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

Investigation of Gesture and Eye-gaze towards Direction-Giving Using AR Embodied Conversational Agent

平松拓馬 [1*] 長谷川大 [2] 佐久田博司 [2]
Takuma Hiramatsu[1] Dai Hasegawa[2] Hiroshi Sakuta[2]

[1] 青山学院大学大学院理工学研究科
[1] Graduate School of Science and Engineering, Aoyama Gakuin University
[2] 青山学院大学理工学部情報テクノロジー学科
[2] Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University

Abstract: We propose a direction-giving system that employs an AR Embodied Conversational Agent for tablet devices, and investigate the use of gesture and eye-gaze in four typical positional relationships among speaker, listener, and direction of destination. Three participants gave directions in our experiments and we found out that the behaviours of eye-gaze changed, depending on the positional relationships.

1 はじめに

人間らしい身体を持つ擬人化エージェントは空間的なジェスチャを利用することで，地図による案内よりも，より直感的で認知的負荷の低い案内が可能であると考えられる．

そのためこれまでに擬人化エージェントを用いた道案内システムの研究が多く行われている．Vilhjalmssonら [1] の研究においては，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からBML(The Behavior Markup Language)[2] を生成し，ジェスチャを自動的に生成するメカニズムを開発した．しかし，本メカニズムはディスプレイに固定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デスクトップ上で動作するシステムとして実装されている．その為，システムを持ち歩いて道案内を行う設計とはなっておらず，ディスプレイ内の世界と現実世界のインタラクションがない．よって，実環境を参照しながらの逐次的な道案内ジェスチャが不可能である．

また，塚本らの研究では，空間的情報とテキストによる言語情報から擬人化エージェントの道案内ジェスチャ決定方式を提案している．ここでは人-人のインタラクションを分析し，立ち位置や言語表現によって手のひらの向きなどのジェスチャ形態が決定しており，セカンドライフ上に道案内アバタシステムを実装している [3]．しかし，これは仮想世界内での道案内を対象としており，また，アバタの言語情報は人手で入力されることを想定している．

実世界での実用性を考慮すると，道案内システムは持ち歩きが可能で，周囲の環境を参照しながら逐次的に案内可能であることが望まれるが，擬人化エージェントを利用したこのようなシステムはこれまでに研究が行われてこなかった．また，コンピュータグラフィクスで描画されたオンスクリーン擬人化エージェントによる空間的ジェスチャ表現が実世界を参照するために最も効果的であると考えられる手段として拡張現実（AR）技術の利用が考えられるが，これまでに着目されていない．

以上のことから，本研究ではタブレット端末上で動作するAR擬人化エージェントによる道案内システムの提案を行う．また，タブレット端末を携帯した逐次的な道案内であるため，ユーザが常に進行方向を向いているとは限らず，また擬人化エージェントは立ち位置を変更することができない．そこで，典型的な４つの位置関係を想定し，人-人のインタラクションを調査することで，本システムで利用可能なジェスチャ・視線の利用および言語情報との関連について調査を行う．

*連絡先：青山学院大学大学院理工学研究科
〒 252-5258 神奈川県相模原市中央区淵野辺 5-10-1
E-mail: c5613161@aoyama.jp

以下，2 章にシステムの概要について述べる．また，3 章に本システムを想定した人-人の道案内インタラクションの調査について述べ，4 章に実験結果および考察を述べる．おわりに，5 章で結論をまとめる．

2　システム概要

システムの構成を図 1 に示す．タブレット端末には Apple 社の iPad（iOS5）を用いた．本システムは，ユーザからのリクエスト（目的地の住所やランドマーク名）に対し，まず Google Geocoding API により緯度，経度を取得する．その後，Geocoding で取得した緯度，経度で現在地からの経路情報を Google Directions API を用いて取得する．次にカメラの映像よりトラッキングを行い，平面認識を行う．ここではマーカーレスでの平面認識に metaio SDK [1] を用いた．次に，生成された経路情報をサーバへ送信し，合成音声を取得する．音声合成にはボイスソムリエ [2] を利用した．カメラ映像内で認識された平面に 3D 擬人化エージェントを描画し，直近の経路情報とタブレット端末の向きに応じて，予め作成されたアニメーションを選択し，道案内を行う．また，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．道案内実行時の動作例を図 2 に示す．

