

AR 擬人化エージェントによる道案内における ジェスチャ生成視点が案内の理解度に与える効果

平松 拓馬^{1,a)} 長谷川 大² 佐久田 博司²

概要: 携帯端末での地図による道案内システムは多くのユーザに使用されているが、地図空間と現実空間のマッピングを必要とするため認知的負荷が高い。本稿では、現実空間に AR 表示された擬人化エージェントがジェスチャを用いて道案内を行うシステムを提案し、エージェントの道案内ジェスチャを生成する際に採用する視点について検討を行う。携帯端末向けに本システムを開発し、擬人化エージェントが Survey Perspective(地図のような上空からの俯瞰視点)によるジェスチャを使用する道案内と Route Perspective(実際にその経路を歩いていくような歩行者的視点)によるジェスチャを使用する道案内に対し、白紙に地図を描くタスクおよびアンケートによりユーザの案内理解度を比較した。参加者 20 名による実験の結果、Survey Perspective によるジェスチャを使用した道案内では、Route Perspective によるジェスチャを使用した道案内に比べ、理解度が向上する傾向にあることを確認した。

The Effect of Gesture Generation Perspective on Route Understanding in a Direction-giving System by an AR Human-like Agent

TAKUMA HIRAMATSU^{1,a)} DAI HASEGAWA² HIROSHI SAKUTA²

Abstract: The map-based direction-giving systems in mobile devices have been widely used, but sometimes it is not easy for users to convert a two dimensional map to surrounding three dimensional space. In this paper, to achieve intuitive understanding of direction-giving, we propose a direction-giving system that employs a human-like agent depicted in Augmented Reality (AR), and investigate the perspectives when the agent give direction in speech along with gestures. We developed the system for tablet devices and compared the direction giving when the agent's gestures were generated in survey perspective and route perspective. Listeners' understanding were measured by map task and questionnaire. The results showed that participants scored higher understanding when the gestures were generated in survey perspective than route perspective.

1. 序論

現在、携帯端末において、地図による道案内システムは多くのユーザに利用されている。しかし、地図情報を理解するためには地図空間と現実空間のマッピングを必要とするため認知的負荷が高い。そこで、地図を利用せず直感的に経路情報が理解可能な道案内システムが求められている。地図を用いない道案内システムの例として拡張現実 (AR)

を用いて端末のカメラ映像上に矢印を表示し、進行方向を指示する「Map Fan eye」*1 が実用化されている。しかし、矢印には進行方向や右左折の指示しか情報が含まれておらず、カメラ映像に映らない直近以降の経路情報の提示が困難なため、ユーザに与える情報量に限りがある。

カメラに映らない遠方までの道案内を直感的に伝える有効な方法として、擬人化エージェントを利用する方法が挙げられる。人間らしい身体を持つ擬人化エージェントのジェスチャを利用することで、直近以降の経路や周辺ランドマークとの位置関係、道路やランドマークの形状などの情報を手や腕の形、動作で表現可能なため、一度の指示

¹ 青山学院大学大学院
Graduate School of Aoyama Gakuin University, 5-10-1 Fuchinobe Sagami-hara-shi Chuo-ku, Kanagawa 252-5258, Japan

² 青山学院大学
Aoyama Gakuin University

a) c5613161@aoyama.jp

*1 <http://www.mapfan.com/iphone/ar/>

で多くの情報をユーザに与えることが可能である。また、ジェスチャによる指示は人間同士で日常的に行われているため、ユーザに受け入れられやすいと期待される。

擬人化エージェントを利用した道案内システムの研究として、伊藤ら [1] は、拡張現実を用いて擬人化エージェントを携帯端末のカメラ映像に重ねて出力し、ジェスチャと音声による道案内システムを開発した。そして地図による道案内システムと比較した結果、地図による道案内システムよりも擬人化エージェントを利用したシステムの方が経路情報の理解度が高いことを確認している。しかし、伊藤らの使用した道案内ジェスチャは 8 方向への直進指示と直進後の右折、または左折指示のみの単純なものであった。また、携帯端末の画面に表示される擬人化エージェントはカメラ映像の上に重ねて表示されているが、AR による表示ではない。そのため、擬人化エージェントが道路上に立っているようには見えず、ユーザは奥行のあるジェスチャの理解が困難である。

