

遠隔調理活動に対する他者の気づきと声掛けの初期分析

益田 岳^{†1*} 酒造 正樹^{†2*} 荒木 英夫^{†3}
 東京女子医科大学 東京電機大学 大阪工業大学

1. はじめに

IoT 行動変容学研究グループ キックオフシンポジウムにおいて、著者らは遠隔ソーシャルダイニングの実施によりもたらされる行動変容について議論を行い、ちょっとした失敗も周りのサポートを受けてポジティブな相互関係性の醸成に繋がりうることを示唆した [1]。これを受けて、調理活動の遠隔共有の場に、意図的な失敗行動を実験的に挿入したならば、周囲の参加者からどのような声掛けを得られるかについて検討を行った。

物理的なサポートができない遠隔の共同作業において、困っている人に対する配慮は、唯一、声掛けといった精神的なサポートに限られている。遠隔で共有可能な情報に制約が伴い、かつ複数人参加の映像情報を複数視点で閲覧する条件下で、困っている状態を検出し、適切な声掛けを行うことは、熟練したファシリテータであっても難しい。将来の画像認識システムやロボットエージェントの活用を狙い、どのような行動要素であれば検出が可能であり、また、対象者にとって意味のある声掛けがどのようなものかをセットで議論することが肝心である。

2. 関連研究

画像認識システムの研究として、橋本らは画像情報から調理シーンの認識手法の開発 [2] を行っている。井上らは人間のモーションデータなどを用いて調理行動を認識するコンテスト [3] を開催し、比較的長時間の要素として扱われるマクロ行動と、短時間の要素として扱われるミクロ行動の双方を同時に認識するチャレンジに取り組んでいる。これらの関連給は、カメラやセンサデバイスによって、リアルタイムに調理状況の認識が現実的になることを予想させる。

さらに、文脈理解に伴うロボットエージェントの文章生成技術も進んでいる。杉山らは車載カメラから見える画像情報をもとに、リアルタイムに最適な文章生成を行い、ロ

ボットが運転手と雑談対話するシステムを開発した [4]。言語生成に当たっては、事前学習済みの言語モデル [5] を活用し、調理や共食のドメインの対話データによってファインチューニングすることで自然な発話文章を生成できると期待される。今後、調理において、こういったシチュエーションでどのような声掛けをなすべきかを定めてデータを収集すれば、ロボットエージェントがファシリテーションを行うことも可能であろう。

また、浅野らは、エージェントからのアドバイスは画一的なものではなく、個人の気質を考慮した手法が有効であると主張している [6]。一人で集中して調理に専念したい人もいれば、料理の出来栄よりも調理プロセスをみんなと楽しむことに重きを置く人もいる。何をどのようにアドバイスできるかできるかの視点のみならず、どんな相手に声掛けするか、またその人の受け止め方を考慮する必要がある。

3. 遠隔調理活動の映像収録

3.1. 条件

2 者間の遠隔調理活動の映像収録を 2022 年 5 月に行った。食材は主としてカット野菜などを指定し、皿に盛り付ける作業を収録対象とした。調理時間は約 5 分とした。顔の表情と調理テーブルを撮影するカメラを 2 点用意し、Zoom のカメラ接続により両者の映像を共有した。モニタの表示は、図 1 に示すようにギャラリービューの全画面表示とした。調理中、両者は自然な会話をして良いとした。相手の進行状況を意識することなく、自らの調理活動を優先する



図 1 Zoom の表示レイアウト

Figure 1 Layout of Zoom application

Preriminarily Analysis of Attention and Considering Utterance on Remote Cooking Activities

^{†1*} GAKU MASUDA, Tokyo Womens Medical University

^{†2*} MASAKI SHUZO, Tokyo Denki University

(* Equal contribution)

