

音声認識失敗時の不快感を緩和する複数人格エージェント

Discomfort Reduction by Multiple Personality Assistant Agent

in Speech Recognition Failure

本戸丈裕¹ 小林一樹²

Takehiro Hondo¹ Kazuki Kobayashi²

¹信州大学 工学部

¹ Faculty of Engineering, Shinshu University

²信州大学 学術研究院

² Academic Assembly, Shinshu University

Abstract: 本研究では、家電操作におけるユーザとエージェントとの音声対話において、複数人格を有するエージェントが対話の印象に与える影響を調査する。複数人格エージェントとは、実体は単一であるが、発色や声質を変えることで複数の人格の内在を表現したエージェントである。音声認識の失敗時には言い直しの必要からユーザに不快感を与えるが、複数人格エージェントの発話により不快感を緩和し、無理のない言い直しへの誘導する手法を提案する。参加者実験の結果、複数人格条件では騒がしく批判的な印象をユーザに与える結果が示唆された。

1 はじめに

近年、音声認識技術の発展により Google 社の Google Home や Apple 社の Siri などの音声による操作が可能なシステムが一般化してきている。音声による操作は機器に接触せずに操作できることから利便性が高く、自然言語による入力であることから操作方法や特殊なコマンドの習得が不要な利点がある。反面、音声認識の失敗時にはユーザには言い直しの必要が生じ、ユーザに不快感を与えてしまう。

音声認識の改善に向け様々な研究がされている。Xiong ら[1]は、深層学習に基づく音響モデルと言語モデルを用いた高精度認識システムを開発している。また、後藤ら[2]は、認識結果をユーザに公開し、誤認識の訂正のためユーザの協力得ることで、認識精度を向上させるシステムを提案している。一方、高度なシステムを構築する手法とは異なり、単純なシステムでユーザを言い直しの生じにくい状況に誘導する研究がされており、小林ら[3]は、明滅光源とビーブ音による ASE を使用し、発話衝突を避ける手法を提案している。

これらのように、音声認識、音声対話の向上に向け多方面のアプローチがとられている。しかし、音声認識は環境音や認識タイミングといったものに影響されるため、技術が高度になったとしても誤認識を完全に排除することは難しいと考えられる。

そこで本研究では、複数人格エージェントを用いた音声認識の失敗時におけるユーザの不快感を緩和する手法を提案する。提案手法では、ユーザの発話に影響を与える要因としてインタラクション時にユーザが抱く印象に着目し、エージェントの振る舞いを変容させることで、ユーザに与える不快感を緩和させ無理のない言い直しに誘導することが目的である。具体的には、主人格エージェントの発言に対する他の副人格エージェントの発言について、音声認識失敗時におけるユーザの心情を代弁するものを採用し、エージェント集団によってユーザの振る舞いを支持するポジティブな印象を与えることを目指す。

2 複数人格エージェント

複数人格エージェントの外見を図 1 に示す。複数人格エージェントとは、発色、声質、口調をひとまとまりにしたものを「人格」とし、これらの設定を変化させることで、単一の外見を持ちながら、複数の異なる人格の内在を

表現するエージェントである。複数人格エージェントは、人格設定の変更により異なるエージェントを表現できることから複数体の実装が容易であり、外見が単一であるため物理的なスペースを必要としないことが利点である。

複数人格エージェントには表 1 に示すように主人格 A、



図 1: 複数人格エージェント

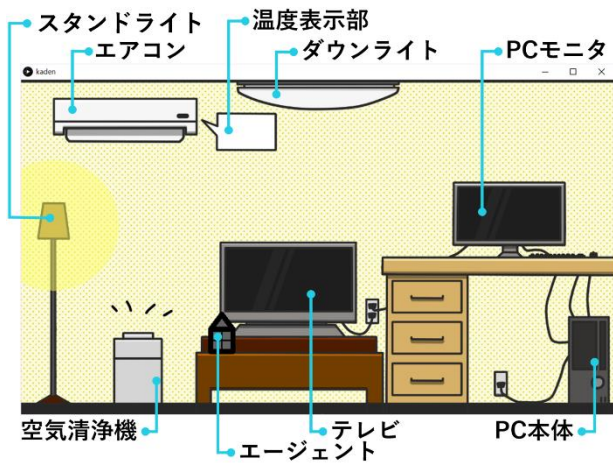


図 2: 家電操作シミュレータ

副人格 B1, B2, B3 の 4 人格が用意され、各人格には発色、声質、口調、役割に応じた発話内容が割り当てられる。

3 ユーザ心情の集団発言化表現

奥森ら[4]は擬人化エージェントによる集団圧力[5]を利用しており、本研究では複数人格エージェントによって集団圧力の形成を狙う。このような手法を利用することで、音声や動作による直接的な内部状態の表出の制限を超えてユーザの心理に働きかけることができる可能性がある。

