コマンド認識失敗時に人格を交替する音声対話エージェント

Development of Personality Changing Spoken Dialogue Agent in Command Recognition Failure

堀立樹 [1*] 小林一樹 [2]
Tatsuki Hori[1] Kazuki Kobayashi[2]

[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，音声対話エージェントがユーザからの音声コマンドの認識に失敗した場合に，ネガティブな印象を与えずに対話体験を向上させることを目的とする．提案手法では，認識失敗時にエージェントの人格を疑似的に交替させる．認識を失敗した人格が退出し，新規に出現する人格が音声認識を引き継ぐことにより，ユーザへ与えるネガティブな印象のリセットを試みる．実験では，家電音声操作シミュレータ上での家電操作タスクにおいて，エージェントの人格交替が参加者に与える影響を調査した．実験後のアンケート調査により，エージェントの人格交替によって，コミュニケーションの円滑さやシステムの責任感が向上したことが示され，提案手法による音声対話におけるユーザ体験の向上が示唆された．

1 はじめに

近年，音声認識技術が広く普及し，人々にとって日常生活の基盤になりつつある．Google 社の Google Assistant や Amazon 社の Amazon Alexa，Apple 社の Siri などの音声認識技術を活用した AI アシスタントが開発され，スマートフォンやスマートスピーカに搭載されている．スマートスピーカの普及率に関する 2018 年の調査[1]では，アメリカでは 21%の家庭にスマートスピーカが設置されているという結果が報告されており，初のスマートスピーカである Amazon Echo の発売から 4 年で，スマートスピーカはセキュリティウェブカメラやスマートサーモスタットといった他のスマートホームデバイスと比肩するほどの普及レベルに達している．この普及の背景には，計算機の処理性能の向上や，クラウドサービスの実現による，スマートデバイス上での音声認識精度の向上などが挙げられる．

音声認識精度の向上に向けた研究として，例えば Thanh-Ha ら [1] の研究がある．Thanh-Ha らは，音声操作デバイスが，他のデバイスのスピーカから再生される録音音声を認識してしまう問題を解決するために，スピーカで再生した発話音声と人間の発話音声をクラス分けしたデータセットを用いて深層学習を行い，最大 90%の精度でスピーカの音声を区別することに成功している．しかし，実環境には周囲の環境音や他者の発話音声といった多種のノイズが存在し，音声認識は頑健ではない．庄境 [2] は，自動車運転時の雑音環境下でのカーナビへの音声入力を行う実験を行い，音声認識精度には発話音声とノイズとのパワー比や発話速度，滑舌の良さが関係することを報告している．

このように，音声認識技術は様々なアプローチにより進展しているものの，複雑な実環境でのノイズの影響により，音声の認識失敗を完全に排除することは難しい．認識失敗が発生した場合，人間はネガティブな印象を音声認識デバイス側に抱き，その後の継続的な利用に支障を与えかねない．人間はネガティブな印象をポジティブな印象よりも持続しやすく，覆しにくいため [3]，認識失敗時には，人間の抱くネガティブな印象を適切に取り除く必要がある．

そこで本研究では，単一のシステム内に，複数の人格を持った音声対話エージェントを提案する．提案手法では，認識失敗の発生時に人格の交替を実施し，失敗したエージェントの人格へのネガティブな印象をリセットできることを期待して，音声対話エージェントとの対話時のユーザ体験の向上をねらう．

*連絡先： 信州大学工学部電子情報システム工学科
〒 380-8553 長野県長野市若里 4-17-1
E-mail: 16t2813h@shinshu-u.ac.jp

[1]Michael Philpott: 2019 Trends to Watch, Smart Home, Ovum- Informa Telecoms & Media (2018)

2　関連研究

2.1　エージェントのリカバリー行動による印象の向上

香山ら [4] は，観光案内を行う音声対話システムにおいて，ユーザの質問に対してシステムから誤応答が発生したときに，システムにはユーザの不満解消を目的としたリカバリー行動を実行する手法を提案している．実験では，誤応答の発生時にエージェントがユーザに謝罪する条件，謝罪しない条件，正しい応答をした際に謝罪する条件の 3 条件で,「対話の自然さ」と「システムの好感度」の 2 項目について，実験参加者による主観的評価を比較した．実験の結果,「対話の自然さ」,「システムの好感度」のどちらの項目でも，誤応答時に謝罪を行う条件の評価が高いことが報告されている．このように，音声対話システムにおいて，システムの誤応答が発生した場合，謝罪というリカバリー行動がユーザにポジティブな印象を与えることが示されている．

