

# インタラクション前の実写アバタの待ち状態の検討

## Considering Wait State before Interaction State with users for Image-based Avatar

大野 礼人<sup>1\*</sup> 宮内 翼<sup>1</sup> 吉村 宏紀<sup>1</sup> 西山 正志<sup>1</sup> 岩井 儀雄<sup>1</sup>  
Ayato Ohno<sup>1</sup> Tsubasa Miyauchi<sup>1</sup> Hiroki Yoshimura<sup>1</sup> Masashi Nishiyama<sup>1</sup> Yoshio Iwai<sup>1</sup>

<sup>1</sup> 鳥取大学大学院 工学研究科  
<sup>1</sup> Graduate School of Engineering, Tottori University

**Abstract:** We consider a system of image-based avatars for naturally communicating with users by smoothly transitioning from the wait state to the action state. To achieve the system, we believe that the avatar attracting the users' interests is a very important factor. In order to evaluate the transition from the wait state to the action state, we conducted two experimental setups by assuming information counters in a public space and applied subjective assessments by asking whether the users felt more easily talked to avatar.

### 1 はじめに

近年、人間と機械とが自然にインタラクションを行うシステムの実現に向けてアバタが開発されており様々な場所で応用が期待されている。開発されたアバタの例として、エントランスで施設の案内をするアバタ [1]、看護実習における患者アバタ [2]、博物館で案内をするアバタ [3]、過去の体験を語るアバタ [4] などが挙げられる。これらのアバタを用いてシステムを構築することで、環境や時間などの制約を受けることなくユーザとインタラクションを行うことができるようになる。本研究ではこのようなシステムの中でも、特に大型ディスプレイ上に実際の人物の映像を映す実写アバタ [4, 5, 6] に注目した。実写アバタの例を図1に示す。実写アバタは既存の大型ディスプレイを積極的に活用できるという利点がある。以下では実写アバタを用いたシステムのうち、空港や駅、大型施設のインフォメーションセンタに設置されるものを想定し、具体的なインタラクションとしてユーザへの施設案内やパンフレット配布などを行うことを想定する。

このようなインタラクションシステムにおいて実写アバタはユーザと対話などを行う「行動状態」と、行動の前後に起こる「待ち状態」を相互に切り替えることで動作している。行動状態に関しては今日に至るまで様々な研究 [5, 6] がなされてきた。また近年はアバタとユーザが対話を開始する前の待ち状態に注目した研究 [4, 7] も行われている。



図 1: 実写アバタの例

しかし従来手法では、それぞれの状態に的を絞ったものが多く、待ち状態から行動状態への遷移中については注目されていない。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場合、対話前の待ち状態から行動状態への遷移が適切に実写アバタに組み込まれていないとユーザが実写アバタへ接近する際、声をかけるまで実写アバタの映像が大きく変化しないため、ユーザが接近することへの反応がなく、興味を持たれていないと感じる恐れがある。また状態の遷移において人間の動きと異なるような映像を表示した場合、ユーザはシステムが対話可能な状態だと判断できず話しかけるタイミングを掴むことができず、インタラクションが円滑に開始されないという問題が発生する。

そこで本研究では実写アバタを用いたシステムにおいて、人間の動きに近く、接近するユーザへ興味を持っていることを伝えられるような待ち状態から行動状態への遷移の実現を目的とし、遷移時のアバタの映像について検討・評価を行い、実写アバタを生成する方法を提案する。具体的な手法としてはじめに空港への案内係への取材を行いその結果より待ち状態と行動状態の間に新たに状態を定義し、各状態における実写アバタ映像

\*連絡先：鳥取大学大学院 工学研究科  
〒 680-8550 鳥取県鳥取市湖山町南 4 丁目 101  
E-mail: nishiyama@eecs.tottori-u.ac.jp  
iwai@ike.tottori-u.ac.jp



図 2: ユーザーの行動とそれに応じた案内係の反応

について検討を行った。次に各状態における実写アバタ映像を生成するシステムを構築した。最後に生成した実写アバタ映像が人間に近い動きを再現できているか主観評価実験で確認した。

