写真を媒介とする人とロボットとの三項関係に基づくコミュニケーションの研究

Research on photo-mediated communication based on the triadic relationship between humans and robots

大内　直[1]*　　西村　駿[1]　　長谷川　孔明[1]　　岡田　美智男[1]
Nao Ouchi[1], Shun Nishimura[1], Komei Hasegawa[1] and Michio Okada[1]

[1] 豊橋技術科学大学 情報・知能工学系
[1] Department of Computer Science and Engineering, Toyohashi University of Technology

Abstract: 人とロボットとのコミュニケーションデザインでは，その間を媒介する話題などの第三項の選択が大切な要素となる．本研究では，マルチモーダル生成 AI を援用し，人と複数のロボット (Muu) との間で写真や絵を媒介とするコミュニケーションの可能性を探ってきた．本稿ではロボット (Muu) たちと人とがゆるく依存しあいながら，コンヴィヴィアルな会話を生み出すための方法論について議論する．

1　はじめに

テレビをみながら食事をし，ついついテレビに話しかけてしまう．その後，なんだか恥ずかしくなる．そしてふと，家族でテレビをみながら，わいわいと話していた頃を思い出す… こうした際に，同じ写真をみながら自分と一緒に話してくれる存在がいてくれたらどうだろうか．少し寂しさが和らぐかもしれない．そして，そんな存在が複数いたら愉しくなるかもしれない…

リビングでよたよたとしたクリーチャーたちがバイクの写真の前で何かお話をしている．よく聞き耳を立てると，なにやら写真に写っているバイクについてのお話をしているようだ．ふむふむとお話している内容を聞いていると，クリーチャーたちに「これすきなの？」と問いかけられた．恐る恐る「すきだよ」と返答すると，「そうなんだ」，「バイクかっこいいよね」，「ぼくもすきだな」と返答がきた．そしてまた「どこにいきたい？」と問いかけられた．また恐る恐る「海に行きたいな」と返答した．すると「ほー」，「うみいいよね」，「おさかないるかな」と返答がきた．今度は「おさかなすきなの？」と問いかけられた．このようなやり取りを何度かしているうちになんだか愉しい気分になってきた．次はどんな写真でお話ししようかな．

このごろ，OpenAI 社の ChatGPT[1] をはじめとする生成 AI の急速な進化に伴い，自然な文章が容易に作成できるようになった．そのような生成 AI を会話エージェントに活用することにより，人とエージェントの間で自然な会話が構築できることが期待されている．しかし，そのまま活用した場合には，スマートホームやスマートフォンに搭載されている会話アシスタントのような事務的な会話が多く生まれてしまうため，愉しさにはつながりにくい．そこで筆者らは，写真をみながらロボットたちと雑談をおこなうという形式にすることで愉しさ，つまりはコンヴィヴィアルな関わりにつながるのではないかと考え，新たな会話システム〈Muu^{++}〉を構築した (図 1)．

本稿では，日本語での会話の場について整理しつつ，写真を見ながら共にコミュニケーションをおこなうプロトタイプ〈Muu^{++}〉のコンセプトと実装手法について述べる．

図 1: 〈Muu^{++}〉

*連絡先： 豊橋技術科学大学　情報・知能工学系
〒 441-8154 愛知県豊橋市天伯町雲雀ケ丘 1-1
E-mail:ouchi.nao.jg@tut.jp

2　研究背景

日本語での会話の場を構築する上での議論を整理しつつ，本研究の位置付けについて述べる．

2.1　人同士の会話における様相

人同士で雑談をおこなう際に，どのような情景がみられるだろうか．人同士の日々の雑談のような，共感しあい相手と心を通わせることを重視した会話形態として「ラポールトーク」がある [2]．このような共感しあうことを志向する会話形態では，話し手の発話を受け止めつつ，話し手からさらに話を引き出すような傾聴的な姿勢が伺える．デボラ・タネンらは，このような「ラポールトーク」を構成する 9 つのグラウンディングのパターンとその効果を表 1 のように整理した [3]．

