

インタラクティブパートナーとしての AI スピーカーに関する検討

Consideration of AI speaker as interactional partner

打谷 拓巳[†], 西崎 友規子[†]
Takumi Uchitani, Yukiko Nishizaki

[†]京都工芸繊維大学
Kyoto Institute of Technology

概要

AI スピーカーのコミュニケーション相手としての実用化を目指し、外見がインタラクティブに与える影響を明らかにするため、AI スピーカーと表情のあるロボットを用いて、共同サイモン効果の生起を比較した。

結果として、意思を持って動いていると教示された場合に共同サイモン効果が有意に大きく生じ、それは AI スピーカーでは顕著な傾向が認められた。これより、AI スピーカーのような無機質な外見でもインタラクティブパートナーとして有用な可能性が示唆された。

キーワード: AI スピーカー, インタラクティブパートナー, 外見差, 言語教示

1. はじめに

近年、人とのコミュニケーションを主軸においたロボット（以下コミュニケーションロボットと呼ぶ）の開発が盛んに行われ、実際に店頭などで稼働しているのを目にすることも多くなってきている。一方で AI スピーカーやスマートスピーカーなどと呼ばれるスピーカー（以下 AI スピーカーと呼ぶ）も徐々に我々の生活に取り入れられ始めている。コミュニケーションロボットと AI スピーカーは、対話によって人と繋がるという点が類似しているが、コミュニケーションロボットとされているものの多くは生物的な外見であるのに対し、AI スピーカーは無機質な外見である。この見かけの違いは、用途の違いを反映しているようである。

最も知名度が高いコミュニケーションロボットの一つである Softbank 社の Pepper は、そのホームページ [1] で「Pepper とのコミュニケーションをもっと」と Pepper との会話を主として宣伝されている。また、利用方法ではなく「Pepper との暮らし方」とまるで家族の一員のように表現されている。一方、AI スピーカーの一つである Google 社の Google Home は、そ

のホームページ [2] で「あらゆる生活シーンをサポート」と人間の援助がその主用途として宣伝されている。また、「音声でスマートホームを管理できます」と AI スピーカーは人間に利用・管理されるものという位置づけがされている。

この差異は多くの研究（例えば、岡ら[3]や不気味の谷現象）で示されているように、人や動物に似せた生物的な外見を持つもののほうが親近感を抱きやすいという知見が基盤となっていると思われる。

しかし、最近 AI スピーカーを子供の発達に必要な会話相手として期待する研究成果[4]や独居高齢者の会話相手としての利用方法を検討する研究[5]がなされ始めている。これらは、AI スピーカーの価値は従来捉えられてきた人間の援助的役割だけでなく、人間のパートナーとして捉える動きである。

AI スピーカーは、コミュニケーションロボットにはない利点を有している。例えば、比較的安価である点や、サイズや形状が固定されていない点である。安価であるため、普及のしやすさが利点となる。サイズや形状が固定されていない点については、様々な場面への適応のしやすさが挙げられる。人間や動物に外見を寄せたロボットに比べ AI スピーカーは外見に縛りがないため使用できる状況や場面の自由度は格段に高い。

そこで本研究は、AI スピーカーをコミュニケーションロボットと同様の存在として活用することを目指し、人は AI スピーカーを、対等にインタラクティブを行うパートナーとして認識するかどうかを検討する。

二人以上で協力して机を持ち上げるような協働行為においては、相手の行動を意識して自分の行動を決定することが必要となる。このように他者とのインタラクティブにおいては、他者の行動を自分の行動であるかのように認識して表現することは重要な要素の一つである[6]。そういった行動の確からしさは、共同サイモン課題（社会的サイモン課題）によって検証されている。

共同サイモン課題とは、Sebanz らによって初めて実施された課題であり[7]、通常のサイモン課題とは違い

二人の実験参加者で分担して課題が行われる。

標準的なサイモン課題は、2色の刺激に対応した左右のボタンを、予め指示されたとおりに、できるだけ早く正確に押し分けるという課題[8]である。サイモン課題の特徴として、課題中に課題とは無関係な視覚刺激（左右への矢印など）が表示される点が挙げられる。この刺激によってボタンの場所（左右）への反応が促進されるため、本来押すべきボタンの場所（左右）と、提示される刺激の向き（左右）が一致すれば反応時間が短くなり、一致しなければ反応時間が長くなるという現象が見られる。これがサイモン効果である。

サイモン効果は、実験参加者がどちらか片方の刺激にのみ反応する場合（個人 go-nogo 条件）では生じない。しかし、二人の実験参加者が個々に片方の刺激とボタンのみを担当して課題を行う場合、それぞれの行っている動作は個人条件と同様であるにも関わらず、サイモン効果が生起する。この効果は、共同サイモン効果と呼ぶ[7]。

