

Talking-Ally: 聞き手性と宛名性に配慮した発話生成システムについて

Talking-Ally: Designing Utterance Generation in considering ‘hearership’ and ‘addressivity’

小田原 雄紀^{1*} 蔵田 洋平¹ 松下 仁美¹ 大島 直樹¹
P. Ravindra S. De Silva¹ 岡田 美智男¹

Yuki Odahara¹, Yohei Kurata¹, Hitomi Matsushita¹, Naoki Ohshima¹,
P. Ravindra S. De Silva¹ and Michio Okada¹

¹ 豊橋技術科学大学 情報・知能工学系

¹Department of Computer Science and Engineering,
Toyohashi University of Technology

Abstract: In this study we developed Talking-Ally by integrating novel mechanism for robot's eyes which are behaving alike to human's eyes. Through this novel mechanism, our robot is capable to enhance the persuasive power of communication through natural eye-gaze behaviors by incorporating with proposed utterance generation approach. However, Talking-Ally is producing utterance (toward addressivity) by utilizing the state of the hearer's resources (eye-gaze behaviors) to persuade the user through the dynamic interactions.

1 はじめに

私たちは大地の上を歩くと同時に、大地が私たちを歩かせている。一度、そうした視点から私たちの身の回りのことを見渡してみるのも面白い。街の中を何気なく歩く時に、「街を歩いているのは、この私だ」と自分を中心に考えてしまいがちだけれど、その歩道や街の看板、駅に向かう人の流れは、私たちの歩くという行為を制約し方向づける。その意味で、なにげない一歩とそれを支える大地や街並みとが一緒になって「歩行」という行為を生み出しているといえる [1]。

これは「歩く」という行為に限られない。人前で何かを話そうとするとき、その相手が無表情なときには、とても話し難いということがある。なにげなく言葉を繰り出そうとするとき、その相手は私たちの発話が向かう対象であると同時に、私たちの発話内容を制約し、方向付ける。つまり、その聞き手のうなずきや視線の動きに支えられながら、一緒になって発話を生み出しているといえる [2]。

一方で、これまでの音声合成システムの場合はどう



図 1: Talking-Ally

だろう。スイッチをオンにすると、そのスピーカーからは確かに合成音は聞こえてくる。しかし、その合成音は「宛名」を伴ったものでない。その発話は聞き手の存在を予定したものでも、その状態に配慮したものでもない。私たちがテキストの読み上げソフトを使用する際には、そうした「宛名のない合成音」の聴取を一方的に強いられてきたともいえる。

本研究では、私たちの発話はむしろ聞き手との相互

*連絡先：豊橋技術科学大学 情報・知能工学系
〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1
E-mail: odahara@icd.cs.tut.ac.jp

行為的調整の結果として組織されたものと捉え直す。これらの観点から、聞き手の存在を予定しつつ、その聞き手との相互行為的調整に基づいて発話の組織化を行う新たな発話生成システム「Talking-Ally」の構築を行い、実験・分析を進めてきた。その結果では、聞き手に配慮し調整的に行われる発話は、被験者の関心を素早く引き戻したり、注目・興味を維持することや、アンケートによる印象評価により、聞き手の状態に配慮しない機械的な発話生成システムと比べて Talking-Ally に生き物らしさを感じさせる可能性を示した [3]。一方で、その身体的な制約から、身体による志向性の表示が弱いことが伺えた。

これらの結果をうけて、視線表出機構や聞き手の位置検出機構など志向を表示する要素を新たに備えたプラットフォームの構築を行った。本稿では、本研究の背景を整理するとともに、プラットフォームの概要、インタラクションデザインについて述べる。発表では、現在計画している実験の概要と、実施し得られた分析結果を、ここでの議論と合わせて紹介する。

