

# 人間とロボットの会話における「間」の検討

## An Examination of the Conversation Timing between Human and Robot

大藤聖菜<sup>1\*</sup> 馮建美<sup>1</sup> 小山大幾<sup>1</sup> 今井倫太<sup>1</sup>  
Kiyona Oto<sup>1</sup>, Kenmi Hyo<sup>1</sup>, Daiki Koyama<sup>1</sup>, Michita Imai<sup>1</sup>

<sup>1</sup> 慶應義塾大学

<sup>1</sup> Keio University

**Abstract:** あらゆる場面においてコミュニケーションロボットの導入が進んでいる現在、人間とロボットの会話の質の向上のために様々な研究がなされている。本研究では、より自然な会話を行うための重要な要素の一つとして、返答のタイミングや沈黙のような発話と発話の「間」に着目し、人間とロボットとの会話中に発生する間について調査実験をした。その結果、人間がロボットに対して使用する間には個人差がみられ、ロボットによる間は人間にとって会話の文脈上において意味のあるものとしては捉えられなかった。この結果をもとに、今後の実験課題や適切に間が使用できるロボットを設計するための方針を検討した。

### 1 はじめに

現在、あらゆる場面においてコミュニケーションロボットの導入が進んでいる。店舗先に置かれたロボットは、客の呼び込みから受付、商品の説明や商品の検索を行うなどの一通りの接客をすることができる。駅や観光地では人々の案内を行い、プレゼンをする場面においてロボットを使用することで人々の注目を得ることも可能だ。かわいらしい見た目をしたロボットは、子供たちやお年寄りの良い話し相手となる。これらは既に現場にて活躍しているロボットの使用例であり、実際にこれらのロボットと会話の経験がある人も多い。このようなロボットを現在よりもさらに実用的なものにしていくためには、人間と会話をするのと変わりのないような会話能力をロボットに付与することが必要である。

このような背景の中、ロボットとより自然な会話を行うための研究が盛んに行われている。その中でもさらに、ジェスチャー、あいづち、ふりむき、視線や表情などを取り上げた研究が広く行われており、これらの非言語的コミュニケーション要素は感情や意志を表していると言われている。ロボットが人と遜色ない会話をするには、上記の非言語コミュニケーションを用いて、感情や意志を持っているように振る舞う必要がある。

本研究では、発話と発話の中に発生する時間間隔を

「間」と称し、会話の質を向上するための非言語的コミュニケーション要素としてこれに着目する。例えば、会話中の沈黙や回答までの遅延などもこれに含む。人間と人間の普通の会話をイメージしても、何か答えづらい質問を投げかけられたときや、断りづらい依頼を受けたときなど、とっさの返答ができずに間において返事をしてしまうだろう。このように、間も感情を表す重要な要素であるが、あまり研究はされていない現状にある。

その中でも本研究に関連する従来研究について、まず第2章で取り上げた。次に、適切な間を設計していくにあたって必要であるのは人間側のロボットに対する振る舞いについての確認である。具体的には、人間はロボットに対してどのような間をつくるのか、そして、人間がロボットの間をどのように捉えるか、の2点を明確にすることであり、第3章でその仮説を立て、そのために行った調査実験について第4、5章でまとめる。そして、第6章において、第4、5章の知見を踏まえ、間を利用した適切な振る舞いをするロボットの設計方針について検討を行う。

### 2 従来研究

#### 2.1 人間同士の会話中の沈黙に関する研究

人間とロボットの会話を設計していくにあたって、まずは人間同士の会話に着目した。間の捉え方や間のもつ

\*連絡先：慶應義塾大学 理工学部情報工学科  
〒223-0061 神奈川県横浜市港北区日吉 3-14-1  
E-mail: oto@ailab.ics.keio.ac.jp

意味は人間同士の会話においても様々であり、これまで社会心理学や言語学などの分野において研究されてきた [1]-[4]。話者交替の視点からは、沈黙をその出現位置をもとに Pause, Gap, Lapse と Significant/Attributable silence の 4 種類に分類することができる [1][2]。このうち、話者交替時に現れるのは、話し手がターンを終えて聞き手がターンを取るまでの Gap、次の発言権を特に求めずその後続く発話の無い沈黙である Lapse、そして、指名された次の話し手が何かしらの理由により発話を開始しない沈黙の Significant/Attributable silence である。また、T.J.Bruneau は心理的沈黙、相互作用的沈黙、社会文化的沈黙の 3 種類の沈黙を定義した [3]。D.Kurzon は、沈黙の語用論的な立場から意図的な沈黙と非意図的な沈黙があるとした [4]。

