

# 表出プロジェクションを用いた対話エージェントの開発 Development of Behavior Expressing Agent by Using Projection Mapping technique

石原義久<sup>1\*</sup> 小林一樹<sup>2</sup> 山田誠二<sup>3</sup>  
Yoshihisa Ishihara<sup>1</sup> Kazuki Kobayashi<sup>2</sup> Seiji Yamada<sup>3</sup>

<sup>1</sup> 信州大学大学院総合理工学研究科

<sup>1</sup> Graduate School of Science and Technology, Shinshu University

<sup>2</sup> 信州大学学術研究院

<sup>2</sup> Academic Assembly, Shinshu University

<sup>3</sup> 国立情報学研究所/ 総合研究大学院大学/ 東京工業大学

<sup>3</sup> National Institute of Informatics/ The Graduate University for Advanced Studies/  
Tokyo Institute of Technology

**Abstract:** This paper proposes a method for manufactured objects such as anime figures to exhibit highly realistic behavioral expressions to improve speech interaction between a user and an object. Using a projection mapping technique, an anime figure provides back-channel feedback to a user by appearing to nod or shake its head. We developed a listener agent based on the anime figure that listens to a user give directions to a specific location. We performed two experiments to investigate the users' impression of the speech interaction. The experimental results suggested that the anime figure with projection mapping made the agent seem more realistic and humanlike voice and humanlike behavior affect the user.

## 1 まえがき

近年, Apple 社の Siri や Google 検索などにおいて自然言語を用いるユーザインタフェースがスマートフォンをはじめとする身近なデバイスに搭載されている. これらのインタフェースは, ユーザがデバイスに話しかけることによって電話をかけたたり Web 検索といった機能を提供する. しかし, ユーザは音声入力を公の場所で使用することをためらう傾向がある. 多くの人が公の場所でスマートフォンを用いて音声通話をしているのに対し, 音声入力をしている姿を見ることは少ない. 音声入力時には, 音声通話に比べて趣味や趣向に関する言葉を使用する可能性が高く, プライバシーを保つために使用していない可能性がある. しかし, 音声通話と音声インターフェースに対する受容の差は, デバイスの人間らしさや知性に起因している可能性もある. たとえば, コンピュータを相手に画面上でしりとりゲームを行うとき, 人間が相手であると知らされた場合には, 相手がコンピュータプログラムであると知らされた場合よりも長い時間ゲームを続けることが報告され

ている [1].

これまで, 機械の人間らしさやユーザと機械との対話体験の向上に関する研究が行われてきた. Breazeal ら [2] は, 非言語情報を利用する人型ロボットが人間とのチームワークを向上させることを示している. Riek ら [3] は, チンパンジーに似せた頭部ロボットを用い, ユーザのうなずきなどの非言語情報を模倣し, ユーザと調和した関係の構築を試みている. Goetz ら [4] は, 社会的なタスクを実行するロボットがより人々に受け入れられることを示している. また, ロボットとソフトウェアエージェントとを比較する研究として, Powers ら [5] は画面に表示されたエージェントと実ロボットとのインタラクションを比較する実験を行い, ロボットが画面に表示されたエージェントよりも, 実ロボットのほうがユーザに対して好印象を与えることを示した. さらに, 原田 [6] は第三者がいる環境において, 画面に話しかけるよりも, ロボットに話しかける方が, ユーザの違和感を和らげると報告している.

上記の先行研究では, 非言語情報を用いたり, 社会的タスクを実行するロボットが人間らしさを向上させること, ソフトウェアエージェントよりも実ロボットのほうがユーザに対して好印象を与えることが示され

\*連絡先: 信州大学大学院総合理工学研究科  
〒380-8553 長野県長野市若里 4-17-1  
E-mail: 16w2003f@shinshu-u.ac.jp

ている。しかし、ロボットの場合、アクチュエータを用いるために、動作にぎこちなさが生じるため、ロボット制御の問題点として指摘されている [7]。

このような問題に対し、本研究では既存の人工物に動きを中心とする非言語情報を付加してユーザと対話する手法を提案する。対話対象として、市販のアニメフィギュアを用い、うなずき動作と首振り動作によるあいづち表現を、プロジェクションマッピング [10][11] を用いて付加する。

提案手法の利点は、対象物の形状に依存せず、モータのようなアクチュエータを用いないため、ぎこちない振動を生じさせることなく自然な表現を可能にし、非言語情報を用いて人工物の人間らしさを向上させる。会話におけるうなずきなどの非言語情報は合意形成や、会話内容、会話の長さに影響を与える重要な要素である [8][9]。これまで、MMDAgent をはじめとして人間らしいソフトウェアエージェントが提案されているが [12][13][14]、実体を持つロボットの方がリアリティが高い [15]。

以降では、提案手法について説明し、開発したエージェントを用いたフィギュア対話実験と行動表現の違いを検証するためのシミュレータ実験、および、それらの実験結果について述べる。

## 2 投影型行動表出エージェント

投影型行動表出エージェントとは、プロジェクションマッピングを用いて、既存の人工物にうなずき動作や首振り動作などの動的な行動表出を付加してユーザと対話を行うエージェントである。本手法は身近な愛着物に動作を加えてエージェント化することが可能になり、ユーザとエージェントの円滑な対話を実現することが期待できる。

