聞き手の表情に合わせて発話を微視的に調整する〈トーキング・アリー〉

星野 翔平[*1]　巽 将司[*1]　岡田 美智男[*1]

Talking-Ally: Spontaneous Utterance Generating System adapting Hearer's Facial Expressions and Gaze-Direction

Hoshino Shohei[*1], Masashi Tatsumi[*1] and Michio Okada[*1]

Abstract - In this study, we are building a robot "Talking-Ally" that adjusts the timing and contents of utterance according to the gaze direction of listener. As a result, utterances from this robot are accompanied by disfluencies such as fillers and turn initials, but you can feel robot's intention to tell hard to the listener. In our demonstration presentation, we introduce the outline of "Talking-Ally" which microscopically adjusts the utterance to the expression of the listener and the movement of the line of sight.

Keywords: Human-Robot Interaction and Disfluencies of Utterance

1.　はじめに

私たちの日常的な会話では，話し手は聞き手の表情を見ながら発話を行っている. 話し手が楽しそうな表情をしているときには, 聞き手もそれにつられて笑顔になったり, 会話を行う中で聞き手が険しい表情になっている時には, 話し手はこれまでの話題を変えたり, フォローを入れたりしようとする.

これまで「サイモンの蟻」[2]という議論がなされているように，私たちは，砂浜の上でアリの足跡が複雑な模様を描いているとき, アリという個体にその複雑さの要因を帰属させて考えてしまいがちである．しかし，アリとそれを取り巻く環境が関係しあうことで複雑な足跡を残していたと捉えることもできる.

これを日常会話に当てはめると，一見，非流暢な発話を伴うような会話も, その話し手によって一方的に繰り出されるだけではなく，話し手と聞き手との間で，お互いに相手の所作や表情を読み取りながら, 適応的に相互調整を行う中で生まれてきたものと考えることができる.

このような観点から，本研究では聞き手の表情の変化や視線の動きをとらえ, 微視的に発話方略と振る舞いを変化させながら発話を行う〈トーキング・アリー〉というロボットを構築している（図 1).

この〈トーキング・アリー〉は発話を行うなかで聞き手の表情や視線の変化をセンシングしながら, 発話及びそれに伴う振る舞いを調整する. 本システムと人とのインタラクションを通じて，聞き手の表情をセンシングし，振る舞いを変えながら発話を行う発話生成システムが人に与える影響について調べること目的としている.

図 1〈トーキング・アリー〉
Fig. 1 Talking-Ally

本発表では，デモンストレーションにより，本研究の概要や〈トーキング・アリー〉のハードウェア，基本動作を紹介する.

2.　研究背景

2.1　コミュニケーションにおける表情

私たちは他者と相対するなかで，自分に良い印象を与えるために笑顔で話しかけたり, すこしまじめな話をするときに真顔で相手を見ることがある．このように，日常的なコミュニケーションのなかでは, 言葉だけでなく表情によって相手に自分の状態を伝えたり, 気持ちを共有しあうことが行われている.

特に，相手の笑顔に誘われて，思わず笑ってしまうように，相手の表情に対して同調し，自分の表情を変えてしまうことが知られている. 感情を模倣する行為とその効果

*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

については，Lipps が心理学に「感情移入」の概念を導入してから今日まで，多くの研究が行われてきている[3].

人の表情に対して同調的な振る舞いをすることが，コミュニケーションを行うなかで重要な一部分を担っていると考えられている一方で, これまでの発話生成システムにおいてこのような観点は十分に考慮されていない. そこで，発話生成システムが人の表情を認識し，その振る舞いを変えることで, 聞き手に対して親密さを感じさせることができるのではないかと考え, そうしたシステムのプロトタイプの構築を試みることにした.

2.2 発話の非流暢性

私たちが何気なく繰り出している発話の中では,「えーと」「あのー」といった発話開始要素(turn-initials)やフィラー，そして言い直し(repair/restart)，言い淀み(hesitation)などの非流暢な振る舞いを数多く観察することができる. これらは総称して発話における「非流暢性(disfluencies)」と呼ばれている.

これまでの発話生成システムにおいて，例えばスマートスピーカーにおいては，一定のリズムで，報告的に発話が行われており, 発話における非流暢性については聞き手に情報を伝える上でネガティブな要素として捉えられてきた．しかし，非流暢性は話し手と聞き手との間で相互作用的な調整の中で生み出されたものであると捉えると, 非流暢な発話は, 聞き手と調整しながら円滑なコミュニケーションを行うための方略でもあり, また非流暢な発話を手がかりとして, 聞き手の注意を引き出しているとも考えられる.

2.3 先行研究

筆者らの研究室では，これまで非流暢性を持つロボットからの発話の会話のなかで果たす役割を明らかにしてきた.

Ohshima ら[4]は，非流暢な発話と身体的な志向の表示が聞き手の志向的な構えに与える影響についての研究を行い，“あなたに対して話しています”という構え（宛名性）を考慮したシステムにおいて，聞き手への配慮を感じさせることができ, 機械的なシステムに比べ聞き手の注目を多く集めることができることを示した.

