

「聞き手性」を配慮した発話生成系における 聞き手状態の把握手法

Grasping Method of the Hearer's States in Utterance Generation while Considering the 'Hearership'

蔵田 洋平^{1*} 松下 仁美¹ P. Ravindra S. De Silva¹ 岡田 美智男¹

Yohei Kurata¹, Hitomi Matsushita¹, 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: Resources of hearer (state of the hearership) are significant in generating the structure of utterance generation mechanism (toward addressivity) in persuading the addressees. Since, we develop a Talking-Ally capable of producing utterance by considering the concept of hearership and addressivity. Talking-Ally tracks the hearer's non-verbal behaviors in order to produce influential utterance by using the Intel Perceptual Computing (PerC).

1 はじめに

これまで自然な発話 (spontaneous speech) における言い直しや言い淀みなどの非流暢な振る舞い (disfluencies) の多くは、その要因を発話プランと行為との乖離、およびその修復プロセスなど、〈話し手〉の内部メカニズムに帰属させ、説明されてきた。同様に、自然な発話生成モデルや発話生成システムも、これらの知見に基づいて構築されていることが多い。

一方で、社会的相互行為論の指摘するのは、〈話し手〉の発話の生成における〈聞き手〉の役割である。〈聞き手〉は、私たちの発話が向かう対象であると同時に、私たちの発話を制約し、方向づけているという。その意味で、発話における非流暢性も、この〈話し手〉と〈聞き手〉との相互行為的な調整の結果として生まれてきたと捉えることができると考えられる [1]。

本研究では、こうした観点から、〈聞き手〉との相互行為的な調整に基づいて発話を組織することを特徴とする〈関係論的な発話生成システム〉の構築を試みている。同時に、自然な発話における非流暢性を生み出すという試みの中で、その現象の背後にある原理を探るといふ、発話の非流暢な振る舞いに対する構成論的な研究を進めてきた。

本研究では、聞き手の「聞き手性 (hearership)」をリソースとする発話生成システムのプロトタイプ (Talking-



図 1: Talking-Ally

Ally)(図 1) の構築を行い、ここで生み出された発話が聞き手に与える影響を検証した。また、聞き手の状態を把握するために、Intel Perceptual Computing (PerC) を使用する Talking-AllyII についても、デモンストレーションで紹介する。

2 研究の背景

発話というのは、本来はだれかに向けられたもので、かならず宛て名を伴うものだといふ。これは対話の哲学者として知られる Bakhtin の「宛て名性 (Addressivity)」に関する指摘 [2] である。

*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: kurata@icd.cs.tut.ac.jp

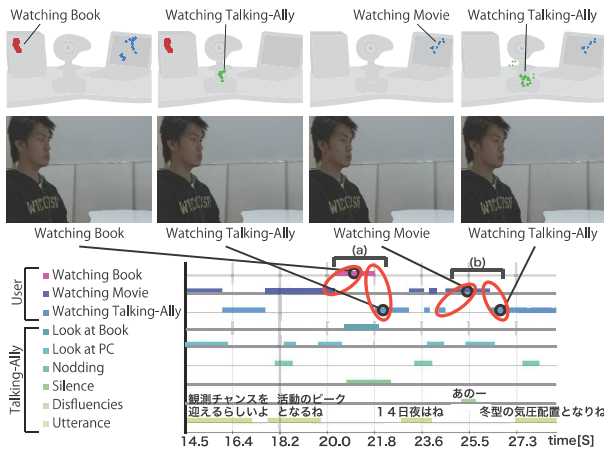


図 2: 実験の様子

一方で、私たちがいま〈話し手〉であるのは、相手がいま〈聞き手〉になってくれるからに他ならない。会話の中での〈話し手〉や〈聞き手〉、〈傍観者〉などの参与役割も、その会話参与者間での相互行為的な調整の結果として組織される。ここでは、Goodwinの指摘 [3]にあるように、〈聞き手〉もまた「いま、あなたの話をちゃんと聞いてますよ」という「聞き手性 (hearer-ship)」の表示が欠かせない。私たちは〈話し手〉の「聞き手性」をリソースの一部として、発話を組織化しているといえる。

