

対話ロボットの複数台化とジェスチャ表出による対話意欲変動の分析

丸山 洸太^{†a)} 大和 淳司^{†b)} 杉山 弘晃^{††c)}

Analysis of How Motivation to Conversation Varies Depending on the Number of Dialogue Robots and Presence/Absence of Gestures

Kota MARUYAMA^{†a)}, Junji YAMATO^{†b)}, and Hiroaki SUGIYAMA^{††c)}

あらまし 本論文では、ユーザと対話ロボットの対話に際して、ロボットの複数台化とジェスチャ表出がユーザの対話意欲に与える影響について分析する。ロボットを複数台化することで、対話が成立していると感じやすくなることや、対話が破綻していないと感じやすくなることが分かっている。そのため、対話ロボットの複数台化がユーザの対話意欲向上にも有効であると仮定し、対話実験を行った。実験の結果、ジェスチャ表出をする対話ロボットを複数台化するとユーザの対話意欲が高くなった。また、複数台ロボットとの対話では、ジェスチャ表出をするロボットと対話したユーザの方が、ジェスチャ表出をしないロボットと対話したユーザよりも対話意欲が高くなった。この結果の原因として、ロボットの複数台化に伴いユーザが扱わねばならない情報も増えるという点があげられる。扱う情報が増えることで、それらの情報を適切かつ分かりやすくユーザに提示する必要がある。しかし、対話ロボットの台数増加に伴いジェスチャ表出などを適切に提示できれば、対話意欲が向上することを示唆している。そのため、対話ロボットの複数台化に応じた、適切な環境の構築が重要であると考えられる。

キーワード 対話意欲, 2 台対話, 音声対話, 対話ロボット, ジェスチャ表出

1. ま え が き

近年は様々な対話システムが登場し、人と対話を行えるシステムやサービス、ロボットに注目が集まっている。例えば、高齢者の認知症予防や、一人暮らしの話し相手として非タスク志向型の雑談対話システムに期待が集まる。しかしこれらの対話システムは注目度に比して未だ課題は多く、実際に人同士で行われるような対話は実現できていないのが現状である。

例えば、長谷川 [1] は Siri や Google Home, Amazon Echo を例に出し、ボットとの会話は「刺激 (命令)」と

「反応 (情報提供)」という関係であると述べている。その上で、「言葉の文脈ではなく、話している相手の文脈 (経験、感情) を考慮すると、私たちの会話はより親密なものになる」と指摘している。そのため、文脈を理解できる対話システムが登場してきてはいるが、感情などを考慮した文脈の理解においては未だ十分とは言いきれない。

文脈を無視した不自然な発話や、ユーザが返答しづらいような発話を行い続けると、ユーザに「もう話したくない」という感情を抱かせる可能性がある。このような感情を引き起こすと対話システムを使用してもらえなくなる恐れがある。

そこで本研究では、ユーザの「対話したい・対話を続けたい」という対話意欲を変動させる要因について調べる。特に、対話ロボットの複数台化やジェスチャ表出に着目し、分析を行う。また、対話意欲はユーザの主観的な評価であるため、客観的な議論が行いづらい。そこで、客観的な分析を行うため、生理指標の利用についても検討を行う。

[†] 工学院大学大学院, 東京都

Graduate School of Kogakuin University, 1-24-2 Nishishinjuku, Shinjuku, Tokyo, 163-8677 Japan

^{††} 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所, 京都府

NTT Communication Science Laboratories, NTT Corporation, Hikaridai, 2-4 Seika-cho, Souraku-gun, Kyoto-fu, 619-0237 Japan

a) E-mail: em19021@ns.kogakuin.ac.jp

b) E-mail: yamato@cc.kogakuin.ac.jp

c) E-mail: h.sugi@ieee.org

DOI: 10.14923/transinfj.2020CNP0004

2. 関連研究

対話が破綻していたり対話が成立していないと感じている中で対話を続けることは困難であり、話者にとって大きな負担となる。そのため、対話を行うには、話者相手に「対話が成立していない」あるいは「対話破綻が発生している」と感じさせないようにする必要がある。このことから、対話ロボットを構築する際には、「対話が成立している」という対話感を向上させたり、対話破綻感を低下させるよう意識しなければならない。飯尾ら[2]や Sugiyama ら[3]の研究から、対話ロボットの複数台化が対話感の向上や対話破綻感の軽減に資することが分かっている。本研究では、対話ロボットの複数台化が対話意欲に及ぼす影響について分析することで、どのような条件であればユーザがよりロボットと対話し続けたいと感じるのかについて考察する。

一方、対話ロボットを複数台化することにより、次に誰が話すかという話者交代が複雑になる。話者交代が適切に行えないと発言が重なってしまったり、話したいのに話させてもらえないなどの状況が発生し、対話意欲が低下する要因になる。そのため対話ロボットを複数台化した場合、話者交代を適切に行う必要性が増す。徳永ら[4]は、対話参加者らが表出する態度とその解釈から、話者が次々と交代する順番交代の仕組みを明らかにした。本研究で扱うロボットは、発話時にジェスチャ表出を行い、話者交代時には質問する相手へ体の向きを変えるなど、話者交代が容易に認識できるように設計している。そのため、ユーザはロボットのジェスチャ表出に対する解釈によって次話者を予測すると考えられる。そのためロボットの複数台化に際し、ジェスチャ表出の有無を比較することでジェスチャ表出が対話意欲に及ぼす影響について分析を行う。湯浅ら[5]は発話交代モデルを提案し、擬人化エージェントのノンバーバル行動によって発話マインドの読み取りと、伝わったマインドによって人が発話行動をしようとするかを確かめる評価実験を行った。この実験で用いた音声は無意味音声であるため、発話内容の影響を取り除いた実験だといえる。本研究では実際に行われる対話環境を想定し、発話内容も考慮した上でジェスチャが対話意欲に影響を及ぼすか分析を行う。