先行研究[6]では、主人格 A がユーザとの対話を行い、副人格 B1 がユーザと主人格 A との対話への感想を述べ、副人格 B2 と副人格 B3 は副人格 B1 の感想に同調するという発話役割が与えられた。先行研究でユーザに批判的な印象を与えた要因として、聞き返し時の副人格 B1 の発言内容がユーザを責めるような内容だったことが考えられる。つまり、ユーザの指示に対し主人格 A が認識失敗の旨を伝え、それに対して副人格 B1 が聞き取れないという感想を述べ、副人格 B2, B3 がそこに同調することで、結果的にユーザに対してネガティブな働きかけを繰り返す形になっている問題がある。

人間には自身に好意的な人に対して友好的な関係を取りたいとする親和動機[7]がある。竹内ら[8]は人間とエージェント間においても、エージェントが自分の意見

表 1: エージェントの人格設定

人格	主人格 A	副人格 B1	副人格 B2	副人格 B3
発色	青	黄	緑	橙
声質	丁寧な女性	明るい女性	静かな女性	明るい男性
口調	~です	~だよ	~だね	~だね
役割	ユーザとの対話	対話への感想 ユーザ心情の代弁	人格 B1 への同調	人格 B1 への同調

表 2: エージェントの発話例

	先行研究[6]	提案手法
聞き返し	A「すみません、エアコンをどうしたらいいですか」	
感想 / 代弁	B1「聞こえなかったね」	B1「ちゃんと聞いてほしいよね」
同調	B2「分からない」 B3「さあ」	B2「そうだね」 B3「うんうん」

に同意した場合、人間はそのエージェントに対して親和的反応を起こしたことを確認している。そこで本研究では、副人格 B1 の台詞を認識失敗時におけるユーザの心情を代弁する台詞に変更することでポジティブな印象の形成を狙い、認識失敗時にユーザに与える不快感の緩和を目指す、ユーザ心情の集団発言化表現を提案する。表 2 に提案手法と先行研究の認識失敗時におけるエージェントの発言の違いを示す。音声認識の失敗時、ユーザには操作指示の言い直しの必要が生じる。提案手法の副人格 B1 の台詞は、言い直しを求められた時のユーザの心情を想定して設定された台詞であり、続く副人格 B2, B3 の台詞は、副人格 B1 の発言内容に対して自然な形で同調する台詞である。このように、認識失敗時のユーザの心情をエージェント集団の発言として表現することで、ユーザの振る舞いを支持する印象形成を狙う。

4 実験

複数人格エージェントが対話印象に与える影響を調査するため、家電操作シミュレータを用いて参加者実験を実施した。

4.1 家電操作シミュレータ

図 2 に家電操作シミュレータの画面を示す。家電操作シミュレータでは、マイクからのユーザの音声入力に応じて、エージェントに画面内の家電操作を命じられる。シミュレータは、フロントエンドは Processing 3.4 によって、バックエンドは JavaScript によって開発されており、音声認識部は Web Speech API を用いた。エージェントは音声合成ソフトウェア Open JTalk と CeVIO を用いて作成された音声によりユーザと対話を行う。音声認識によって得られたユーザの発話データは、構築した辞書データを参照し、該当項目があればその内容に従って応答

表 3: 各条件でのエージェントの発話例

認識の可否	単一人格条件	複数人格条件
成功	A「はい、テレビをつけますね」 A「つけました。今日の番組どうなってるんでしょうか、気になりますね」	A「はい、テレビをつけますね」 A「つけました」 B1「今日の番組どうなっているのかな」 B2「気になる」 B3「そうだね」
失敗 (意図)	A「すみません、テレビをどうしたらいいですか」 A「うまく聞き取れませんでした」	A「すみません、テレビをどうしたらいいですか」 B1「ぼーっとしないでほしいよね」 B2「うんうん」 B3「本当にね」
失敗 (非意図)	A「すみません、もう一度お願いします」	

する。

4.2 実験条件

実験条件は、単一人格条件と複数人格条件の2つを設定した。表3にテレビをつけるタスクでの各条件でのエージェントの発話例を示す。単一人格条件では人格Aのみが発話を行い、複数人格条件では主人格A、副人格B1, B2, B3が発話を行う。