以下では、2章で新しく定義した状態について紹介し、その状態のアバタ映像と状態遷移条件について検討を行い、3章で実写アバタ映像の生成方法とそのシステムについて述べ、4章で生成した映像の主観評価実験を行った。最後に5章でまとめる。

## 2 実写アバタの状態遷移の検討

### 2.1 実在の案内係の観察

本研究では、実写アバタを用いたシステムにおいてより人間の動きに近い待ち状態から行動状態への遷移を実現するために、実在の案内係（以下、案内係とよぶ）の行動・反応を実写アバタに実装することが好ましいと考えた。そこで案内係の行動・反応を確認するため、空港の案内係へ取材を行った結果、ユーザーと案内係の行動は図2に示す流れで行われていることが判明した。ただし、ユーザーがアバタに興味がない場合は対象外とした。この観察結果より、ユーザーの行動に対する案内係の行動・反応は段階ごとに変化しており、既存手法で取り扱っている「待ち状態」「行動状態」のみではユーザーの行動に応じた行動・反応が行えないことが判明した。そのため、本研究では案内係の行動・反応を忠実に実写アバタに反映するために「待ち状態」と「行動状態」の間に新しく状態の定義を行い、その状態への遷移条件についても定義を行う。

### 2.2 状態の定義

図3に示すように既存手法の「状態 S1：待ち状態」「状態 S3：行動状態」の間に新しく「状態 S2：準備状

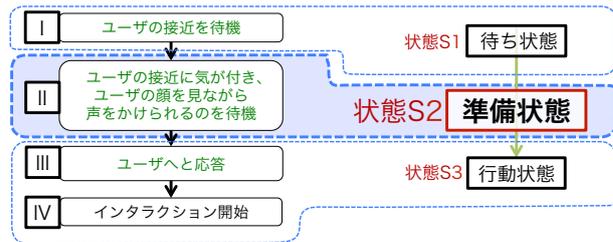


図 3: 待ちから行動への状態遷移

態」を定義した。各状態の詳細について以下に示す。

#### 状態 S1：待ち状態

図2におけるI部分に該当する。この状態では、案内係はユーザーが接近することを待機する。

#### 状態 S2：準備状態

図2におけるII部分に該当し、新しく定義した状態である。この状態では案内係はユーザーの接近に気が付き、現在行っている行動をキャンセルし何らかの反応を行う。それにより案内係がユーザーの接近に気がついていることを明示し、声をかけやすい環境を作る。

#### 状態 S3：行動状態

図2におけるIII部分に該当する。この状態では案内係はユーザーとの会話の内容やユーザーの表情・現在時刻などの周辺の環境や情報などに応じた反応を行う。

本研究では特にS2の「準備状態」に注目し、待ち状態から行動状態へ遷移する際にS2を挿入することでより人間に近い動きの実写アバタ映像の生成を試みる。

### 2.3 各状態における実写アバタ映像の検討

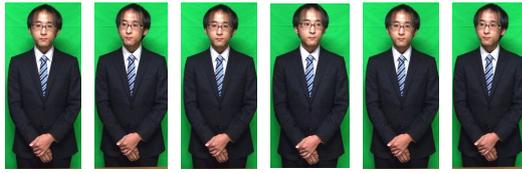
定義した状態を実写アバタを用いたシステムに実装する際に、それぞれの状態に最適と思われる実写アバタの映像について検討を行う。映像の検討は2.1章で観察した案内係の行動を参考に行う。

#### 状態 S1 における実写アバタ映像

案内係は、接客時に接近するユーザーを早期に発見することなどを目的に、直立姿勢にて両手を下に伸ばした状態で体の前で組み、案内係の周囲を確認している（以下、周囲を確認と呼ぶ）。本研究では案内係の行動を模倣し、周囲を確認する映像を実写アバタ映像として採用する。採用した実写アバタ映像の例を図4に示す。