表 1: ラポールトークを構成する 9 つのグラウンディングのパターン

Shadowing	直後に同じ発話を続ける
Echoing	相手の発話を繰り返して自分たちに向けて発話する
Expanding	相手の発話に同意しつつ確認する
Ratifying	相手の発話の一部を繰り返して承認する
Self-repetition	自分で自らの発話を繰り返す
Tense changed	時制を変えて同じ発話を繰り返す
Rhyme	相手の発話の韻を踏む
Savouring	キーワードを繰り返す
Participate	相手の発話に対し、短い質問を繰り返す

また,「ラポールトーク」に関連した会話形態のうち，日本人によくみられるものとして「共話」がある.「共話」とは，ひとつの発話を必ずしもひとりの話し手が完結させるのではなく，話し手と聞き手で作っていくというものである [4]．ひとつの発話で完結させないことで，相手の理解にゆだねる余地を残すことができる．そうすることで，ひとつの発話を相手と一緒になって作り上げるというような，相補的な会話の場が構築される．例えば，普段の雑談では次のような会話が度々みられる.「今日寝不足なんだよね」,「ゲームしてたんでしょどうせ」,「そうそうきいてよ」．このように「共話」は，相手と分かり合っているといった安心感を与えるような機能を持つ．

このような「ラポールトーク」や「共話」がおこなわれる雑談の会話の場においては，コンヴィヴィアルな関わりがうまれると考える．コンヴィヴィアル (Convivial) とはイヴァン・イリイチによって着目された概念であり，自立共生的な社会関係において個人の自由や自己決定を基本とした概念である [5]．本稿においては，会話の場の参加者同士の主体性が入り交じることで参加者全員が生き生きとした状態であることをコンヴィヴィアルな関わりとする．

このように，相手と心を通わせ，和気あいあいとした関わりを志向した会話形態を，人は何気なく実現している．しかし，人とロボットの間では途端に難しくなる．それは何故だろう．

2.2　人とロボットの間で実現するためには

人同士の会話でみられるような，コンヴィヴィアルな関わりを志向した会話形態を，人とロボットの間で実現するためにはどのようにすれば良いか．コンヴィヴィアルな関わりを実現するためには，話題の共有は必須である．会話の場において，話者が見当違いのことを言った途端に会話の場が壊れてしまう．人同士であれば，会話の話題を適切に追いかけることができるが，人とロボットではどうにもうまくいかない．それならば，会話の話題を第三項として，人とロボットの会話の場に置いてみてはどうだろうか．そうすることで，話題を明示することができ，人とロボットの間でのコンヴィヴィアルな関わりを志向した会話の場の実現に一歩近づくのではないだろうか．

実際に，対面的な言葉のみのコミュニケーションよりもむしろ，積み木のような第三項を介した共同行為の方が，子どもとロボットとのコミュニケーションの可能性を開きやすいことが示唆されている [6]．また，ショッピングモールにおいてのロボットが，単純な挨拶から対話をはじめるよりも，ショッピングモール内の店舗を第三項に客引きをおこなうことから対話をはじめるほうが有効であることも示唆されている [7]．

このように，人とロボットの会話の場に第三項を置き，話題を共有した会話をおこなうことは，従来の言葉のみのコミュニケーションよりも有効であると考えられる．しかし，第三項で話題を共有しただけで，コンヴィヴィアルな関わりを志向した会話の場を構築することはできるのだろうか．人とロボットの間でのコンヴィヴィアルな関わりを志向した会話の場の構築には，まだ何かが足りない．

2.3　コンヴィヴィアルな関わりを志向した会話をするために必要なことは

人とロボットの間でコンヴィヴィアルな関わりを志向した会話をするために足りないものは何だろうか．それは人とロボットの間のゆるいつながり感なのではないだろうか．一対一での会話は，発話義務や応答責任の強い会話となりがちである．一方，複数人での会話ではどうだろうか．複数人での会話では，会話の参加者に話し手と聞き手以外の役割が出現する [10][11][12]．会話への「参与を承認された者」(ratified participant) と「承認されていない者」(unratified participant) とに分