人同士の対話は言語情報だけでなく、非言語情報の共有も行われる。非言語情報が話者交代において重要な役割を果たしていることが示唆されているように、

ジェスチャなどの身体表現は、ロボットがユーザと対話する上で欠かせない要素である。小野ら[6]は人同士が対話を行う際に同調的な身体表現が見られることを利用し、人とロボットが対話するときロボットに身体表現が人の身体表現に近づくほど、被験者の同調的な身体表現が多くなり、被験者のロボットに対する評価が高くなることを明らかにした。この実験は道案内を行う状況を想定したものであり、指さしを行うなど同調的な身体動作が取りやすい環境であるといえる。本研究では、ロボットがユーザに顔を向けるといった身体表現を取り入れることにより、ユーザが話者交代をスムーズに行えるようになり、対話意欲が高くなる要因となるか分析を行う。神田ら[7]は、ロボット同士の指差し会話を被験者にあらかじめ観察させることが、ロボットの発話理解を容易にさせることを示した。本研究ではユーザが対話ロボットと対話を行う際に事前に対話を見る機会がない状況を想定し、ロボット同士の対話を事前に提示することなく、対話の最中にユーザに提示することで、対話意欲の向上につながることを分析する。

また、渡辺[8]は、「ウンウン」という頷きの動作をロボットに行わせることによって、ユーザは話を聞いてもらえていると認識しやすくなった事例を紹介している。松原ら[9]は愚痴を聞くロボットに着目し、ユーザが発する愚痴に対して頷きの回数を変えることで、ユーザの愚痴満足度が向上し、愚痴をこぼしやすくなったと報告している。頷きを取り入れることで話を聞いてもらえると認識しやすくなるという事例や、愚痴満足度が向上するという研究は、対話中の頷きがユーザの対話意欲を高めていることを示唆している。本研究ではユーザとロボットを対話させる際に頷きを含めたロボットのジェスチャ表出の有無を比較することで、ユーザの対話意欲がどのように変動するか分析を行う。

ユーザの対話意欲減退には複数の原因が挙げられ、そのうちの一つにストレスがある。そのため、対話ロボットを構築する上で対話中のユーザにストレスをかけないことも考慮する必要がある。ストレス評価には萩野谷ら[10]が唾液アミラーゼ (salivary α -amylase: sAA) の有効性を示しており、本研究でも sAA をストレス指標として用いた。対話意欲は主観評価であるため、生理指標を取り入れることで客観性を示す目的がある。また、sAA は他のストレス指標に比べて測定が容易であり、対話の合間に計測ができるため、sAA を

評価指標として選択した。しかしこれまでのロボット対話において、ストレス指標として sAA が用いられた例はない。そのため、本研究では sAA を用いることで対話意欲とストレスの関係についても分析を行うが、その妥当性についても検討していく。

これらの研究から、ユーザが対話ロボットと対話する際に対話意欲が変動する要因として、

- 対話ロボットの台数を増やす
- 頷きを含むジェスチャ表出を行う

などが挙げられる。そこで本研究では、1 台のロボットと対話したユーザと 2 台のロボットと対話したユーザの対話意欲の比較、及び複数台化に際してジェスチャ表出の有無を比較することで、それぞれの要因が対話意欲に及ぼす影響について分析を行った。

先行研究に対する本研究の新規性は以下の三点にある。一つ目は完全に独立した雑談対話システムを使用した点である。本実験で用いた対話システムは音声入力から発話文生成、音声合成までを自動で行うことができ、WoZ 実験のように人の手を借りることなく対話を行うことができる。そのため、完結した一つのシステムを評価することが可能となっている。また、多様な話題に対応可能で、更にシステムとして独立しているため、雑談対話ロボットが応用される状況に即した状況で実験を行った。二つ目はジェスチャ表出の有効性が示されたことである。対話ロボットの複数台化が、対話感の向上や対話破綻感の軽減に資することが先行研究で明らかになっている。一方で、話者交代の複雑化といったデメリットについては調査されていない。本研究ではそのような点に着目した。ジェスチャ表出がユーザの発話ロボットの識別を容易にし、複雑化する話者交代において有効であることが示された。これは 3 台以上のロボットにも有効であると考えられる。三つ目は、sAA という客観的な指標を取り入れた点である。これまで多くの研究は被験者に対して 5 件法といった主観評価がなされてきた。本研究では、そのような主観評価に加え、より客観性をもたせるために客観的な指標である sAA を取り入れた。

3. 対話実験

3.1 実験目的

ユーザと対話ロボットを対話させたとき、1 台のロボットと対話するよりも 2 台のロボットと対話する方がユーザの対話意欲が向上することを検証する。また、ジェスチャ表出を行うロボットと対話したユーザ

の方が、ジェスチャ表出を行わないロボットと対話したユーザよりも対話意欲が高くなることを検証する。

3.2 評価指標

本実験で用いる評価指標を表 1 にまとめた。本実験は本学に在学する 24 名 (20–23 歳) の男女に、実験協力者として対話実験を行った。実験協力者は対話ロボットと複数回対話を行い、各対話中手元に設置したスイッチで「もっと対話したい」か「もう対話したくない」かを入力させた。対話実験中に実験協力者が入力した「もっと対話したい」と「もう対話したくない」の評価数のうち、「もっと対話したい」の評価数の割合を意欲率とし、対話意欲の評価指標とした。実際の対話で対話意欲の評価は行われないため、なるべくユーザに負担がかからないよう配慮する必要がある。一方で、対話意欲はユーザの主観による評価が最も信頼できるため、ユーザが直接対話意欲を評価する手法が望ましい。そのため本研究では、簡単に評価ができるようボタンを押下することで対話意欲を評価させた。より負担の少ない、ペダルを足で踏むことで評価させる手法や、対話実験後に撮影した動画を見せることで、発話時点での対話意欲を評価させる手法も考えられるが、現在検討中である。足での評価は、足でペダルを操作するため、少ないミスで確実に評価が行えるか、そして本当に負担が軽減するのか、という点を考慮する必要がある。また、対話実験終了後に動画を見せることで評価する手法は、過去の気持ちを思い出しながら評価するため、正しく評価が行われているかを考慮する必要がある。徳永ら [4] は、話したい、話したくないの評価を第三者評価で行っていたが、これは第三者が表情やジェスチャなどから推定しているため、必ずしもその評価が正しいとは限らない。例えば、対話者が感情をあまり表情に出さない人だとしたら、「対話したい」と評価される回数が減る可能性が高い。逆に、普段から笑顔な人が対話者だった場合、「対話したい」と評価される回数が増える。発話者自信が評価を行えばより正確な対話意欲の評価が行えるものと考えたため、本実験では対話者自身に対話意欲の評価を行