Cassell ら [2] の研究では、道案内において参照されるランドマークの形状や道路の曲がり具合などの画像的特徴を元に、MURML(Multimodal Utterance Representation Markup Language)[3] によって記述されたジェスチャ仕様を生成し、道案内ジェスチャを自動的に生成する道案内システムを開発した。しかし、このシステムで使用されている擬人化エージェントの NUMACK はデスクトップ PC 上で動作するスタンドアロンアプリケーションとして実装されており、地図と並列的に表示された擬人化エージェントが現在地から目的地までの案内を行うシステムとなっている。そのため、システムを持ち歩き、実環境を参照しながらの道案内を逐次的に行うことは実現されていない。また、経路やランドマーク形状が事前に人手によって登録されている必要があるため、使用区域が限定的なシステムとなっている。

そこで、本稿では AR 擬人化エージェントを用いた地域を限定せず、実環境を参照しながらの道案内が逐次的に可能な携帯端末向けの道案内システムを提案する。また、本システムで擬人化エージェントが効果的な道案内を行うために、道案内ジェスチャを生成する際に採用する視点について検討を行う。

以下、2 章で道案内ジェスチャを生成する視点について述べる。次に 3 章で案内者-被案内者の位置関係について述べる。4 章に提案システム概要について述べ、5 章に評価実験概要、6 章では実験結果および考察を述べる。最後に 7 章で結論をまとめる。

2. 道案内ジェスチャを生成する視点

Taylor ら [4] は案内者が説明の際に採用する視点として、Survey Perspective(地図のような上空からの俯瞰視点)と Route Perspective(実際にその経路を歩いていくような歩

行者的視点)の 2 種類が採用されうることを実際のジェスチャの分析により確認している。伊藤ら [1] や Cassell ら [2] のシステムではそのうち Route Perspective のみで道案内ジェスチャが生成されている。

しかし、Taylor ら [4] は、複数の経路や特徴のあるランドマークから構成される環境を表現する場合には、Survey Perspective でのジェスチャが採用されやすく、逆に単一の経路で同じ大きさのランドマークから構成される経路では、Route Perspective のジェスチャが採用されやすいことを示した。また、Striegnitz ら [5] は案内者が経路の見通しや建造物の特徴、被案内者の理解度によって上記の 2 種類の視点を使い分けることで認知的負荷の軽減に繋がる可能性があることを示している。同様に、杉本ら [6] は Survey Perspective と Route Perspective が被案内者の理解度に与える影響について日本語での調査を行っており、道案内情報を提示する手段としてテキストを用いた。道案内のテキストとして Route Perspective では、方向を表現する際に「前後左右」などの自己中心的参照枠に基づいたテキストが用いられ、他にも「歩く」「見える」といった動的な動詞が用いられる。しかし、Survey Perspective では「東西南北」といった環境中心的な参照枠となり、「ある」「いる」といった静的な動詞が用いられる。以上を踏まえて、実験では 2 つの視点のうち、片方の視点での道案内テキストを学習し、学習後は経路情報の正誤問題と地図描写のタスクを行う。正誤問題では学習した視点とは違う視点での想起問題も容易されており、視点の切り替えコストについても検証している。実験結果より Survey Perspective の方が Route Perspective よりも正解率が高く、回答時間が短いことが示されており、道案内の認知負荷が低いことを示している。しかし、ジェスチャを用いた比較については調査されておらず、ジェスチャにおいても同様な結果が得られるか調査を行う必要がある。よって Route Perspective のみの道案内ジェスチャの自動生成だけではなく、Survey Perspective での道案内ジェスチャの自動生成も必要と考えられる。しかし、Survey Perspective による道案内ジェスチャを自動生成する既存のシステムは存在していない。

3. 案内者-被案内者の位置関係

道案内ジェスチャを自動生成するシステムを提案するにあたって、案内者と被案内者の位置関係について検討を行う必要がある。

Hasegawa ら [7] は、擬人化エージェントによる道案内システムにおいて、案内者視点でのジェスチャと被案内者視点で行うジェスチャの比較を行い、被案内者視点に合わせよう、姿勢を傾けてジェスチャを行うことで理解が容易になることを明らかにしている。