2 研究の背景

2.1 「協応関係」

私が何かを行うとき、その行為を自身のみで生み出したものと捉えやすい。それは、私たちの身体が個として閉じている系であるかのように観察がされやすい [4]。ことや、私たちが「主体的に行動しているモノ」を見たとき、そのモノが「自発的に、合目的に意思決定をしている」と解釈してしまう、「志向的な構え」 [5] をとってしまうことによる。

このような、行為主体の内部メカニズムに完全に依存し、その行動は行為主体から一方的に生み出されているという考え方は「個体能力主義的」と呼ばれている。これまでのロボット研究や、発話生成システム開発は、この「個体能力主義的」な考えに基づいて、構築される傾向が強かった。これらの内部を作り込まれたロボットやシステムは、想定された限定的な環境空間において有用であるが、生活空間の多様な環境で活動するには十分とはいえない。

一方で、私たちの行為は、環境と行為主体が相互に調整・協応しながら、その結果として顕在化するという見方もある [6]。例えば、「歩く」という行為は、自身の何気ない一歩と大地の支えによって達成されており、また、その行き先は街の看板や歩道、駅に向かう人の流れ等に左右されながら行われているだろう。

生態心理学的な見方によれば、環境と行為主体は互いに協応しながら、一つのシステムを作り、行為を生み出すという [2]。そして、この環境と行為主体で形成

された「協応関係」こそが、実世界の多様な変化に対して、柔軟で頑健なシステムに繋がるという。

また、この協応関係を会話領域に展開すると新たな視点で会話における発話現象について再考することができる。「あの一」、「えっと」、「それでね」といった言い淀みや言い直しなどの、発話の「非流暢性 (disfluencies)」は、その要因を発話のプランと行為との乖離と、その修復プロセスとして、「話し手」である発話者の内部のメカニズムのみに帰属させて説明されることが多かった [7]。発話の非流暢性を話し手と聞き手による相互の調整・協応により、創出されるものと捉え直した時、非流暢性は会話の場において、環境 (他者) と行為主体の間でその時々にかかる変化に応じた相互の「柔軟な調整過程」が顕在化したものだといえる [8]。

本研究では、言語的・身体的行為は、環境 (他者) と行為主体が協応関係を形成することにより、その柔軟な調整過程を経て生み出されていると考えられる。

2.2 「聞き手性 (hearership)」

私たちがいま「話し手」であるのは、その発話を向けた相手が「聞き手」となってくれているという単純な事実を支えられている。

会話に参加している聞き手は、視線、うなずきなどを用い、「いま、あなたの話を聞いていますよ」という「聞き手性 (hearership)」の表示を行う。その表示により、自身が話し手の発話に応じる聞き手であるということを示すと同時に、話し手の発話に関する興味や理解度など様々な聞き手の状態をリソースとして話し手に提供する。

Goodwin は、「話し手は発話の冒頭で聞き手の視線を獲得できていないならば発話を一時中断し、聞き手の視線が得られた時点で発話を再開する」という振る舞いを観察できるとし、聞き手が話し手を見るという行為は会話に参加しようとする態度 (engagement) を示す一つの方法であり、もし参加者が互いに視線を合わさないでいる場合は、一時的に会話に参加しようとしなかった態度 (disengagement) の表示になるという [9]。

また、メイナードは会話における頭部の動きには会話の調子を調整するなどの会話管理の側面があるとし、聞き手のうなずきは、「話の内容を理解したので、話を続けて」という、話し手へメッセージを伝える役割を持つとする [10]。これらから、「聞き手性」の表示を捉えることは次話者の選択や、聞き手の参加態度や話題に対する興味の度合いを知ることなどの手掛かりになると考えられる。

すなわち「聞き手性」の表示をリソースとして、発話を構築することで、聞き手の状態を反映させた、相互調整的な発話を組織化することが可能になると考えられる。

2.3 「宛名性 (addressivity)」

バフチンは、「言葉というのは、本来は誰かに向けられたもので、必ず宛名を伴ったものだ」とし [11]、発話を主体的に行う話者は、必ず「いまこの発話を聞いてほしい相手」へと宛先を向け、その相手との関係や状況に合わせて適切に発話を構築するという [12]。