かれ，話し手の発話の宛先となる「受け手」(addressee) と宛先とならない「傍参与者」(side-participant)，その存在が会話参与者に認識されている「傍観者」(overhearer）と存在自体が認識されていない「盗み聞き者」(eavesdropper）にそれぞれ分化する．多人数会話に出現するこれらの役割は，社会学における会話研究では「参与役割」(participation status）と呼ばれる [11]．このように，複数人での会話の場は，参与役割が多岐にわたることにより，発話者たちがゆるくつながり，発話義務や応答責任の低減されたコンヴィヴィアルな関わりを志向した会話の場の構築につながると考えられる．そして，そうした複数人での会話の場においては，多くの情報を伝えるような発話を避けることが有用であると示唆されている [14]．これらより，複数の幼児がわいわいとお話するように，複数のロボットたちがたどたどしくシンプルな発話をおこなうことが，人の興味を惹き，人とロボットたちの間でのコンヴィヴィアルな関わりを志向した会話の場の構築につながるものと考えられる．

そこで，本研究では，人とロボットの一対一の会話ではなく，人とロボットたちとの会話に着目する．さらに，人とロボットたちの間に第三項をおき,「ラポールトーク」や「共話」と呼ばれるような相補的な会話の場を構築できないかと考え，OpenAI 社の ChatGPT[1] などの生成 AI を活用し，システムを構築した．

3　会話システム〈Muu^{++}〉

本システムではプラットフォームとして,〈Muu〉を用いた．その外観を図 2 に示す.〈Muu〉は，目がひとつであること，丸みのある身体，シンプルな内部構造，といったミニマルデザインをもとに設計されている [8]．また，Google 社の Speech-to-Text AI (音声認識 API)[9] を用いた音声認識が可能であり，ATR-Promotions 製音声合成エンジン Wizard Voice SDK を使用することで，子どものような声で発話をすることができる．そして，クリーチャーのような見た目をして，子どものような声で発話をすることができる〈Muu〉は，どのような能力か期待させすぎることがない [13]．本研究では，このような特徴を備えた〈Muu〉三体と第三項である写真選択用のタブレット PC を用い，人とロボットたちの会話の場の構築をおこなった．この会話システムを〈Muu^{++}〉とよんでいる．

3.1　ハードウェア構成

〈Muu〉のハードウェア構成を図 3 に示す.〈Muu〉は，2 つのサーボモータとカメラを備えた頭部，PC とスピーカとマイクを備えた土台，頭部と土台を繋ぐばねで構成されている．2 つのサーボモータ，カメラ，スピーカ，マイクは PC によって制御されている．そして，頭部と土台を繋いだばねは，よたよたとしたやわらかい動きを実現している．クリーチャーのような見た目の〈Muu〉がよたよたとしたやわらかい動きをすることによって，傍で会話を聞いている人の興味を惹くことにつながる．

図 2: プラットフォーム〈Muu〉

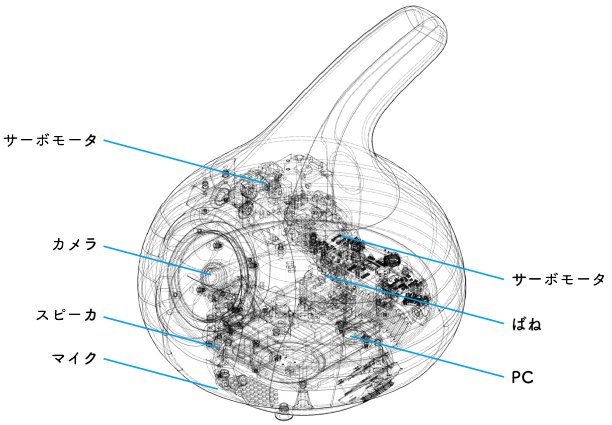

図 3:　〈Muu〉のハードウェア構成

3.2　ソフトウェア構成

本システムは，ウェブアプリケーションと〈Muu〉から構成されている.〈Muu^{++}〉のソフトウェア構成を図 4 に示す.