他にも案内者-被案内者の位置関係に着目した研究としては、塚本ら [8] が仮想空間内におけるアバターを想定し

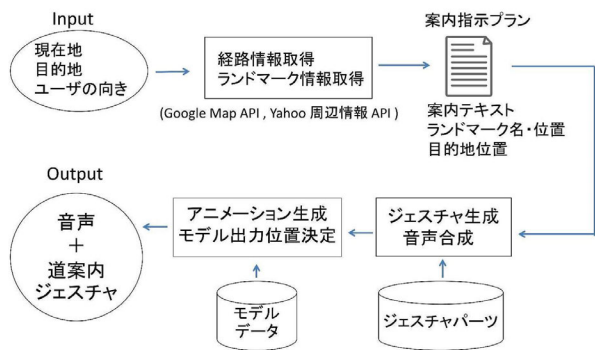


図 1 道案内生成処理の流れ

た、案内者と被案内者の位置関係によるジェスチャの変化を調査しており、位置関係によりジェスチャの可動範囲に制限が生じ、ジェスチャが変化することを明らかにした。それを踏まえ、平松ら [9] は実空間において案内者と被案内者の位置関係の条件が変化した際のジェスチャについて調査を行い、案内者と被案内者がハの字型で向き合っている際に案内者のジェスチャは利き手によらず、被案内者の立ち位置より遠い腕を使って行われる傾向があることを明らかにした。

以上の従来研究の成果より本システムにおいても Survey Perspective ではハの字型の位置関係を採用し、被案内者から遠い側の腕を使用したジェスチャ生成を行う。

4. 提案システム概要

道案内生成処理の流れを図 1 に示す。携帯端末には Apple 社の iPhone5 (iOS6) を用いた。以下にシステム構成を述べる。

4.1 経路情報・ランドマーク情報取得

ユーザが端末に目的地を入力すると、経路情報と交差点周辺のランドマーク情報を取得し、その結果をジェスチャ指定データとしてサーバに送信する。

はじめに、Google Geocoding API^{*2} により緯度経度を取得する。その後、Geocoding により取得した緯度経度を用いて現在地から目的地までの経路情報を Google Directions API^{*3} を用いて取得する。そして、Google Directions API により取得した経路情報の中から右折または左折する交差点の緯度経度を抽出する。その後、Yahoo 場所情報 API^{*4} を利用し、交差点の緯度経度の周辺に駅やコンビニなどのランドマークが存在するか判定を行う。ランドマークが交差点に面しているかの基準は Yahoo 場所情報 API から取得した score の値から判断する。本システムにおいて、こ

*2 <https://developers.google.com/maps/documentation/geocoding/>

*3 <https://developers.google.com/maps/documentation/directions/>

*4 <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/placeinfo.html>

表 1 事前生成したジェスチャデータ数

指示ジェスチャ	パターン数
(a) : はじめの進行方向指示	1
(b) : ランドマークの場所指示・交差点指示	5
(c) : 右折, 左折指示	2
(d) : 目的地の場所指示	2
(I) : エージェント初期姿勢 : 右向き・左向き	2
(II) : 進行方向指示に対するユーザの向き : 8 方位	8
計 $\{(a)+(b)+(c)+(d)\} * (I) * (II)$	160

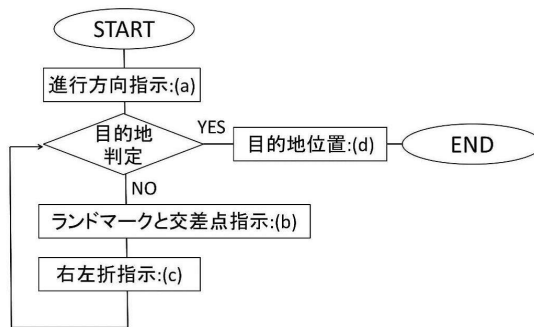


図 2 2 つ以下の交差点を経由する道案内手順

の score の値が 80 点を超えるものを交差点に面しているランドマークとみなしている。

また、ランドマークが交差点に対してどの位置に存在するかについても判定を行う。交差点の一つ前の緯度経度、交差点の緯度経度、ランドマークの緯度経度の 3 点からランドマークが交差点に対して右手前、右奥、左手前、左奥の 4 箇所いずれかに面しているか判定をしている。