本研究では、これらの観点を踏まえ、聞き手の存在を予定しつつ、その聞き手との相互行為的な調整に基づいて発話の組織化を行う発話生成システムのプロトタイプ (Talking-Ally) を構築した。

この Talking-Ally は、聞き手の状態を把握しながら発話内容や「ターン開始要素」などの発話調整要素、ポーズやリスタートなどでタイミングを調整し、聞き手に対して「宛て名性」を表示しながら発話を組織する発話生成システムとして構築している。

特に、聞き手の状態にあわせた発話のデザインに加え、聞き手に志向を向けながら発話を行うため姿勢や視線の表出機構、聞き手の状態を特定するためのカメラや LRF、視線追跡のためのセンシングシステムを備えている。

3 実験

Talking-Ally を用いて、宛名性と聞き手性に配慮し、聞き手と相互調整的に組織される発話がその聞き手どのような影響を与えるかを調べる実験を行った。

3.1 実験内容

実験では、被験者である聞き手の状態を伺いながら、顔や視線を向けながら発話を行う条件と、聞き手との調整なしに顔や視線を一定の間隔で左右に振るなどの

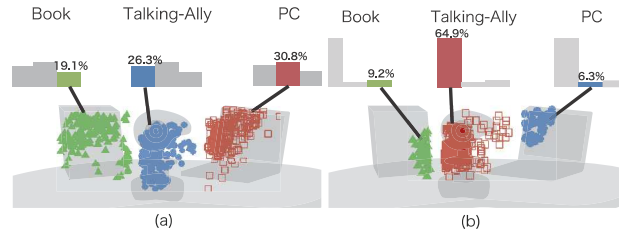


図 3: 代表的な被験者の視線注目率

機械的な振る舞いを行う条件を設定し、両者の比較を行った。

3.2 実験の様子

実験の際の Talking-Ally の振る舞いと被験者の振る舞いを図 2 に示す。この図で示した Talking-Ally の振る舞いは、聞き手の様子を伺いながら、聞き手に対して志向を向け発話を行う条件時のものである。実験では、被験者の注意を分散させるため、Talking-Ally の左隣に、ムービーを流すための PC を設置し、Talking-Ally の右隣には本棚を設置した。

実験には、被験者として計 17 名 (20~34 歳、男性 15 名、女性 2 名) が参加した。

3.3 実験結果

3.3.1 Talking-Ally の振る舞いについて

図 2 中の (a) では、被験者の視線が得られていない状態で、視線や顔を被験者の視線先に向けながら沈黙し、被験者の注目を得ようとする様子を示している。Talking-Ally は「活動のピークとなるね」という発話を行うものの、被験者の視線は PC へ向けられているため、Talking-Ally も PC に顔と視線を向け、被験者の視線を得ようと試みる。しかし、次に被験者は Talking-Ally の右側に配置されている本へと視線を移してしまったため、被験者の視線を獲得できなかった。そこで、Talking-Ally は被験者の視線先である本に自身の顔や視線を向けながら、沈黙することによって、先ほどよりも強く被験者に対して興味を引くよう振る舞い、被験者の視線獲得を試みている。結果、被験者の視線を獲得することができ、Talking-Ally は「14 日夜はね」と発話を続けている。図 2 中の (b) は、ターン開始要素発話と、視線・顔の向きを調整することにより、被験者の視線を得た例である。