表 1 評価指標一覧

評価指標	説明
意欲率	「もっと対話したい」「もう対話したくない」の評価数のうち「もっと対話したい」の評価数の割合
sAA	1 対話ごとに計測した唾液アミラーゼ活性の値
sAA 比	対話終了後 sAA/対話開始前 sAA を 1 対話ごとに計測
発話長	1 対話中に発した 1 発話当りの平均文字数

わせた。

また、ストレス指標として有効性が示されている sAA を対話前後のストレス変化の指標として用いた。sAA とは、唾液中に含まれるアミラーゼの触媒能力の尺度のことで、単位は主に気温や湿度を考慮した Unit が用いられる。文献[10]によれば、暗算後に計測した sAA と心拍・スキンコンダクタンス・緊張に対する主観評価 (VAS; Visual Analogue Scale) のいずれの項目も有意に上昇し、sAA とそれぞれの評価項目間に有意に相関が見られたとの報告があり、ストレス増大が sAA の上昇につながる事が分かっている。しかし sAA の値そのものは個人差が大きいので、対話前後の sAA 比を測ることとした。sAA の差を比較する手法も考えられたが、sAA の個人ごとの標準偏差が各人の平均値とおおむね比例傾向にあることから、sAA 比を用いた。sAA 比から、各対話でどれだけ sAA が変化したかが分かる。文献[10]によると、暗算開始前から暗算終了後において sAA が 18kU/L から 39kU/L になったため、約 2.17 倍になったことが分かる。そのため、ストレス上昇時の sAA 比として 2 倍程度が目安になるものと考えられる。

また、ユーザが発した 1 発話の平均文字数を発話長とした。このときの発話長には漢字仮名交じり文を使用した。発話長が長ければその対話で多くの言葉が発したものとする。

3.3 実験環境

本実験では以下の器具を用いて対話実験を行った。

- 対話ロボット Sota … 2 台
- PC (Windows10)
- モニタ
- 指向性マイク
- スピーカ … 2 台
- USB オーディオインタフェース
- 評価 Box (ボタン ×2, レバー ×1)
- 唾液アミラーゼモニタ

対話実験を行った実験室の概観図を図 1 に示す。残響の少ない防音室で対話実験を行った。実験者用の机と実験協力者用の机を用意し、実験協力者用の机にコントローラ (Raspberry Pi に Python で実装) を入れた評価 Box を設置した。評価 Box には対話意欲を評価するためのボタン 2 個 (「もっと対話したい」ボタンと「もう対話したくない」ボタン) と、直前に質問したロボットは左右のどちらだったかを回答するためのレバーを評価 Box 中央に取り付けた。なお、ボタン押下

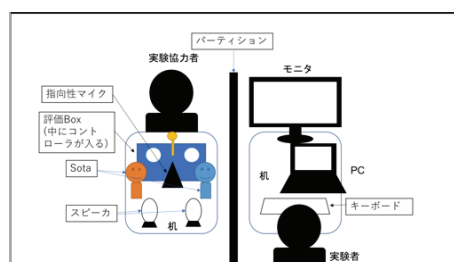


図1 実験室の概観図

のタイミングはロボットが発話を終えた直後に行い、その後レバー操作を行わせた。このときの詳細な手順については、3.6 実験手順を参照されたい。評価 Box の上に音源分離機能を有する指向性マイクをユーザに向けて設置した。実験協力者側から見て評価 Box の奥側に対話ロボット 2 台を設置し、更に各対話ロボットの奥にそれぞれスピーカーを設置した。

実験者用の机には、対話システムを構成する PC を設置した。また、コントローラを制御するためのキーボード、マウスと、コントローラからの出力を表示するためのモニターを設置した。

本実験は実際にユーザが自然に対話する状況を想定しているため、通常の対話シーンではユーザが見るはずのない配線や PC などを見せないよう、実験者用の机と実験協力者用の机の間にパーティションを設置した。

対話ロボットのジェスチャには、腕を上下に振る、体の向きを発話者の方に合わせる、質問するとき首を横に傾ける、ユーザの発話に対して「ウンウン」と頷く、がある。その中でも、ユーザ発話直後に見られる「ウンウン」という頷きは、対話が成立しているとユーザに感じさせやすくするため、単純な質問応答であってもユーザの対話意欲が高くなったと考えられる。(文献[8]) 体の向きは約 1 秒かけて 30 度ほどもう一台のロボットの方へ傾ける。腕のジェスチャも約 1 秒かけて腕を 45 度近く上下に動かす。これを 3 回ほど繰り返す。

対話システムについては、Sugiyama ら [3] の対話システムを使用した。

対話内容については、自由な対話を想定して行った実験であったため、ユーザには自由に発話させた。制約のない対話を実現するために、システム側はユーザの多岐にわたる発話内容に対応できるよう、様々なトピックを用意した。そのため、実験協力者ごとに対話

内容はばらついている。対話破綻をきたしたり、対話感が著しく低い状態であると判断するような対話があれば評価から取り除く予定であったが、全ての対話を通して明確に対話破綻を起こすような事例は発生しなかったため、評価から取り除いた対話はなかった。

以上の手法を用いることで、対話がある程度の長さであっても、文脈を考慮した自然な対話が実現できていることを確認した。

3.4 実験条件

本実験では対話ロボットの台数とジェスチャ表出の有無を変化させて対話意欲を比較する。表2のように、実験協力者を1台のロボットと対話させる条件（以降条件1（1台対話））、2台のロボットと対話させる条件（以降条件2（2台対話 G 有））、ジェスチャ表出を行わない2台のロボットと対話させる条件（以降条件3（2台対話 G 無））に分けて対話実験を行った。