3　実験概要

道案内に使用されるジェスチャの調査では，Cassell らが，人が方向や建物の画像的特徴を表現する場合のジェスチャ形態を分析し，参照項の形状とジェスチャ形態のパラメータとの対応付けを行っている [4]．また，Cassell らは案内のジェスチャをする場合の視点について，被案内者の視点，地図のように上空の俯瞰視点，目的地や周辺の建造物を基準とした相対的な視点がとれることを確認しており，案内者は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これらの視点を使い分けることで認知的負荷を軽減可能であると考えられる．[5]

また Hasegawa らは，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，被案内者視点で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 [6]．

また，擬人化エージェントの視線利用についての研究としては，深山ら [7] が道案内を行うタスクにおいて，擬人化エージェントの視線によってユーザの印象

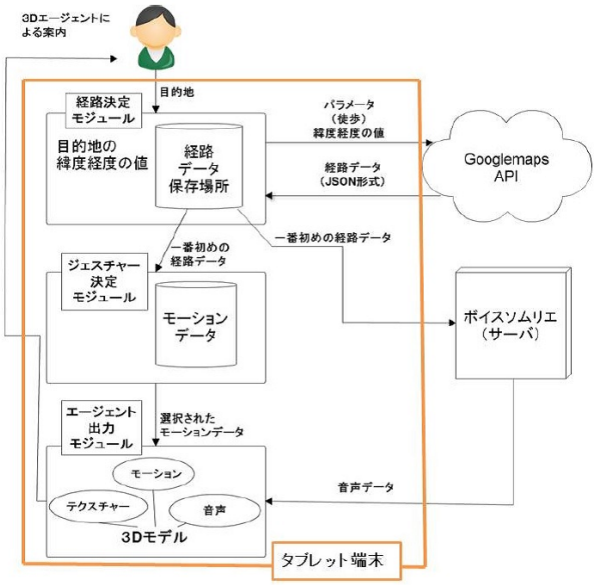

図 1: システム構成

図 2: ジェスチャ付 3D モデルの出力例

を操作できることを明らかにしており，注視持続時間が 2 秒程度のときに話し手を「好き」と評価する [8][9] というこれまでの知見が道案内においても同様であることを確認している．

このように，道案内に使用されるジェスチャや視線についての研究は既に行われているが，案内者-被案内者の位置関係との関連はあまり調査されていない．位置関係に着目した研究としては，塚本らが仮想空間内におけるアバターを想定した，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 [10]．しかしながら，ジェスチャの利用頻度や視線の利用およびこれらと言語情報との関連については調査されていない．

そこで，本調査ではタブレット端末上での擬人化エー

[1] http://www.metaio.com/sdk/
[2] http://www.hitachi-solutions-business.co.jp/products/package/sound/voice/

ジェントによる案内インタラクションのデザインを行うことを目的とし，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，視線の移動，凝視時間，および言語情報との関連について調査を行う．

3.1 実験方法

実験は図 3 に示すような配置において，典型的であると思われる 4 つの位置関係から中央の被案内者（実験者）に対して案内を行った．スクリーンには Google Maps のストリートビューの景色を表示しておき，案内役の実験参加者は事前に渡された地図のあるポイントからの風景であると説明を受ける．本実験において案内場所として選んだ場所の地図とストリートビューを図 4 で示す．

- 目的地 A：直進のみで到着
- 目的地 B：一度右折か左折が必要
- 目的地 C：一度ずつ左折と右折が必要

大学院生の男性 3 名が実験に参加し，参加者は全員地図の場所には一度も訪れたことがないと回答した．また，実験にあたって足元のマーカ枠内から出ないよう指示を受けた．4 つの位置関係において 3 つのうち 2 つの目的地をランダムに選択し経路案内を行い，これを 2 回繰り返し計 8 回の道案内が行われ，その様子を 2 方向からビデオカメラで記録した．

