀をする行動に使う．提案手法では，映像を滑らかに切り替えて結合し，映像の再生速度をユーザの動きに合わせて制御する．

A.3.2　映像の結合と制御

ここでは，実写アバタの映像を結合する方法について述べる．映像を単純に連結して結合する場合，各映像の初期姿勢には僅かに違いがあるため不連続な映像

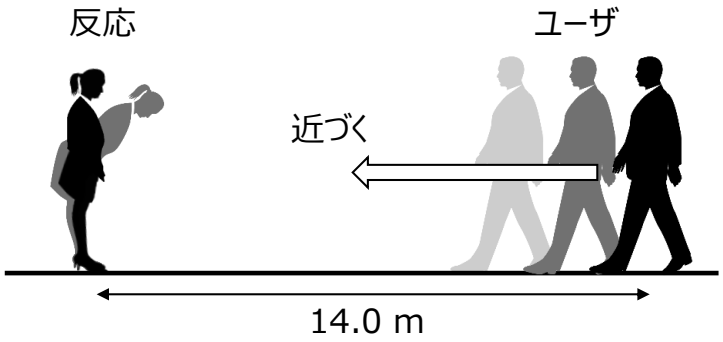


図 A.3: パラメータを計測するための実験環境.

になってしまう．そこで提案手法では，オプティカルフローとノイズ除去を用いて初期姿勢の間の補間を行う．

次に再生速度を制御する方法について述べる．S1 の「周りを見る」という行動を表現するために，提案手法では時間 T_1 の周期内で C2 と C3 を繰り返す．S2 でユーザに顔を向ける行動を表現するために，提案手法では現在再生されている映像を早送りまたは巻き戻しする．この操作は，$T_2 = (D_1 - D_2)/v$(v はユーザの歩行速度) 以内の時間で終了するように制御する．深度センサでリアルタイムに計測された v を用いて，毎回自動的に時間 T_2 が計算される．また，S3 でユーザにお辞儀をする動作を表現するために，時間 T_3 内で C4 を再生する．その他の行動を表現するためには C1 の映像を連続再生する．時間 T_1，T_3 は案内者の行動の時間の長さを観察することで決定する．

A.4　実写アバタへの行動モデル再現の評価実験

A.4.1　閾値決定のためのパラメータ計測

提案手法の閾値を決定するために，インタラクションの開始時における行動のパラメータを計測した．2 人の案内者（平均年齢 23.5 ± 0.5，男性 1 人，女性 1 人）と 5 人のユーザ（平均年齢 23.5 ± 0.5，男性 4 人，女性 1 人）で計測実験を行った．案内者がユーザを案内することを前提として，ユーザに空港の搭乗ゲートの場所を案内者に尋ねる状況であることを伝えた．ユーザは案内者の正面から近づいて

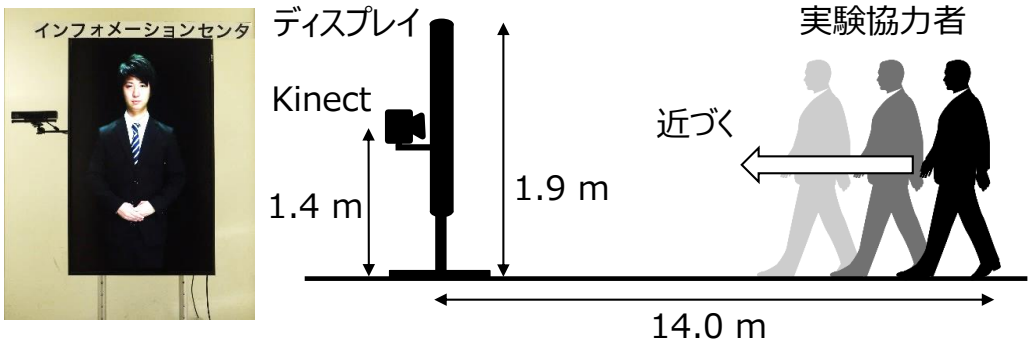


図 A.4: 実写アバタを用いた主観評価の実験環境.