### 2.1 エージェントの振る舞い

図 1 にプロジェクションマッピングを用いたうなずき動作の様子を示す。エージェントはユーザの発話に対して肯定的表現としてうなずき動作、否定的表現として首振り動作を行い、ユーザの発話に対して応答を行う。また、肯定的表現時には「はい」と「うん」、否定的表現時には「ううん」という音声出力を行う。音声は女性の声を録音し、ピッチを変更したものを使用した。エージェントはユーザの発話を常に監視しており、一定以上の音量の発話が 200ms 以上継続した場合にあいづち動作を行う。

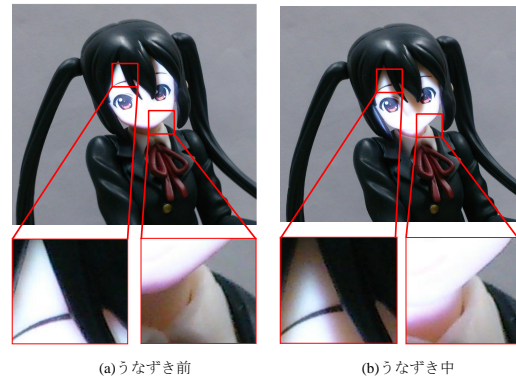


図 1: プロジェクションマッピングによるうなずき表現

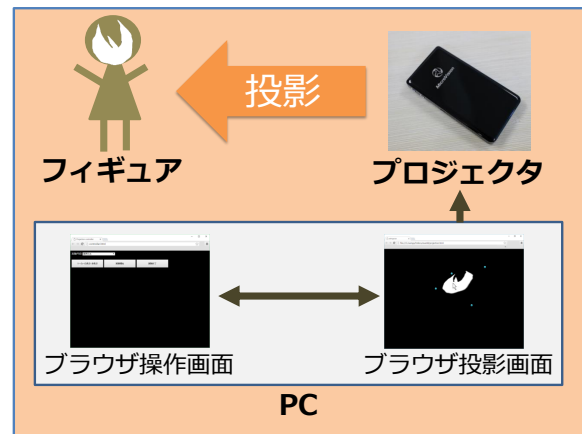


図 2: システム構成

### 2.2 エージェントの動作

エージェントの動作表現に用いる投影画像は、エージェントの顔の輪郭に合わせた白黒の図形を用いる。図 1 に示すうなずき動作は投影画像を 150ms で 20px 下に動かし、150ms で 20px 上に動かす。首振り動作は投影画像を 150ms で 10px 右に動かし、150ms で 20px 左に、150ms で 10px 右に動かす。システム構成を図 2 に示す。投影画像は Web ブラウザで動作し、操作画面と投影画面を有する。投影画面に表示された表現アニメーションは PC に接続したプロジェクタから出力している。投影ソフトウェアは JavaScript と PHP を用いて開発した。

## 3 実験

提案した行動表現手法の有効性を検証するために、開発したエージェントを用いたフィギュア対話実験と行



図 3: 投影型行動表出エージェントの外観

動表現の違いを検証するためのシミュレータ実験の2つの実験を行った。

### 3.1 フィギュア対話実験

プロジェクションマッピングを用いた行動表出エージェントの効果を検証するために、アニメフィギュアを用いた投影型行動表出エージェントを開発し対話実験を行った。

#### 3.1.1 投影型行動表出エージェントシステム

開発した投影型行動表出エージェントを図3に示す。エージェントはアニメフィギュアと台座(310×240×110mm)、プロジェクタで構成され、台座の中に小さなプロジェクタを配置し、その上に人型のアニメフィギュアを置く。プロジェクタはフィギュアの顔にアニメーションを投影できるように固定した。

#### 3.1.2 実験タスク

参加者には、エージェントに信州大学工学部正門からJR長野駅までの道案内を行うように指示した。また、会話時間が3分間であること、エージェントの反応を見ながら会話すること、エージェントに伝わるようにわかりやすくはっきりとした声量で発話することが指示された。対話終了後、参加者に対してアンケート調査を実施した。アンケート調査の内容を表1に示す。

表 1: フィギュア対話実験のアンケート内容

リアリティ	
Q1	フィギュア自体が動いているように見えた
Q2	フィギュアが生きているように感じた
Q3	フィギュアに意思があるように感じた
表現の自然さ	
Q4	フィギュアのうなずきかたは自然だった
Q5	フィギュアの声に違和感がなかった
Q6	フィギュアは人間のような振る舞いをした
Q7	フィギュアの応答は機械的だった
満足度	
Q8	フィギュアと楽しく話せた
Q9	もっとフィギュアと会話を継続してもよい
Q10	フィギュアは可愛らしかった
Q11	フィギュアに話すのは苦痛だった
Q12	フィギュアに対して話しやすかった
ユーザの発話戦略	
Q13	事務的に話しかけた
Q14	優しく話しかけた
Q15	フィギュアが理解できるように気をつけた
Q16	フィギュアの振る舞いにあわせて話した
Q17	意図的にゆっくり話した
意図や意思の認識程度	
Q18	フィギュアの考えがわかる気がする
Q19	フィギュアはあなたに好感を抱いているように感じた
Q20	フィギュアは誰かに操作されているように感じた
生物らしさ	
Q21	フィギュアを触ってみたかった
Q22	フィギュアに無理な要求をしてみたかった
Q23	あなたの考えをフィギュアが探っているかのように感じた
ユーザの発話に対する理解度	
Q24	フィギュアはあなたの話を真剣に聞いていた
Q25	フィギュアはあなたの話を理解しているように感じた
Q26	フィギュアはあなたの話に集中していないように感じた

#### 3.1.3 実験環境

図4に実験環境を示す。参加者はエージェントの前にあるイスに着席し、ヘッドフォンとマイクを着用する。そして、実験開始の合図ののち、参加者はエージェントへ話しかけるものとした。