3.2 分析方法

録画した 2 つビデオデータを同期し，ビデオアノテーションツール ELAN [3] により視線利用に関する注釈付および書き起こしを行った．注釈は「被案内者注視」「スクリーン注視」「その他」の 3 つのタグを付け，それぞれ注視時間を調査した．また，案内中に使用した単語をリストアップし，3 つのタグとの共起回数の分析を行った．

4 実験結果および考察

本章では，実験結果およびタブレット端末上での擬人化エージェントによる道案内に利用可能なジェスチャ，視線の利用，言語情報との関連について考察を行う．

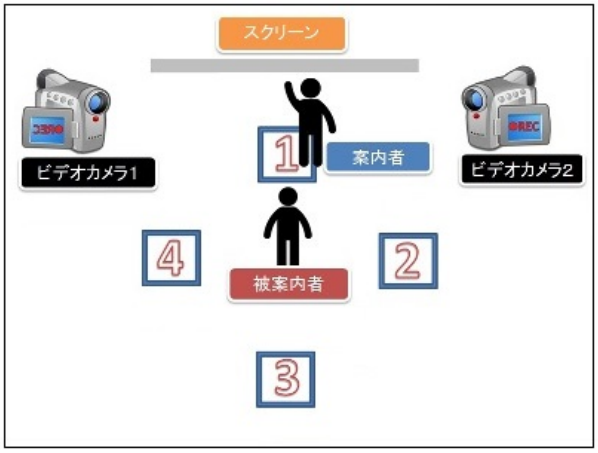

図 3: 実験室見取り図

図 4: 実験の地図と案内開始地点のストリートビュー

4.1 ジェスチャ

案内に使用されたジェスチャは，指さし（ランドマークを指示）や，手のひらを右左折の方向に向けるなど，先行研究による知見と一致していた．また，位置関係との関連では，エリア 2 では右腕を用い，エリア 4 では左腕を用いてジェスチャする傾向があった．

[3] http://www.lat-mpi.eu/tools/elan/

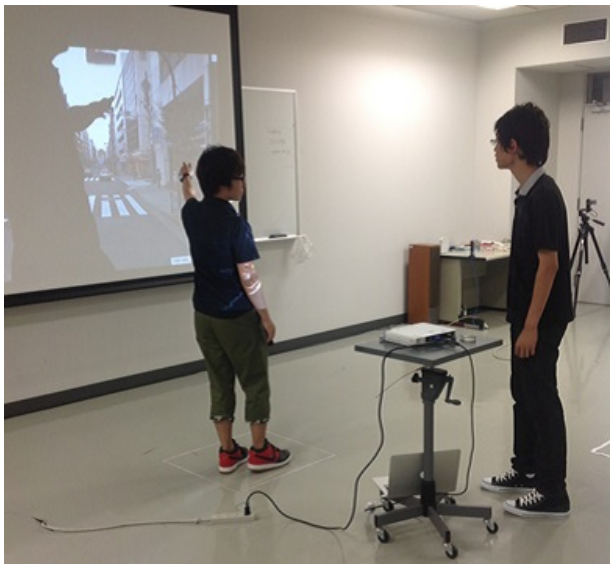

図 5: 実験の様子

4.2 注視先および平均注視時間

「被案内者注視」「スクリーン注視」「その他」の注視時間の平均とその割合を表 1 に示す．被案内者を注視する時間は全体の約 3 割，スクリーンの注視時間は約 5 割，それ以外は約 2 割となっていた．

被案内者注視	スクリーン注視	その他
149 秒 (29.7%)	263 秒 (52.2%)	91 秒 (18.1%)