くるものと仮定した．ユーザが案内者に近づく状況を図 A.3 に示す．ユーザは案内者から 14m 離れた位置から歩き始め，自分が話しやすいと思う範囲に近づき，案内者に話しかけた．パラメータの計測はユーザごとに 2 回行った．計測結果から，それぞれのパラメータは D_1 = 5.6 ± 1.6m．D_2=3.9 ± 0.4m，A=20dB in 0.08s，T_1=2.0 ± 0.3s であった． なお，ユーザの平均歩行速度は v=1.4 ± 0.1m/s であった．これらのパラメータの平均値を提案手法の閾値に用いた．

A.4.2 主観評価の実験環境

インタラクション開始の主観的な評価を行うためにパラメータ計測実験に参加していない 14 人の実験協力者（平均年齢 23.0 ± 1.3，男性 12 人，女性 2 人）を集めて，インタラクション開始に関する主観的な評価を行った．実験協力者には A.4.1 の実験と同様の状況であることを伝えた．図 A.4 は実写アバタを使って主観的な評価を行うための実験環境を示している．実写アバタの映像は縦に配置された 80 インチのディスプレイ（Sharp PN-A601）で再生した．遠くからでも実写アバタを見つけやすいように，ディスプレイの上部に「インフォメーションセンター」という看板をつけた．実写アバタまでの距離と実験協力者の声の大きさを測るために，深度センサとマイク（Microsoft Kinect v2）を使った．各実験協力者は，実写アバタから 14m 離れた位置から歩き始め，実写アバタに近づき話しかけた．各実験協力者には男性と女性の実写アバタを使用して実験を行った．

以下の実写アバタを比較するために，主観評価実験を行った．

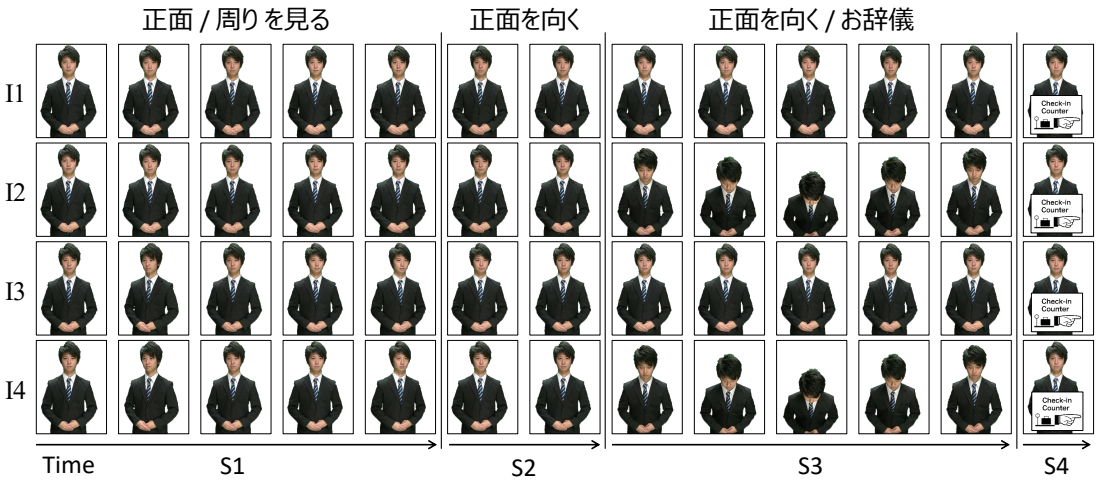


図 A.5: 主観評価における実写アバタ映像.

- **I1**: 常に正面を向いている.
- **I2**: ユーザが実写アバタに近づくと正面を向いてお辞儀をする.
- **I3**: 周囲を見て，ユーザが近づくと顔をユーザの方に向ける.
- **I4**: 周囲を見て，ユーザが近づくと顔をユーザの方に向け，さらにユーザが実写アバタに近づいた時にお辞儀をする.