#### 3.1.4 水準

実験には以下の4水準を設け、各水準に5名ずつ割り当てる。参加者間配置とし、参加者一人あたり1水準を経験する。プロジェクションマッピングと音声出力によるあいづち動作を行う(PV)水準、音声出力のみによるあいづち動作を行う(V)水準、プロジェクションマッピングのみによるあいづち動作を行う(P)水準、プロジェクションマッピングと音声出力、音声出力のみ、プロジェクションマッピングのみの3つの水準を織り交ぜ、それぞれ同確率でランダムにあいづち動作を行う(C3)水準を設定する。これらの水準間はエージェントのあいづち動作の表現以外の条件は全て同じとする。

#### 3.1.5 評価指標

評価指標は、アンケート調査による主観評価、エージェントとの対話における参加者の発話回数、参加者



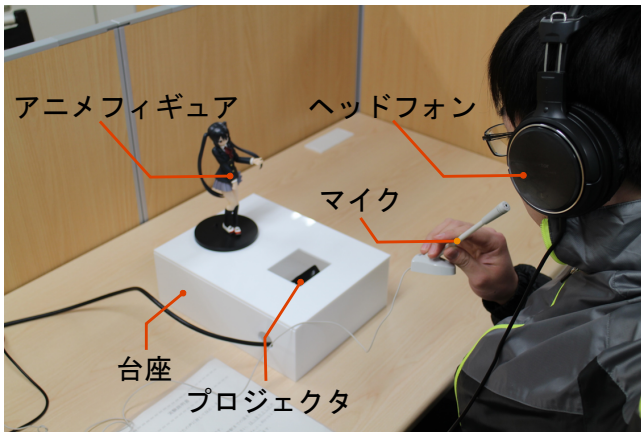


図 4: フィギュア対話実験の環境

の平均発話時間，参加者の平均無音時間，参加者が発したフィラーの数とした．会話中のフィラーは話者の発話権の保持や深く考えていることを表すと報告されており [16] [17][18]，集中の度合いを比較するために採用した．

### 3.1.6 実験参加者

実験参加者は信州大学の学生 20 名（男性：16 名，女性：4 名）で，平均年齢は 21.8 歳であった．

### 3.1.7 フィギュア対話実験の結果

フィギュア対話実験のアンケート調査結果および，参加者の行動評価結果を以下に示す．

**アンケート調査結果** 表 2 にアンケート調査による主観評価の結果を示す．各水準間の評価値を比較するために，クラスカルウォリスの検定を行ったところ，Q3「フィギュアに意思があるように感じた」( $p=0.03$ )，Q6「フィギュアは人間のような振る舞いをした」( $p=0.05$ )，Q24「フィギュアはあなたの話を真剣に聞いていた」( $p=0.08$ )において有意差があることが認められた．そこで下位検定として，ボンフェローニの調整を行ったマンホイットニーの U 検定を行ったところ，Q3 について，PV 水準の参加者は V 水準の参加者よりも有意にエージェントに意思があるように感じたことが示された．また，Q6 について，PV 水準の参加者は P 水準の参加者よりも有意にエージェントをより人間らしいと評価していることが示された．Q24 について，C3 水準の参加者は P 水準の参加者よりも有意にエージェントが話を真剣に聞いていると評価していることが示された．

**参加者の行動結果** 表 3 に参加者の発話回数，平均発話時間，平均無音時間，平均フィラー数を示す．各水準間の値を比較するために，クラスカルウォリスの検定を行ったところ，平均無音時間において有意傾向が認められた ( $p=0.085$ )．そこで下位検定として，ボンフェローニの調整を行ったマンホイットニーの U 検定を行ったところ，P 水準に参加した参加者は V 水準に参加した参加者よりも対話における無音時間が有意に短いことが認められた．このことからプロジェクションマッピングによる行動表出は無音時間を短くすることが示唆された．

### 3.1.8 フィギュア対話実験の考察

**アンケート調査による主観評価** アンケート調査による主観評価から有意差が認められたものは Q3, Q6, Q24 であった．PV 水準は，V 水準よりもエージェントに意思があるという印象を強く与えることが示唆された．また，PV 水準のエージェントはプロジェクションマッピングのみの P 水準よりもより人間らしい動作をしているという印象を与えることが示唆された．さらに，プロジェクションマッピングと音声出力，プロジェクションマッピングのみ，音声出力のみを組み合わせた C3 水準は P 水準のエージェントよりも，参加者の話をより真剣に聞いているように感じるという印象を与えることが示唆された．

以上の結果から，PV 水準と C3 水準のエージェントは参加者に満足度の高いリアリティを与えることができたと考えられる．また，有意差は認められていないものの，C3 水準のエージェントについて Q10「フィギュアは可愛らしかった」，Q15「フィギュアが理解できるように気をつけた」，Q24「フィギュアはあなたの話を真剣に聞いていた」における評価値が 6 を超えることから，様々な種類のあいづちを行うエージェントはユーザの満足度を高められる可能性がある．

**参加者の行動分析** 参加者の発話行動計測の結果から，発話回数は各水準間で有意差が認められなかった．有意差は認められていないが P 水準では V 水準の 2 倍近くの発話回数が見られ，また，PV 水準や C3 を超える発話回数になっていることがわかる．音声によるあいづちを含む水準では，参加者がエージェントの発話と自分の発話が重なることを避けるために発話回数が少なくなった可能性がある．