表 1: 注視時間の平均および割合

対面会話において視線を向ける割合は，5 割程度がもっとも好印象である [8][9] ことが確認されているが，今回の実験には案内者が全く知らない場所を記憶してもらい案内を行った為，実験中に道順を思い出しながら案内をしている素振りが度々見られた．これは，一般に発言内容に自信が無いときには注視の時間が減少すること [10] が，道案内でも同様であったことを示している．本システムにおいても，注視時間を変化させることでエージェントの印象や案内内容の信頼などを制御可能であると考えられる．

4.3 目的地毎の平均注視時間

次に表 2 に目的地ごとの被案内者注視時間の平均とその割合を示す．直進のみで到着する目的地 A の場合では注視の平均持続時間が 1.51 秒と小さく，右折や左折を行う目的地 B や C の案内では 1.84 秒，1.88 秒と，被案内者を見ている平均持続時間が増加していることが示されている．これは目的地 B，C ではスクリーン注視の持続時間が減少していることから，目的地がスクリーン上では確認できない案内の困難性と関連があると考えられる．本システムでは現在地から直接確認できる直近の案内のみ行っているが，現在地から不可視な案内を行う場合に注視時間を変更することも可能である．

目的地	被案内者注視	スクリーン注視
A	3.38 秒 (23.0%, 1.51 秒/回)	11.55 秒 (66.2%, 6.28 秒/回)
B	5.79 秒 (39.9%, 1.84 秒/回)	16.20 秒 (48.5%, 3.70 秒/回)
C	9.51 秒 (32.0%, 1.88 秒/回)	5.24 秒 (35.7%, 2.87 秒/回)

表 2: 目的地毎の注視時間

4.4 各位置関係における注視時間

表 3 に位置関係ごとの注視時間合計および注視持続時間を示す．位置関係により注視時間，注視持続時間が異なることが示されている．まず，4 つのエリアの中ではエリア 3 が被案内者を注視する時間および注視持続時間が長かったが，これはスクリーンと被案内者の視線移動が容易であったためと考えられる．また，被案内者注視が全体に占める割合に着目すると，スクリーンと被案内者，案内者が直線的に並ぶ配置になると被案内者注視を頻繁に行うことが示されている．一方、スクリーンに対して案内者と被案内者が横並びの配置になると被案内者よりもスクリーンを見ながら道案内を行う時間が長く，被案内者を見る注視持続時間も表 1 で示した平均値を下回った．

	被案内者注視	スクリーン注視
エリア 1	5.45 秒 (41.1%, 1.87 秒/回)	5.82 秒 (29.9%, 2.34 秒/回)
エリア 2	5.90 秒 (24.8%, 1.09 秒/回)	12.01 秒 (51.2%, 3.67 秒/回)
エリア 3	11.15 秒 (43.3%, 3.13 秒/回)	9.09 秒 (38.%, 5.15 秒/回)
エリア 4	4.31 秒 (19.5%, 1.16 秒/回)	15.86 秒 (71.0%, 4.82 秒/回)

表 3: 各位置関係における注視時間と注視持続時間

4.5 言語情報との視線の関連

表 4 に被案内者に視線を向ける際に出現した単語の全リストと視線利用の関連について示す．これより，左，右などの位置関係を示す単語や右左折，直進などの行動を指す単語，および小学校やローソンなどのランドマークを示す単語の出現と被案内者注視が共起しやすい結果となった．一方，まずや次のなど順序を示す単語の際はスクリーンを見ている頻度が高かった．従って，擬人化エージェントにも同様の傾向で被案内者注視，スクリーン注視を実装すれば自然な道案内に近づけることが可能であると考えられる．