^{†3} HIDEO ARAKI, Osaka Institute of Technology

表 1 意図的行動の一覧

Table 1 Intentional actions list

ID	意図的行動	開始	終了	時間
A01	蓋を開けるのに苦労する	1:15	1:47	0:32
A02	調味料が手について困る	3:18	3:27	0:09
A03	混ぜている葉っぱが溢れる	4:24	4:40	0:16
A04	おやつを自慢する	6:04	6:15	0:11
B01	サラダを豪快にいれる	1:01	1:07	0:06
B02	箸がなくスプーンを使う	1:35	2:00	0:25
B03	パストラミを飛ばす	2:51	3:13	0:22
B04	「間違った」と発言する	3:30	3:40	0:10
B05	別皿に移しかえる	5:41	6:00	0:19

こととした。

この収録中に、事前に用意した意図的行動を設けた。例えば、食材が溢れるという状態は、他者の気づきを得やすく、また声掛けに繋がりがやすいものと思われる。両者は互いに用意した意図的行動を収録前に情報共有していない。また、第三者評価に影響を与えないよう、収録中に気づいたとしても言及することを求めなかった。

3.2. 収録データ

パイロットスタディとして著者らが実施した 6 分 24 秒の映像に対して、事前に用意した意図的行動の一覧を表 1 に示す。それぞれを代表するシーンを図 2, 3 に示す。

4. 意図的行動に対する第三者評価

4.1. 評価方法

2 者間の遠隔調理映像に対して、第三者の評価実験を行った。評価は、前提として、評価者が「他人に対して優しく振る舞う思いやりある人」であるとして、15 秒単位に切り出した映像を視聴し、(1) どんな気づきを得たか、(2) それに対してどんな声掛けをしたいかとした。Google フォームのアンケート機能を用いて実施した。評価者は、共食の研究プロジェクトに従事している研究者の 6 名とした。

4.2. 気づきに関する結果

評価者 6 名による気づきの有無を表 2 にまとめた。

A03 および B03 の食材が溢れてしまうという事象は視覚的变化が大きく認知しやすいため、それぞれ 3 人、6 人の気づきを得た。同様に、A04, B01, B05 の事象も視覚的变化が大きく気づきを得やすいと言える (5 人, 3 人, 4 人)。

一方で、A01 の遅延を伴うのみで視覚的变化に乏しい事象は、即座の気づきを得ることが難しい。最終的に蓋を開

表 2 6 名の評価者による気づきの有無

Table 2 Attention by 6 reviewers

ID	1	2	3	4	5	6
A01	○	-	○	○	-	○
A02	-	-	-	-	-	-
A03	-	-	○	○	○	-
A04	○	○	-	○	○	○
B01	○	-	○	○	-	-
B02	-	-	○	-	-	○
B03	○	○	○	○	○	○
B04	○	-	○	-	-	○
B05	○	○	-	○	○	-

けるのに困っているという気づきを得たのは 4 人であった。また、A02 のように短時間で終わってしまうような事象も気づきを得にくく、0 人であった。

また、B02 は A01 に並行して生じたものであるが、ともに気づきを得たのが 2 人であった。

アンケートの際のコメントとして、同時に進行する 2 者の共同調理に対して、複数のカメラ映像のどこを注視してよいか困るという意見が上がった。このことは、参加者の増加とともに、表示領域が小さくなり、見落とす可能性が上がることを示唆している。

4.3. 声掛けに関する結果

意図的行動に対して何らかの気づきを得た評価者が、どのような声掛けを行えばよいか、自由回答で得た結果を事象ごとに以下に示す。

4.3.1 A01

- かわりに開けましょうか？ お手伝いしますよ。大丈夫ですか？
- 蓋がすごくかたいの？ 結構、がっちりとかされているみたいですね。
- それ開きにくいです？
- 開けづらいんですか？ そのバック。テープが結構しっかりくっ付いてる？