発話時間については，有意差は認められていないが V 水準ではほかのプロジェクションマッピングを含む水準と比べて 1 秒長い発話時間が見られた．無音時間については，P 水準と V 水準の間に有意傾向が認められ，P 水準に参加した参加者は V 水準に参加した参加者よりも無音時間が有意に短いことが認められた．ま

表 2: フィギュア対話実験のアンケート結果

問	水準	PV		V		P		C3		Kruskal-Wallis test	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	p-value	備考
Q1	フィギュア自体が動いているように見えた	4.40	0.80	3.20	1.17	4.60	1.02	3.40	1.36	0.21	
Q2	フィギュアが生きているように感じた	4.00	1.10	2.80	1.17	2.20	0.75	2.80	1.17	0.15	
Q3	フィギュアに意思があるように感じた	5.40	0.49	3.40	0.80	3.60	1.50	4.60	1.02	0.03	※ 1
Q4	フィギュアのうなずきかたは自然だった	4.60	1.02	3.80	0.75	4.00	1.79	3.60	1.02	0.56	
Q5	フィギュアの声に違和感がなかった	4.20	1.47	4.00	1.10	2.80	1.47	4.60	1.50	0.45	
Q6	フィギュアは人間のような振る舞いをした	4.00	0.63	3.40	1.02	2.40	0.49	4.20	1.60	0.05	※ 2
Q7	フィギュアの応答は機械的だった	4.20	1.94	5.40	0.80	4.80	1.33	5.00	1.10	0.76	
Q8	フィギュアと楽しく話せた	4.00	1.67	3.80	1.17	2.60	1.02	4.40	0.80	0.17	
Q9	もっとフィギュアと会話を継続してもよい	3.60	1.50	3.20	1.72	2.80	1.17	4.60	1.36	0.37	
Q10	フィギュアは可愛らしかった	5.20	1.60	5.00	1.67	5.00	2.19	6.20	0.40	0.61	
Q11	フィギュアに話すのは苦痛だった	3.60	1.74	3.60	1.85	4.60	1.50	2.40	0.80	0.27	
Q12	フィギュアに対して話しやすかった	5.20	0.40	3.80	0.75	3.20	1.83	3.60	1.85	0.14	
Q13	事務的に話しかけた	4.40	1.20	5.80	0.75	3.80	2.32	4.20	2.04	0.42	
Q14	優しく話しかけた	5.00	1.79	4.80	1.17	5.20	2.23	3.80	1.83	0.61	
Q15	フィギュアが理解できるように気をつけた	5.80	0.75	5.00	1.67	4.40	2.06	6.00	0.63	0.59	
Q16	フィギュアの振る舞いにあわせて話した	4.60	1.02	4.00	2.10	3.60	1.02	5.20	1.33	0.38	
Q17	意図的にゆっくり話した	5.40	0.80	5.20	1.72	4.80	2.40	5.20	1.33	0.99	
Q18	フィギュアの考えがわかる気がする	3.20	1.33	2.60	1.02	2.20	1.17	3.00	1.10	0.64	
Q19	フィギュアはあなたに好感を抱いているように感じた	4.40	0.80	3.60	1.50	2.60	1.36	3.60	1.02	0.23	
Q20	フィギュアは誰かに操作されているように感じた	3.60	1.02	3.80	1.17	4.60	1.96	2.20	0.98	0.11	
Q21	フィギュアを触ってみたかった	2.20	0.98	2.00	1.26	3.60	2.15	3.00	1.41	0.49	
Q22	フィギュアに無理な要求をしてみた	2.20	1.94	4.20	1.94	3.20	2.04	3.00	1.55	0.27	
Q23	あなたの考えをフィギュアが探っているかのように感じた	3.00	1.41	5.00	1.10	4.40	1.74	3.80	1.60	0.23	
Q24	フィギュアはあなたの話を真剣に聞いていた	5.20	0.75	5.60	1.50	3.80	1.60	6.00	0.63	0.08	※ 3
Q25	フィギュアはあなたの話を理解しているように感じた	4.00	1.10	5.00	2.10	3.00	1.41	3.80	1.94	0.50	
Q26	フィギュアはあなたの話に集中していないように感じた	3.20	1.60	3.20	1.83	3.40	1.62	3.40	1.74	0.99	

※ 1 (Q3): PV>V (p=.048, Mann-Whitney U test with Bonferroni correction)

※ 2 (Q6): PV>P (p=.096, Mann-Whitney U test with Bonferroni correction)

※ 3 (Q24): C3>P (p=.096, Mann-Whitney U test with Bonferroni correction)

た, V 水準における無音時間はほかの水準と比べると 2 秒程度長いことがわかる. 音声のみのあいづちでは, エージェントの表情に変化がなく, 動作と音声の一致が取れなくなり, 参加者が困惑したために無音時間が長くなった可能性がある. このことから, プロジェクションマッピングは対話の円滑さを向上させる効果があると考えられる.

フィラー数において各水準間での有意差は見られなかったが, P 水準と C3 水準のフィラー数がほぼ等しく,ほかの水準のおよそ 2 倍近くの回数になっていることがわかる. P 水準と C3 水準では, エージェントがほかの水準に比べ参加者に発話内容を考えさせるような効果を与えている可能性がある. 音声が無いあいづち行動を含む P 水準と C3 水準は音声を含む PV 水準と

V 水準と比べてフィラー数が多いことから, 対話エージェントの開発において, 適切な音声によるフィードバックが効果的かもしれない.

### 3.2 シミュレータ実験

プロジェクションマッピングを用いた行動表現の違いを検証するために, 投影型行動表出エージェントシミュレータを開発し, 動画視聴実験を行った.

