

Pocketable-Bones: 絵本をながめながら 会話を愉しむクリーチャ

Pocketable-Bones: A Creature That Enjoys Conversation With a Child While Looking at Picture Books

松浦 涼太^{1*} 西村 駿¹ 天野 僚¹ 長谷川 孔明¹ 岡田 美智男¹

Ryota Matsuura¹, Shun Nishimura¹, Ryo Amano¹, Komei Hasegawa¹ and Michio Okada¹

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: 親子で一緒に絵本を眺めおしゃべりを愉しむように、子どもとロボットとが目の前の絵本を読むことができれば、どんなに嬉しいことだろう。本研究では、ロボットと子どものコミュニケーションにおける絵本を媒介とした三項関係に着目し、マルチモーダル生成 AI を援用した実現手法を探っている。本発表では、子どもと一緒に絵本を愉しむ〈Pocketable-Bones〉のコンセプトとそのプロトタイプを紹介する。

1 はじめに

夜、親が絵本を手に取り、子どもと一緒にソファへ腰を下ろす。ページをめくるたびに、物語の世界が広がり、子どもは目を輝かせながら登場人物の冒険に心を躍らせる。「これはどうして起こったの?」「次はどうなるの?」という子どもの問いかけに、親は優しく答えたり、共に考えたりする。その瞬間、絵本は単なる紙の上の物語ではなく、親子が会話を通じて共に創り上げる特別な体験となる。このような時間は、親子の絆を深めるだけでなく、子どもの想像力や好奇心を育む貴重なひとときである。

本研究では、このような親子の絵本体験に触発され、人とロボットが視線を重ねながら絵本を読み進め、お互いに愉しく会話ができる仕組みを実現したいと考えている。そこで、胸ポケットに収まる大きさのモバイル型ロボット〈Pocketable-Bones〉をベースに目の前の絵本について愉しく会話する仕組みを構築した。

本稿では、構築したシステムと、そのインタラクションの詳細について述べる。

2 研究背景

2.1 絵本と親子関係

親子のコミュニケーションといえば何が思い浮かぶだろう。おもちゃで一緒に遊んだり、学校で起きたことを話したり、多様な形でコミュニケーションが考えられる。その中でも、絵本遊びを介したコミュニケーションは養育者の子どもに対する働きかけについて、他の遊びよりも有利であることが指摘されている [1]。

絵本遊びを遂行するにあたって、養育者は子どもに合わせて自らの行動を調整する必要がある。これは、養育者の働きかけが子どもの意図を無視して養育者本位で行われた場合、子どもの興味はたちまち減退し、絵本遊びが終了してしまうからである。養育者はこのような行動調節をするために常に子どもに気を配るようになるという [1]。

では、人とロボット間で、一緒に絵本を読み愉しむ関係を構成してみたらどうだろう (図 1)。人とロボットがお互いに相手の様子に気を配りながら自身の行動を調節するような関係が築くためのヒントが得られるかもしれない。

そこで本研究では、人とロボットと一緒に絵本を愉しむ仕組みの構築を試みた。

*連絡先: 豊橋技術科学大学 情報・知能工学系
〒441-8122 愛知県豊橋市天伯町雲雀ヶ丘 1-1
E-mail: matsuura.ryota.mm@tut.jp

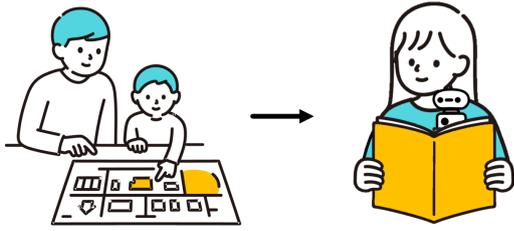


図 1: 人とロボットと一緒に絵本を読むイメージ

2.2 マルチモーダル生成 AI

ロボットの発話生成や人とのインタラクションに利用できる技術の一つとして、マルチモーダル生成 AI がある。マルチモーダル生成 AI は、人からの問いかけ（入力テキスト）に回答するだけでなく、目の前にある写真や絵、テキストなどの視覚的・言語的情報を組み合わせて、それに基づいた発話（テキスト）を生成する点において優れている。これにより、人とロボットがその場の状況や絵本の内容を共有しながら、自然な会話を展開できる可能性が高まると考えられる。