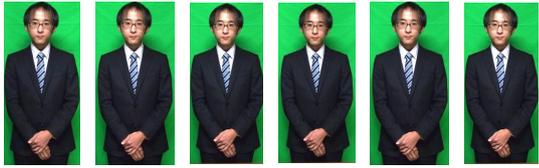
#### 状態 S2 における実写アバタ映像

ユーザーの接近に気がついた案内係は、現在の行動を中止し視線をユーザーの顔に向ける（以下、ユーザーを注視）。案内係が視線をユーザーに向けること



顔の向き

図 4: 状態 S1 における実写アバタ映像の例



顔の向き

図 5: 状態 S2 における実写アバタ映像の例

で、案内係がユーザの接近に気が付いていることをユーザに示すことができる。本研究では S1 と同様に案内係の行動を模倣し、ユーザを注視する映像を実写アバタ映像として採用する。採用した実写アバタ映像の例を図 5 に示す。

#### 状態 S3 における実写アバタ映像

状態 S3 ではユーザの行動や声かけの際の会話内容、案内係の周辺環境によって実写アバタ映像が異なる。本研究では実写アバタ映像の検討は行わない。

## 2.4 状態遷移の条件

#### 状態 S1 から状態 S2 への遷移

エドワード・ホールらの著書によるパーソナルスペースの定義 [8] のうち、特に社会距離に注目し、S1 から S2 へ遷移する条件である距離  $D$  をユーザと実写アバタまでの距離が 10 フィート (3m) 以内であることと定める。

#### 状態 S2 から状態 S3 への遷移

状態 S2 から状態 S3 への遷移はユーザから実写アバタへ何らかの声がかけられたときと定める。

すべての状態遷移条件を網羅したフローチャートを図 6 に示す。

## 3 実写アバタの準備状態の実装

本章では 2 章にて検討した実写アバタ映像を実写アバタを用いたシステムへ実装し準備状態を再現する。な

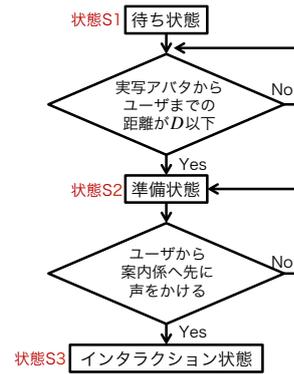


図 6: 状態遷移条件を網羅したフローチャート図

お、本章で実装するアバタ映像の動きはすべて単純な繰り返し映像を使用する。

## 3.1 実写アバタ映像の撮影

#### 撮影環境

インタラクションシステムで使用する実写アバタの映像の撮影を行う。撮影では空港でのインフォメーションセンタを想定し、被撮影者の前にパンフレットなどを設置する机を設置して撮影を行った。実写アバタとしてディスプレイに表示して使用する際には映像の机部分に同じ机を設置することで実写アバタ映像と実在の世界の差による違和感を減少させるねらいがある。

#### 撮影方法

撮影は以下の順番で行った。撮影時の被撮影者の姿勢は両手を下に伸ばし体の前で手を組んだ状態を使用した。撮影には SONY 社製の HDR-CX560 を使用した。

1. 指定された姿勢でカメラのレンズを見た状態で  $N$  秒間待機。
2. 続けて頭部を動かして視線を左方向に移し、その後カメラのレンズを見た状態まで戻ってくる。
3. 続けて頭部を動かして視線を右方向に移し、その後カメラのレンズを見た状態まで戻ってくる。
4. カメラのレンズを見た状態で  $N$  秒待機。