表 3: エージェントとの対話における参加者の発話回数, 発話時間, 無音時間, フィラー数結果

行動内容	PV		V		P		C3		Kruskal-Wallis test	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	p-value	備考
発話回数	47.00	22.06	34.60	23.69	65.40	11.91	52.2	25.09	0.400	
平均発話時間 [s]	1.97	0.49	2.84	2.14	1.65	0.37	1.77	0.31	0.476	
平均無音時間 [s]	1.72	0.35	4.42	2.59	1.35	0.60	2.45	2.03	0.085	※ 4
フィラー数	7.20	5.31	6.60	9.26	13.60	5.57	14.60	4.84	0.106	

※ 4: V>P (p=.096, Mann-Whitney U test with Bonferroni correction)

### 3.2.1 投影型行動表出エージェントシミュレータシステム

シミュレータは Web ブラウザで動作し, 操作画面とシミュレーション画面を有する. 画面上で, 光の当たり方をシミュレートするアニメーションによって行動表出を行う. シミュレータソフトウェアは JavaScript と PHP を用いて開発した.

### 3.2.2 シミュレータ上のエージェントの振る舞い

シミュレータにおけるエージェントの振る舞いは, 音声と動作を組み合わせた表現が可能である. 音声については人間的音声と人工的音声の2種類を, 動作についても同様に人間的動作と人工的動作の2つを組み込んだ. 人間的音声はフィギュア対話実験と同様の「はい」と「うん」の音声とした, 人工的音声は, 独自に開発した人工的音声生成システムを用いて生成した. 生成システムでは, 入力語の母音のみを用いて波形を生成する. 母音の波形は, 合成音声アプリケーション OpenJTalk で生成した音声「あ, い, う, え, お」から, それぞれ第1~3フォルマントのみを抽出し, 合成したものである. 肯定的な人工的音声として「はい」を生成システムに入力し, 否定的な人工的音声として「うん」を生成システムに入力して波形を生成した. これらの音声のスペクトログラムを図5に示す. 人間的動作はフィギュア対話実験と同様のうなずき, 首振り動作とした. 人工的動作は誇張された動作を行う, 肯定的表現時にはうなずき動作で影を模した帯が顔の上を縦に動くアニメーションを行い, 否定的表現時には首振り動作で影を模した帯が顔の上を横に動くアニメーションを行う. 顔の上を影が動くことは人間の通常動作ではありえないため人工的表現として採用した. 人工的なうなずき動作の様子を図6に, 人工的な首振り動作の様子を図7に示す.

### 3.2.3 実験タスク

実験はオンラインで実施し, 参加者は自身が所有する PC のブラウザからアンケートページへアクセスすることで開始された. オンライン実験では, 動画を視聴

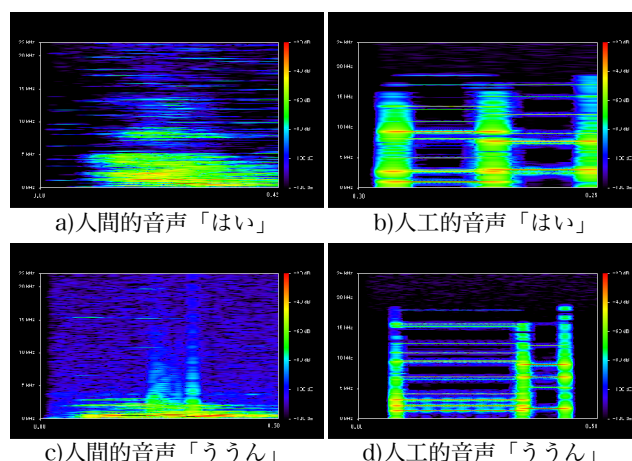


図 5: 音声のスペクトログラム

表 4: シミュレータ実験のアンケート内容

Q1	システムに意思があるように感じた
Q2	人間同士の話のように感じた
Q3	システムは話者の話を真剣に聞いていた
Q4	システムの反応はわかりやすかった
Q5	システムの反応に違和感があった
Q6	話者は話しやすそうだった
Q7	親しみやすい
Q8	会話が成立していた

してからアンケートに答える形式とした. 参加者は図8に示すようなブラウザ画面上で投影型行動表出エージェントシミュレータを用いて作られた4つの動画を見る. 動画の内容については以降の節で説明する. アンケート調査の内容を表4に示す.

### 3.2.4 動画の内容

実験動画には音声要因と動作要因の2要因を設定し, それぞれ人間的表現と人工的表現の2水準で構成した. 参加者内配置とし, 参加者は全ての水準を経験する. 人間的動作と人間的音声によるあいづち動作を行う動画(HH)条件, 人間的動作と人工的音声によるあいづち動作を行う動画(HM)条件, 人工的動作と人間的音声によるあいづち動作を行う動画(MH)条件, 人工的動作