実験協力者として日本語話者24名（男18人、女6人、20-23歳、口腔内疾患が認められない人）に協力していただいた。各条件における実験協力者数は、条件1（1台対話）が5人、条件2（2台対話 G 有）が12人、条件3（2台対話 G 無）が7人であった（表3）。実験開始前の食事制限を満たせなかった者や、被験者都合により実験が行えなかった者がいたため、条件1及び条件3の人数は当初の予定よりも少なくなった。本実験は一人の実験協力者に対して5対話（1対話につき10発話）・計50発話させるため、1条件でも実験の所要時間が1時間程度に及ぶことになった。そのため、過度の疲労を避けるため被験者間実験とし、各条件ごとに被験者を集めることにした。

2台の対話ロボットの音声はそれぞれ別の音声をを使用したため、ユーザは音声を聞くことで話者交代を認識することができる。音声の発話モデルはNTT製の合成器^(注1)を使用した。子供の声をベースに話速とピッチを調整した。結果として、2台のロボット間でピッチは平均すると1割程度異なっている。

また、条件1（1台対話）や条件2（2台対話 G 有）ではジェスチャ表出を行う。そのため、条件2（2台対話 G 有）は音声とは別に対話ロボットのジェスチャから話者交代を認識することができる。しかし条件3（2台対話 G 無）ではジェスチャ表出を行わないため、条件2（2台対話 G 有）と比べると話者交代を認識するための情報が少ない条件であるといえる。

表2 実験条件

		
条件1	条件2	条件3
1台対話	2台対話	2台対話
ジェスチャ有	ジェスチャ有	ジェスチャ無

表3 実験協力者詳細

項目	詳細
人数	条件1（1台対話）：5人 条件2（2台対話 G 有）：12人 条件3（2台対話 G 無）：7人 合計：24人（男性18人、女性6人）
年齢	20-23歳
その他備考	口腔内に異常が見られない 実験開始1時間前から食事禁止

3.5 仮説

分析を行うにあたり、各対話を意欲率の平均値を基準として対話意欲が高いセッションと低いセッションに分けた。また、発話長の平均値を基準として発話長が長いグループと短いグループに分けた。ユーザが行った全ての対話を、対話意欲の高低と発話長の長短の4通りに分けることで分析を行った。以下に本実験の仮説についてまとめる。

- 仮説1: 条件1（1台対話）よりも条件2（2台対話 G 有）の方が対話意欲が高い
- 仮説2: 条件3（2台対話 G 無）よりも条件2（2台対話 G 有）の方が対話意欲が高い
- 仮説3: 対話意欲が高いセッションよりも対話意欲が低いセッションの方が sAA 比が大きい
- 仮説4: 対話意欲が高いセッションでは、発話長が短いグループ（以降グループ2（高短））よりも発話長が長いグループ（以降グループ1（高長））の方が sAA 比が小さい
- 仮説5: 対話意欲が低いセッションでは、発話長が長いグループ（以降グループ3（低長））よりも発話長が短いグループ（以降グループ4（低短））の方が sAA 比が小さい

条件1（1台対話）では1台のロボットと対話する

(注1) : <https://www.futurevoice.jp/>

が、条件 2 (2 台対話 G 有) では 2 台のロボットと対話を行う。そのため、ユーザは条件 2 (2 台対話 G 有) の場合はロボット同士の対話を見聞きすることができる。ロボット同士の会話はユーザの発話理解を容易にする効果が見込まれる [7] ため、条件 2 (2 台対話 G 有) の方が条件 1 (1 台対話) よりも対話が行いやすく、対話意欲も高まると仮説を立てた (仮説 1)。

条件 3 (2 台対話 G 無) では 2 台のロボットがジェスチャ表出を行わないため、条件 2 (2 台対話 G 有) に比べてロボットがユーザに提示する情報が少ない。そのため、条件 2 (2 台対話 G 有) は条件 3 (2 台対話 G 無) に比べて次話者予測が容易になると考え、対話意欲が高まると仮説を立てた (仮説 2)。

対話実験では少なからず何らかの発言を要求されるため、平均値よりも高い意欲率となった対話はストレスを感じにくく、平均値よりも低い意欲率となった対話はストレスを感じやすくなると考えた。そのため、対話意欲が低いセッションの方が対話意欲が高いセッションよりも、sAA 比が大きくなると仮説を立てた (仮説 3)。

また、対話意欲が高いセッションでは、ユーザが発言したがっていると考えられるため、よりたくさん話すことで欲求が満たされると考えた。そのため、発話長が長いグループ (たくさん発話したグループ) の方が、発話長が短いグループ (あまり発話できなかったグループ) よりも sAA 比が小さくなると仮説を立てた (仮説 4)。仮説 4 は、対話意欲が高いセッション内で比較を行う。発話長が長いグループは対話意欲が満たされたため、発話長の短い対話意欲が満たされなかったグループよりも sAA 比が小さくなるとの考えに基づく仮説である。

同様に、対話意欲が低いセッションでは、ユーザが発言をしつづける状態と考えられるため、よりたくさん話すすとストレスの増加につながると考えた。そのため、発話長が短いグループの方が、発話長が長いグループよりも sAA 比が小さくなると仮説を立てた (仮説 5)。仮説 5 は対話意欲が低いセッション内で比較を行う。仮説 4 とは逆に、対話意欲が低いセッション内では、発話長が長いグループは短いグループに比べて sAA 比が大きくなるとの考えに基づく仮説である。

sAA 比を比較するため、各対話で得られた意欲率と発話長の平均値を境界値としてグループ分けし、対応する sAA 比の大小関係をまとめると、仮説 4 と仮説 5 は表 4 のような関係になると考えた。

表 4 発話長と sAA 比の関係 (仮説)

	発話長が長い	発話長が短い
意欲率が高い	グループ 1 (高長) sAA 比 小	グループ 2 (高短) sAA 比 大
意欲率が小さい	グループ 3 (低長) sAA 比 大	グループ 4 (低短) sAA 比 小