この指摘から、発話には、必ず「宛名」が伴っているべきだと考えられる。これまでの発話生成システムは、その状況や聞き手が異なる場合でも、同じ発話内容を機械的に繰り返すものが多かった。このような発話は宛名性を欠いており、私たちは応答責任を感じ得ない。聞き手へ向けた発話を行うには、聞き手の様子に配慮しながら、その状況を把握し発話を組織する必要がある。聞き手に向けた発話を行うことで、聞き手のあいづちなどの発話に対する支えを引き出せると考えられる。その意味で、聞き手に発話の組織化への参加余地を提供することに繋がる。

また、聞き手の参加余地を残し、相互的に発話を組織化することで、バフチンのいう「対話性 (dialogicality)」を考慮できると考える。対話性とは、「発話は、発話連鎖の一環の中で行われ、決して独立して行われるものではない。話者の声は必ず、相手の声を内包する形で行われる。」というものである。つまり、発話はボトムアップ的に相互調整的に構築され则认为られ、発話を構築する際には、聞き手の反応とそのシチュエーションが重要であるといえる。

2.3.1 話者の「視線」

発話において、その音声的な側面に限らず、それに伴う視線や頭部の動き、身振り手振りなど身体的な振る舞いも重要な意味を持つ。このことから、発話を構築する上で、発話における身体的な側面について十分に考慮する必要がある。特に視線は、強い志向性を示すことができることから、聞き手を見ることで、聞き手の領きを引き出す [13]、話者交替のサインとして機能する [14]、また話し手と聞き手の共同注視により、会話が円滑に進行していることを示す手掛かりになり得る [15] など、会話において重要な役割を担う。

本研究では、視線の志向性に着目し、聞き手に対する宛名の手掛かりとして、「Talking-Ally」に視線表出機構を実装した。

3 プラットフォームのデザイン

3.1 Talking-Ally の構築

2章での議論を考慮し、発話生成システムのプラットフォームである「Talking-Ally」を構築した。Talking-

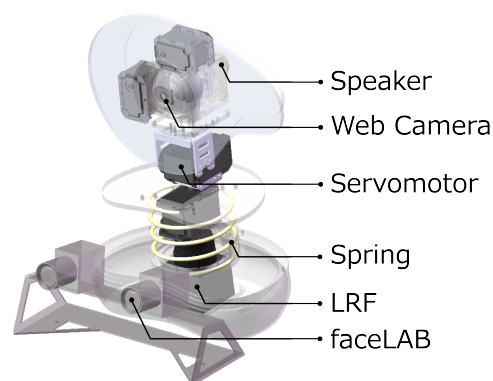


図 2: Talking-Ally のハードウェア構成

Ally は、より強い志向を表示できるようデザインされている。また、ハードウェア構成は図 2 のようになっており、システム構成の概要を図 3 に示す。

Talking-Ally は以下の特徴を持つ。

3.1.1 丸みを持ったデザイン

乳幼児や小動物など一般的に可愛いと言われるものは、丸みを持ったデザイン上に成り立つことが多く、そのような存在に対して、私たちは親しみを感じたり、自身から積極的に関わりを持とうとする。Talking-Ally を丸みを持ったデザインにすることにより、コミュニケーションを取る上で、馴染みやすく、積極的に関わりを引き出すことを狙いとしている。

3.1.2 おぼつかない動き

乳幼児のおぼつかない振る舞いを見ると、思わず手を差し伸べたくなったり、危険が及ばないように近くで見守ったりするなど、思わずその行為をアシストしてしまう傾向がある。Talking-Ally は下部に備えたバネにより、おぼつかない動きを表現することで、自身の動きの不安定さを示し、結果として聞き手のアシストを積極的に引き出そうとする。