従来のシステムでは、ロボットの発話は事前に定義されたスクリプトに依存することが多く、柔軟な対応が難しい場合がある [2]。一方で、マルチモーダル生成 AI の導入により、ロボットは絵本の絵やテキストを読み取り、それを基にした独自の解釈や応答を生成することができる。こうした特性から、人の問いかけに応じるだけでなく、ロボット自身が新しい視点や意見を提案する余地を生み出し、会話に深みを与えることが期待できる。

この技術の活用は、人とロボットの間でより豊かな対話体験を創出するための基盤となるかもしれない。たとえば、ロボットが絵本の中のキャラクターや場面について意見を述べたり、人の反応に応じて適切なコメントを加えたりすることで、会話はより双方向的かつ動的なものになる。このような発話生成の質の向上により、ロボットは単なるツール以上の存在として、人より深い相互作用を築くことができるのではないだろうか。

そこで、本システムの発話生成にマルチモーダル生成 AI を援用することとした。

2.3 Pocketable-Bones

本研究ではプラットフォームとして〈Pocketable-Bones〉 [3][4] を用いた (図 2)。この〈Pocketable-Bones〉を採用したのは、本システムのコンセプトでもある一緒に絵本を読むという部分において、〈Pocketable-Bones〉

が他のロボットよりも人に寄り添った振る舞いができるためである。

友人や兄弟と一緒に漫画を読んでいるところを想像してみたい。お互いに漫画を快適に読むために、本の正面から見ようとして、顔がくっつきそうなくらい近づいている。このように、人とロボットが絵本の正面に位置し身を寄せ合うようにすることで、人同士と一緒に絵本を読む態勢と近い状況が作れると考え、〈Pocketable-Bones〉の胸ポケットに収まるデザインは、まさにこの条件に沿ったものである。

このような、人とロボットが同じ方向を向いて行うインタラクションは、会話における〈並ぶ関係〉 [5] にも繋がる場所がある。人とロボットとのかかわりにおいて、〈並ぶ関係〉でのコミュニケーションは、人とロボットが共生する社会を目指す中で重要な役割を果たすものと言われており [6]、そうした意味でも〈Pocketable-Bones〉を用いることで、人とロボットの楽しいインタラクションが構築できると考えられる。



図 2: Pocketable-Bones の外観

3 本研究の位置付け

前章で述べたように絵本は、親子のコミュニケーションにおいて、親の子どもに対する気配りを引き出す効果がある。本研究ではこの絵本が介入した関わりに注目し、人とロボットでも同様に一緒に絵本を読み、コミュニケーションをとる枠組みを構築した。この枠組みのインタラクションの詳細を調査することで、人とロボットの間でお互いに気を配りながらも、相手に強制されることもない楽しい会話を構築するためのアイデアが得られないかを探るものである。

4 研究プラットフォーム

4.1 ハードウェア〈Pocketable-Bones〉

〈Pocketable-Bones〉は胸ポケットに差し込むことのできる、ウェアラブルロボットである。〈Pocketable-Bones〉の外観と内部構造を図3に示す。〈Pocketable-Bones〉はArduino Nano 33 BLE, エンコーダ付きDCモータ, バッテリーが取り付けられている。Arduino Nano 33 BLEにより通信, モータ制御を行い, エンコーダ付きDCモータによって縦横2軸の自由度まで, 頭部の向きを動かすことができる。〈Pocketable-Bones〉の下部にはスマートフォンに取り付け可能な治具がついており, 〈Pocketable-Bones〉を取り付けたスマートフォンを胸ポケットに差し込むことができる。