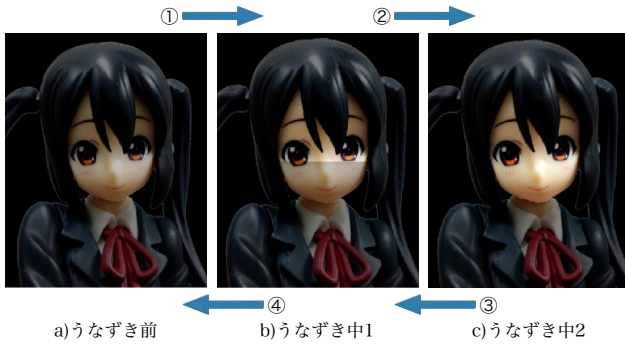


図 6: 人工的なうなずき動作

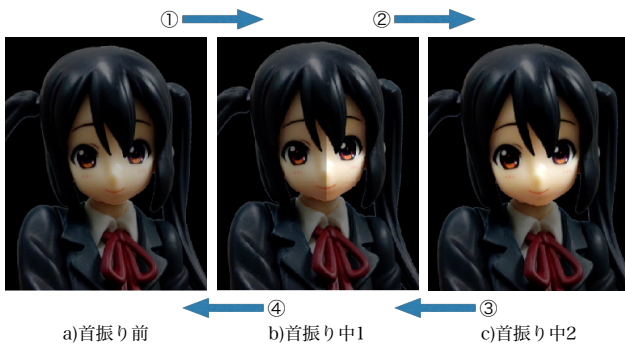


図 7: 人工的な首振り動作

と人工的音声によるあいづち動作を行う動画 (MM) 条件を設定した。各動画はエージェントのあいづち動作の表現以外の条件は全て同じとする。

### 3.2.5 実験参加者

実験参加者は信州大学の男子学生 10 名で、平均年齢は 21.5 歳であった。

### 3.2.6 シミュレータ実験の結果

表 5 にアンケート調査による主観評価の結果を示す。各条件間の評価値を比較するために、2 要因参加者内配置の分散分析を行ったところ、Q1「システムに意思があるように感じた」において音声要因の主効果が認められた ( $p < .10$ )。Q2「人間同士の話のように感じた」においては交互作用が有意であった ( $p < .01$ ) ため、下位検定として単純主効果の検定を行った。その結果、人間的音声において動作の単純主効果が有意であった ( $p < .05$ : 人工的動作 < 人間的動作)。人間的動作において音声の単純主効果が有意であった ( $p < .01$ : 人工的音声 < 人間的音声)。人工的動作において音声の単純主効

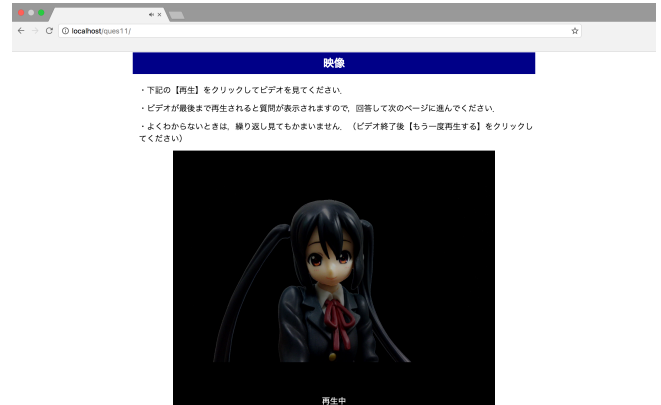


図 8: オンライン実験の画面

果が有意傾向であった ( $p < .10$ : 人工的音声 < 人間的音声)。Q7「親しみやすい」においては、音声要因の主効果が有意であった ( $p < .10$ )。

以上をまとめると、Q1 については、人間的音声の方が人工的音声よりもシステムに意思があるように感じると評価されていることが示された。Q2 については、人間的音声において人間的動作の方が人工的動作よりも人間同士の話のように感じると評価されていること、人間的動作において人間的音声の方が人工的音声よりも人間同士の話のように感じると評価されていること、また、人工的動作において人間的音声の方が人工的音声よりも人間同士の話のように感じると評価されていることが示された。Q7 については、人間的音声の方が人工的音声よりも親しみやすいと評価されていることが示された。

### 3.2.7 シミュレータ実験の考察

アンケート調査による主観評価から有意差が認められたものは Q1, Q2, Q7 であった。また、Q1, Q2, Q7 のいずれにおいても人間的音声の評価値が高いことから、人間的音声がリアリティの高い対話を実現する上でユーザに与える影響が大きい要因であることを示唆する結果が得られた。人間的音声は我々が普段使用している言葉であり意図が通じやすいため Q1「システムに意思があるように感じた」、Q2「人間同士の話のように感じた」、Q7「親しみやすい」という印象を持ちやすかったと考えられる。人工的音声は母音だけの表現であったため何を話者に伝えたいのかわからず参加者を混乱させた可能性がある。また人間的音声の際に人間的動作の評価値が高かったことは、人間同士の会話のように感じやすかったためと考えられる。

エージェントを開発する際には人間的動作表現と人間的音声表現の両方を用いることが有効である可能性



表 5: シミュレータ実験のアンケート結果

問	動作要因 音声要因	人間的動作				人工的動作				分散分析	
		人間的音声		人工的音声		人間的音声		人工的音声			
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	p-value	備考
Q1	システムに意思があるように感じた	4.70	0.78	4.00	1.61	4.30	0.78	3.70	1.35	p<.10	音声要因※ 5
Q2	人間同士の話のように感じた	4.40	0.92	1.90	0.70	3.00	0.89	2.00	1.34	p<.01	交互作用※ 6
Q3	システムは話者の話を真剣に聞いていた	5.00	1.10	3.90	1.30	3.50	1.43	3.90	1.45	n.s.	
Q4	システムの反応はわかりやすかった	5.10	1.45	4.20	1.72	4.30	1.42	3.90	1.58	n.s.	
Q5	システムの反応に違和感があった	4.40	1.11	3.70	1.27	5.10	1.30	4.50	1.50	n.s.	
Q6	話者は話しやすそうだった	3.90	0.94	3.00	1.34	3.20	1.47	2.80	1.17	n.s.	
Q7	親しみやすい	4.30	0.90	3.10	1.51	3.90	1.58	2.90	1.04	p<.10	音声要因※ 7
Q8	会話が成立していた	4.00	1.10	3.20	1.17	3.40	1.02	3.10	1.30	n.s.	