3.1.3 視線・頭部認識機構

聞き手の状態を把握する機構の一つとして、聞き手の「視線」を利用する。Goodwin は、聞き手の視線は会話への参加態度を示す手掛かりや、話し手に発話を続けるよう促す役割があるとする。このことから、「聞き手性」を示す一つの手掛かりとして、視線を捉えることは、現在の話題に対して興味があるのか、今まで

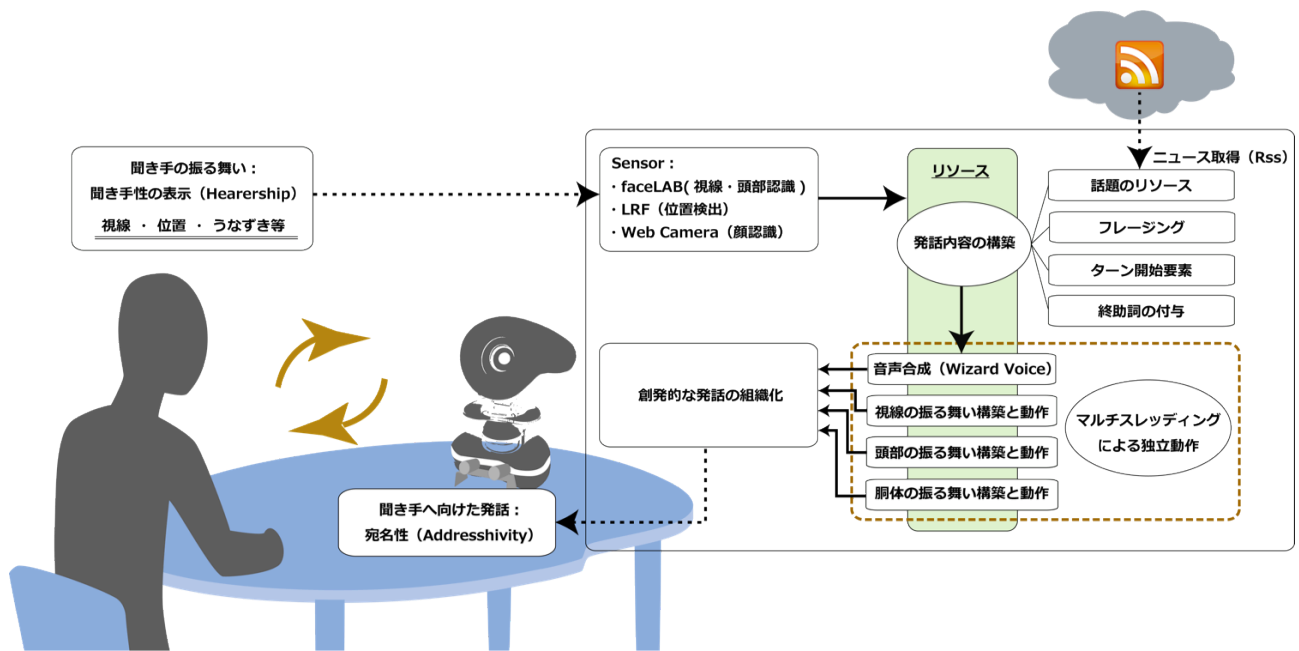


図 3: システム構成

の会話に対する理解度はどの程度かなど、会話を円滑に進める上で重要といえる。

また、会話におけるうなずきや、顔の向きによる志向の表示など、頭部の振り舞いや位置関係は会話の管理的な側面を持つことから、頭部の動きも有用な手掛かりとなる。

そこで、聞き手の視線・頭部を認識するシステムとして視線・頭部追跡システム faceLAB5.0（SeeingMachines 社）を利用する。聞き手の視線や顔の向きから得られる注目の度合いなど、状況に合わせて、リアルタイムに発話方略を選択することで、聞き手の関心を反映させた発話の組織化を行う。