香山らの研究では，リカバリー行動は謝罪行為に絞っており，謝罪行動による印象向上は示されたが，他の行為がユーザへ与える印象についても調査の余地がある．そこで本研究では，謝罪とは異なる手法でのリカバリー行動をとることによって，ユーザに与えるネガティブな印象の改善を目指す．

2.2　複数のエージェントによる協調性の効果

藤堂ら [5] は，1 画面につき 1 体のエージェントを表示する実験環境において，実験参加者 1 人とエージェント 1 体による二者音声対話と，実験参加者 1 人とエージェント 2 体による三者音声対話の 2 条件で実験参加者が抱く印象の調査を行っている．三者音声対話では，二者音声対話に比べ「対話の弾み」や「雑談らしさ」をより実験参加者へ印象付けることが示された．また,「雑談らしさ」と音声認識率の相関係数が，二者音声対話に比べ三者音声対話で小さくなったことが示され，三者音声対話では音声認識率に関係なく，雑談対話の印象をユーザへ与えることが示唆された．

一方，二者音声対話を支持する意見として,「エージェント間のやりとりを待ってしまう」,「(三者対話は) 画面を視線が行き来するので大変だった」というものも挙げられた．三者音声対話では，複数体のエージェントを用いることで，発話権がなかなか回ってこないことへのストレスや，視線の移動による会話の煩雑さといったネガティブな印象をユーザへ与える可能性が考えられる．そこで，本研究では，多人数エージェントの存在によって発生するネガティブな印象を低減しつつ，多人数対話を行うことによる対話体験の向上をねらう．

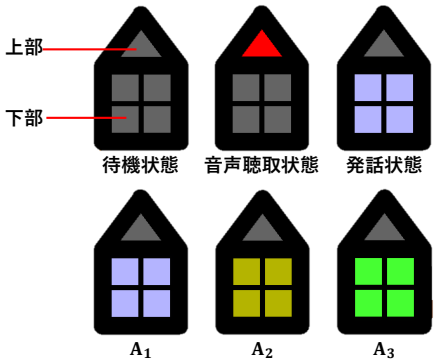

図 1: エージェントの外観

表 1: エージェントの状態

状態名・人格名	発色			発色箇所
	R	G	B	
待機状態	-	-	-	なし
音声聴取状態	255	0	0	上部
発話状態・A_1	180	180	255	下部
発話状態・A_2	180	180	0	下部
発話状態・A_3	70	255	50	下部

3　人格交替エージェント

3.1　人格交替エージェントの構成

本研究では，A_1，A_2，A_3 の 3 人格を持つ人格交替エージェントによる音声対話システムを提案する．複数人格エージェントでは，エージェントの外観の発光色，音声の音色，口調という要素をひとまとめにしたものを「人格」と定義し，それらの要素を変化させることで，疑似的に異なる人格を表現する．

本研究ではエージェントとの音声対話をシミュレータ上で再現するが，スマートスピーカといったデバイスとしての使用を考慮した場合，人格交替エージェントは単一のデバイスで実装され，ユーザとの対話は音声ベースで行うため，多人数のエージェントを実装する開発コストを節約でき，多人数のエージェントを使用することによるスペースを圧迫しない．先行研究 [6] では，本研究と同様な手法に対して，参加者がエージェントの人格交替によって表出する複数体の人格を認識していることが示されている．図 1 に音声対話エージェントの外観を，表 1 にエージェントの各状態に関する説明を示す．エージェントは，待機状態，音声聴取状態，発話状態の 3 状態をもつ．ユーザからの音声入力がなく，エージェントが発話を行わない状態を待機状態とし，待機状態ではエージェントは発光しない．ユーザの音声を聞き取っている音声聴取状態では，エージェントの上部を発光させ，ユーザの発話音声が入力され