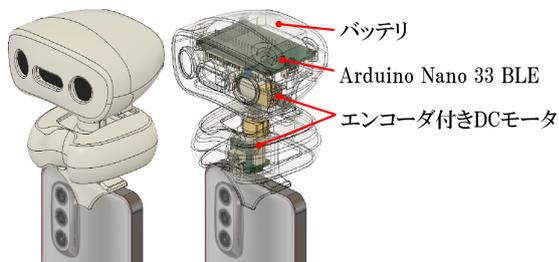


図3: 〈Pocketable-Bones〉の外観と内部構造

4.2 システム構成

本システムの構成を図4に示す。本システムはAndroidスマートフォンとミニPC (NUC), カメラ, 〈Pocketable-Bones〉で構成されており, 図4のスピーカーとマイクはスマートフォン内蔵のものを利用している。

本システムのインタラクションは人が〈Pocketable-Bones〉に話しかけることから始まる。システムはまず, マイクから音声認識を介して人からの発話内容 (発話テキスト) を受け取る。発話テキストを受け取ったタイミングで, カメラから絵本の画像を取り込み, マルチモーダル生成AIに人の発話テキストと絵本の画像を送信する。マルチモーダル生成AIにロボットの発話テキストを生成させ, システムを介して, 〈Pocketable-Bones〉の振る舞いと共に合成音声を用いてスピーカーで再生する。人は再生されたロボットの発話を聞いた後, 再びロボットに話しかけ, 以後, 交互に発話することで会話が展開されていく。

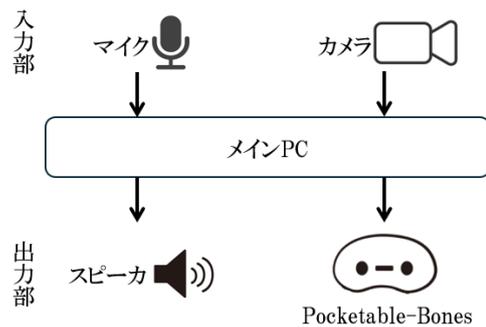


図4: システム構成

4.3 ソフトウェア構成

システムの実装には, ROS2 (Robot Operation System 2) を利用した。ROS2 の特徴として, センサーなどを制御するプログラムがノードという単位で独立している。そのため, 疎結合な設計や非同期処理などを容易におこなうことができる。

本システムのソフトウェア構成を図5に示す。本システムは複数のノードから構成されており, それぞれの役割として音声認識, 画像 (画像取り込み), マルチモーダル生成AI, 制御部 (システムの状態と遷移の管理), 振る舞い (モータ制御), 発話 (合成音声) が割り当てられている。

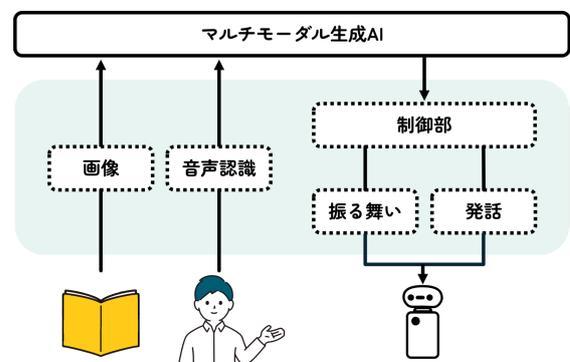


図5: ソフトウェア構成

4.3.1 制御部

本システムでは, インタラクションのシチュエーションごとに状態を設定しており, 各ノードは現在の状態をもとに動作する。制御部ではシステムの状態と状態遷移の管理を行う。図6にシステムの状態遷移図を示す。