I1 は認識状態や反応状態が実装されていないベースライン手法である．I2，I3，I4 は両方の状態を組み込んだ方法であるが，それぞれの状態での行動は異なる．各実験協力者は，それぞれの実写アバタに近づき話しかける．図 A.5 は I1 から I4 の映像を示したものである．映像は C1，C2，C3，C4 の動作を使用した．C1 を I1 の S1 から S3，I2 の S1 と S2，I3 の S3 で使用した．C2 と C3 を I3 と I4 の S1 と S2 で使用した．C4 を I2 と I4 の S3 で使用した．なお，全ての映像における S4 には，単純にメッセージを表示した．インタラクションの開始が完了した後に，実験協力者に以下のような質問をした.

- **Q1**: 実写アバタにインタラクションする時にタイミングは分かりやすかったか？
- **Q2**: 実写アバタは実際の案内者の行動に似ていたか？

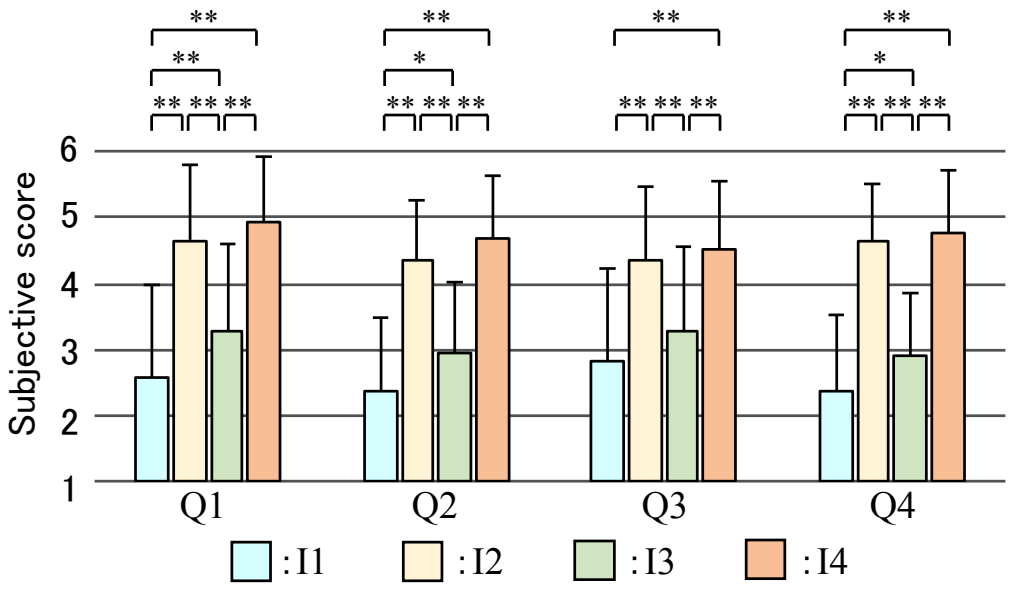


図 A.6: 実写アバタを用いた主観評価の結果（** : p ＜ 0.01，* : p ＜ 0.05）.

- **Q3**: 実写アバタとインタラクションを円滑に開始することができたか？
- **Q4**: 実写アバタとインタラクションを開始する時に丁寧な振る舞いをしていると感じたか？

各実験協力者は各質問に対して 5 段階（1 : そう思った，5 : そう思わない）で評価を行った．また，Q1 から Q4 の逆質問も行った．実写アバタの映像を再生する順番は，各実験協力者毎にランダムに選択した．

A.4.3　主観評価結果

図 A.6 は各質問に対する点数の平均値と標準偏差を示したものである．逆質問の点数は反転させて，Q1 から Q4 の対応する点数に加えた．本実験では，各質問の点数の正規性を確認した後に Friedman 検定，ウィルコクソンの符号付順位検定，Bonferroni 補正を行った.「前を見る」行動（I1）と「周りを見る」行動（I3）を比較すると，Q2 と Q4 ではわずかに有意差（p¡0.05）が出たが，Q3 では差が出なかった．S1 では実写アバタがこれらの行動を行うことはあまり効果的ではないと考えられる.「ユーザにお辞儀をする」（I2，I4）と「ユーザを見る」（I1，I3）を比較す