3.1.4 対人距離・方向認識機構

2 章で議論したように、発話には必ず宛名が伴っている。宛名を伴わせるためには、聞き手に対するアドレスが重要であり、話し手と聞き手の身体的配置がこれを表現すると考えられる [16]。

Talking-Ally は Laser Range Finder (LRF) と呼ばれる測域センサを下部に備えており、Talking-Ally の前面を 0 度として左右に 90 度、計 180 度以内に入った人の位置を検出することが可能となっている。検出した聞き手との距離や、自身と聞き手の相対角度をリソースとし、Talking-Ally の身体を聞き手のいる方向へ調整することで、聞き手に対する動的なアドレス付けが可能である。

3.1.5 視線表出機構

Schefflen や Kendon の指摘によれば、身体は志向の表示に関して階層化されており、最も基本的な志向を表示する腰の前面方向から、上体の前面方向、手、指あるいは顔、目の順でより先鋭な志向を表示する [17][18]。これをうけて、Talking-Ally には視線表出機構を実装した。この機構は 2 つのサーボモーターで構築され、上下左右に視線を移動させることが可能であり、聞き手の顔をリアルタイムに追従できる。加えて、Web Camera を内蔵する特徴的な「大きな一つ目」により視線を表現することで、より強い志向を表示する。

faceLAB の視線・頭部認識や内蔵された Web Camera による顔認識を利用し、大きな眼球を聞き手に向けたり、聞き手の志向先に視線を向けるといった振り舞いを行うことで、聞き手との身体的な協調を引き出す。

3.1.6 社会的表示機構

視線だけではなく、頷く、顔を向ける、上半身の前面を向けるといった社会的表示も、聞き手へ向けた発話をデザインしたり、会話を円滑に進める上で重要な意味を持つ。Talking-Ally は 4 つのサーボモーターを内蔵し、これらの表示を可能としている。

3.2 発話のデザイン

相互調整的な発話の組織化を実現するために、「他者からのアシストを引き出ししながら、結果として合目的

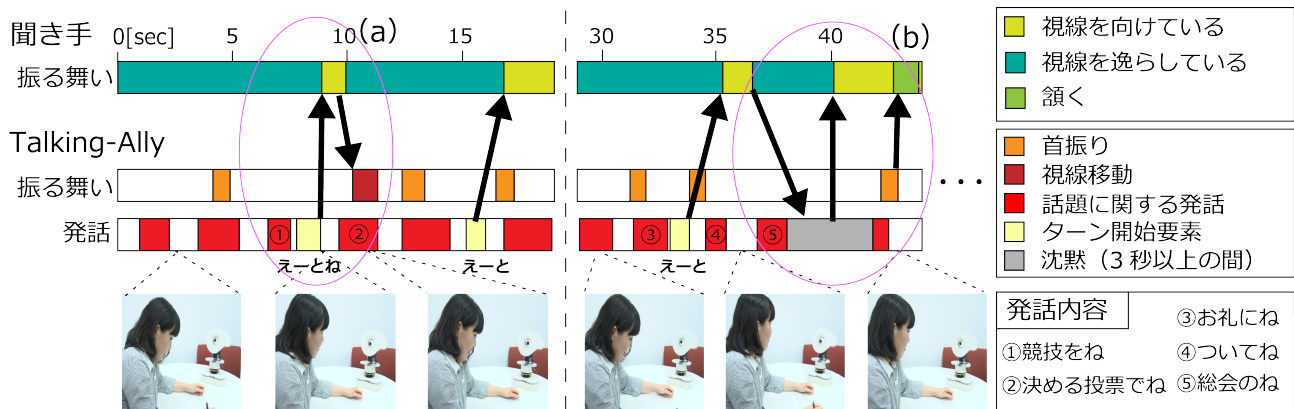


図 4: インタラクション例 (a), (b)

的な行為を実現する」という関係論的な行為形成を前提としている [19].