蔵田ら[5]では宛名性に注目し，宛名性を持った発話を行うことで聞き手の関心を引きやすいこと．松下[6]では，発話の中で聞き手に適した方略を探ることで, 聞き手の視線を再獲得しやすくなることが挙げられている.

先行研究を踏まえ，聞き手の視線情報だけでなく，表情もリソースに含めたシステムを構築することで, より聞き手に適した発話が行えるのではないだろうか.

本研究では，聞き手の表情をリソースに含めて発話と振る舞いを調整するシステムを構築することにした.

3. 〈トーキング・アリー〉の構築

3.1 概要

本研究に用いた研究プラットフォームである〈トーキング・アリー〉は，先行研究である Ohshima ら[4]や蔵田ら[5]によって構築されたものに，本研究の意図に沿うように，聞き手の表情認識などの新たな構成要素を追加したものである.

3.2 ハードウェア構成

〈トーキング・アリー〉のハードウェア構成を図 2 に示す.〈トーキング・アリー〉はこれまで，視線追跡システムやレーザーレンジファインダーなどのセンサを用いて聞き手の状態を把握しながら，リアルタイムに発話を生み出すロボットとして構築されてきた.

本研究に用いる〈トーキング・アリー〉は，視線の動きや頭部の動作生成に用いる 4 つのアクチュエータと，聞き手の表情や視線方向などの認識を可能とする画像センサ HVC-P2 を使用している.

3.2.1 社会的表示の生成機構

〈トーキング・アリー〉は頭部のサーボモータを用いて，顔と目をそれぞれ上下左右方向に動作させることができる．この機構により，何かに注意を向ける(Attention)，聞き手に姿勢を向ける（Addressing)，聞き手に視線を向ける（Gazing）などの振る舞いが可能である．また〈トーキング・アリー〉は画像センサで捉えた情報をもとに社会的表示となる多様な振る舞いを生成する.

3.2.2 聞き手の状態に対するセンシング機構

聞き手の表情や視線等のセンシングを行うため，画像内の顔，表情，視線検出機能等を有する画像センサ HVC-P2*2 を用いた．HVC-P2 は画像センシング技術 OKAOVision を備えた画像センサであり，人体検出，手検出，顔検出，顔向き推定，年齢推定，性別推定，目つむり推定，表情推定，視線推定の 10 通りの画像処理を行うことができる．本研究では，これらの機能のうち，顔検出，表情推定, 視線検出を用いることで聞き手の状態をセンシングする.

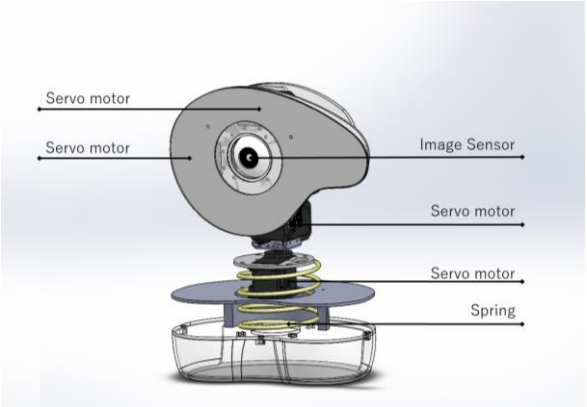

図 2 〈トーキング・アリー〉ハードウェア構成図

Fig. 2 Hardware configuration of Talking-Ally

*2: オムロン株式会社による画像センサ

*2: Image sensor manufactured by OMRON Corporation

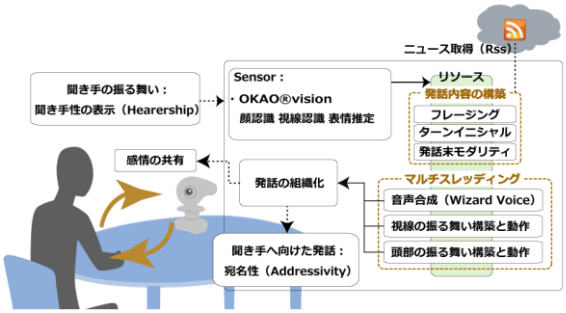

図 3 〈トーキング・アリー〉システム構成図

Fig. 3 System Configuration of Talking-Ally

3.3 システム構成

発話生成システム〈トーキング・アリー〉は，表情を含めた聞き手の振る舞いをリソースとし，ニュース等からリソースを含めた発話と振る舞いの生成を行い，聞き手に対して宛名性を高めた発話を行うものである．

〈トーキング・アリー〉システム構成図を図 3 に示す．

3.3.1 非流暢な発話の生成

本システムでは，ニュース等のリソースの発話末にモダリティ「ね」を加えた”Normal Utterance”と非流暢要素として”言い淀み”と”Silence”を繰り出しながら発話を行う．〈トーキング･アリー〉の視界内に人がいる状態において，人の視線が獲得できている場合は” Normal Utterance”を行い，人の視線が獲得できていない場合においては非流暢な要素である”言い淀み”または”Silence”を行う．