4.3 実験手順

実験は1人ずつ、PCの置かれた机の前に着席して実施された。参加者には家電操作指示項目の一覧が記されたチェックリストが渡される。実験に関する説明がされた後、シミュレータに慣れるための練習として、練習用画面において「電子レンジをつける」、「洗濯を始める」の2項目の操作を音声で行う。その後、実験用画面において、表4に示す12項目の家電操作指示項目を順に音声で行う。調査のための意図的な誤認識として、表4の聞き返し欄に示すように、全12項目中8項目に操作の対象を聞き返す名称欠如項目、または操作の方法を聞き返す操作内容欠如項目を設定した。また、参加者への教示内容として、「家電の名称と動作内容を含めた内容を発話すること」、「指示が正しく実行されない場合は、もう一度同じ内容の発話すること」、「どうしても指示が認識されないと感じた際は、スキップキーにより次の項目に移ること」が教示された。

4.4 評価指標

評価指標として、アンケートによる主観評価と自由記述として感想の記述を採用し、すべての家電操作指示項目が終了した後記入が行われた。アンケートは27項

表 4: 家電操作項目と聞き返し設定

No.	操作内容	聞き返し
1	天井の明かりをつける	なし
2	スタンドライトを消す	操作対象
3	テレビをつける	操作方法
4	空気清浄機を消す	操作対象
5	パソコンをつける	なし
6	エアコンのスイッチを入れる	操作方法
7	天井の明かりを消す	操作対象
8	スタンドライトをつける	なし
9	テレビを消す	なし
10	空気清浄機をつける	操作対象
11	パソコンを消す	操作方法
12	エアコンのスイッチを切る	操作方法

目であり、Q1を除く26項目について7段階のリッカートスケールを採用し、質問内容にあてはまるほど高い値をとるものとした。Q1については1~10体の選択式とした。また、実験中はユーザの発話音声を記録した。

4.5 実験参加者

参加者数は単一人格条件、複数人格条件に各12名の計24名とし、信州大学の大学生24名(男性21名、女性3名)が参加した。平均年齢は22.0歳であった。

5 結果

表5にアンケート内容とその結果、および単一人格条件と複数人格条件での統計的有意差検定の結果を示す。統計的有意差検定にはマンホイットニーのU検定を用いた。ここでは有意水準を5%とした。複数人格条件において男性1名がシミュレータの不具合により実験を中断したため、得られた結果は単一人格条件12名、複数人格条件11名のものとなっている。

Q1について、単一人格条件では12人全員が1体と回答したのに対し、複数人格条件では3体、4体に5名ずつ、5体に1名が回答し、11人全員が複数体であることを認識していたことが確認できる。Q3では有意差が認められ、複数人格条件の方が責められているような感覚が強いことがわかる。Q8においても有意差が認められ、複数人格条件での発言の方が回りくどいという認識であった。また、Q12でも有意差が認められ、複数人格条件の方が騒がしいと感じていることが示唆された。さらに、Q19でも有意差が認められ、単一人格条件の方が発言の意味が理解できたとなっている。

表 5: アンケート結果

No.	質問項目	単一人格条件		複数人格条件		p 値
		Mean	S.D.	Mean	S.D.	
Q1	システムは何体いるように感じましたか	1.00	0.00	3.64	0.64	-
Q2	システムに感情があるように感じた	4.17	1.56	4.82	1.40	0.358
Q3	システムから責められているように感じた	1.92	1.31	3.64	1.67	0.008
Q4	システムにうまく指示が伝わったと感じた	4.42	1.48	4.36	1.72	0.975
Q5	自分の指示の仕方は適切だった	5.25	1.11	4.82	1.75	0.750
Q6	このシステムは指示を聞こうとしていた	5.58	1.62	5.09	1.31	0.231
Q7	このシステムは安定していると感じた	4.58	1.37	4.64	1.23	0.868
Q8	このシステムの発言は回りくどいと感じた	3.00	1.76	4.27	1.21	0.045
Q9	このシステムから聞き返された回数は多いと感じた	5.08	0.72	4.64	1.30	0.371
Q10	このシステムと上手くコミュニケーションが取れた	4.50	0.99	3.73	1.29	0.159
Q11	このシステムは友好的だと感じた	5.58	0.48	4.82	1.27	0.111
Q12	このシステムは騒がしいと感じた	2.83	1.47	4.36	1.37	0.035
Q13	このシステムにはイライラを感じた	3.00	1.73	4.09	1.88	0.198
Q14	このシステムは煩雑だと感じた	2.33	1.14	3.55	1.44	0.051
Q15	このシステムはあなたの指示を理解していると感じた	5.67	0.88	5.18	1.27	0.193
Q16	このシステムは指示に従順であったと感じた	6.25	0.62	5.27	1.42	0.081
Q17	このシステムは性能が良くないと感じた	3.00	1.78	2.82	0.94	0.796
Q18	このシステムにどのように言えばいいか素早く理解できた	5.67	1.37	4.82	1.40	0.164
Q19	このシステムの発言の意味が理解できた	6.58	0.50	4.64	1.49	0.001
Q20	システムの使い方が理解できた	6.33	0.75	6.09	0.67	0.405
Q21	実験中は落ち着いた雰囲気だった	6.67	0.64	5.55	1.62	0.051
Q22	実験中は楽しい雰囲気だった	5.25	1.16	5.36	0.64	0.974
Q23	自分自身で家電を操作している感じがした	5.75	0.74	5.73	1.14	1.000
Q24	このシステムはあなたに興味があるように感じた	3.33	1.49	4.64	1.30	0.061
Q25	もう一度同じ実験をすればスムーズに指示できると思う	4.25	1.71	4.55	1.37	0.754
Q26	このシステムは自分に寄り添ってくれているように感じた	4.25	1.50	4.82	1.53	0.427
Q27	このシステムを日常的に使いたいと思う	5.33	1.59	4.82	1.59	0.448