4.2 ジェスチャ生成・音声合成

本システムは 2 つ先の交差点までの道案内ジェスチャを自動生成するシステムとして実装を行う。なお、3 つ先の交差点を曲がる前に目的地がある場合には目的地が左右のどちらに存在するか指示を行う。はじめに、一連のジェスチャを自動生成するにあたって、道案内のプロセスを以下のように 4 つのパーツに分ける。本システムでは表 1 に示す通り、プロセス毎に 8 方向分ジェスチャデータを事前に 160 個用意した。上記のプロセスを組み合わせることによって、一連の道案内手順を生成することが可能である。具体的な道案内手順を図 2 に示す。ジェスチャデータの事前生成については、基本的に右折や直進などの行動指示はユーザから遠い腕を動かして表現し、ランドマークの位置を指示する場合はユーザに近い腕でジェスチャを行えるものとする。それに合わせ、エージェントの向きは初めの交差点が右折の場合はユーザから見て斜め右向き、左折の際は斜め左向きで出力するように指定した。

また、ジェスチャ生成に使用するプロセス毎のジェスチャデータは案内指示プランニングにより取得した以下の要素によって決定される。この要素を数値化し、サーバ内

Survey Perspective



Route Perspective



図 3 Survey Perspective・Route Perspective のジェスチャ

のジェスチャ生成モジュールに送信する。

- (1) 端末から取得したユーザの向き
- (2) Google Directions API より取得した進行方向
- (3) 3つ先の交差点までに目的地が存在するか
- (4) 目的地が左右どちらに存在するか
- (5) 交差点の右折左折判定
- (6) 交差点周辺のランドマークの有無・位置

以上よりジェスチャのプランニングを行い、ジェスチャを生成する際に必要な複数のジェスチャデータを1つに結合する。

また、案内指示プランニングによって決定された経路情報やランドマーク名、ランドマークの位置から案内テキストを生成する。そしてそのデータをサーバへ送信し、サーバ上で案内音声ファイル(wav形式)を生成する。音声合成にはボイスソムリエ*5を使用した。

4.3 アニメーション生成モジュール

結合したジェスチャデータをMikuMikuDance*6用のジェスチャファイル(vmd形式)に変換し、3Dモデルデータ(pmd形式)と合わせ、アニメーション付き3Dモデルデータ(fbx形式)に変換する。

擬人化エージェントが道路上に立っているように出力するためにはカメラの映像よりトラッキングを行い、平面認識を行う必要がある。ARマーカーを使用して平面認識を行うことも可能だが、屋外においてARマーカーを使用するシステムは手間が掛かり現実的ではない。よって、本システムでは平面認識を行うことが可能なマーカーレスARであるmetaio SDK*7を用いた。トラッキングを終了する

とサーバ内で3Dモデルファイルと案内音声を自動生成し、認識された平面に擬人化エージェントを表示させ、動きに合わせて案内音声を再生する。

本システムはユーザがリクエストする度に現在地からのルート検索を行うため、逐次的な道案内が可能である。

5. 評価実験概要

本実験では携帯端末上動作するAR擬人化エージェントを利用した道案内システムを用いて、Survey PerspectiveとRoute Perspectiveのジェスチャによる道案内の理解度をそれぞれ調査し、比較を行う。

5.1 実験方法

本実験では被験者間実験によってSurvey PerspectiveとRoute Perspectiveの理解度の比較を行う。システムの動作環境を一定にするため、被験者は本システムの動作が録画された動画を視聴する。動画は3回流し、被験者は案内開始地点から目的地までの経路地図を動画が1回視聴する度に白紙へ描く(マップタスク)。その経路地図の採点結果と実験後のアンケートによって経路の理解度や記憶定着度の検証を行う。道案内実行時の動作例を図3に示す。

5.2 方向感覚調査

道案内情報の理解度や記憶定着度を調査するにあたり、得られる実験データは被案内者の空間把握能力に左右されることが予測される。よって、本実験を行う前に空間把握能力を測るテストを実施し(44名)、能力が高い評価群(10名)と低い評価群(10名)の被験者に本実験の協力を依頼した。空間把握能力については性別による差が存在すると松田ら[10]も示しているが、本調査は空間把握能力の高い評価群と低い評価群を区別する目的で行っているため、男女差に関しては考慮していない。

*5 <http://www.hitachi-solutions-business.co.jp/products/package/sound/voice/>

*6 <http://www.geocities.jp/higuchuu4/>

*7 <http://www.metaio.com/sdk/>

空間把握能力を計測するテストとして、竹内 [11] が作成した方向感覚質問紙を用いた。この質問紙は 5 段階のアンケート形式となっており、空間把握能力を主観的に求めることが可能である。全 50 問のうち、竹内の実験において方向感覚の自己評価 (5 段階) とアンケート結果の相関が有意であった 38 問を今回の計測テストとして用いる。