各状態の遷移は, 人の発話からの入力テキスト (speech) と, マルチモーダル生成AIから受け取った発話テキス

トのリストによって分岐している。以下に各状態の説明を述べる。

reset はシステムの初期化を行う。この状態はシステム起動時か、人から「バイバイ」と入力を受けた場合に遷移する。

listen は人からの入力を待っている状態である。「バイバイ」以外の入力があれば call_gpt に遷移する。

call_gpt はマルチモーダル生成 AI に、入力テキストと画像データを送信して生成された発話テキストを受け取るまで待機する状態である。

speak は受け取った発話テキストを読み上げている状態である。テキストをすべて読み上げると end に遷移する。

end は次のプロセスのために、入力テキストと画像データを初期化する状態である。

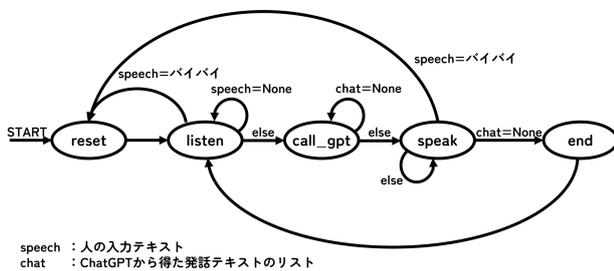


図 6: システムの状態遷移図

4.3.2 音声認識

音声認識ノードでは人の発話を認識し、call_gpt 状態時に、音声認識結果をマルチモーダル生成 AI に送信する。音声認識には Google 社の Google Cloud Speech To Text API を使用している。

4.3.3 画像

画像ノードでは、カメラから画像データを取り込み、マルチモーダル生成 AI に送信できるようにデータ加工を行ったうえで、上記の音声認識の結果とともに送信する。

4.3.4 マルチモーダル生成 AI

マルチモーダル生成 AI はテキストと画像データからテキストを生成し、結果をシステムに送信する。

発話生成には OpenAI 社の GPT-4o mini を使用している。

4.3.5 振る舞い

振る舞いノードは〈Pocketable-Bones〉の首のモータを制御する。制御部から送られてきた振る舞いの情報に基づいて、首を縦に動かしたり左右を向いたりするように DC モータを制御する。

4.3.6 発話

発話ノードは生成されたテキストを合成音声を用いてスピーカで再生する。合成音声ソフトには、ATR-Promotions 社の Wizard Voice を使用した。

4.4 インタクションデザイン

人と〈Pocketable-Bones〉が並び、目線を重ねながら目の前の絵本を読み進める。読み進めるなかで「この料理おいしそう!」、「食べてみたいなあ」と、互いに絵本から会話を広げる。このように絵本の内容を話し合うことで、お互いに相手の返答を意識しつつも縛られないような、楽しい会話の構成を目指す。



図 7: 〈Pocketable-Bones〉のインタクション

5 フィールドワーク

本研究で構築したシステムと子どもたちとの関わりを見るために実際にフィールドワークを行った。

5.1 概要

5.1.1 日時、場所

2024年8月20日(火)の10:00から15:00の時間帯に愛知県にある豊橋市立天伯小学校にてフィールドワークを実施した。

5.1.2 参加者

天伯夏休み限定児童クラブの小学1, 2年生の計22名に参加してもらった。

5.1.3 内容

「ぶたさんのおかいもの [7]」という絵本を用意し、〈Pocketable-Bones〉と一緒に読んでもらった。カメラ、絵本、〈Pocketable-Bones〉と人の位置関係を図8に示す。デスクライトにカメラが真下を向くように取り付け、カメラの画角に収まるようデスク上にまっすぐ絵本を置く。人と〈Pocketable-Bones〉はカメラを邪魔しないような位置から絵本を読むようにしてもらった。



図 8: フィールドワークでの位置関係

5.2 観察されたこと

フィールドワークの様子を図9, 10に示す。フィールドワークを始めてすぐは、インタラクション方法に慣れていないため、絵本を動かしてしまったり、逆に絵本に顔を近づけてカメラを遮ってしまったりなど、思うようにインタラクションをしてくれない様子が目立った。また、〈Pocketable-Bones〉の返答の遅れから、人が話すタイミングと被ってしまう場面も見受けられた。

しかし、〈Pocketable-Bones〉とのインタラクションに慣れてくると、〈Pocketable-Bones〉と絵本の内容について話し合う姿が見られた。また、あらかじめ用意した絵本の「ぶたさんのおかいもの」以外の本も持ってきていいことを伝えると、数人の子どもたちが次々に本棚から絵本を持ち出すようになり、積極的に〈Pocketable-Bones〉と会話を試みる様子が見られた。

以上のことから本システムは人とロボットが絵本を読みながら愉しく会話する枠組みとして機能していることがわかる。



図 9: フィールドワークでの様子 1



図 10: フィールドワークでの様子 2

5.3 考察

フィールドワークで観察されたことから、本システムが人とロボットが絵本を読みながら愉しく会話する枠組みとして機能している可能性がある。では具体的にどのような要素が関連して、人とロボットと一緒に絵本を愉しむインタラクションを生み出したのだろうか。