実際に撮影する映像例を図 7 に示す。このような撮影順を使用した理由として、後ほど行う状態遷移時の映像の作成が容易になるという利点がある。

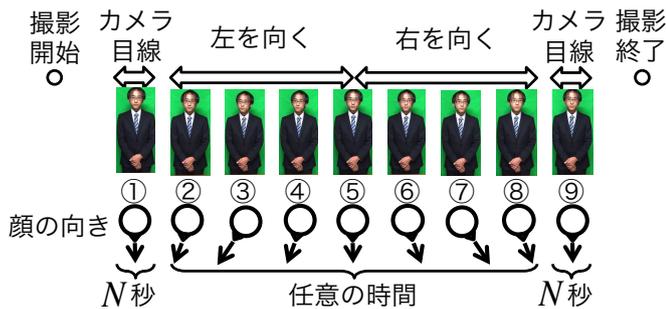


図 7: 実際に撮影する映像例

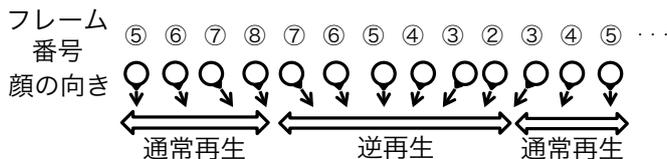


図 8: 状態 S1 で使用する違和感のない繰り返し再生

### 3.2 状態遷移の再現方法

本章では撮影した実写アバタ映像を使用して、任意のタイミングで任意の状態の映像へと遷移する方法について考える。実写アバタを使用したインタラクションシステムではユーザとディスプレイの距離によって動的に状態遷移が発生し、それに伴い表示する実写アバタ映像が変化する。また撮影された動画の時間より長い間状態が遷移しなければ違和感のないように繰り返し再生を行わなければならない。ここでの違和感とは、フレーム飛び、処理の遅延、再生スピードの低下など案内係では起こり得ない状況を指す。各状態間の違和感のない遷移および状態内の繰り返し再生の実現方法に関して以下に示す。

#### 状態 S1

図 7 に示すような撮影データに対し、図 8 のように逆再生と通常再生を組み合わせることで違和感のない繰り返し再生を実現する。

#### 状態 S1 から状態 S2 への遷移

図 7 に示すような撮影データに対し、S1 と同様に逆再生と通常再生を組み合わせることで違和感のない繰り返し再生を実現する。状態 S1 から状態 S2 への遷移のタイミングによって遷移の方法が変化する。状態 S1 において実写アバタが左を向いているときに S2 へ遷移する方法を図 9 に、状態 S1 において実写アバタが右を向いているときに S2 へ遷移する方法を図 10 に示す。

#### 状態 S2

図 7 に示すような撮影データをそのまま使用した。

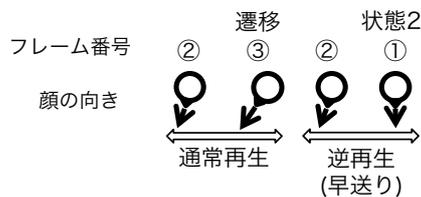


図 9: 状態 S1(右を向いているとき) から状態 S2 への違和感のない状態遷移

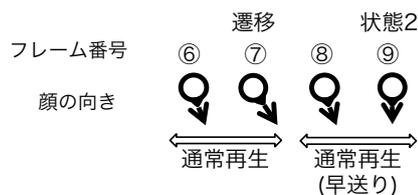


図 10: 状態 S1(左を向いているとき) から状態 S2 への違和感のない状態遷移

## 4 評価実験

準備状態の実装を行ったインタラクションシステムを使用して、状態 S1 と S3 の間に S2 を挿入することで、実写アバタの映像がより人間に近い動きを行えているか主観評価を行った。

### 4.1 実験目的

本実験では以下に挙げる 2 つの仮説の検証を目的とする

**仮説 1:** 状態 S2 を挿入することで従来手法に比べ実写アバタ映像がユーザへの興味を持っていることをより伝えられるようになる

状態 S2 を挿入したことで、ユーザが実写アバタへ接近している状態で、実写アバタが反応することが可能になった。これにより実写アバタがユーザへ興味を持っていることをより伝えられるようになると思われる。