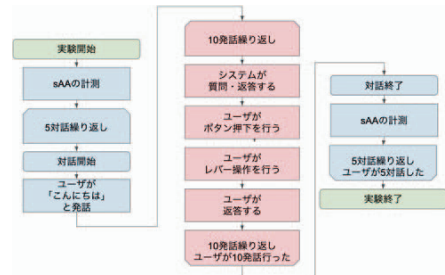


図 2 実験手順フローチャート



(a) 1 台対話 (b) 2 台対話 G 有 (c) 2 台対話 G 無

図 3 対話中の様子

3.6 実験手順

実験手順のフローチャートを図 2 に、実験中の様子を図 3 に示す。

まず、実験協力者に対話ロボットと対話することを教示した。教示後に口腔内に異常がないか、実験開始 1 時間前から食事を摂取していないか確認した。その後 sAA を計測し、計測が終わってから実験協力者を 5 分間安静にさせた。その後に再び sAA を計測した。計測が終わったら、対話中の対話意欲の評価と、質問したロボットの推定に使うレバー操作に慣れさせるため、1 対話を通して本実験と同じように評価させた。訓練で取得した評価データは結果に含まない。訓練が終了したら対話実験を開始する。

対話はロボットが実験協力者に質問する形式で進行するため、実験協力者は質問に返答していく。このとき、実験協力者の発話 (以降ユーザ発話と呼ぶ) 前に、毎回対話意欲評価のため「もっと対話したい」か「もう対話したくない」かの評価を手元のボタンで入力させた。「もっと対話したい」場合は左手側にある青い

ボタンを、「もう対話したくない」場合は右手側にある赤いボタンを入力させた。その後、条件2（2台対話G有）と条件3（2台対話G無）の場合は、最後に質問したロボットが実験協力者から見て右側のロボットか、左側のロボットかを手元のレバーを傾けることで入力させた。条件1（1台対話）の場合はそのまま次の発話に移る。次の発話に移ると、再びロボットが発話を始め、実験協力者へ質問を行う。実験で行った対話例を付録に記載した。

4. 結果と各条件間の比較

4.1 実験結果

対話実験で得た意欲率の結果を図4に示す。条件1（1台対話）、条件2（2台対話G有）、条件3（2台対話G無）の意欲率の平均と標準偏差はそれぞれ表5のようになり、条件2（2台対話G有）は平均値が最も高く、標準偏差が最も小さいことが分かる。条件1（1台対話）、条件2（2台対話G有）、条件3（2台対話G無）でクラスカル=ウォリス検定を行った結果、 $p < 0.01$ となり有意差が認められた。そのため、被験者数の偏りによる多少の影響は考えられるが、3条件の関係性が変わる可能性は低いと考えられる。

以上のことから、条件2（2台対話G有）は他の条件と比べて意欲率が有意に高いことが分かり、意欲率が比較的高い範囲に集中していることが分かる。一方条件3（2台対話G無）は意欲率が最も低く、意欲率

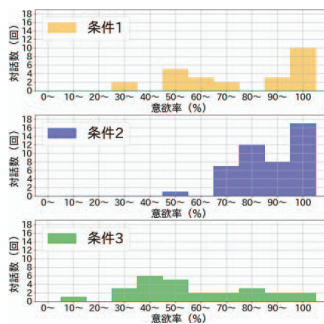


図4 各条件における意欲率のヒストグラム。横軸は意欲率、縦軸はその対話が発生した回数。

表5 条件1（1台対話）、条件2（2台対話G有）、条件3（2台対話G無）の意欲率の平均と標準偏差

	平均値	標準偏差
条件1（1台対話）	0.80	0.22
条件2（2台対話G有）	0.90	0.11
条件3（2台対話G無）	0.61	0.23

の低い対話から高い対話まで広く分散していることが分かる。

次に、sAA比の結果について述べる。対話意欲が低いユーザはあまり話したがらないと考えられるため、長い発話を行うことはストレスとなって、短い発話を行ったユーザに比べてsAA比が高くなると仮説を立てていたが不成立となった。本実験の結果が得られた原因として、ユーザが感じる「長く話した」という基準がユーザ個人ごとに異なることが挙げられる。本実験では全ユーザの発話長の平均より大きい小さいかで発話長の大小関係を論じてきた。しかし発話長が長いかどうかはユーザ個人ごとに異なるため、ユーザに合わせた基準を設ける必要がある。更に、sAAに関しても個人差があり、sAAの平均値が大きい人ほど標準偏差も大きくなる傾向が見られた。そのためsAA比を用いたが、それでもなお個人差が大きく影響していた。例えば、sAA値の平均が45となったユーザは45, 92, 40, 3と推移し、sAA比は0.49, 2.3, 13.3となった。一方で、sAA値の平均が5となったユーザは5, 7, 5, 3と推移したため、sAA比は0.71, 1.4, 1.7となった。

4.2 条件1（1台対話）と条件2（2台対話G有）の比較

対話実験の結果、条件1（1台対話）の意欲率の平均値が0.77だったのに対して、条件2（2台対話G有）の意欲率の平均値は0.89となり、条件2（2台対話G有）の方が意欲率の平均値が高いという結果が得られた。条件1（1台対話）での対話のうち、79.9%（発話数）の評価が「もっと対話したい」だったのに対して、条件2（2台対話G有）での対話では88.3%の評価が「もっと対話したい」であった。条件1（1台対話）と条件2（2台対話G有）間で χ^2 検定を行ったところ、 $p < 0.01$ となり有意差が認められた。

4.3 条件2（2台対話G有）と条件3（2台対話G無）の比較結果

対話実験の結果、条件3（2台対話G無）の意欲率の平均値が0.58であったのに対して、条件2（2台対話G有）の意欲率の平均値は0.89となり、条件2（2台対話G有）の方が意欲率が高くなるという結果が得られた。条件3（2台対話G無）での対話のうち、59.8%（発話数）の評価が「もっと対話したい」だったのに対して、条件2（2台対話G有）での対話では88.3%の評価が「もっと対話したい」であった。条件3（2台対話G無）と条件2（2台対話G有）の間で χ^2 検定を

表6 発話長と sAA 比の関係 (結果)

	発話長が長い	発話長が短い
意欲率が高い	1.17	1.21
意欲率が低い	0.92	1.03

行ったところ、 $p < 0.01$ となり有意差が認められた。

また、2 台あるロボットのうち、ユーザへ質問を投げかけたロボットがどちらだったかを推定させたときの正解率は、条件 2 が 100% であったのに対し、条件 3 は 84.4% となった。この結果から、条件 2 (2 台対話 G 有) では全ての質問でどちらのロボットが質問を投げかけたかを正しく認識できていたのに対して、条件 3 (2 台対話 G 無) では全対話中約 16% の質問で発話ロボットを誤って認識していたことが分かった。