ここでは、聞き手の状態を反映する視線やうなずきといった「聞き手性の表示」を、話し手が発話を構築する上で重要な手掛かりになるという意味から、聞き手のアシストと捉える。また、音声合成システムとして、ATR-P 社の「Wizard Voice」採用した。これは子供の声を構築する音声合成エンジンであり、柔らかく可愛らしい声を表現することが可能である。これを頭部に内蔵されたスピーカーで発話することにより、子供のような印象を与え聞き手のアシストをより強く引き出す狙いがある。

発話をデザインする上で、以下の 3 つの方略を取り入れている。

- (a) フレージング: 発話内容を構成単位 (フレーズ) 毎に分割して、発話を行うことで聞き手の話題に対する理解を容易にする。また、聞き手との間で発話を組織化する際の基本的な単位として利用する。
- (b) ターン開始要素 (turn-initials): 発話の非流暢性の一つとして、ターン開始要素 (turn-initials) がある。ターン開始要素には、「いまから私が発話しますよ」ということを聞き手へ伝える効果がある。特にターン開始要素は聞き手の注意が話者ではなく、他の対象物に移っているときに、注目を引き戻そうとする場面で見られる傾向が強い。そのため、「あのね」、「えーとね」、「んっ」といったターン開始要素を聞き手に注目度合いに合わせて発話を行うことで、Talking-Ally の発話内容に対する聞き手の興味を維持できるよう発話を組織化していく。
- (c) 終助詞の付与: フレージングされた発話末に「～でね」、「～らしいよ」といった、終助詞を付与することで、聞き手に呼びかけるような発話を構築する。呼びかけるような発話は聞き手の応答責任を

引き出し、あいづちなどの振る舞いを喚起することに繋がる。これにより、結果として、聞き手の参加態度を維持すると考えられる。その意味で、発話を Talking-Ally の内部で完結させるのではなく、一部を聞き手に委ねることで、発話の組織化への参加を促す狙いがある。

これらの他に言い直しや沈黙などを利用し、聞き手の視線、うなずき、位置関係をリソースとしながら、発話の組織化を行う。

4 インタラクション例

ここで構築した発話生成システム「Talking-Ally」とユーザとのインタラクション例を示す。なお、ユーザとのインタラクションにおいて、ここではユーザ側からの発話による介入を想定していない。話者としての Talking-Ally に対する聞き手の身体的な協調に関して着眼しているため、聞き手のあいづちといった音声的な反応については今後の課題とする。

また、発話内容は、web から収集した News 記事をもとにしている。今回の例では、東京オリンピックの話題を取り上げた。

4.1 インタラクション (a)

図 4 の (a) では、Talking-Ally がターン開始要素や、視線を聞き手に向けることにより、聞き手の注目を引き戻そうとする場面を示している。①で「競技をね」という発話を行うが、聞き手の視線が得られていないため、ターン開始要素である「えーとね」と発話し、視線を引き戻そうとしている。この時、少しの間、視線を引き戻せたが、継続して注目を得ることはできなかった。

たため、聞き手へ視線を向けながら話を続けることで、引き続き聞き手の注目を得ようとしている。

4.2 インタラクション (b)

(b) は、聞き手の注意を沈黙により得て、その後の発話で聞き手のあいづち (うなずき) を引き出した例である。③の「お礼にね」という発話の後に、ターン開始要素の「えーと」という発話を行うことで、聞き手の視線を引き戻している。視線が再び Talking-Ally から逸れると、沈黙を行うことで、聞き手の注意を引き戻し、多少の間を置いて発話を再開している。この時、発話に合わせて Talking-Ally の頭部を上下に動かすことで、聞き手に対する語りかけを強めている。この発話によって、聞き手はあいづち (うなずき) を Talking-Ally へと返している。