Stenzel らは、共同サイモン課題を人とヒューマノイドロボットが共同で実施する課題として用い、二者間での協働行為がどのように行われるか検討した[9]。その際、ヒューマノイドロボットの動作について、2種類の条件を設定した。一方は人工知能をもった人間のようなロボットと教示した条件（human-like robot 条件）であり、他方は予めプログラムされた通りに動作する意図のないロボットと教示した条件（machine-like robot 条件）である。共同サイモン効果の生起を確認した結果、前者では生起したが、後者では生起しなかった。これより、ロボットが自律した存在と認識した場合、ヒューマノイドロボットを人と同様に協働行為の相手として認識することが明らかとなった。

本研究は AI スピーカーのインタラクションパートナーとしての有用性を検討することを目的とし、Stenzel et al. [9]に倣い、共同サイモン課題を用いた検討を行う。

AI スピーカーにヒューマノイドロボットと同様の結果が認められれば、AI スピーカーが協働行為の相手として有用であることが示される。AI スピーカーとは異なり、外見に生物的特徴を有するロボットとの比較を行うことによって、本課題を検討する。

2. 実験

2.1 実験参加者

工学系以外の理系学部・または理系研究室に所属する大学生、大学院生（男性 21 名、女性 10 名、19-25 歳、 $M=21.25$, $SD=1.37$ ）の計 40 名であった。後述する教示の効果を最大化するために人工知能に対して初学者程度以上の知見を有すると考えられる工学系以外の学生および日常的に AI スピーカーを使用していない学生を対象とした。

実験参加者には、事前に十分な説明を行い、実験は著者らの所属する大学倫理委員会の承認を得て実施した。

2.2 実験計画

実験参加者内でのエージェント要因（AI スピーカー、ロボット）×実験参加者間での教示要因（human-like agent, machine-like agent）の 2 要因混合計画とした。エージェント要因は実験参加者内、教示要因は実験参加者間計画であった。

2.3 エージェント要因

共同で課題を行うエージェントとして、AI スピーカーとロボットを設定した。

AI スピーカー条件では Amazon Echo (amazon 社製)、ロボット条件では cozmo (Anki 社製) を用いた。cozmo は、AI スピーカーと同様に動作がなく、音のみで返答することに違和感がない形状のロボットとして選定した。さらに、cozmo は操作の自由度が高く、表情で感情表現が可能な点が外見に生物性付与されているロボットとして適切であると考えた。2つのエージェントを図 2.1 に示す。

実験課題音声の読み上げにはボイスロイド（結月ゆかり、<https://www.ah-soft.com/vocaloid/yukari/>) を使用した。また、AI スピーカーの反応音声には合成音声読み上げソフト（棒読みちゃん、<https://chi.usamimi.info/Program/Application/BouyomiChan/>) を使用した。cozmo の反応音声は、エージェント本体からの発声という点を重視したため、本体音声を使用した。



図 2.1 実験エージェント

(左：ロボット条件で使用した **cozmo**,
右：AI スピーカー条件で使用した **Amazon echo**)

2.4 教示要因

human-like agent 条件では、「このエージェントが人工知能デバイスとして動作し、自身で学習して課題を行っている」と教示した。その後、エージェントが教示通りの動作をする様子を見せた。

machine-like agent 条件では、「実験者が設計したプログラムで動作している」と教示した。その後、**human-like agent** 条件と同様、それぞれのエージェントが教示通りの動作をする様子を見せた。

2.5 実験課題

先行研究 [7] [10] を参考にして、VisualStudio で Microsoft Visual C#を用いて実験者が作成した。この際、Kim-Phoung ら[10]では課題とは無関係の空間刺激を音の高低としていたが、本実験ではスピーカーの位置（左右）で区別するように修正した。

実験課題はできるだけ早く正確に行うという教示のもと、「あ”あるいは”い”という2音が左右いずれかのスピーカーから聞こえ、実験参加者とエージェントが分担して反応するというものである。例えば、エージェントが”い”に対して”青”と発声することで反応する場合、実験参加者が”あ”に対して赤いシールが貼られたキーを押して反応するということが求められた