4.4 発話長と sAA 比の関係

sAA 比と発話長の関係を表 6 に示した。この結果から、対話意欲が高い場合では、発話長が短かった対話よりも発話長が長い対話の方が sAA の変化が小さいことが分かった。また、対話意欲が低い場合では、発話長が短くなるような対話よりも発話長が長くなるような対話の方が sAA の変化が小さいことが分かった。そのため、グループ 1 (高長) よりもグループ 2 (高短) の方が、また、グループ 3 (低長) よりもグループ 4 (低短) の方がストレスが高い結果となった。

5. 考 察

5.1 ロボットの台数増加に伴う対話意欲の変動要因

条件 1 (1 台対話) と条件 2 (2 台対話 G 有) を比較すると、条件 2 (2 台対話 G 有) の方が条件 1 (1 台対話) よりも意欲率の平均値が高くなる結果となった。これは、1 台対話に比べて 2 台対話の方がユーザへの負担が小さく、対話意欲を減ずる要因が少なかったためだと考えられる。条件 1 (1 台対話) では 1 対話中に占めるユーザ発話の割合が大きく、条件 2 (2 台対話 G 有) に比べてユーザへの負担が大きい。また、条件 2 (2 台対話 G 有) ではユーザ発話の直前にロボット同士の対話が行われるが、条件 1 (1 台対話) では対話ロボットが 1 台しかなくロボット同士の対話がないため、ユーザ発話後にロボットが一言話し、すぐユーザへの質問に移る。そのためユーザは休む間もなく質問責めされていると感じやすい。条件 1 (1 台対話) のロボットが 1~2 発話した後ユーザに質問している例を付録 1. に示す。一方で条件 2 (2 台対話 G 有) は、ユーザ発話の直後にロボット同士が対話をしている。

付録 2. に条件 2 (2 台対話 G 有) におけるロボット同士の対話例を示した。付録 2. の 1 発話目と 2 発話目、5 発話目から 8 発話目などから、Sota1 と Sota2 が対話していることが分かる。そのためユーザは発話後に時間的余裕が生まれ、質問責めされていると感じにくかったと考えられる。

また、2 台対話では、ロボットがユーザへ質問する前から質問の返答を考えることができたことも、2 台対話の方が対話意欲が高い要因としてあげられる。ロボット同士の対話はユーザへの質問の直前に行われるため、この対話に係る質問をユーザに投げかけることが多い。そのため、条件 2 (2 台対話 G 有) の実験協力者は次に尋ねられる質問を前もって予測することができ、その質問に対する返答にロボット同士の対話を参考にすることができた。例えば、付録 2. の 7 発話目以降は、Sota1 が Sota2 に向かって好きな食べ物について質問している。このやりとりを聞いたユーザは好きな食べ物について考えるだろう。その後、Sota1 がユーザに「あなたはどうですか?」と質問している。ロボット同士の一連の対話を見ていたユーザは心理的余裕が生まれ、対話を続けたいと思う要因になったと考えられる。もし心理的余裕がない状態であれば、ユーザは他のことについて考えようと思わなくなり、対話意欲が低下すると考えられる。そのため、心理的余裕を与えることで対話意欲の低下が避けられると考えられる。

他の要因として、ロボット同士の対話が面白いと感じたことも挙げられる。実験協力者の中にはユーザ発話を正しく聞き取れていることに感心する者や、対話ロボット同士の対話を聞いて楽しんでいる者がいた。特に、ロボット同士の対話を聞いて楽しむという感情は、「今発話した内容がロボット達の対話にどのように影響を及ぼすか」というようなワクワク感につながりやすく、このような感情がポジティブにはたらくことで、もっと対話したいと思わせる要因になったと考えられる。

以上をまとめると、対話ロボットの台数を 1 台から 2 台に増やすことで、ユーザは質問されるまでに時間的余裕が生まれ、その時間にロボット同士の対話が見られるため、心理的な余裕も生まれると考察できる。

5.2 ジェスチャ表出に伴う対話意欲の変動要因

条件 2 (2 台対話 G 有) と条件 3 (2 台対話 G 無) を比較すると、条件 2 (2 台対話 G 有) の方が条件 3 (2 台対話 G 無) よりも意欲率の平均値が高い結果となっ

た。理由の一つとして、実験協力者が発話ロボットがどちらかを特定する際にジェスチャを判断材料としていることが関係していると考えられる。ジェスチャがない条件3（2台対話 G 無）では、実験協力者は音声でどちらから聞こえてくるかを判断する必要があるが、約 16% の発話で判断を誤っている。一方で条件2（2台対話 G 有）の場合はジェスチャがあるため、ロボットを一目見ればどちらが発話しているか判断できるので、特定が容易である。実際、条件3（2台対話 G 無）で対話を行った実験協力者が、実験終了後に「どちらのロボットが話しているのか分からなかった」という感想を述べており、発話ロボットの特定が困難であった可能性がある。発話者の特定は、ターンテイキングを円滑に行うためには欠かせない要素であるため、話者特定のしづらい条件3（2台対話 G 無）より話者特定のしやすい条件2（2台対話 G 有）の方が対話意欲が高くなったと考えられる。ロボットからユーザへターンテイキングを行うとき、ユーザはロボットの質問が自分に向けられたものなのか、もう一台のロボットに向けられたものなのかを判断しなければならない。ここで、もう一台のロボットに向けられた質問にもかかわらず、ユーザに向けられた質問だと誤認識してしまうと、ユーザ発話とシステム発話が重なってしまい、ユーザに不快な感情を与えてしまう。特に、条件3（2台対話 G 無）では4人のユーザで発話がシステム発話と衝突したことを確認した。このような状況がユーザの対話意欲低下につながったと考えられる。一方条件2（2台対話 G 有）はジェスチャがあるため、ターンテイキングがうまく行われ、条件3（2台対話 G 無）に比べて不快な感情を与えなかったと考えられる。