3.3.2 視線注目に関する分析

聞き手の状態を伺い、聞き手と調整しながら顔や視線を向けながら発話を行う条件と、聞き手との調整なし

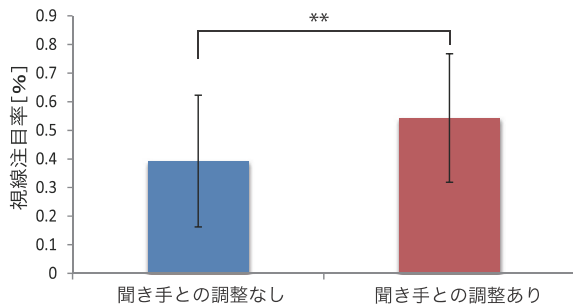


図 4: 全被験者の視線注目率

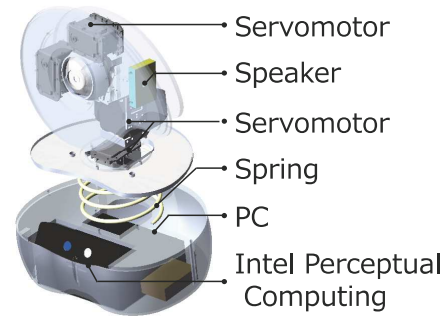


図 5: ハードウェア構成

しに顔や視線を一定の間隔で左右に振るなどの機械的な振る舞いを行う条件についての比較を行った。

Goodwin によると、聞き手の視線は聞き手の会話への参加態度の推定を可能とする。そのため、評価指標としては、視線の遷移や分布などの視線データを用いた。

図 3 はある代表的な被験者の各オブジェクトに対する聞き手の視線注目率を示したものである。ここで視線注目率とは、「聞き手が各インタラクション時間の中で、そのオブジェクトをどれくらい見ていたかを表す割合」である。図 3 の (a) は聞き手との調整なしの条件で、図 3 の (b) は聞き手との調整ありの条件である。(a) では聞き手の視線注目率の内、一番高い注目率を得たものは「PC」であった。PC では動物に関するムービーが再生されており、実験教示の際にムービーに関する簡単な質問に答えてもらうよう被験者に知らせていたため、PC のムービーへ視線が集中したものと考えられる。

次に視線を多く獲得できたものは、「Talking-Ally」であり、一番低かったのは、「Book」となった。設計された通りの順序で振る舞いながら、機械的な発話を行う発話生成システムの発話の聞き手への宛名性が、PC のムービーに対する質問への義務的な影響力に対して弱かったと考えられる。

それに対して、(b) で一番注目を集めたのは、「Talking-Ally」であり、次に「Book」、「PC」の順になった。Talking-Ally と Book、PC との視線注目率には大きな差があり、聞き手の視線が Talking-Ally へ積極的に集められたことが伺える。

また、全被験者の Talking-Ally に対して向けた視線注目率の平均値を示したのが図 4 である。聞き手との調整なしの条件は 39.2%、聞き手との調整ありの条件は 51.6% となり、統計的な有意な差がみられた (** $p < 0.01$)。このデータから全被験者において聞き手との調整なしの条件よりも聞き手との調整ありの条件のほうが Talking-Ally への視線注目率が高くなる傾向が得られた。このことから、聞き手と調整しながら発話を組織化することで、視線注目の維持に繋がり、結果的に

聞き手の会話への参加態度を引き出す可能性がある。

4 Talking-AllyII の構築

Talking-Ally の実験によって、聞き手の状態を特定しながら発話を行う事の重要性を確認することができた。しかし、Talking-Ally の振る舞いの様子をデモで行った際、得られたコメントや、気づいた点から (1) 聞き手が Talking-Ally の方向を向いていなくても発話を続ける (2) 聞き手の顔の高さに顔を向ける事ができない (3) 聞き手の発話に対応できないといった問題点があることが分かった。この問題を解決するために Talking-Ally に Intel 社の「Intel Perceptual Computing」(以下 PerC) を内蔵し、Talking-AllyII の構築を行った。Talking-AllyII のハードウェア構成を図 5 に示す。

4.1 聞き手のセンシング

これまで Talking-Ally においては、測域センサ (LRF) による人との距離感や方向の把握と、Talking-Ally 眼球内の Web カメラによる顔認識で、聞き手の様子の把握を行っていた。しかし、Talking-Ally 眼球内の Web カメラでは顔認識が行われる頻度が少なく、ほとんど測域センサによる位置検出を主として、聞き手のセンシングを行っていた。そのため、聞き手が Talking-Ally を向いているかどうかの認識が難しく、測域センサは人までの距離と方向の 2 次元空間のデータしか認識できず、聞き手の顔の高さに顔を向ける事ができなかった。