抽出単語	出現回数	被案内者注視	スクリーン注視
左	24	5(20.8%)	7(29.2%)
右	6	5(83.3%)	0(0.0%)
まっすぐ	44	8(18.2%)	7(15.9%)
左折	18	7(38.9%)	2(11.1%)
右折	20	13(65.0%)	0(0.0%)
直進	5	5(100%)	0(0.0%)
小学校	15	8(53.3%)	1(6.7%)
学校	1	1(100%)	0(0.0%)
河合塾	11	5(45.5%)	2(18.2%)
通り	15	7(46.7%)	0(0.0%)
ローソン	7	4(57.1%)	1(14.3%)
ファミリーマート	6	2(33.3%)	2(33.3%)
交差点	23	7(57.1%)	5(21.7%)
見えてくる	10	3(30.0%)	0(0.0%)
こっちの	1	1(100%)	0(0.0%)
まず	5	1(20.0%)	3(60.0%)
つぎの	1	1(100%)	4(80.0%)
角に	2	1(50.0%)	0(0.0%)
そこが	4	3(75.0%)	1(25.0%)
手前に	3	1(33.3%)	1(33.3%)
曲がる	7	2(28.6%)	0(0.0%)
合計	228	90(0.39%)	39(0.17%)

表 4: 被案内者注視と共起する単語一覧

5 結論

本稿ではタブレット端末上で動作する AR 擬人化エージェントによる道案内システムの提案を行い，タブレット端末でのエージェントによる道案内を想定したジェスチャ，視線の利用および言語情報との関連について調査を行った．

3 名の実験参加者に 4 つの異なる位置関係から道案内を行ってもらい，案内者の注視・持続時間の変化を調査した．その結果，案内する目的地までの経路が複雑なほど被案内者に視線を向ける平均時間が長くなることが示された．また，被案内者が進行方向に対し、逆向きに立っていた場合には被案内者注視が最も長くなり，左右を向いていた際には被案内者よりも案内方向を注視していた．さらに，方向を表す単語，行動を指示する単語，ランドマークとなる建物の名称を説明する場合に被案内者注視が行われ,「次に」などの順序を示す単語では案内方向へ視線を向けるという結果が得られた．

今後の展望としては，視線に関しての得られた結果を本システムの擬人化エージェントに反映させ，検証実験を行う．また被案内者側の振る舞いの調査も行いたい．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，タブレット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，これをトリガーに案内が進行するような，より自然な道案内システムの実現が可能となる．

参考文献

[1] 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: Recentdevelopments and challenges, Lecture Notes in Computer Sci- ence（IVA2007）, Vol. 4722,pp. 99-111 (2007).

[2] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thorisson, Hannes Vilhjalmsson:Towards a Common Frameworkfor Multimodal Generation:The Behavior Markup Language,INTERNATIONAL CONFERENCE ON INTELLIGENT VIRTUAL AGENTS,pp.21-23(2006)

[3] 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への道案内ジェスチャの自動付与, TVRSJ, Vol. 17, No. 2, pp. 79-89 (2012)

[4] Kopp, S.; Tepper, P. A.; Ferriman, K.; and Cassell, J. 2007. Trading spaces: How humans and humanoids use speech and gesture to give directions. In in T. Nishida (ed.), Conversational Informatics (John Wiley, 2007), chap. 8, 133?160.

[5] Kristina Striegnitz, Paul Tepper, Andrew Lovett, Justine Cassell:Knowledge Representation for Generating Locating Gestures in Route Directions,In:Coventry,K.,Tenbrink,T.,Bateman,J.(eds.) Special Language in Dialogue,pp.133-160.Oxford University Press,Oxford(2009)

[6] 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).

[7] 深山篤，大野健彦，武川直樹，澤木美奈子，萩田紀博: 擬人化エージェントの印象操作のための視線制御方法, 情報処理学会論文誌 43(12), 3596-3606, (2002)

[8] Cook,M.and Smith,M.C:The Role of Gaze in Impression Formation,Br.J Clin.Psych,Vol.14,pp.19-25(1975)

[9] Argyle,M,Lefebvre,L. and Cook,M:The Meaning of Five Patterns of Gaze,Eur.j Soc.Psych,Vol.4,No.2,pp.125-136(1974)

[10] 塚本剛生, 室谷優実, 岡本雅史, 中野有紀子: 道案内対話におけるマルチモーダルインタラクションの収集と分析—メタバースアバタのためのジェスチャ自動決定にむけて—, HAI2010, 1A-2 (2010)