**仮説 2:** 状態 S2 を挿入することでユーザが実写アバタへ何かを尋ねやすくなる

状態 S2 を挿入したことで、ユーザが実写アバタへ何かを尋ねやすくなると思われる。

### 4.2 実験条件

#### 4.2.1 想定環境

実験は空港のインフォメーションセンタに実写アバタを表示するディスプレイが設置された環境を想定し

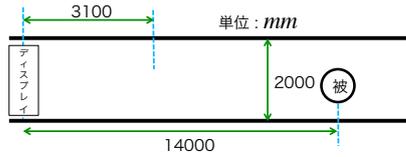


図 11: 想定環境 1 を想定した実験環境

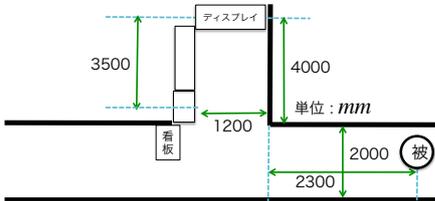


図 12: 想定環境 2 を想定した実験環境

て行った。また、インフォメーションセンタの環境に関して以下の2つを想定し実験を行った。

環境 1: インフォメーションセンタが通路などに設置してあり、ユーザからインフォメーションセンタに設置されたディスプレイが直接目視できる環境

環境 2: インフォメーションセンタが扉などを隔てた先に設置してあり、ユーザがインフォメーションセンタの場所はわかるものの、そこに設置されたディスプレイが直接目視できない環境

環境 1 を想定して構築した実験環境 1 を図 11 に、環境 2 を想定して構築した実験環境 2 を図 12 に示す。図内の数字は各所の寸法である。また丸で囲まれた「被」は被験者を表す。図 12 における「看板」の位置には「インフォメーションセンタ」と書かれた紙が設置されており、被験者にインフォメーションセンタの場所を明示するために設置した。なお図 11 には記載されていないが、こちらの環境にも同様の紙がディスプレイの上に設置されている。

さらに設置されたディスプレイへの接近の方向も以下の2つの場合を想定し実験を行った。

- ユーザがディスプレイ正面から接近
- ユーザがディスプレイに対して斜め 45 度から接近

ユーザのディスプレイへの接近の方向については、設置したディスプレイを通路の進行方向に対して 90 度の向きに設置するか 135 度の向きに設置するかで再現した。角度ごとのディスプレイの設置の様子を図 13 に示す。

実写アバタのディスプレイ表示時の寸法を図 14 に示す。使用したディスプレイは SHARP 社製 PN-A601 である。表示の際は実写アバタの顔の大きさが被撮影者の顔の大きさと同じになるように、肩幅を基準に正規化を行った。

ディスプレイ設置角度: 90度    ディスプレイ設置角度: 135度

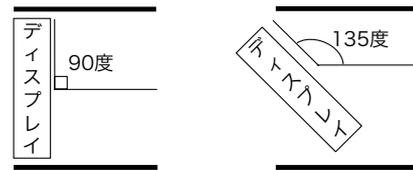


図 13: 角度ごとのディスプレイの設置の様子

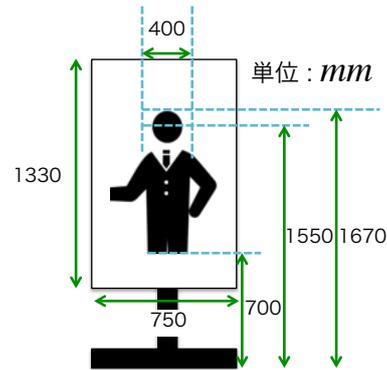


図 14: アバタ映像のディスプレイ表示時の寸法

また、実写アバタを撮影した際の撮影環境を図 15 に示す。撮影は 3 章に記載の方法を使用し、 $N = 30$  で行った。状態 S2, S3 においてはモナリザ視線効果 (Mona Lisa Gaze Effect)[9] を使用することで実写アバタ映像が視線をユーザの顔に向けている状況を再現する。モナリザ視線効果を使用することで、被撮影者がカメラ目線の状態で実写アバタ映像を収録し、それをディスプレイ上に表示するだけで、実写アバタ映像がユーザの方に視線を向けることができる。