5 まとめと今後の展望

これまでの発話生成システムの多くは、「個体能力主義的」な視点を背景に構築されてきた。本研究では、発話の在り方を再考し、「発話は聞き手と相互調整的に組織化される」という着想のもと、「聞き手性」、「宛名性」という概念を用い、発話生成システムに対する新たなアプローチを提案した。また、これをコンセプトとし、聞き手の視線や、うなずき、聞き手との距離感などの「聞き手性」の表示をリソースとする、発話生成システム「Talking-Ally」を構築した。Talking-Ally は視線表出機構を備え、より強い志向性を表示することが可能となっている。

今後は、適応学習機構を実装し、聞き手の性格に適応して発話を組織化する過程の解析や会話の関わり合いの中で起こると考えられる創発的なプロトコルの生成過程の分析、Talking-Ally と聞き手の間で見られる身体の空間的配置などの身体協調に関する側面での分析、非流暢性などの発話現象の生成メカニズムを構成論的アプローチにより検証することを考えている。

謝辞

本研究の一部は、科研費補助金(挑戦的萌芽研究 24650053)の助成による。ここに記して、感謝の意を表する。

参考文献

[1] 岡田美智男: 『弱いロボット』, シリーズ ケアをひらく, 医学書院 (2012).

- [2] 岡田美智男: 『口ごもるコンピュータ』, 情報処理学会編 情報フロンティアシリーズ (9), 共立出版 (2002).
- [3] Yuki Odahara, Naoki Ohshima et al.: *Talking Ally: Toward Persuasive Communication in Everyday Life*, 『Proc. of The 15th HCII』, pp. 394-403 (2013).
- [4] 浜田寿美男: 『身体から表象へ』, ミネルバ書房 (2002).
- [5] Daniel C. Dennett (土屋俊 訳): 『心はどこにあるのか』, 草想社 (1997).
- [6] 三嶋博之: 『エコロジカル・マインド』, NHK ブックス (2000).
- [7] Levelt, W.: *Speaking, From Intention to Articulation*, The MIT Press (1989).
- [8] エドワード・S・リード著, 細木直哉訳・佐々木正人監修: 『アフォーダンスの心理学 生態心理学への道』, 新曜社 (2000).
- [9] Goodwin, C.: *Conversational Organization: Interaction between speakers and hearers*, Academic Press (1981).
- [10] 泉子・K・メイナード: 『会話分析』, 日米語対照研究シリーズ (2), くろしお出版 (1992).
- [11] ミハエル・バフチン (伊東 訳): 『小説の言葉』, 平凡社ライブラリー, 平凡社 (1996).
- [12] ミハエル・バフチン (桑野隆・小林潔 編訳): 『バフチン言語論入門』, せりか書房 (2002).
- [13] 坊農真弓, 片桐恭弘: 対面コミュニケーションにおける相互行為的視点: ジェスチャー・視線・発話の協調, 『社会言語科学』, Vol. 7, No. 2, pp.3-13, 社会言語学会 (2005).
- [14] 榎本美香, 伝康晴: 3人会話における参与役割の交替に関わる非言語的行動の分析, 『人工知能学会研究会資料』, SIG-SLUD-A301, pp.25-30 (2003).
- [15] 山田誠二: 『人とロボットの<間>をデザインする』, 東京電機大学出版社 (2007).
- [16] 西坂 仰: 活動の空間的および連鎖的な組織 一話し手と聞き手の相互行為再考一, *Cognitive Studies*, Vol. 16, No. 1, pp.65-77 (2009).
- [17] Scheflen, A. E.: *Communicational Structure*, Bloomington: Indiana University Press (1973).
- [18] Kendon, A.: *Conducting Interaction*, Cambridge: Cambridge University Press (1990).
- [19] 岡田美智男: ゴミ箱ロボット 一関係論的なロボットの目指すもの, 『計測と制御』, Vol. 51, No. 8, pp.753-758 (2012).