視線が獲得できておらず，”言い淀み”が行われる場合，”えーと”，”んーと”，”あの”，さらにそれらの発話末にモダリティ「ね」を付与した発話”えーとね”，”んーとね”，”あのね”の 6 種類から方略を選択する．表 1 に各状態における方略の一覧を示す．

3.3.2 音声合成エンジン

〈トーキング・アリー〉の発話生成には音声合成システム WizardVoice*3 を用いている．WizardVoiceSDK に含まれるサンプル音声の中から，子供の声のライブラリを用いて発話の生成を行うことで，たどたどしい発話を生成している．このたどたどしい発話は，聞き手に対しての参加を促すことに効果的に機能する．

3.3.3 表情認識

〈トーキング・アリー〉は，目に内蔵された画像センサを用いることで視界内の人間の表情及び目線を認識することができる．図 4 における右上の画像が，〈トーキング・アリー〉が認識している映像である．このように画像センサで捉えた人の顔について常に表情推定を行っており，図 4 の場面においては聞き手の表情を”真顔”と判定している．

搭載している画像センサは，人間の表情において”真顔”，”怒り”，”悲しみ”，”喜び”，”驚き”の 5 つの表情について推定を行うことができる．これらの表情情報をもとにして発話の速度や振る舞いを変更する．

3.3.4 聞き手の表情による振る舞いの調整

〈トーキング・アリー〉は，発話を行うなかで聞き手の表情を常にセンシングし，聞き手の表情に合わせた振る舞いを行う．

発話の中で〈トーキング・アリー〉は聞き手の 5 つの表情（”真顔”，”怒り”，”悲しみ”，”喜び”，”驚き”）を認識する．認識した結果において，”喜び”の値が大きいほど，聞き手が〈トーキング・アリー〉の発話や話題に好感を示しているであると判断する．

表 1 状態と行動一覧

Table 1 List of states and actions

状態	ロボットの振る舞い	
GazeIn	UT	通常発話 （ニュース等の発話+発話片末のモダリティ）
GazeOut	UT	通常発話 （ニュース等の発話+発話片末のモダリティ）
	TI	"えーと" ”えーとね” "んーと" ”んーとね” "あの" ”あのね”
	SI	沈黙（1500ミリ秒の無音）

図 4 聞き手の表情センシング

Fig. 4 Sensing of listener's expression

*3: 株式会社国際電気通信基礎技術研究所が開発した音声合成システム

図 5 インタラクション例
Fig. 5 Interaction example

一方で”怒り”，”悲しみ”の値が大きいほど，発話や話題に好感を示していないと判断し，この場合の方略として，話題の変更や聞き手に対して発話する際の振る舞いの変更を行う．

インタラクション例として図 5 を示す. 図 5 では聞き手が怒っている表情を認識しており，〈トーキング・アリー〉は聞き手の顔を窺うような振る舞いを行いながら発話を行う．

4. まとめ

本研究では，日常的な会話のなかで，話し手と聞き手がお互いの表情をもとに表情を変え，それによって相互作用的に発話と振る舞いが調整されることに着眼し，人に対して発話と振る舞いを行うなかで，聞き手の表情を含めた状態をリソースとして，聞き手と相互調整を行うなかで発話を行うロボット〈トーキング・アリー〉を提案し，構築を行った．

今後の課題として，インタラクション実験を通じて，聞き手に対して振る舞いがどのような印象を与えているかについて印象評価実験を行うとともに，〈トーキング・アリー〉が聞き手の表情をリソースに含めた振る舞いを行うことで，人とロボットの発話における効果について検証を行っていく．

また，今回実装した振る舞いや発話の妥当性についても検証を行う必要があると考えている．ロボットのふるまいによって聞き手に与えたい印象を与えられているかを確認する必要がある．

謝辞

本研究の一部は，科学研究費補助金(基盤研究(B): 18H03322)の助成を受けて行われている．ここに記して感謝申し上げたい．

参考文献

[1] 岡田美智男：『弱いロボット』，医学書院(2012).

[2] H. A. サイモン（稲葉元吉，吉原英樹訳）:『システムの科学』，パーソナルメディア（1999）.

[3] 澤田瑞也：『共感の心理学』，世界思想社(1992) .

[4] Naoki Ohshima , Yasuke Ohyama, Yuki Odahara, P. Ravindra S. De Silva, and Michio Okada: Talking-Ally: The Influence of Robot Utterance Generation Mechanism on Hearer Behaviors, International Journal of Social Robotics, Volume 7, Issue 1, Pages 51-62 (2015).

[5] 蔵田洋平、松下仁美、小田原雄紀、大島直樹、デシルバ ラビンドラ、岡田美智男 : Talking-Ally: 聞き手性をリソースとする発話生成システムの実現にむけて、ヒューマンインタフェース学会論文誌、Vol.17, No.2, pp.159-170 (2015).

[6] 松下仁美、香川真人、山村祐之、岡田美智男：非流暢性を伴うロボット(Talking-Ally)の発話調整方略とその聞き手に対する適応に関する研究、ヒューマンインタフェース学会論文誌 Vol.20, No.2, pp.255-268 (2018).