実験参加者はキーを利き手の人差し指で押す。このとき、スピーカーから発せられる音が“あ””い”どちらであるかとは無関係に、発声されるスピーカーの左右が決定される。

どちらの教示要因でも、実験参加者にはエージェント自身が反応すると伝えているが、実際には実験者が

パーティションの外側で操作した。そのため、エージェントの動作は人間と同程度の反応速度と正答率に設定していると説明した。

AI スピーカーの反応はスマートフォン内音声再生アプリを使用した。一方、**cozmo** の反応は **cozmo** アプリを使用した。

2.6 実験手順

手順に関する説明の後、エージェント要因に応じた教示を行った。聴覚共同サイモン課題の動作を数回見せたのち、100回を1セットとして実施した。エージェントを替えて2セットずつ、合計4セット行った。実験エージェントの順番は各条件で偏りが出ないように順序効果を考慮した。また、100回ごとに3分の休憩を与えた。最後に2つの設問(①それぞれのエージェントに対する印象、②どちらのエージェントがより親しみやすいと感じたか)に対して記述による回答を求めた。実験の様子を図 2.2 に示す。



図 2.2 実験の様子

3. 結果

3.1 共同サイモン課題

反応時間として、スピーカーから音が発せられてから実験参加者がキーを押すまでの時間を計測した。ここで、キーボードにはDキー（左）とKキー（右）を押すキーとして設定している。キーの左右と発音されたスピーカーの左右が一致している場合を一致条件、一致していない場合を不一致条件とする。このとき、共同サイモン効果は一致条件における反応時間の平均値から不一致条件における反応時間の平均値の減算によって測定される。

エージェント要因×教示要因での各群の結果を図3に示す。

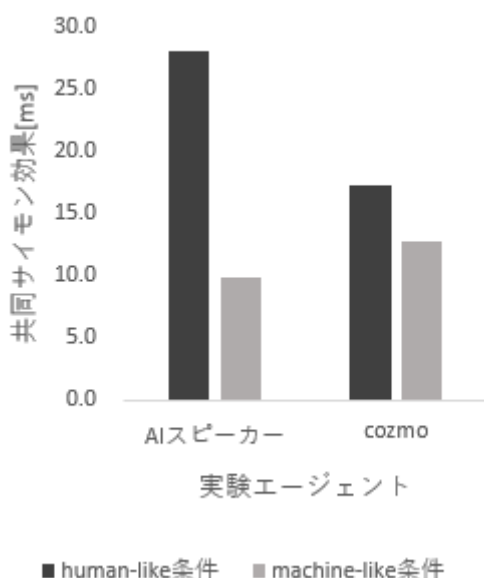


図3 教示要因×エージェント要因の共同サイモン効果

3.2 教示とエージェントの関連に関する分析

教示とエージェントの関連に関して、教示要因×エージェント要因における対応のない2要因分散分析を行った。その結果、教示要因において、有意な主効果の傾向が見られた($F(1, 76) = 3.83, p = .054, p < .10$)。一方で、エージェント要因に有意な主効果は得られなかった ($F(1, 76) = 0.39, p = .530, n. s.$)。また、有意な交互作用は得られなかった。 ($F(1, 76) = 1.43, p = .235, n. s.$)

エージェント要因の主効果は得られなかったが、エージェントの外観が及ぼす影響を確認するために、それぞれのエージェントで教示要因における共同サイモン効果の値についてt検定を行った。AIスピーカーをパートナーとした場合に human-like 条件では machine-like 条件よりも共同サイモン効果が有意に高くなることが確認された ($p = .027, p < .05$)。一方で cozmo においては教示間に有意な差は得られなかった ($p = .604, n. s.$)。

4. 考察

教示要因において、有意な主効果の傾向が認められたことから、human-like 条件の方が machine-like 条件よりも共同サイモン効果が大きく生起する傾向にあるこ

とが示された。

この結果は、Stenzel らの結果[9]と一致しておる。自律した存在であるという教示を付加することで、従来のヒューマノイドロボットとは程遠い外見を有するエージェントであっても、他者存在として認識され、人のインタラクションパートナーになりうると言える。

AIスピーカーがインタラクションパートナーとして認識されることは、人や動物など身近なものに寄せられているとインタラクションをしやすいという知見[3]には合致しない結果である。この理由として、AIスピーカーに対する認知の変化が考えられる。従来、コミュニケーションロボットというものは非日常的な存在であるため、親近感を得るためには身近なものに似せる必要があった。これは、犬を飼っている方が犬型エージェントに親近感を抱きやすいという研究結果[11]からも想定される。ただコミュニケーションロボットやAIスピーカーの存在が身近なものとなりつつある現代、AIスピーカーという自律的存在自体に、動物などと同様に親近感をもつようになってきている可能性が考えられる。これを検証するために、無機質な外見であるが、身近ではない存在を自律していると教示した場合との比較が必要である。