表 2: 音声対話の流れ

認識	人格交替条件	非人格交替条件
成功	例) ユーザ「テレビを消して.」 A_N「はい，テレビを消します．... 消しました.」	
失敗	A_N「すみません，上手く聞き取れませんでした.」 (人格を交替する) A_{N+1}「次は，私にご命令をどうぞ.」	A_1「すみません，上手く聞き取れませんでした.」

表 3: 人格交替エージェントの 3 人格の音声設定

設定	人格		
	A_1	A_2	A_3
キャラ名	タカハシ	すずきつづみ	さとうささら
性別	男性	女性	女性
大きさ	＋ 0.00	＋ 0.00	＋ 0.00
速さ	1.00	1.21	1.00
高さ	＋ 0	＋ 0	＋ 0
声質	＋ 0.00	＋ 0.00	＋ 0.00
抑揚	1.00	1.00	1.00
個別項目	元気：0.00 普通：1.00 へこみ：0.00	クール：0.47 照れ：0.53	元気：1.00 普通：0.00 怒り：0.00 悲しみ：0.00

ていることを可視化する．エージェントがユーザに対して発話を行う発話状態では，エージェントの下部を人格固有の色で発光させる．各状態でのエージェントの発光は，エージェント自身の内部状態の表出や，発話衝突の回避などを目的とするほか，エージェントの人格の差異を表現するために用いている．

3.2 エージェントの人格交替

ユーザからエージェントへの音声コマンド入力と，音声認識の可否によってエージェントが行う発話の内容とを表 2 に示す．エージェントによる誤認識が発生した場合には，エージェントの人格を交替させ，特定の人格に対するネガティブな印象のリセットを狙う．人格の交替は，$A_1 \rightarrow A_2$，$A_2 \rightarrow A_3$，$A_3 \rightarrow A_1$ とし，一方向に循環させる．人格交替エージェントの人格交替タイミングは，音声認識が失敗し，エージェントがその旨をユーザに伝え，謝罪を行った直後とした．

エージェントの各人格 A_1，A_2，A_3 の発話音声は，音声合成ソフトウェアである CeVIO Creative Studio Ver.6.1 を用いて作成し，CeVIO から提供されるキャラクターボイス「タカハシ」,「すずきつづみ」,「さとうささら」を使用した．CeVIO Creative Studio 上で，A_1，A_2，A_3 における発話音声の作成に用いた設定値および性別情報を表 3 に示す．

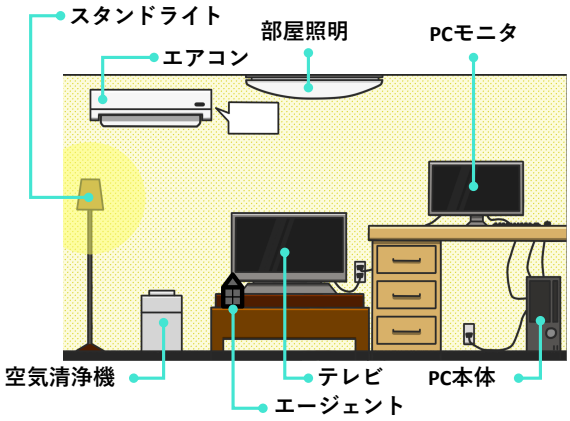

図 2: 音声対話シミュレータ

3.3 家電音声操作シミュレータ

音声コマンドによる家電操作を疑似的に行う家電音声操作シミュレータを開発した．本システムのフロントエンド部には Processing3.5.3 を使用し，音声認識部には HTML5 と JavaScript，Web Speech API を用いた．Web Speech API は，Speech API Community Group によって策定された，ユーザの発話音声認識機能を Web アプリケーションに組み込むことが可能な JavaScript API である．Web Speech API による音声認識を行うため，Processing 上で Web サーバを構築し，Web ブラウザから Web サーバへアクセスすることで，JavaScript を呼び出して実行する．このときの Web ブラウザには，Google Chrome を使用した．また，WebSocket サーバを Processing 上で構築し，Web ブラウザとの WebSocket 通信によって，WebSocket サーバからは音声認識に対する制御信号を送信し，Web ブラウザからはユーザの発話音声から生成したテキストデータを送信する．Processing 上の WebSocket サーバへ送られたテキストデータは，事前に用意された辞書データと照合し，登録されている単語がテキストデータ内に含まれていた場合には命令に応じた家電操作を実行させ，登録されている単語が含まれていなかった場合には音声認識が失敗したとみなし，人格の交替を行う．図 2 に示す通り，シミュレータ上では，7 種類の家電とエージェントが配置された部屋が表示される．