5.3.1 共働的なかわり

まず、人と〈Pocketable-Bones〉が絵本について話し合っていた様子から、人と〈Pocketable-Bones〉との間に共働的な関係が生まれていたのではないかと考えた。共働 (Conviviality) とは、「お互いに自立しつつも、ゆるく制限 (依存) し合うことで生まれる生き生きとした情動や喜び」のことで、Illich(1973) によって提唱された、自立共働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 [8][9]。

この共働の概念を今回の会話に当てはめて考えると、お互いに自分の話したい事を話しつつも、相手の話に耳を傾け、時には自身の意見が変わったりするような、お互いに緩やかに相手の影響を受けた会話が共働的と

言えるだろう。相手の話から自身の話が影響を受けるのであれば、発話がゆるく制限された会話を実現するためには、会話する両者間に共通の話題やテーマが存在しなければならない。

5.3.2 第三項

共通の話題やテーマといった会話者同士が共通で参照している物事を第三項と呼び、二者の間で何らかの物や事象が共有されている関係は三項関係と呼ばれている [10]。第三項は前述した並ぶ関係を構成するうえで重要な役割を担っており、本システムにおいて、絵本がその第三項の役割を果たしていたのではないかと考えられる。

6 おわりに

本研究では人とロボットが目の中の絵本について愉しく会話するための枠組みを、〈Pocketable-Bones〉をベースに、マルチモーダル生成 AI を援用して構築した。ここで構築したシステムでフィールドワークを行ったところ、人とロボット間で共働的な会話が構成されていることが示唆されたため、共働的な会話にどのような要素が関連しているのかを考察した。

今後の課題として、フィールドワークでカメラの位置や会話の間の長さからインタラクション方法に戸惑う様子が見られたことから、よりスムーズに会話できるようにシステムのブラッシュアップを行いたい。また今後の研究として、本システムを用いることで共働的な会話が構成されることについては実証実験を行い、共働的な会話に関連していると思われる要素を軸に、調査を行いたい。

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IV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)」により行われた。ここに記して感謝の意を示す。

参考文献

- [1] 佐藤鮎美:絵本遊びが親子関係に良い効果をもたらすのは本当か?, 日本赤ちゃん学会, Vol.16, pp.18-27 (2017)
- [2] 呉剣明, 楊博, 服部元, 滝嶋康弘:マルチモーダル認識技術を活用した対話 AI「KACTUS」, 情報処理学会, Vol.2021-AVM-112, No.7 (2021)
- [3] Naoki Ohshima, Katsuya Iwasaki, Ryosuke Mayumi, Komei Hasegawa, Michio Okada: Pocketable-Bones: Self-Augment Mobile Robot that Mediates Our Sociality, Journal of Robotics and Mechatronics, Vol.35, No.3, Rb35-3-12644 (2023)
- [4] 真弓凌輔, 長谷川勇輝, 岡田美智男: Pocketable-Bones: お互いの関心を共有しながら一緒に街のなかを歩くモバイルなロボット, Human-Agent Interaction シンポジウム 2018, P-28 (2018)
- [5] 真弓凌輔, 岡田美智男, 長谷川孔明: PoKeBo と生み出すインタラクション空間に関する研究, 豊橋技術科学大学 修士論文 (2021)
- [6] 長谷川孔明, 林直樹, 岡田美智男: マコにて: 並ぶ関係に基づく原初的コミュニケーションの研究, ヒューマンインタフェース学会論文誌, Vol.21, No.3, pp.279-292 (2019)
- [7] 杉田香利: 『ぶたさんのおかいもの』, ひかりのくに (2019)
- [8] Illich, I. *Tools for conviviality*. Harper & Row.(1973)
- [9] 荷方邦夫: 共働的活動における共働的な協働: プロジェクト型デザイン教育実践での社会的相互作用分析, 認知科学, Vol.31, No.3, pp.459-473 (2024)
- [10] 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の成立の考察-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)