ると，全ての質問で有意差（$p < 0.01$）が出た．したがって，S3 で「ユーザにお辞儀をする」ことは効果的であることが分かった．このことから，認識状態 (S2) と反応状態 (S3) を実写アバタに実装することは重要であると考えられる．

A.5　付録まとめ

実写アバタとユーザとのインタラクションを円滑に開始するために，実写アバタに認識状態と反応状態を実装する手法を提案した．インフォメーションセンターでの案内者を観察することで，案内者の行動モデルを設計した． 提案手法では，案内者の行動モデルを実写アバタに再現するために，行動モデルに基づいた行動の映像を結合して制御した．実験結果は，認識状態と反応状態を実装した実写アバタがユーザとインタラクションを開始する上で効果的であることを確認した．今後は，インタラクションシステムにおける行動状態の行動について注目して開発を行う．

研究業績

論文誌

1. 宮内 翼, 吉村 宏紀, 西山 正志, 岩井 儀雄, 身体動揺の計測による待ち状態の実写アバタ生成, 電子情報通信学会論文誌 D, Vol. J100-D, No. 3, pp. 365-375, March 2017.

2. Tsubasa Miyauchi, Ayato Ono, Hiroki Yoshimura, Masashi Nishiyama, Yoshio Iwai, Embedding the awareness state and response state in an image-based avatar to start natural user interaction, IEICE Transactions on Information and Systems, Vol. E100-D, No. 12, pp. 3045-3049, December 2017.

3. 宮内 翼, 西山 正志, 岩井 儀雄, 実写アバタ映像における動き表現を用いた対象者の指定, ヒューマンインタフェース学会論文誌, Vol. 22, No. 2, pp. 77-88, May 2020.

査読付き国際学会

4. Masashi Nishiyama, Tsubasa Miyauchi, Hiroki Yoshimura, and Yoshio Iwai, Synthesizing Realistic Image-based Avatars by Body Sway Analysis, Proceedings of 4th International Conference on Human-Agent Interaction (HAI), pp. 155-162, October 2016.

5. Tsubasa Miyauchi, Masashi Nishiyama, Yoshio Iwai, Directing a Target Person among Multiple Users using the Motion Effects of an Image-based Avatar,

Proceedings of 21st International Conference on Human-Computer Interaction (HCII), vol. 3, pp. 341-352, July 2019.

6. Tsubasa Miyauchi, Wataru Ganaha, Masashi Nishiyama, Yoshio Iwai, Investigation of Motion Video Enhancement for Image-based Avatars on Small Displays, Proceedings of 23rd International Conference on Human-Computer Interaction (HCII), LNCS 12763, pp. 44 - 55, July 2021.

国内学会

7. 宮内 翼, 吉村 宏紀, 西山 正志, 岩井 儀雄, 実写アバタ生成に向けた部位毎の重心動揺の計測, IEEE 広島支部学生シンポジウム (HISS), pp. 464-467, November 2015.

8. 宮内 翼, 吉村 宏紀, 西山 正志, 岩井 儀雄, 自然な実写アバタに向けた身体動揺の計測と再現, HAI シンポジウム, pp. 317-324, December 2015.

9. 大野 礼人, 宮内 翼, 吉村 宏紀, 西山 正志, 岩井 儀雄, インタラクション前の実写アバタの待ち状態の検討, HAI シンポジウム, G-3, pp. 1-7, December 2016.

10. 宮内 翼, 井上 路子, 吉村 宏紀, 西山 正志, 岩井 儀雄, インタラクション開始直前の実写アバタによるユーザ指定方法の検討, HAI シンポジウム, G-20, pp. 1-7, December 2017.