5.3 分析方法

経路地図の採点方法は加点方式とし、正しい経路が 1 本描けるごとに 1 点追加する (最大 4 点)。また経路上の交差点周辺に存在するランドマークや目的地を正しい位置に描けるごとに 1 点追加する (最大 4 点)。したがって 1 枚の地図の最大の得点は 8 点となる。この得点が 2 つの視点によって変化するのか、また方向感覚質問紙の結果により変化するのか調査を行う。また、実験後にジェスチャに関するアンケートを 2 問実施しており、この結果についても視点やテスト結果により変化するのか調査する。アンケート項目は以下の 2 問である。

- (1) 道案内は理解しやすかったか
- (2) 道順が記憶に残りやすかったか

6. 実験結果および考察

本章では、実験結果より視点の違いによる理解度及び、空間把握能力の違いによる理解度について考察を行う。

6.1 視点の違いにおける実験結果

加点項目である経路情報とランドマーク情報において「視点別のテスト結果」の平均値を表 2、表 3 に示す。2 視点の平均は経路得点、ランドマーク得点ともに 1 回目から 3 回目まで Route Perspective よりも Survey Perspective の方が高い結果となった。特に Survey Perspective の経路得点においては空間把握能力に依存せず、全員が 1 回目から満点である 4 点を獲得している。

Route Perspective の結果が Survey Perspective に比べ低い理由として考えられることは、擬人化エージェントが提示する道案内情報を被案内者の向きに変換する必要があるためである。Route Perspective は案内者が実際に説明する経路を辿るようにジェスチャを行う。そのためジェスチャは話し手目線となり、案内者に対面している被験者は

表 2 視点毎のテスト結果 (経路得点の平均点)

	1 回目	2 回目	3 回目
survey	4.000	4.000	4.000
route	2.5	3.8	4.000

表 3 視点毎のテスト結果 (ランドマーク得点の平均点)

	1 回目	2 回目	3 回目
survey	1.9	3.3	3.7
route	1.1	2.6	3.1

表 4 視点毎のアンケート結果 (平均点)

	survey	route	t 検定
道案内は理解し易かったか	6.1	3.5	**
道順が記憶に残り易かったか	6	4.5	**

** : $p < 0.01$

表 5 Survey Perspective の方向感覚質問紙結果及びテスト結果

方向感覚質問紙得点	マップタスク結果		
	1 回目	2 回目	3 回目
10	7	8	8
3	5	7	7
2	6	6	8
-9	7	8	8
-23	5	8	8
-53	7	8	8
-63	4	6	8
-66	7	8	8
-75	5	8	8
-78	5	8	8
相関係数	0.194	0.010	0.221

表 6 Route Perspective の方向感覚質問紙結果及びテスト結果

方向感覚質問紙得点	マップタスク結果		
	1 回目	2 回目	3 回目
4	1	7	8
4	5	7	7
-2	7	8	8
-14	4	8	8
-24	4	7	8
-55	7	8	7
-62	3	6	8
-70	5	6	7
-72	1	5	5
-93	0	2	5
相関係数	0.360	0.738	0.697

180 度回転した説明を受けることになる。右と左が逆転するため、心的回転をリアルタイムで行う必要があり、情報の整理がつかなくなることが大きな要因であると考えられる。案内情報が提示される媒体が地図の場合に、この現象は整列効果と呼ばれ、松井ら [12] は進行方向と地図の向きを合わせることで理解が容易となるという知見と一致する。この結果より経路情報及びランドマーク情報を説明する際には Survey Perspective のジェスチャの方が被案内者の理解度が高いといえる。

6.2 視点の違いにおけるアンケート結果

実験後に実施したジェスチャに関するアンケート結果の平均を表 4 に示す。テスト結果と同様に、Survey Perspective の結果が Route Perspective より良い結果となった。主観的な結果も高いことから、Survey Perspective のジェスチャの方が被案内者の理解度・記憶定着度が高いことをより証明可能な結果となった。

6.3 空間把握能力とテスト結果の相関

2 視点のアンケート結果及びテスト結果を表 5 と表 6 に示す。2 視点を比較した際に、空間把握能力が低い被験者の 3 回目の得点が大きく異なっている。Survey Perspective では空間把握能力に関係なく、殆どの被験者が 8 点を獲得しているが、Route Perspective では低評価群の被験者は 8 点に到達していない。また、空間把握能力が低いほど 2 回目・3 回目のテスト結果が低くなる相関が出ている。道案内ジェスチャの理解度の差が顕著に現れており、Route Perspective での道案内の理解度は空間把握能力に依存していると考えられる。よって、道案内ジェスチャには空間把握能力に依存しない Survey Perspective を採用することが望ましい。