ジェスチャ表出を行わないと発話が重なるだけでなく、話者交代が行われないために誰も発話しなくなる状況も発生する。条件2（2台対話 G 有）では発話ロボットを視覚的に判断できるが、条件3（2台対話 G 無）だと視覚的に判断できる材料がない。2台の対話ロボットの音声は異なるため、音声から発話ロボットを識別することは可能であるが、2台とも声域が比較的高く、同一個体の変化の範囲内として誤認される可能性もあった。そのため、ロボットが2台で発話していても1台のロボットが発話しているように感じる可能性がある。ロボット A → ロボット B → ロボット A という順番でターンテイキングが行われたとする。このときにロボット B の発話をユーザが誤ってロボット A の発話だと誤認識すると、ロボット A が話し続け

ていると解釈するため、ロボットが自問自答しているように聞こえる。そのような状態になってしまうと、ユーザはどのタイミングで発話をすれば良いか分からなくなり、対話が始まらなくなってしまう。実際に条件3（2台対話 G 無）で対話した7人のユーザ全てにおいて、1対話中少なくとも1回はロボットの質問→ユーザ発話の間に5秒以上の沈黙があったことを確認した。ユーザは自分自身に質問されていることにすぐに気づかなかったため、沈黙状況から現在の発言権がユーザ自身にあると解釈し、慌てて喋り出していた。そのため、ユーザは「いつ発言権が自分に移るのか」と常に意識しなければならず、ユーザの負担が増えたものと推察できる。一方、条件2（2台対話 G 有）では、ジェスチャ表出を行うことで視覚的に非言語情報の共有が行えたといえる。ロボットが質問者に向けて体や手の方向を変えるなど話者交代が把握しやすい状態であり、対話意欲が高いという結果が得られた。したがって、ロボットは誰に質問をして、いつ話し終えたのか、という非言語情報をユーザと共有する必要があるといえる。

以上のとおり、対話しているときに今誰が話しているのか、誰に向けて話しているのかという非言語情報の共有がおろそかになると、対話意欲が低減することにつながる可能性が考えられる。

5.3 対話意欲と sAA 比の関係

仮説3が不成立であった原因として次の2点が考えられる。一つ目は、対話意欲の低いセッションの対話前 sAA 値が既にか高かったため、sAA 比が小さくなったという点。二つ目は、対話意欲が低く気楽に実験に臨んだため、実験を通してユーザへの負担が少なかったという点である。実際に、対話前の sAA 値を計測したところ、対話意欲が高いセッションは平均 16.8、標準偏差 14.0 だったのに対して、対話意欲が低いセッションは平均 36.4、標準偏差 19.6 となった。そのため、対話前の sAA 値は対話意欲が低いセッションの方が高いことが分かり、仮説3が不成立となった原因でもあると考えられる。

5.4 発話長と sAA 比の関係

発話意欲が高いとたくさん話したいという欲求があるため、対話意欲が高く発話長が長いグループ1（高長）は欲求が満たされた状態といえる。逆に、対話意欲は高いが発話長が短いグループ2（高短）は欲求が満たされなかったといえる。そのため、グループ1（高長）はグループ2（高短）に比べて sAA 比が小さく

なったと考えられる。これはグループ 1（高長）の方がストレスを溜めにくいグループだといえる。そのため、対話意欲が高いユーザに対しては Yes/No question のような簡単な質問ではなく、理由を尋ねるような、回答が長文になると期待できる質問を投げかける方がユーザのストレスが増加しにくい可能性がある。

しかし本実験では、仮説とは異なった結果が得られた。対話意欲が低いユーザはあまり話したがりなないと考えられるため、長い発話を行うことはストレスとなって sAA 比が短い発話を行ったユーザに比べて sAA 比が高くなると仮説を立てていた。本実験の結果が得られた原因として、ユーザが感じる「長く話した」という基準がユーザ個人ごとに異なることが挙げられる。本実験では全ユーザの発話長の平均より大きい小さいかで発話長の大小関係を論じてきた。そのため、どのユーザも同一の基準で「発話長が長いグループ」と「発話長が短いグループ」に分けられた。しかし発話長を長いと感じるかどうかはユーザ個人ごとに異なるため、ユーザに合わせた基準を設ける必要がある。比較的長く話したユーザであっても、その人が「話しすぎた」や「全然話せなかった」と感じなければ、sAA 値の増減として現れない可能性が高い。同様に、本実験の手法では、相対的に長く発話したため、「発話長が長いグループ」に割り振られたにもかかわらず、「あまり発話したと感じなかった」というユーザや、その逆となるユーザが含まれてしまう可能性がある。また、sAA は個人差が大きく、各人の平均を比較すると 4 から 78 と最大で 18 倍の差が見られた。そこで sAA 比を用いたが、それでもなお個人差の正規化が不十分で発話長と sAA 比との関係が適切に示せなかった可能性がある。今後の課題として、発話長の長短となる基準をユーザごとに設けるといった点や、sAA の個人差による影響をより小さなものにするため、個人ごとの適切な正規化を行うことが挙げられる。

6. む す び

本論文ではユーザと対話ロボットに雑談対話を行わせることで、ユーザの対話意欲の変動要因について分析を行った。1 台対話と 2 台対話ジェスチャ表出有り、2 台対話ジェスチャ表出なしという 3 条件間での意欲率を比較することで、ユーザが自由に発話できる雑談対話という環境においても、ジェスチャや傾き、対話ロボットの複数台化が対話意欲向上に寄与していることを確認した。特に、2 台対話ジェスチャ表出有

りの条件が最も対話意欲が高くなった。これはユーザが「対話できていない」と感じるような発話をロボットがしても、もう 1 台のロボットがフォローすることでユーザが対話を続けられたと感じたためだと考えられる。また、ロボットが傾く場面をユーザが見ることで、発話を理解しているとユーザが感じたためだとも考えられる。話者交代を行う際にロボットが体の向きを変えたり、身振り手振りすることで発話者であることを視覚的に伝えたことも、対話意欲が高まった要因となった。ジェスチャ表出を行わない条件では、ユーザ発話とロボットの発話が重なり、話者交代が適切に行われない場面が見られた。更に、ロボット同士の対話をユーザに見せることで、ユーザは次の発話までの時間に心理的余裕をもつことができたことも挙げられる。

本実験を通して対話意欲と sAA の間に明確な関係性は示されなかった。しかし、sAA の正規化や発話長のグループ分けに不十分な点があることから、今後の実験次第で対話意欲との関係を示せる可能性がある。sAA はストレス指標の中でも計測時間が短いという特徴がある。対話中での計測こそできないものの、対話実験での利用は妥当であると考えられる。