実験エージェントを分けて分析したところ、AIスピーカーでは教示間の有意差が認められ、cozmo には有意差が認められなかった。cozmo は外見に生物性を有したロボットとして設定しており、先行研究[9]と同様に、教示間において共同サイモン効果に差がみられることが期待されていた。仮説と異なった結果の理由として、以下の2つが考えられる。

第一に、今回の課題においてエージェントが音声によって反応していたことが挙げられる。今回の実験では Stenzel ら[9]で実施されたような視覚刺激ではなく聴覚刺激による課題であった。また、cozmo は本体の音声で反応した。この本体音声に関して事後質問紙で「声がわかりにくい」という意見が多く見られた。内田[11]に示されているように、音声の明瞭性は話者印象に影響する。これらから cozmo の音声共同サイモン効果の生起に影響を与えた可能性が考えられる。上記を明らかにするためには今後 cozmo の反応音声を、AIスピーカーと同様に人間に近い音声にしたうえで再度検討が必要である。

第二に、cozmo の外見の影響が考えられる。cozmo は表情を有するものの、ロボット然とした外見を有している。中川ら[12]では、ロボット不安尺度の高群で

ロボットの機械的な外見が親近性に影響したことが報告されている。そのため、無機質な外見である AI スピーカーよりも、意志を持っているという教示の影響を受けにくかった可能性が考えられる。上記を明らかにするためには、人型、動物型などの外見のロボットとの比較が必要である。

5. 結論

本研究では、AI スピーカーが人とのインタラクションに与える影響について、共同サイモン課題を用いて検討した。結果として、性能についての教示の違いによる共同サイモン効果に有意差が認められた。これより、AI スピーカーのような無機質な外見であっても、インタラクションの相手として役割が果たせる可能性が示唆された。これは、AI スピーカーに対する認識の変化によるものであると考えられた。一方、先行研究 [9] と反して *cozmo* において教示の違いによる有意差が生じなかったのは、ロボット然とした音声や外見が影響した可能性が考えられた。

引用文献

- [1] “Pepper (ペッパー) — ロボット — ソフトバンク,” <https://www.softbank.jp/robot/pepper/>. (Accessed on 01/14/2020).
- [2] “Google home - スマートスピーカー - google ストア,” https://store.google.com/jp/product/google_home. (Accessed on 01/14/2020).
- [3] 岡夏樹, 山田誠二他,(2006) “適応のためのインタラクション設計 (i 特集, hai: ヒューマンエージェントインタラクションの最先端), ” 人工知能学会誌, vol. 21, no. 6, pp. 642-647
- [4] S. B. Lovato, A. M. Piper, and E. A. Wartella,(2019) “Hey google, do unicorns exist?: Conversational agents as a path to answers to children’s questions,” Proceedings of the 18th ACM International Conference on Interaction Design and Children, pp. 301-313, ACM
- [5] 王晶, (2019) “独居高齢者の自立性を向上させるスマートスピーカーを利用した見守りシステムの提案,” 修士論文, 北陸先端科学技術大学院大学先端科学技術研究科
- [6] N. Sebanz, H. Bekkering, G. Knoblich,(2006) “Joint action: Bodies and minds moving together”, *Trends in Cognitive Sciences*, Vol. 10, pp. 70-76.
- [7] N. SEBANZ,(2003) “Representing others’ actions : just like one’s own?”, ” *Cognition*, vol. 88, pp. B11-B21
- [8] J. R. Simon,(1990) “The effects of an irrelevant directional cue on human information processing,” *Advances in psychology*, vol. 65, pp. 31-86, Elsevier
- [9] A. Stenzel, E. Chinellato, M. A. T. Bou, A. P. delPobil, M. Lappe, and R. Liepelt,(2012) “When humanoid robots become human-like interaction partners: corepresentation of robotic actions.,” ” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 38, no. 5, p. 1073
- [10] K. -P. L. Vu, R. W. Proctor, and P. Urcuioli(2003), “Transfer effects of incompatible locationrelevant mappings on a subsequent visual or auditory simon task,” ” *Memory & Cognition*, vol. 31, no. 7, pp. 1146-1152
- [11] S. Kiesler, L. Sproull, K. Waters(1996), “A prisoner’s Dilemma Experiment on Cooperation With People and Human-Like Computers”, *Journal of Personality and Social Psychology*, Vol. 70, No. 1, pp. 47-65
- [12] 内田照久,(2011) “音声中の母音の明瞭性が話者の性格印象と話し方の評価に与える影響”, *The Japanese Journal of Psychology*, Vol. 82, No. 5, pp. 433-441
- [13] 中川弘也, 神田智子,(2017) “受付ロボットの積極性及び外見が与える印象の ユーザ属性別分析”, *HAI シンポジウム 2017*