7. 結論

本稿では携帯端末上で動作する AR 擬人化エージェントを用いた道案内システムの提案を行い、Survey Perspective と Route Perspective の 2 視点のジェスチャにおいて、被案内者の理解度や記憶定着度に差が現れるのか調査を行った。44 名の被験者に空間把握能力を調べるアンケートを実施し、そのうち高評価群 10 名と低評価群 10 名に本システムを用いた本実験を行った。その結果、道案内ジェスチャにおいて Route Perspective より Survey Perspective の方が道案内情報である経路情報やランドマーク情報の理解度が高いことが明らかとなった。今後の展望としては、実際に被験者に本システムを屋外で利用してもらい、他システムと比較した際の目的地に到達する時間や誤った経路に進む回数を記録し、本システム自体の評価を行いたい。また、本システムでは事前生成したジェスチャデータを組み合わせ一連のジェスチャを生成した。自動生成システムを改善するためにはジェスチャデータ自体を自動で生成可能なことが望ましく、今後の課題とする。最後に、被案内者側の振る舞いや視点の切り替えに関する調査も行いたい。頷きや視線の移動など、案内者へのシグナルとなる動作が判明すれば、携帯端末の加速度センサー等を用いてユーザの道案内の理解度を検知し、案内が進行するような対話型道案内システムの実現が可能となる。視点の切り替えについても、経路やランドマークの状況によって Survey Perspective と Route Perspective のいずれかを選択することが可能なシステムを開発し、Survey Perspective のジェスチャのみのシステムと比較実験を行い、評価を行いたい。

参考文献

- [1] 伊藤かほ, 小林裕貴, 平松拓馬, 長谷川大, 佐久田博司 : 道案内システムにおける AR 擬人化エージェントの効果, 第 75 回全国大会講演論文集 (2013).
- [2] Justine Cassell, Stefan Kopp, Paul A. Tepper, Kim Ferriman, Kristina Striegnitz : Trading Spaces: How Humans and Humanoids use Speech and Gesture to Give Directions, Conversational Informatics (John Wiley, 2007).

- [3] N.Cantelmo, J.Cassell, H.Vilhjalmsson, N. E.Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall and C. Pelachaudet: The behavior markup language: Recent developments and challenges, Lecture Notes in Computer Science (IVA2007), Vol. 4722, pp. 99-111 (2007).
- [4] Holly A. Taylor, Barbara Tversky: Perspective in Spatial Descriptions, JOURNAL OF MEMORY AND LANGUAGE 35, 371-391 (1996).
- [5] Kristina Striegnitz, Paul Tepper, Andrew Lovett, Justine Cassell.: knowledge Representation for Generating Locating Gestures in Route Directions, Spatial Language in Dialogue. Oxford University Press. (2008).
- [6] 杉本匡史: 空間メンタルモデルにおける個人差の要因, 京都大学教育学部卒業論文 (2010).
- [7] Dai Hasegawa, Justine Cassell and Kenji Araki: The Role of Embodiment and Perspective in Direction-Giving Systems, In Proceedings of the 2010 AAAI Fall Symposium on Dialog with Robots, pp.26-31, Arlington, VA, USA (2010).
- [8] 塚本剛生, 中野有紀子: メタバースにおける言語・空間情報に基づくアバターへの道案内ジェスチャの自動付与, TVRSJ, Vol. 17, No. 2, pp. 79-89 (2012)
- [9] 平松拓馬, 長谷川大, 佐久田博司: AR 擬人化エージェントを利用した道案内におけるジェスチャ及び視線の調査, HAI シンポジウム 2013, P26 (Online).
- [10] 松田浩一郎, 塚口 博司, 竹上 直也: 歩行者の空間認知能力と経路選択行動に関する分析, 土木学会第 58 回年次学術講演会 (2003).
- [11] 竹内謙彰: 空間認知の発達・個人差・性差と環境要因, 風間書房 (1992).
- [12] 松井孝雄: 空間認知の異方性と参照枠一整理効果はなぜ生じるのか?, 慶応義塾大学大学院社会学研究科紀要 (1992), 34, 51-58.