これまではある特定環境下での実験を行うことで、非言語情報などの有効性が示されてきた。しかし本研究から、雑談対話というユーザにとって自由度の高い発話が行える環境下においても先行研究と同様の結果が得られることが明らかとなった。また、対話ロボットの複数台化についてポジティブな要因について議論されることが多かったが、非言語情報を適切に扱えなければ、1 台の対話ロボットよりも対話意欲が低くなるという新たな結果が得られた。対話システムの面で見ても、WoZ 実験など人の手を介することなくロボットが独立して対話を行う実験であった。そのため、今後対話システムが導入される実環境に近い条件下での知見を得ることができた。

謝辞 本研究を行うにあたり、ご協力して下さったヒューマンインタラクション研究室の皆様、実験協力者の皆様に御礼申し上げます。

文 献

- [1] 長谷川恭久, “AI の進化から学ぶ会話型 UI の課題,” <https://yasuhisa.com/could/article/ai-and-conversational-ui/> (2020 年 3 月 13 日)
- [2] 飯尾尊優, 吉川雄一郎, 石黒 浩, “ロボットの複数体化が対話感に及ぼす影響: 展示会におけるボタン入力対話体験の評価,” The 30th Annual Conference of the Japanese Society

for Artificial Intelligence, 2016.

- [3] H. Sugiyama, T. Meguro, Y. Yoshikawa, and J. Yamato, "Improving dialogue continuity using inter-robot interaction," Proc. 27th IEEE Int. Symp. Robot and Human Interactive Communication, MoBT3.4, 2018.
- [4] 徳永弘子, 武川直樹, 寺井 仁, 湯浅将英, 大和淳司, "会話における参与役割志向態度に基づく順番交代分析," 知能情報ファジィ学会誌, vol.25, no.5, pp.889-900, 2013.
- [5] 湯浅将英, 徳永弘子, 武川直樹, "発話マインドに基づく発話交代モデル," 情処学シンポジウム論文集, vol.2008, no.4, pp.1-8, 2008.
- [6] 小野哲雄, 今井倫太, 石黒 浩, 中津良平, "身体表現を用いた人とロボットの共創対話," 情処学論, vol.42, no.6, pp.1348-1358, 2001.
- [7] 神田崇行, 石黒 浩, 小野哲雄, 今井倫太, 中津良平, "人-ロボットの対話におけるロボット同士の対話観察の効果," 信学論 (D-I), vol.J85-D-I, no.7, pp.691-700, July 2002.
- [8] 渡辺富夫, "うなずきロボット InterRobot," 日本ロボット学会誌, vol.24, no.6, pp.692-695, 2006.
- [9] 松原 大, 上田博唯, "愚痴を聞くロボットの提案," 信学技報, MVE2011-5, 2011.
- [10] 萩野谷浩美, 佐伯由香, "ストレス指標における唾液 α アミラーゼ活用の有用性," 日本看護技術学会誌, vol.10, no.3, pp.19-28, 2012.

付 録

1. 条件 1 (1 台対話) 対話ログ例

Sota1: テレビはよく観ますか？

user:あまり見ません

Sota1: そうなんですねー。うんうん。

Sota1: テレビは録画がたまるのです

Sota1: 好きなお笑い番組は何ですか？

user:わかりません

Sota1: あー、なるほど

Sota1: 家事は好きですか？

user:好きな方です

Sota1: そうなんですかー。

Sota1: 家事は女性がやらないんですって

Sota1: 育児をどう考えますか？

user:大事なことです

Sota1: そうなんですかー。うんうん。

Sota1: 育児は体力が大事ですね

Sota1: ペットを飼ってらっしゃいますか？

user:飼っていません

2. 条件 2 (2 台対話 G 有) 対話ログ例

Sota1: ランチでよく食べるのは何？

Sota2: パスタ料理です。

Sota2: ランチでよく食べるものは何ですか？

user:サンドイッチです

Sota2: サンドイッチなんですね

Sota2: サンドイッチはチーズが入ります

Sota1: 好きな食べ物？

Sota2: 蕎麦です。

Sota1: あなたはどうですか？

user:ハンバーグ好きです

Sota2: ハンバーグですかー

Sota1: 好きな果物は

Sota2: みかんです。

Sota1: 肉と魚のどちらが好きですか。

user:肉です

Sota2: 肉ですかー

Sota1: お肉とお魚はどちらが好きですか？

Sota2: 肉の方が好きです。

3. 条件 3 (2 台対話 G 無) 対話ログ例

user:こんにちは

Sota1&2: こんにちは

Sota1: テレビはよく観ますか？

user:よく見ません

Sota2: よくなんですね

Sota2: テレビは録画が出来るのです

Sota1: お見舞いに持っていく花のコーディネートと注文をやれませんか？

Sota2: やれます。

(2020 年 3 月 25 日受付, 7 月 22 日再受付,
9 月 8 日早期公開)



丸山 洸太 (学生員)

2019 年工学院大学情報学部コンピュータ科学科卒業。現在、同大学院情報学専攻修士課程 2 年に在籍中。ロボットを用いたインタラクションの研究に従事。



大和 淳司 （正員：フェロー）

1988 年東京大学工学部精密機械工学科卒業。1990 年同大学院工学系研究科精密機械工学専攻修士課程了。同年日本電信電話（株）入社。1998 年 MIT 電気工学及びコンピュータ科学科修士課程了。NTT ヒューマンインタフェース研究所、NTT コミュニケーション科学基礎研究所等を経て、2016 年工学院大学情報学部教授。博士（工学）。画像認識、HRI などの研究に従事。2010 年本会論文賞、喜安善市賞受賞。IEEE、ACM、情報処理学会、人工知能学会各会員。



杉山 弘晃

2007 年東京大学工学部機械情報工学科卒業。2009 年同大学院情報理工学系研究科知能機械情報学専攻修士課程了。同年日本電信電話（株）入社。人と自然な対話を行う雑談対話システムの研究に従事。博士（工学）。IEEE、情報処理学会、人工知能学会、言語処理学会各会員。