以上のような人間同士の会話でみられるような沈黙が、人間とロボットの会話においてもみられると予想する。従って、これらの沈黙に関する知見をもとにして、人間とロボットの間について分析していくことが可能であると考えられる。

## 2.2 ロボットの間に関する研究

間を対ロボットにおける非言語的コミュニケーションの要素として取り扱った研究も、数は多くないが行われたきた。

ロボットの反応の適切なタイミングについての研究から、即座の返事よりも 1 秒間ほど間を置いた返事の方が好まれたという結果が出ている [5]。早く返答するのが必ずしも最適であるとは限らないことが分かる。一方で、ユーザからシステムに対してインプットがあったからシステム側が反応するまでに 2 秒以上かけるべきではないという「2 秒ルール」が、コミュニケーションロボットにおいても有効であることもわかっている [5]。また、ロボットの発する「えっとー」などの会話的フィラーは人間に対する印象を向上させた [5][6]。ロボットに発話のターンが回ってきているという発話義務が明確な条件下においては、ロボットが音声・動作フィラーを見せることで沈黙の気まずさが軽減した [7]。しかし、人間とロボットは同等の立場であるとは言えず、発話ターンがどちらにあるか分からない時には人間の方に沈黙を破る義務が働く [7]。

ロボット側の沈黙も何かしらの意味を持ち、気まずさのようなネガティブなイメージを与えることは分かる。しかし、さらに踏み込んで間の意味付けをしたとき、2.1 で挙げたような人間同士の沈黙と同じような意味付けを人間がするかどうかは分かっていない。具体的には、人間はロボットの沈黙を、「考えている」、「答えたくない」、「嫌そう」、などと解釈するか、ということである。

また、従来研究において、ロボットのつくる沈黙にフィラーを加える会話デザインについては言及されているが、人間とロボットの双方がつくる間そのものを利用させるデザインまでには至っていない。つまり、ロボット自身が間を間として取得、使用できるようにさせることが本研究の狙いである。

## 3 仮説

人間はロボットに対してどのような間をつくるのか、そして、人間がロボットの間をどのように捉えるか、という問題に対して次のような仮説を立てた。

**仮説 1:** 会話相手がロボットの場合、人間はロボットに対して多くの間を使用しない。

人間側として、返答するまでにロボットを待たせてはいけないという意志がはたらくと考えられるからだ。

**仮説 2:** ロボットの使用する間は、文脈上において意味のある間として捉えられる場合と処理時間などの機械的な間として捉えられる場合がある。

## 4 実験

仮説を検証するための実験を実施した。一つの実験であらゆる種類の間について扱うことは難しいため、今回は隣接ペア [8] が明確になるような設定で実験をした。隣接ペアとは、基本的な相互作用を達成する発話対のことであり、「質問—答え」のように 2 つの部分から構成され、この場合であれば「質問」が第 1 部分、「答え」が第 2 部分となる。隣接ペアの第 2 部分にくる沈黙に着目できるように、一問一答形式をベースにした会話スクリプトを設定して会話を行った。第 1 部分のパートと第 2 部分のパートを人間とロボットに交互に行わせた。

### 4.1 実験の設定

今回、被験者と会話を行うコミュニケーションロボットとして pepper を使用した。pepper には会話に必要な台詞をあらかじめ設定しておき (表 1, 2)、間を置いた反応や即座の反応にも対応できるようにした。その発話実行命令は Wizard of Oz (WOZ) 法で行った。

ロボットは、発話時にはランダムなジェスチャー (腕の動き) をし、それ以外のときは瞬きを模倣した目の光の動作と呼吸を模倣した腕のわずかな上下運動を繰り返している。間の意味合いに影響を与えず、かつ自然な会話をさせるため、最低限の動きだけをやる狙いである。また、顔認識により相手の顔を追跡して、常に相手の方向を向くように設定してある。