4 実験

本実験は，音声対話を行う人格交替エージェント使用時のユーザの印象を評価し，提案手法がユーザの印象に与える影響を明らかにすることが目的である

表 4: 実験時の音声コマンド命令項目

No.	命令内容	登録キーワード	意図的な認識失敗
1	部屋照明をつける	天井/電気/明かり/照明/部屋/蛍光灯/ライト/つけ/いれ/入れ/on/動か/オン/起動/開始/始め/投入/初め/始め/灯/明るくし/明るくす	なし
2	スタンドライトを消す	スタンド/間接/立/ライト/消/止/切/off/オフ/落/ストップ/終/暗/stop/ストップ	あり
3	テレビをつける	TV/テレビ/つけ/いれ/入れ/on/動か/オン/起動/開始/始め/投入/初め/スイッチ/映	なし
4	空気清浄機を止める	空気/清浄機/消/止/切/off/オフ/落/ストップ/終	あり
5	PC を起動させる	パソコン/PC/つけ/いれ/入れ/on/動か/オン/起動/開始/始め/投入/初め/始め/立ち	あり
6	エアコンをつける	エアコン/クーラー/暖房/冷房/空調/つけ/いれ/入れ/on/動か/オン/起動/開始/始め/投入/初め/始め	なし
7	部屋照明を消す	天井/電気/明かり/照明/部屋/蛍光灯/ライト/消/止/切/off/オフ/落/ストップ/終/暗	あり
8	スタンドライトをつける	スタンド/間接/立/ライト/つけ/いれ/入れ/on/動か/オン/起動/開始/始め/投入/初め/灯/明るく	なし
9	テレビを消す	TV/テレビ/消/止/切/off/オフ/落/ストップ/終	あり
10	空気清浄機を動かす	空気/清浄機/つけ/いれ/入れ/on/動か/オン/起動/開始/始め/投入/初め	なし
11	PC を止める	PC/パソコン/消/止/切/off/オフ/落/ストップ/終/シャットダウン/ログオフ	なし
12	エアコンを止める	エアコン/クーラー/空調/暖房/冷房/消/止/切/off/オフ/落/ストップ/終	あり

4.1 実験条件

本実験では，音声認識失敗時にエージェントの人格を交替させる人格交替条件と，同じ人格が継続する非人格交替条件の 2 条件を設けた．実験評価は，上記の 2 条件において，参加者間配置とした．実験時の各条件下でのエージェントの振る舞いは，表 2 に示した通りである．エージェントの音声認識が成功した際の挙動は両条件で統一し，音声認識が失敗した際にも，両条件でエージェントは「すみません，上手く聴き取れませんでした」という台詞によって謝罪を行い，音声認識が失敗した旨を実験参加者に伝える．その後，人格交替条件では失敗するたびに，エージェントの人格 A_1，A_2，A_3 を循環的に交替し，交替後の人格は「次は，私にご命令をどうぞ」という台詞によって，交替を行ったという旨を実験参加者に伝える．

本実験では，音声認識が成功している場合であっても，特定のタイミングで必ず音声認識が失敗するよう，計画的にシミュレータの動作を設定した．12 項目の音声認識タスクのうち半数の 6 項目で，計画的に認識の失敗を挟み込むことで，人格交替条件における人格交替機会の確保をねらう．

4.2 実験手順

実験は，実験参加者の他に誰もいない個室で行い，実験中には実験者が部屋から退出し，外的要因を極力排除した．音声の入出力にはマイク付きのヘッドホンを用いて，音声入力およびエージェントの音声認識の安定化を図った．実験参加者には，シミュレータ上での音声認識システムの振る舞い調査を行うという旨を事前に説明したうえで，実験を開始した．