#### 4.2.2 比較する手法

2.3 章にて検討を行った各状態の映像を 2.4 章の条件にて順次遷移させたものを提案手法、Nishiyama ら [7] の論文の手法を使用し、ユーザから声がかげられるまで直立姿勢で両手を下に伸ばし体の前で手を組んだ状態で正面を見る映像 (以下、正面を見る) を従来手法として評価を行った。表 1 に手法ごとに各状態でどのような映像を使用したか示す。なお、今回は状態の遷移タイミングを、検討した遷移条件をもとに手動で与えて実験を行った。

#### 4.2.3 実験の流れ

実験は以下の流れで行った。

1. 実験の概要を説明する。
2. 被験者を指定された場所に待機させる。

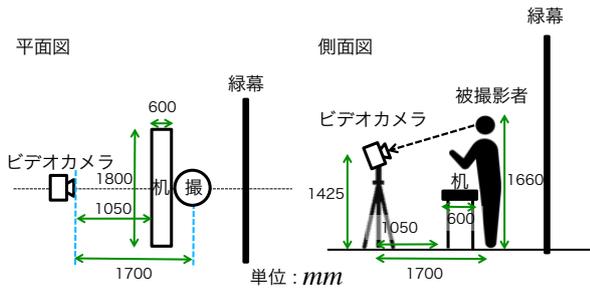


図 15: 実写アバタ映像の撮影環境

表 1: 手法・状態ごとの使用映像

	S1	S1 -> S2	S2	S3
提案手法	周りを確認する	S1の動作を止め顔を正面に向ける	正面を見る	正面を見たま画面に質問の回答文を表示
従来手法	正面を見る			正面を見たま画面に質問の回答文を表示

3. 被験者はディスプレイに接近し、任意の点で停止。表示されている実写アバタに任意の質問を行う。被験者が任意の質問を行い、それに対し実写アバタから何らかの応答があった時点で実験を終了。
4. 提案手法と従来手法を入れ替え再度同様の実験。
5. 再度実験が終了した時点でアンケートを配布し記入してもらう。

この実験が終了したあと、ディスプレイの設置角度を変更し同様の流れを行った。さらに実験環境を変更し同様の実験を行った。

#### 4.2.4 主観評価の方法

提案手法と従来手法の2つの映像を再生順番をランダムにして実験を行い、その後被験者に以下の項目を11段階評価で回答してもらった。

質問1: どちらの実写アバタが自分に興味を持ってくれたように感じましたか。

質問2: どちらの実写アバタが尋ねる際に尋ねやすかったですか。

質問1は仮説1の検証に対応し、質問2は仮説2の検証に対応する。評価は11段階評価のうち中央が手法による差なし、1回目の手法のほうが優位だと感じた場合は右の目盛りに丸を、2回目の手法のほうが優位だと感じた場合は左の目盛りに丸を記入してもらった。なお、中央から丸を記入した目盛りまでの距離が優位度合いを示す。また評価アンケートの他に自由記述欄を設け、被験者がどのように感じたかを任意で記述してもらった。実験は男子大学生・大学院生の8名で行った。平均年齢は22.8(±0.94)歳である。

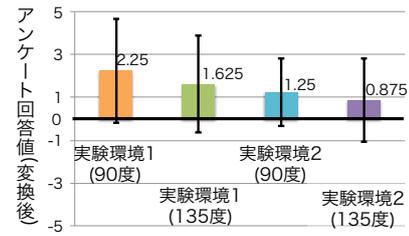


図 16: 被験者質問1に対する実験環境ごとの回答平均値

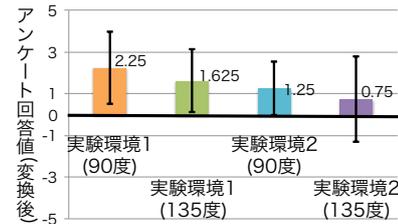


図 17: 被験者質問2に対する実験環境ごとの回答平均値

### 4.3 結果と考察

アンケートより得られた11段階評価を-5(従来手法が最も優位)から5(提案手法が最も優位)の値に変換を行った。その結果を被験者質問1に対する実験環境ごとの回答平均値は図16、被験者質問2に対する実験環境ごとの回答平均値は図17に示す。またアンケート項目の括弧内の角度はディスプレイの設置角度を示す。