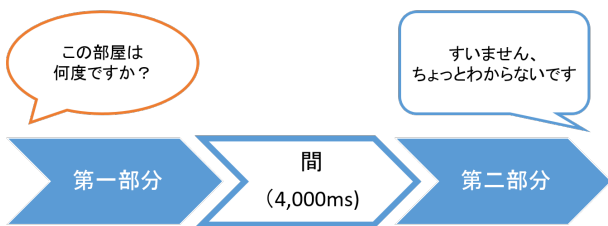


図 1: 実験 2 の様子

#### 4.1.1 実験 1 の設定：ロボットが第 1 部分、人間が第 2 部分

ロボットが挨拶、質問や依頼をし、それに対して被験者が返答していくという設定で短い会話を行った。ロボットは答えが返ってきたときに、ランダムであいづちも返した。隣接ペアのやりとりは 1 人の被験者に対しておよそ 14 往復ほど行った。被験者がどのようなタイミングで返答するかを確認した。

表 1: 実験 1 のロボットの台詞例  
第 1 部分 (ロボット)

---

こんにちは。  
 名前を教えてください。  
 誕生日はいつですか。  
 出身はどちらですか。  
 趣味はなんですか。  
 僕のことどう思っていますか。  
 あなた自身についてどう思っていますか。  
 嫌いな人っていますか。  
 この実験って面倒じゃないですか。  
 何か秘密を教えてくださいませんか。

---

#### 4.1.2 実験 2 の設定：人間が第 1 部分、ロボットが第 2 部分

被験者が挨拶、質問や依頼をし、それに対してロボットが返答していくという設定で短い会話を行った。被験者には質問リストを渡し、その中から好きな順に第 1 部分の呼びかけをしてもらった。6 問ある質問中、3 問の返答時にロボットは 4,000ms の間をとってその後に発話した。この 4,000ms というのは、「2 秒ルール」[5] の倍の時間である。この間に対して被験者がどのような反応をするかを確認した。

## 4.2 実験の流れ

学生 3 人に被験者として実験に参加してもらった。それぞれの被験者には、まず、ロボットが第 1 部分、人間が第 2 部分の場合の会話を行ってもらった。その後、第 1 部分と第 2 部分の役割を入れ替えて会話を行った。実験終了後、ロボットとの会話におけるアンケート調査とインタビューを行い、第 2 部分の発話までに発生した間についての印象を確認した。また、実験時に撮影した動画と音声をもとに、被験者の動作と回答までの時間を分析した。例として図 1 に実験 2 の様子を示した。

## 5 結果

### 5.1 実験 1：ロボットが第 1 部分、人間が第 2 部分の場合

被験者 3 人の間には大きな差があり、人間のつくる間にはかなりの個人差があると言える。実験結果は図 2 にまとめた。ある人は、ほとんどの質問に間をあげず、おおよそ 500ms 以内には返事をした。その理由として、ロボットからの質問が答えるのにそこまで難しいものでなかったこと、ロボットに対してまともに考えて答えても仕方のない内容だと思ったことがインタビューからわかった。

その他の被験者は質問によっては返答までに間があった。今回のやり取りの中では、第 2 部分の発話まで平均して 2,300ms かかっており、完全な回答をするまで最長で 7,000ms を超えることもあった。けれども、この回答までには「うーん」などの発話付きであり、完全な沈黙ではなかった。質問によっては、答えるのが難しかったとの感想を抱かれたものもあり、やはりその質問では回答までに時間がかかっていた。

したがって、人がロボットに対して使用する間には個人差があり、全く間を使用しないわけではなかった。間の発生理由までははっきりしなかったが、少なくとも

表 2: 実験 2 の台詞例

第 1 部分 (人間)	第 2 部分 (ロボット)
こんにちは。	こんにちは。
あなたは何かできるロボットなんですか。	基本的なお喋りができますよ。あとは矢上キャンパスの案内などです。
矢上キャンパスの好きな場所は？	(問) やっぱり図書館ですね、静かで落ち着きますよ。
好きな食べ物は何ですか？	僕の好きな食べ物は白米だと思います。
スマートに見えますね。	あ、ありがとうございます。嬉しいです。
緊急停止ボタンはどこにありますか？	(問) それはもちろん秘密です。
この部屋は何度ですか？	(問) すいません、ちょっとわからないです。