4.3.2 A02

なし。

4.3.3 A03

- 結構飛び出しちゃって、大丈夫ですか。
- いや、そうなりますって
- 菜箸など使ったほうがよいですよ

4.3.4 A04

- わぁ、それ何？ おいしそう。

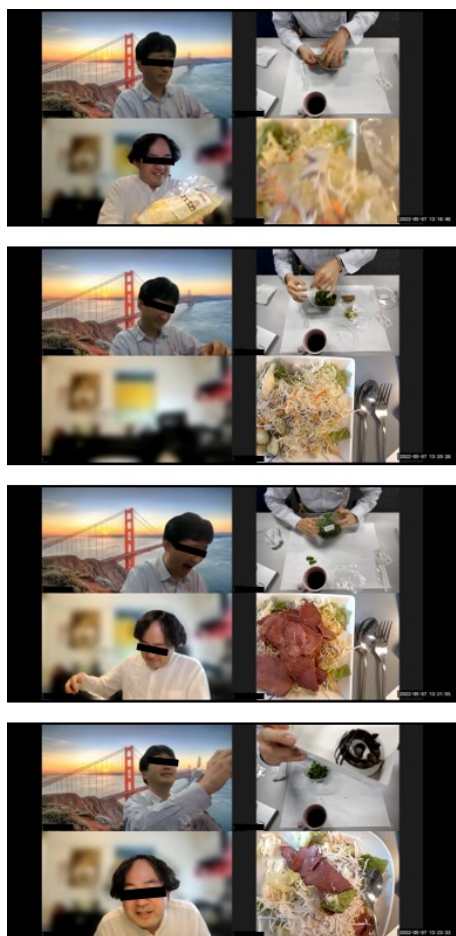


図2 被験者 A (上段) の意図的行動 (上から順に A01～A04)

Figure2 Scene samples of intentional actions by subject A (A01～A04)

- それってドーナツですか？
- ドーナツの箱大きくないですか。
- え、何味？

4.3.5 B01

- ゆっくり作業してくださいね。
- 一気に盛り付けられましたけど、なんかうまく行きますね
- 結構豪快にサラダを入れられるんですね。
- どうしました？何かありました？

4.3.6 B02

- お箸と言っていて、スプーンなんですか？
- スプーン…？



図3 被験者 B (下段) の意図的行動 (上から順に B01～B05)

Figure3 Scene samples of intentional actions by subject B (B01～B05)

4.3.7 B03

- 大丈夫ですか？ 三秒ルールですね
- 大丈夫ですか。
- 袋を開けるときは、慎重に開けましょう
- 大丈夫ですか？どこに飛びました？ ん？軽く洗ったりしてこなくても大丈夫？

4.3.8 B04

- なんか間違えました？
- ん？どうしました？

4.3.9 B05

- お皿かえると、洗い物増えちゃいますよ。
- 豪快に移しますね
- せっかくきれいにに入れていたのに移し替えちゃうんですか。
- え、間違えてます？どこが？ お、そのままひっくり返します？パストラミどけといた方が良くない？

4.4. 第三者評価の考察

調理に手間取り遅延を生じる事象の A01 に関して、心配の声かけが多くみられた。蓋が開かなくて困っていることに対して、気づくもののリモート環境においてはサポートする具体的策が見つからない。また、30 秒を超えた最も長い意図的行動であるが、全体に対して冒頭に位置付けられているため、これが遅延であるとは認識されなかった。目標調理時間の 5 分に対して遅れと見なすか否かは、全行程を知る前にはなかなか気づけない。

次に、食材が溢れてしまうという突発的事象の A03 と B03 については、反射的に「大丈夫ですか？」という問いかけがやすい。また、失敗行動に対して、冷静に考えれば防げたであろうという指摘を行い、代案の提案もあった。

また、文脈上の理解が難しく突発的に出た「間違えた」という音声発話 (B04) に対しては、評価者は意図をとらえることができずに、「どうかしましたか？」と確認の問いかけを行う。

また、同じく文脈理解が難しい事象の B02 は、サラダという食べ物に対してスプーンを求め出す奇怪なシーンであり、適切ではない行動サンプルとして実験的に設けたものである。「箸がない」とネガティブな発話しながら立ち去るシーンに気づきを得ながらも、突飛な行動に閉口気味な対応しか取れない。

B01, B05 に関しては、意図的なオーバーアクションであり、そもそも困っている事象ではない。多くの評価者が気づきを得ながらも、サポートを申し出るコメントはなかった。