※ 5(Q1):交互作用なし, 音声要因の主効果あり (p<.10: 人工的音声<人間的音声).  
 ※ 6(Q2):交互作用あり (p<.01), 単純主効果の検定を行なった.  
 (人間的音声時に p<.05: 人工的動作<人間的動作)  
 (人間的動作時に p<.01: 人工的音声<人間的音声)  
 (人工的動作時に p<.10: 人工的音声<人間的音声)  
 ※ 7(Q7):交互作用なし, 音声要因の主効果あり (p<.10: 人工的音声<人間的音声).

がある。

## 4 考察

### 4.1 フィギュア対話実験とシミュレータ実験の比較

フィギュア対話実験におけるアンケート項目 Q3 とシミュレータ実験におけるアンケート項目 Q1 には共通する問「(フィギュア/システム) に意思があるように感じた」がある。フィギュア対話実験では、プロジェクションマッピングと音声出力によるあいづち動作を行う (PV) 水準が音声出力のみによるあいづち動作を行う (V) 水準よりも意思があるように感じたと評価されている。シミュレータ実験では、音声要因の主効果が認められ、人間的音声の方が人工的音声よりも意思があるように感じたと評価されている。両実験の結果から、人間的音声による表現に加え、プロジェクションマッピングを用いることで、エージェントが意思を持っている感覚をユーザに与えられる可能性がある。

フィギュア対話実験におけるアンケート項目 Q6「フィギュアは人間のような振る舞いをした」とシミュレータ実験におけるアンケート項目 Q2「人間同士の話のように感じた」について、人間らしさについて検討を行う。フィギュア対話実験では、プロジェクションマッピングと音声出力によるあいづち動作を行う (PV) 水準がプロジェクションマッピングのみによるあいづち動作を行う (P) 水準よりも人間のような振る舞いをしたと評価されている。これに対し、シミュレータ実験では、人間的音声を発するとき、人間的動作の方が人工的動作よりも人間同士の話のように感じ、人間的動作を行うとき、人間的音声の方が人工的音声よりも人間同士の話のように感じると評価されている。また、人工的動作を行う場合には、人間的音声の方が人工的音声よりも人間同士の話のように感じると評価されてい

る。両実験の結果から、人間的音声による表現に加え、人間的動作を行うことでエージェントの人間らしさを高めることができると考えられる。

フィギュア対話実験におけるアンケート項目 Q24 とシミュレータ実験におけるアンケート項目 Q3 には共有する問「(フィギュア/システム) は (あなた/話者) の話を真剣に聞いていた」がある。フィギュア対話実験では、プロジェクションマッピングと音声出力、音声出力のみ、プロジェクションマッピングのみの3つの水準を織り交ぜ、それぞれ同確率でランダムにあいづち動作を行う (C3) 水準がプロジェクションマッピングのみによるあいづち動作を行う (P) 水準よりも話を真剣に聞いていたと評価され、シミュレータ実験では、有意差が認められなかった。これは、C3 水準はランダムなあいづち動作を行うため、参加者にとって予想外の表現があり、シミュレータ実験では一貫して同じ動作と音声のあいづち動作を行う点の違いが影響したと考えられる。両実験の結果から、エージェントの予想外なあいづち動作が、真剣にエージェントがユーザの話を聞いていると感じさせることができる可能性がある。

### 4.2 制限事項

本研究には現段階において以下のような制限事項がある。

1つ目は、実験で用いたタスクが日常会話の一部を切り出した形式の対話となっている点である。参加者の発話はエージェントに対し道案内を行うもので、エージェントの動作はあいづち行動のみに制限されていた。そのため、実験結果の一般性には限りがあり、エージェントからの質問などあいづち以外の発話を考える必要がある。

2つ目は、プロジェクションマッピングによる表現が人工的である点が挙げられる。提案手法では、エージェントの頭部画像を動かす人工的なもので、人間の



表情を表現しきれなかった可能性がある。プロジェクションマッピング技術に関する河邊らの技術 [19] を用いればより細かい表現が可能になり，リアリティを向上させることができるかもしれない。

最後に，実験の各水準に参加した参加者の数が十分でない点が挙げられる。本研究の次の段階では，プロジェクションマッピングを用いた行動表出エージェントの効果をより明確に示すため，より多くの参加者を集める実験を行う予定である。

## 5 まとめ

本稿では，プロジェクションマッピングを用いて，人工物に高いリアリティを付与する手法を提案した。提案手法の有効性を調査するために，アニメフィギュアに対して，プロジェクションマッピングによるうなずき動作と首振り動作のあいづちを行うエージェントを開発した。フィギュア対話実験ではエージェントに対して参加者は道案内を行った。アンケート調査による主観評価，発話回数，平均発話時間，平均無音時間，フレーズ数について集計し，分析を行ったところ，プロジェクションマッピングと音声出力を同時に用いたあいづち動作を行うエージェントは，そのエージェントの意思の存在や人間らしさを向上させることを示唆する結果を得た。また，音声のみやプロジェクションマッピングのみであいづちを行うエージェントは十分な対話の円滑さを確保できないことが示唆された。