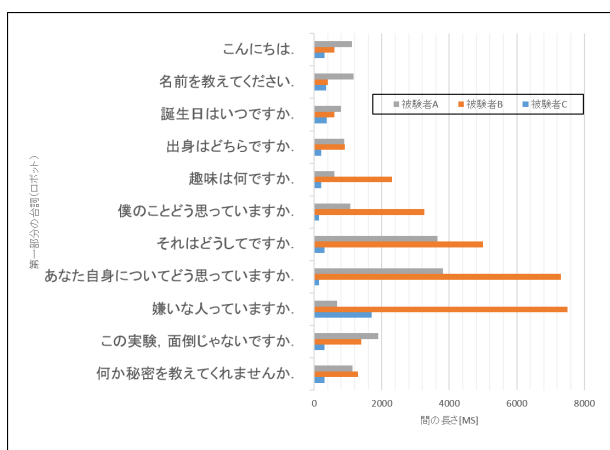


図 2: 実験 1 の結果

も人間側にも返答するまでの処理時間があり、それがロボットに対しても反応として表れていた。

## 5.2 実験 2: 人間が第 1 部分, ロボットが第 2 部分の場合

今回の実験でのロボットの設定, 会話の設定だと, ロボットのつくる間は, 文脈上において意味をもつものとして被験者に捉えられることはなかった。ロボットが間をおいているとき被験者は, 質問を 2 度繰り返す, ロボットをのぞき込む, 首をかしげるなどの行動をみせた。ロボットが返答するまでの 4,000ms の間は, ロボットの音声認識に問題がある, あるいはロボットの処理時間に時間がかかっているように捉えられた。

## 6 「間」の設計に関する考察

今回の調査実験を通して, 人間のつくる間にはかなりの個人差が発生していることがわかった。それは人間のロボットに対する慣れや会話内容にも左右されるだろうが, ロボットが相手の会話に対する態度ごとに間を合わせていく仕様にするようなデザインには大き

な意義がありそうだ。例えば, 会話の中で発生していく間の長さを時系列データとして扱い, ロボットに学習させ, 適切な間を生成させていくことが可能になる。

ロボットがつくる間に関しては, 間をとっていてもただ黙っているだけでは人間にその意図を汲み取ってもらうことは難しいため, 音声・動作フィラーのようなものとの組み合わせがどうしても必要となってくるようだ。これに対しては, 人間がロボットに対して間を使用した際に見せていた発話や動作をもとに設計して, 適切な意味合いを持たせることが可能である。

さらなる具体的な間の分類, 意味付け, 生成のためには, 一問一答形式にとられない会話での調査実験も今後必要である。隣接ペアをもとにした今回の一問一答形式の SCRIPT では, 人間とロボット双方において, 間に文脈上の意味を持たせるためには不十分なものであったと考えられる。間が意味をもつような場面設定や, より深い議題での話し合いの中ではどのような結果になるのか調べることも今後の課題である。

## 7 まとめ

人間とロボットの会話の質を上げるための非言語的コミュニケーション要素の一つとして, 「間」に着目した。隣接ペアをもとにした一問一答形式の会話を人間とロボットに行ってもらった調査実験を実施し, 次の 2 点を分析した。

- 人間がロボットに対してどのような間の使い方をするのか
- ロボットの使う間を人間はどのように捉えるのか

人間の使用する間には大きな個人差があるが, 相手がロボットであっても答えるのが難しい問題には間をとってしまう。ロボットによる間は, 今回の実験において, 文脈上において意味をもつと人間から捉えられることはなかった。これらの結果から, ロボットが相手の間の使い方を学習しながら間を分類・識別し, 間

の生成を行えるような仕様の意義が見いだせた。

## 参考文献

- [1] Sacks, H., Schegloff, E. A., and Jefferson, G.: A Simplest Systematic for the Organization of Turn-Taking for Conversation, *Language*, Vol. 50, No. 4, pp. 696–735 (1974)
- [2] Levinson, S. C.: *Pragmatics*, Cambridge: Cambridge university Press, (1983)
- [3] T.J. Bruneau,: Communicative Silences: Forms and Functions, *J.Communication*, Vol. 23, No. 1, pp. 17–46 (1973)
- [4] D. Kurzon,: *Discourse of Silence*, John Benjamins Publishing Co., March (1998)
- [5] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita,: How Quickly Should Communication Robots Respond?, *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI 2008)*, pp. 153–160 (2008)
- [6] N. Mukawa, H. Sasaki, and A. Kimura,: How do verbal/bodily fillers ease embarrassing situations during silences in conversations?, *Robot and Human Interactive Communication (RO-MAN) 2014 23th IEEE International Symposium on*, pp. 30–35 (2014)
- [7] N. Ohshima, K. Kimijima, J. Yamato, and N. Mukawa,: A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings, *Robot and Human Interactive Communication (RO-MAN) 2015 24th IEEE International Symposium on*, pp. 325–330 (2015)
- [8] Schegloff, E. A. and Sacks, H.: Opening up closings., *Semiotica*, 8 pp. 289–327 (1973)