実験開始時には，実験参加者へ音声操作の順序と操作内容が書かれた命令表を手渡し，命令表に書かれた順序に沿って実験を進行させた．表 4 に，参加者に手渡した命令表の内容と，エージェントが音声認識の可否を判定するためのキーワード，計画的な認識失敗が発生する命令箇所を示す．命令表の内容は，人格交替条件，非人格交替条件の両条件で，全く同じ内容のものを使用した．12 項目の音声コマンド命令が終了した時点で実験を終了し，その後，アンケートへの回答に誘導した．

表 5: アンケート内容と結果

$No.$	アンケート項目	人格交替条件		非人格交替条件		U	df	p
		$Mean$	$S.D.$	$Mean$	$S.D.$			
1	システムは何体いるように感じたか	3.10	0.57	1.20	0.63	-	-	-
2	システムに感情があるように感じた	3.90	2.02	2.20	1.62	28.50	18.0	0.089†
3	システムから責められているように感じた	1.20	0.63	1.20	0.63	50.00	18.0	1.000
4	システムにうまく指示が伝わったと感じた	4.70	1.25	4.50	1.65	48.00	18.0	0.905
5	自分の指示の仕方は適切だった	5.30	1.06	4.50	1.65	38.50	18.0	0.394
6	このシステムは安定していると感じた	4.30	1.49	4.50	1.08	47.50	18.0	0.876
7	このシステムは自信を持っているように感じた	3.90	1.10	4.00	1.56	47.50	18.0	0.877
8	このシステムから聞き返された回数は多いと感じた	4.50	1.35	3.90	1.52	37.50	18.0	0.353
9	このシステムと上手くコミュニケーションが取れた	5.30	1.34	4.10	1.29	23.50	18.0	0.040††
10	このシステムは友好的だと感じた	5.40	1.35	4.60	1.51	34.50	18.0	0.246
11	このシステムは騒がしいと感じた	3.40	1.90	2.10	1.29	29.00	18.0	0.109
12	このシステムにはイライラを感じた	2.40	1.51	3.30	1.70	33.50	18.0	0.216
13	このシステムは煩雑だと感じた	2.70	1.70	2.60	0.97	49.50	18.0	1.000
14	このシステムはあなたの指示を理解していると感じた	5.70	1.16	5.00	1.25	33.50	18.0	0.216
15	このシステムは指示に従順であったと感じた	5.70	1.34	6.00	1.05	44.50	18.0	0.692
16	このシステムは性能が良くないように感じた	3.30	1.77	2.90	1.29	45.50	18.0	0.757
17	このシステムは責任感が強いと感じた	4.50	1.90	2.60	1.71	23.00	18.0	0.041††
18	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	6.30	0.95	5.90	1.60	45.00	18.0	0.707
19	このシステムの発言の意味が理解できた	6.70	0.67	6.60	0.70	45.50	18.0	0.690
20	システムの使い方が理解できた	6.70	0.67	6.60	0.52	42.00	18.0	0.480
21	実験中は落ち着いた雰囲気だった	6.00	1.33	6.80	0.42	32.00	18.0	0.118
22	実験中は楽しい雰囲気だった	5.00	1.63	4.40	1.17	36.50	18.0	0.308
23	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	4.70	1.77	5.20	1.69	40.50	18.0	0.485
24	このシステムは信頼できる	4.70	1.25	5.00	1.15	40.00	18.0	0.452
25	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	4.30	1.64	3.70	1.89	41.50	18.0	0.539
26	このシステムを日常的に使いたいと思う	4.50	1.65	5.00	1.33	38.00	18.0	0.367

†† $p < 0.05$
† $0.05 \leqq p < 0.10$

4.3 評価指標

評価指標として，表 5 に示す 26 項目からなるアンケートによる主観評価を採用する．No.1 は，1〜10 体の範囲での選択肢とした．No.2〜No.26 については，7 段階のリッカート尺度に基づき，参加者は設問の内容について，どれほど当てはまるかを 1〜7 のいずれかの数値を選択して評価する．このとき，設問の内容にあてはまる場合は大きな数字に，あてはまらない場合は小さな数字になる．また，アンケートには実験への感想を記述する自由記述欄を設けた．

4.4 実験参加者

実験参加者数は，人格交替条件，非人格交替条件に各 10 名の計 20 名とし，実験には，信州大学および大学院の学生 20 名 (男性 19 名，女性 1 名，平均年齢 22.2 歳，標準偏差 0.73 歳) が参加した．