6 考察

先行研究と同様に、参加者は単一人格条件よりも複数人格条件の方が騒がしく批判的であると感じていることが実験結果から示唆された。認識失敗時の台詞をユーザの振る舞いを支持するポジティブなものに変更したにも関わらず、ネガティブな印象をユーザに与えた。この要因として、複数人格条件での参加者の感想の中に、エージェントの「ぼーっとしないでほしいよね」という発言に対し「家電から煽られることがあった」という記述があったことから、参加者はエージェントの発言をユーザに対する批判として受け取った可能性がある。また、複数人格条件における認識失敗時のエージェント 4 人格の総発話時間が単一人格条件と比べ平均して 3.46 秒長かったことから、ユーザは言い直しの意欲を持ちながらも発話権を与えられないことでストレスを感じていた可能

性もある。

一方、Q26 の結果では単一人格条件に比べ複数人格条件の方がわずかながら高い値となったことから、提案手法は言い直し時の不快感の緩和につながる可能性がある。

7 まとめ

本研究では、複数人格エージェントがユーザとの対話印象に与える影響を調査した。単一の外見を持ちながら、複数の人格が内在する複数人格エージェントでは、主人格がユーザとの対話を行い、音声認識の失敗時には副人格が認識失敗時におけるユーザの心情を代弁し、さらに別の 2 人格が副人格に同調する。単一人格条件と複数人格条件とを比較するために参加者実験を行ったところ、先行研究と同様に、複数人格条件の参

加者はエージェントに複数の人格を認識し、騒がしく責められている印象を受けることが示唆された。一方、複数人格条件ではユーザはシステムを自身に寄り添っていると感じるということが示唆された。

今後は複数の人格による適切な対話構成を検討し、それらがユーザの印象に与える影響を調査する予定である。

謝辞

本研究の一部は、JSPS 科研費(No.JP26118005)の助成を受けました。ここに謝意を表します。

参考文献

- [1] W.Xiong, L.Wu, J.Droppo, X.Huang, and A.Strolcke: The Microsoft 2017 Conversational Speech Recognition System, In Proc. IEEE ICASSP, pp.5934-5938. IEEE (2018)
- [2] 後藤藤真孝, 緒方淳, 江渡浩一郎: PodCastle:ユーザ貢献により性能が向上する音声情報検索システム, 人工知能学会論文誌, Vol.25, No.1, pp.104-113 (2010)
- [3] 小林一樹, 船越孝太郎, 小松孝徳, 山田誠二, 中野幹生: ASE に基づく相槌によるロボットとの対話体験の向上, 人工知能学会論文誌, Vol.30, No.4, pp.604-612 (2015)
- [4] 奥森拓也, 片上大輔: 擬人化エージェントによる集団圧力体験, 日本知能情報ファジィ学会 第 27 回ファジィシステムシンポジウム, pp941-946 (2011)
- [5] S.E.Asch: Opinions and social pressure, scientific American. Vol.193, No.5, pp.31-35 (1955)
- [6] 小林一樹, 吉村貴大: 複数人格を持つ家電制御エージェントによる対話雰囲気形成, 電子情報通信学会技術研究報告, Vol.117, No.72, AI2017-2, pp.7-12 (2017)
- [7] H.A.Murray: Explorations in personality, Oxford University Press (1938)
- [8] 竹内勇剛, 片上恭弘: ユーザの社会性に基づくエージェントに対する同調反応の誘発, 情報処理学会論文誌, Vol.41, No.5, pp.1257-1266 (2000)