各実験環境における回答について  $\mu = 0$  とした際の一方方向t検定の結果を表2に示す。

被験者へ行った質問1、質問2ともに実験環境2でディスプレイ角度を135度に設置した以外の3環境で提案手法の有意差があることが判明した。この結果より有意差が発生した3環境において仮説1および仮説2が正しいことが証明された。実験環境2でディスプレイ角度を135度に設置した環境において有意差が発生しなかった原因として以下が考えられる。

- 実験環境2ではユーザが状態S1を見ている時間が短く、S1からS2への遷移によってアバタ表示が大きく変わったという印象を受けにくいため。
- ディスプレイを135度に設置することで、S1の周りを確認するという動作がユーザから確認しづらくなったため。
- ディスプレイを135度に設置することで、被験者ごとに何か尋ねる際の立ち位置が変化し、受ける印象が個人ごとに大きく異なったため。

また今回の実験ではS2の挿入の有無による影響を評価するため、状態S2からS3への遷移およびS3における

表 2: 各環境における一方向 t 検定結果

	実験環境	環境1	環境1	環境2	環境2
	ディスプレイ角度 [度]	90	135	90	135
1標本t検定 片側	被験者質問1	0.0174	0.04094	0.03021	0.1235
	被験者質問2	0.0042	0.00927	0.01408	0.1679

: p < 0.01

: p < 0.05

実写アバタ映像については評価を行わなかったが、今後はこの部分も含めた評価が必要である。

## 5 まとめ

本研究では実写アバタを用いたインタラクションシステムにおいて、アバタ1体とユーザ1名の環境でインタラクションを行う場合において待ち状態から行動状態への人間に近い遷移の実現を目的とし、遷移時の実写アバタの映像について検討・評価を行い、評価の結果を実写アバタを用いたインタラクションシステムへ実装する方法を提案した。目標の達成のために、初めに空港の案内係への取材を行った。その結果より待ち状態と行動状態の間に新しい状態を定義し、各状態における実写アバタ映像について検討を行った。その結果をもとに各状態における実写アバタ映像を自動で生成されるシステムを作成した。最後に実装したシステムを使用し、人間の動きに近い映像を生成できたか主観評価実験を行った結果、従来手法に比べより人間に近い動きの待ち状態から行動状態への遷移が可能になった。今後の課題としてユーザ行動の自動センシングと、その情報を使用し実写アバタの状態を自動で遷移するシステムの構築が挙げられる。

## 参考文献

- [1] 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.
- [2] 高橋範子, 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.
- [3] George Veletsianos. How do learners respond to pedagogical agents that deliver social-oriented non-task messages? impact on student learning, perceptions, and experiences. *Computers in Human Behavior*, Vol. 28, No. 1, pp. 275 – 283, 2012.
- [4] 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.
- [5] A. Jones, J. Unger, K. Nagano, J. Busch, X. Yu, H. I. Peng, O. Alexander, M. Bolas, and P. Debevec. An automultiscopic projector array for interactive digital humans. In *ACM SIGGRAPH 2015 Emerging Technologies*, p. 6:1, 2015.
- [6] 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. 4, pp. 102–111, 2013.
- [7] M. Nishiyama, T. Miyauchi, H. Yoshimura, and Y. Iwai. Synthesizing realistic image-based avatars by body sway analysis. *Proceedings of 4th International Conference on Human-Agent Interaction*, pp. 155–162, October 2016.
- [8] エドワード・ホール. かくれた次元. 日高敏隆, 佐藤信行 訳 みすず書房, 1970.
- [9] E Bruce Goldstein. Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 13, No. 2, pp. 256–266, 1987.