5 実験結果

表 5 に，実験後のアンケート項目ならびに，各条件での 7 段階評価値の平均値および標準偏差と，2 条件間での有意差検定の結果を示す．アンケート結果の有意差検定には，マンホイットニーの U 検定を採用した．

表 5 の結果より，No.9 の「このシステムと上手くコミュニケーションが取れた」の項目で，人格交替条件での評価の平均値が非人格交替条件よりも大きく，2 条件間で有意差が認められた ($U = 23.50, df = 18.00, p = 0.040$)．同様に，No.17 の「このシステムは責任感が強いと感じた」の項目でも，人格交替条件での平均値が非人格交替条件より大きく，2 条件間で有意差が認められた ($U = 23.00, df = 18.00, p = 0.041$)．また，No.2 の「システムに感情があるように感じた」の項目では，人格交替条件で評価値平均が大きく，2 条件の差が有意傾向にあることがわかった ($U = 28.50, df = 18.00, p = 0.089$)．

6　考察

アンケート結果の有意差検定により，No.9 の「このシステムと上手くコミュニケーションが取れた」と，No.17 の「このシステムは責任感が強いと感じた」の 2 項目で，2 条件間に有意差が認められ，No.2 の「システムに感情があるように感じた」の 1 項目では，2 条件間で有意傾向にある差が認められた．人格交替条件において，実験参加者の自由記述からは，エージェントの人格交替行動に関して「システムに対する親近感がわいた」「聞き返しがしつこいといった感情を感じにくかった」「交代によって上手くカバーしあっている気がした」といった意見が得られ，エージェントの人格交替が音声認識失敗時の印象を向上させ，ユーザの対話体験を向上させたことが示唆された．

また，藤堂ら [5] の研究で多人数対話の問題として示唆されていた「発話機会の損失」や「対話の煩雑さ」については，No.3 の「システムから責められているように感じた」や，No.12 の「このシステムにはイライラを感じた」，No.13 の「このシステムは煩雑だと感じた」の 3 項目では，アンケート結果の各条件の平均値はすべて 4 を下回っており，設問に対して「あてはまらない」という結果となっている．これらの 3 項目では，2 条件間の差に有意傾向は認められなかったため，人格交替条件において，音声認識失敗時に人格を交替させるという特殊な条件下であっても，人格を交替させない非人格交替条件と比較して問題とならない可能性が示唆された．

今後，定期的に提案手法によるスマートスピーカを使う場合においても，このようなポジティブな効果が認められるかについて調査する必要がある．

7　まとめ

本研究では，音声認識失敗時に人格を交替させることで，ユーザが抱くネガティブな印象のリセットを試みた．実験では，提案手法によってユーザがエージェントへ抱く印象の評価を目的とし，家電音声操作シミュレータを用いた実験を行った．ユーザの印象をアンケートにより調査したところ，提案手法により，コミュニケーションの円滑さや，エージェントの責任感の向上が認められた．今後，定期的にスマートスピーカを使う場合において，提案手法の有効性を調査する予定である．

参考文献

[1] Thanh-Ha Le, Philippe Gilberton, Ngoc Q. K. Duong: Discriminate Natural versus Loudspeaker Emitted Speech, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 501–505 (2019)

[2] 庄境 誠: 使い勝手のいい音声インタフェースの実現, 情報処理 *:* 情報処理学会誌, Vol.51，No.11，pp.1401–1409 (2010)

[3] 吉川肇子: 悪印象は残りやすいか？, 実験社会心理学研究，Vol.29，No.1，pp.45–54 (1989)

[4] 香山健太郎, 小林亮博, 水上悦雄, 翠輝久, 柏岡秀紀, 河井恒: 音声対話型観光案内システムにおける誤応答リカバリー効果の評価, 情報処理学会研究報告, Vol.2011–ICS–162，No.5 (2011)

[5] 藤堂祐樹, 西村良太, 山本一公, 中川聖一: 複数の対話エージェントを用いた雑談指向の音声対話システム, 電子情報通信学会論文誌 *D*, Vol.J99–D, No.2, pp.188–200(2016)

[6] 本戸丈祐, 小林一樹: 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, *HAI* シンポジウム *2018*, P–37 (2019)