3.2.1　ウェブアプリケーション

ウェブアプリケーションの画面を図 5 に示す．ウェブアプリケーションは flutter[15] で構成されており，

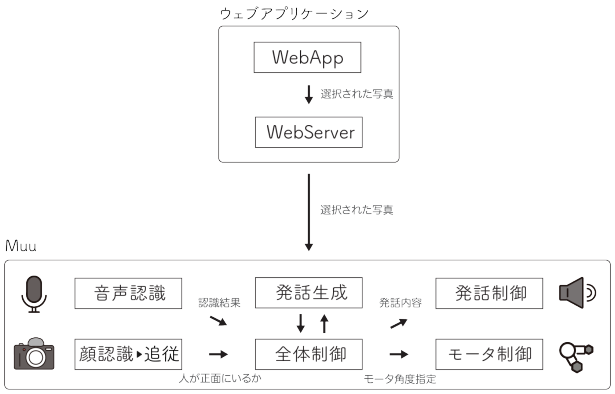

図 4: 〈Muu^{++}〉のソフトウェア構成

〈Muu〉のノードへの通信には Robot Operating System 2 (ROS2)[16] を用いている．表示される写真の中から一つを選択することで,〈Muu〉に画像情報を送信する．写真選択画面はスクロールが可能になっており,画面表示外の写真も選択することが可能である．写真選択後の画面を図 6 に示す．画面に映る写真について〈Muu〉たちは会話をおこなう．

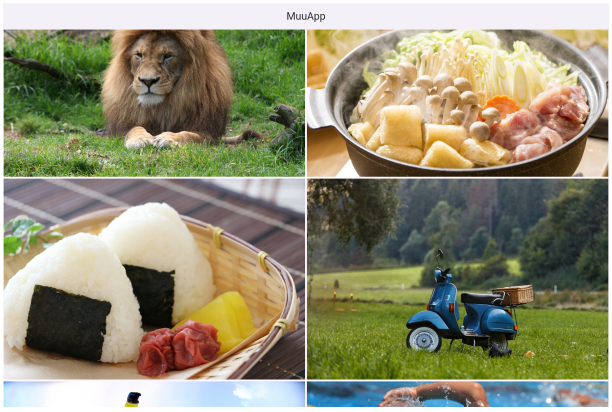

図 5: ウェブアプリケーション画面

3.2.2 〈Muu〉

〈Muu〉は各ノード間の通信に，Robot Operating System 2 (ROS2)[16] を用いている．全体制御ノード,音声認識ノード，顔認識ノード，顔追従ノード，発話生成ノード，発話制御ノード，モータ制御ノードがある．全体制御ノードでは，音声認識やふるまい，発話などのタイミングを管理している．音声認識ノードでは,Google 社の Speech-to-Text AI (音声認識 API)[9] を用いた音声認識が可能である．また，周囲の音圧を計測することで認識可能な閾値を設定している．その後，音声認識をおこなった結果を全体制御ノードに送信している．顔認識ノードでは，カメラを入力として OpenCV により人の顔を認識し，認識結果を全体制御ノードと顔追従ノードに送信している．顔追従ノードでは，顔認識ノードから受け取った情報をもとに人の顔を追従するための情報を生成し，モータ制御ノードに送信している．発話生成ノードでは，ウェブアプリケーションから受け取った画像情報や全体制御ノードで精査した音声認識結果をもとに，OpenAI 社の ChatGPT[1] を用いて，発話を生成している．生成された発話内容を全体制御ノードに送信している．発話制御ノードでは，全体制御ノードから受け取った発話内容を Wizard Voice SDK のサーバを介して発話をおこなっている．モータ制御ノードでは，全体制御ノードから受け取ったふるまい情報をサーボモータのサーバを介してモータを動作させている．

図 6: 写真選択後の画面

4 インタラクションデザイン

〈Muu^{++}〉のインタラクションのイメージを図 7 に示す．

人がタブレット上の写真を選択することで会話が始まり,〈Muu〉同士が写真についてラポールトークや共話をすることで人の興味を引き出し，人に関わる余地を持たせる．そして，人に質問をすることで会話の場に引き込み，参与を促す．このように，写真という第三項を話題に会話ができるようにデザインしている．また,〈Muu〉三体で多人数の会話の場を構築することで,わいわいとコンヴィヴィアルな雰囲気へと繋がるようにしている．