同様に、A04 に関して、困っている事象と異なり、人々の注目を浴びたいシーンである。自らアピールした様に対して、単純な肯定的なレスポンスを得やすいといえる。

4.5. 声掛けの受容可能性

以上は初期検討の結果であり、今後、対象者に対して受容可能な最適な声掛けが何であるか、さらなる調査が必要である。対象区間のビデオに対する声掛け案として、4.3 項のリストを候補として、別途評価者に自らがそのような声掛けを受けると嬉しいか否かなどの評価を行うことを検討している。

4.6. その他のシーンに対する気づきと声掛け

ここまで意図的行動として定義済みの箇所について、第三者の気づきと声掛けについて検討を行ってきた。その他のシーンについても多数の気づきと声掛けの案を得ているが、本稿では詳説しない。他者が好意を持って声掛けをしてあげたいと思ったとしても、それが本人にとって良い印象を与えるとは限らないことを示唆している。

5. おわりに

リモート調理活動を題材に、意図的に実施した行動にどれだけ気づきを得るか、またどのような声掛けがあり得るかを調査した。他者の声掛けが有用に思われる、食材の開封に困る、食材が溢れる、時間的遅延が生じる、手が汚れるなどといった事象において、気づきの得やすさに差があることを示した。短時間で大きな変化を生むコンテキストに比べて、文脈理解を前提とした高次コンテキストは気づきを得にくい。

気づきを得た事象に対して、具体的にどのような声掛けが可能であるか具体例を集めた。いずれのコメントも相手の状況を慮ったものであり、将来的にロボットエージェントなどがシステム発話を行うときの参考データとなる。なお、対象者に対して受容可能な最適な声掛けが何であるかの追加検討が必要である。

従来、自然な形で収録したビデオデータに対して、事後アノテーションを行いケーススタディとして有用なシーンの抽出を行う方法論を取っていた。しかしながら、この手法において偶発的に生じる困った事象を量的に捉えることが難しく、十分量の学習データを求めることができなかった。本稿で提案する実験的手法は、声掛けを必要とするシーンの定義から始まり、再現性のある映像収録が可能となるため、有用な手法と思われる。

謝辞

本稿は JST 未来社会想像事業 探索加速型「世界一安全・安心社会の実現」領域課題「QOL 計測とハートフルネス実践による食体験競争システム」JPMJMI20D9 の支援を受けた。

参考文献

- [1] 益田 岳, 酒造 正樹, “遠隔ソーシャルダイニングはどんな行動変容を起こすのか,” IoT 行動変容学研究グループキックオフシンポジウム, P24, 2022.
- [2] Shinya Michibata, Katsufumi Inoue, Michifumi Yoshioka, Atsushi Hashimoto, “Cooking activity recognition in egocentric videos with a hand mask image branch in the multi-stream CNN,” Proc. of the 12th Workshop on Multimedia for Cook-

- ing and Eating Activities (CEA), pp. 1-6, 2020.
- [3] Paula Lago, Shingo Takeda, Sayeda Shamma Alia, Kohei Adachi, Brahim Bennai, Francois Charpillet, Sozo Inoue, “A dataset for complex activity recognition with micro and macro activities in a cooking scenario,” arXiv preprint, arXiv:2006.10681, 2020.
 - [4] 杉山 弘見, 古賀 光, 西島 敏文, “移動体から見える風景を話題とする雑談対話システム,” 2022 年度人工知能学会全国大会 (第 36 回), 2N5-OS-7a-04, 2022.
 - [5] Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, Toyomi Meguro, “Empirical analysis of training strategies of transformer-based Japanese chit-chat systems,” arXiv preprint, arXiv:2109.05217, 2021.
 - [6] 浅野 倫子, 横澤 一彦, 酒造 正樹, 山田 一郎, “気質分類に基づく適切な情報提示法の心理学的検討,” 信学技報, vol. 109, no. 471, HIP2009-175, pp. 335–340, 2010.