以上を踏まえ、PerC を用いた聞き手の状態の把握の向上を図った。PerC は、一般的な RGB カメラの他に、カメラ前の 3 次元空間の距離を認識できる深度カメラを備えている。この深度カメラにより、聞き手の顔の高さまで、顔を向ける動作を実現した。また、RGB カメラの精度も向上したため、聞き手の顔を認識したときに発話を行う事によって、聞き手が、Talking-Ally の

方向を向いていなくても、発話を続けるといった問題を解消することができた。その他に PerC は音声認識が実装されているため、聞き手の発話に対応することが可能となった。

4.2 発話のデザイン

音声合成エンジンとして、ATR-P 社の「Wizard Voice」を採用した。これは、子供の声を発話するものであり、その声は非常にかわいらしいものとなっている。これにより、聞き手に「弱い存在」として Talking-AllyII を解釈させながら、聞き手からのアシストを引き出すことを狙いとする。また、発話内容についても、聞き手との状態に応じて、「えーと」「あの一」「それでね」といったターン開始要素 (Turn-initials) や、「ね」「だって」などの発話末モダリティの付与を行い、言い直しや言い淀みを伴う非流暢な発話を生成する。先行研究により、このような非流暢な発話には、聞き手の興味を引き出す効果があることを確認している [4]。

4.3 視線表出機構

発話においては、身体の動きが言葉に随伴して起こり、その発話が相手に伝える上で大きな役割を持つという [5]。特に、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 [6][7]。また、「目には向き、つまり志向性があり、そこにはつきり相手が主体として私の方に向かっている力を感じる」という [8]。

このように会話において話し手の視線は、その志向を示す役割が特に大きいものとなっている。そのため、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、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。

ただし、これまでの Talking-Ally では、発話中常に聞き手の方向に向かって視線を向けようとしながら発話を行っていた。一方で、実際の発話では、非流暢な発話を行っている時は、話し手の視線は聞き手からそれやすいため [9]。また、話し手は発話の開始時に聞き手から視線をそらす [10]。ともいう。そのため、Talking-AllyII の視線移動についても、非流暢な発話を行う時は、聞き手から視線を逸らしたり、発話開始時に視線を逸らすなど視線の動かし方を調整する。

5 まとめと今後の展望

本研究では実験によって、聞き手性や宛名性に配慮し、聞き手の様子を伺いながら、聞き手に対して志向しながら発話を行うことで、一定の規則に従って機械的に発話を行うシステムと比べて、聞き手の興味を引き、多くの視線を集めることを確認した。また、PerC を用いた聞き手の状態の把握の向上について検討し、実装を行った。

今後の展望としては、今回新たに再構築したシステムが、聞き手に対し有効か調べるために、実験やデモンストレーションによる検証を行う。また、視線の動かし方についても、常に聞き手の方向を志向するのではなく、言い淀んだ発話を行う時に、視線をそらすなど振る舞いを実装し、その効果を検討したい。

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] 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).
- [3] Goodwin, C.: *Conversational Organization: Interaction between speakers and hearers*, Academic Press (1981).
- [4] 小田原雄紀, 蔵田洋平, 大島直樹, デシルバラビンドラ, 岡田美智男: *Talking-Ally* 聞き手と一緒に発話を組織する発話生成システムについて, ヒューマンインタフェースシンポジウム 2012 論文集, 1411L, pp. 161-166, (2012).
- [5] 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).
- [6] Schefflen, A. E.: *Communicational Structure*,
- [7] Kendon, A.: *Conducting Interaction*, Cambridge: Cambridge University Press (1990).
- [8] 浜田寿美男: 『身体から表象へ』, ミネルバ書房 (2002).
- [9] Kendon, A.: *Some functions of gaze-direction in social interaction*, *Psychologica*, 26, 2263 (1967).
- [10] Duncan, S.: *On the structure of speaker-auditor interaction during speaking turns*, *Language in Society* 3: 161-180(1974).