シミュレータ実験ではプロジェクションマッピングを用いたエージェントの振る舞いにおいて人間的表現と人工的表現を設定し行動表現の違いを検証した。アンケート調査による主観評価を集計し，分析を行なったところ表現が誇張された人工的表現よりも，人間的表現の方がユーザにエージェントの意思の存在，人間らしさの向上，親しみやすさに影響を与える要因であることが示唆された。

提案したプロジェクションマッピングを用いた行動表出エージェントは，人間に対して発話している感覚と同等の対話が可能な音声インターフェースへの応用が期待できる。今後は，図9に示すように河邊ら [19] の研究を参考にして，補色表現を用いた行動表出プロジェクションに取り組む予定である。

## 謝辞

本研究の一部は，JSPS 科研費新学術領域研究「認知的デザイン学」(No.JP26118005) の助成を受けたものである。

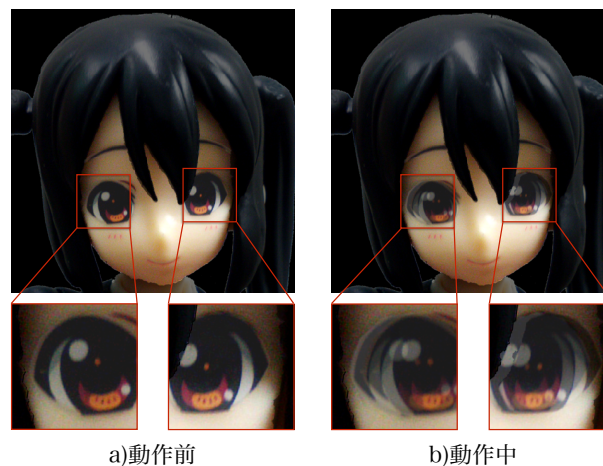


図 9: 補色表現で目を動かす様子

## 参考文献

- [1] 山本吉伸, 安西祐一郎: 特集 人間とシステムのインタラクションインタラクションの楽しさ, 人間工学, Vol.30, No.1, pp.15-25 (1994)
- [2] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman and M. Berlin: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork, In Proceedings of the International Conference on Intelligent Robots and Systems (IROS'05), pp.708-713 (2005)
- [3] L. D. Riek, P. C. Paul and P. Robinson: When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry, Journal on Multimodal User Interfaces, Vol.3, No.1-2, pp.99-108 (2010)
- [4] J. Goetz, S. Kiesler and A. Powers: Matching robot appearance and behavior to tasks to improve human-robot cooperation, In Proceedings of the International Workshop on Robot and Human Interactive Communication (ROMAN'03), pp.55-60 (2003)
- [5] A. Powers, S. Kiesler, S. Fussell and C. Torrey: Comparing a computer agent with a humanoid robot, In Proceedings of the International Conference on Human-Robot Interaction (HRI'07), pp.145-152 (2007)
- [6] 原田悦子: ロボットの「使いやすさ」?: インタフェースとしてのコミュニケーション型ロボットを考え

- る, 日本ロボット学会誌, Vol.24, No.6, pp.720–724 (2006)
- [7] 石黒章夫: きびきびと動くロボットを目指して「心身一元論」に立脚した人工知能システムの構築に向けて, 電気学会論文誌 C, Vol.121, No.1, pp.48–49 (2001)
- [8] H. H. Clark: Managing problems in speaking, *Speech communication*, Vol.15, No.3, pp.243–250 (1994)
- [9] B. Schuller, D. Seppi, A. Batliner, A. Maier and S. Steidl: Towards more reality in the recognition of emotional speech, In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, Vol. 4, pp.941–944 (2007)
- [10] D. G. Aliaga, Y. H. Yeung, A. Law, B. Sajadi and A. Majumder: Fast high-resolution appearance editing using superimposed projections, *ACM Transactions on Graphics*, Vol.31, No.2, pp.13:1–13:13 (2012)
- [11] T. Amano: Projection based real-time material appearance manipulation, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'13)*, pp.918–923 (2013)
- [12] A. Lee, K. Oura and K. Tokuda: Mmdagent—a fully open-source toolkit for voice interaction systems, In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP'13)*, pp.8382–8385 (2013)
- [13] E. André, T. Rist and J. Müller: Integrating reactive and scripted behaviors in a life-like presentation agent, In *Proceedings of the international conference on Autonomous agents (AGENTS'98)*, pp.261–268 (1998)
- [14] T. Bickmore and J. Cassell: Relational agents: a model and implementation of building user trust, In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'01)*, pp.396–403 (2001)
- [15] S. Kiesler, A. Powers, S. R. Fussell and C. Torrey: Anthropomorphic interactions with a robot and robot-like agent, *Social Cognition*, Vol.26, No.2, pp.169–181 (2008)
- [16] M. Candea, I. Vasilescu and M. Adda-Decker: Inter-and intra-language acoustic analysis of autonomous fillers, In *Proceedings of Disfluency in Spontaneous Speech Workshop (DiSS'05)*, pp.47–52 (2005)
- [17] 水上悦雄, 山下耕二: 対話におけるフィラーの発話権保持機能の検証, *認知科学*, Vol.14, No.4, pp.588–603 (2007)
- [18] H. H. Clark and J. E. F. Tree: Using uh and um in spontaneous speaking, *Cognition*, Vol.84, No.1, pp.73–111 (2002)
- [19] T. Kawabe, T. Fukiage, M. Sawayama and S. Nishida: Deformation lamps: a projection technique to make static objects perceptually dynamic, *ACM Transactions on Applied Perception*, Vol.13, No.2, pp.10:1–10:17 (2016)