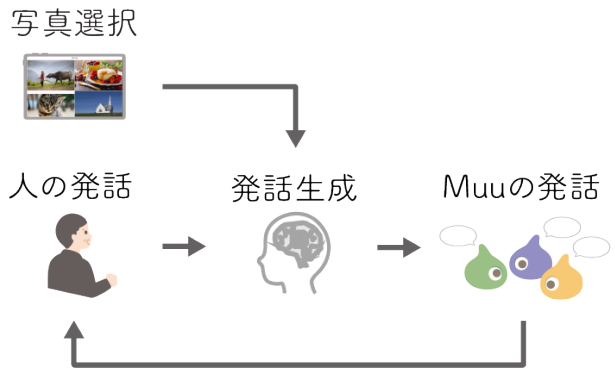

図 7: インタラクションデザイン

5　おわりに

本稿では，日本語での会話の場を構成する上での議論に触れ，人とロボットたちの間でのコミュニケーションをおこなうためのプロトタイプを提案した．本稿で提案したように，写真という第三項を置き，複数体のロボットと会話の場を構築することで，コンヴィヴィアルな関わりを志向した会話ができるのではないだろうか．今後は，実証実験の結果から会話分析などをおこなうことで，実際にこのようなアプローチが有効であるのかを明らかにしてきたい．

謝辞

本研究の一部は，愛知県が公益財団法人科学技術交流財団に委託し実施している「知の拠点あいち重点研究プロジェクト第 4 期（第 4 次産業革命をもたらすデジタル・トランスメーション（DX）の加速）」によりおこなわれた．ここに記して感謝の意を示す．

参考文献

[1] “Models，OpenAI Platform”: [Online]. Available: https://platform.openai.com/docs/models, [Accessed 2.11.2025].

[2] Deborah Tannen: You Just Don ’ t Understand: Women and Men Conversation，New York Quill, (2001).

[3] Deborah Tannen: Talking voices: Repetition，dialogue，and imagery in conversational discourse, Vol. 26，Cambridge University Press，(2007).

[4] 水谷信子:「共話」から「対話」へ，日本語学，Vol. 12，No. 4，pp. 4-10，(1993).

[5] イヴァン・イリイチ (著)：渡辺京二，渡辺梨佐 (訳); 『コンヴィヴィアリティのための道具』; ちくま学芸文庫，(2015).

[6] 礪波朋子，藤井洋之，岡田美智男，麻生武: 子どもとロボットとのコミュニケーション成立の考察 - モノを媒介とした共同行為，ヒューマンインタフェース学会論文誌，Vol. 7，No. 1，pp. 141-148，(2005).

[7] 佐竹聡，神田崇行，Dylan F. Glas，今井倫太，石黒浩，萩田紀博： 対話ロボットの人間へのアプローチ方法 - 対話ロボットの対話開始に対する戦略 - ，日本ロボット学会誌，Vol. 28，No. 3，pp. 327-337，(2010).

[8] 岡田美智男，松本信義，塩瀬隆之，藤井洋之，李銘義，三嶋博之： ロボットとのコミュニケーションにおけるミニマルデザイン，ヒューマンインタフェース学会論文誌，Vol. 7，No. 2，pp. 189-197，(2005).

[9] “Speech-to-Text，Google Cloud”: [Online]. Available: https://cloud.google.com/speech-to-text，[Accessed 2.12.2025].

[10] Clark，H. H. : Using language. Cambridge，UK: Cambridge University Press，(1996).

[11] Goffman，E. : Forms of talk. Philadelphia: University of Pennsylvania Press，(1981).

[12] 小磯 花絵: 坊農 真弓・高梨 克也 (編)『多人数インタラクションの分析手法』，認知科学，Vol. 17，No. 2，pp. 377-379，(2010).

[13] 西脇裕作，板敷尚，岡田美智男: ロボットの言葉足らずな発話が生み出す協調的インタラクションについて，ヒューマンインタフェース学会論文誌，Vol. 21，No. 1，pp. 1-12，(2019).

[14] 西脇裕作，大島直樹，岡田美智男： 多人数会話を構成するロボットの言葉足らずな発話が人の会話への参加態度に及ぼす影響，人工知能学会論文誌，Vol. 36，No. 2，pp. B-K44-1-12，(2021).

[15] “Flutter”: [Online]. Available: https://flutter.dev，[Accessed 2.12.2025].

[16] “Humble，ROS2 Documentation”: [Online]. Available: https://docs.ros.org/en/humble/index.html，[Accessed 2.12.2025].