自律型アンドロイドのキャラクタ表現のための対話の振る舞い制御モデルの構築と評価

山本 賢太　　井上 昂治　　中村 静　　高梨 克也　　河原 達也

京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

人と自然なインタラクションを行う自律型アンドロイドの研究開発が行われている [1]．自律型アンドロイドは外見，振る舞い共に非常に人間に酷似させたものである．そのため，ユーザはアンドロイドに対して，何らかのキャラクタを感じとることが期待される．自律型アンドロイドには，利用されるタスクの種類に応じてカウンセラーや研究室ガイドなどの社会的役割が与えられる．このような社会的役割に応じたキャラクタを表現することで，落ち着いているカウンセラーの方が話しやすいといったように，ユーザに良い印象を与えるだけでなく，対話によるタスクの遂行にもよい影響があると考えられる．

これまでに，対話システムにおけるキャラクタ表現方法として，PERSONAGE という表現したい性格に合わせた応答文を生成するシステム [2] や，言語パターンによるキャラクタの表現法 [3,4] などが研究されている．一方で，音声対話では話し方などの要因も印象に影響すると考えられる．また，ユーザの話速に同調する音声対話システムの研究もされている [5]．そこで，本研究では，アンドロイドの対話の振る舞いを制御することでキャラクタを表現するモデルを構築する．

2. キャラクタ表現

心理学の分野での研究やアンドロイドの社会での利用方法などを考慮して，本研究では，外向性，情緒不安定性，丁寧さの 3 特性を用いて，キャラクタを定義する．これらの 3 特性は，心理学における性格分類 [6] やエージェントの性格表現に用いられている [7,8]．

対話における発話者の印象に影響を与えると考えられる音声特徴量を検討する．発話量は，たくさん話す人ほど外向的と感じられるなど，対話における印象への影響があると考えられる．相槌の頻度と種類は，外向性，情緒不安定性の印象に対して影響することが示されている [7]．フィラーが多いと情緒不安定に感じされるなど，対話における印象への影響があると考えられる．交替潜時の長さは，話者についての印象評定に影響することが確認されている [9]．以上をふまえて，音声の特徴量 (以降，音声特徴量) としては，発話量，相槌の頻度，相槌の種類，フィラーの頻度，交替潜時の長さを用いる．

3. 予備実験

振る舞い制御モデルの構築に際して，音声特徴量のキャラクタに対する影響を調べる．

3.1 実験内容

音声試料は，WOZ 法により収録した 2 者対話を参考に，2 つのシナリオを用意した．このシナリオ内のアンドロイドの発話部分を，音声合成ソフトを用いて合成した．ユーザの発話部分は，実験実施者の声を用いた．

表 1: 音声特徴量の制御内容

音声特徴量	条件	制御内容
発話量	多	アンドロイド：49.2 秒, ユーザ：25.3 秒
	少	アンドロイド：25.5 秒, ユーザ：38.8 秒
相槌の頻度	高	ユーザ発話中の節境界及び文末に挿入
	低	相槌を削除
相槌の種類	多	「えー」,「ふん」,「あー」,「はい」の 4 種類
	少	「はい」のみ
フィラーの頻度	高	アンドロイド発話中の節境界及び文頭に挿入
	低	フィラーを削除
交替潜時の長さ	長	3 秒
	短	0.5 秒のオーバーラップ

2 つのシナリオに対して，1 分程度の基準対話を作成した．相槌とフィラーは，参考とした対話に現れたものを使用し，交替潜時の長さを 0.5 秒とした．各実験条件で基準対話と比較する対話は，基準対話の対応する 1 つの音声特徴量のみを調整し，残りの音声特徴量は基準対話と同様にした．音声特徴量の制御内容は，表 1 に示す．

実験には，男性 28 名，女性 18 名の計 46 名の大学生 (18 ～ 23 歳) が参加した．各実験参加者は，20 個の対話音声を聞きアンドロイドの印象についてアンケートに回答を行った．キャラクタの印象に関するアンケートは，Big Five 尺度の短縮版 [6] から，外向性，情緒不安定性に関する項目を採用し，丁寧さに関する項目をこのアンケートに追加した．各実験参加者は，各特性に対応する複数の項目に対して 7 段階で評定し，さらに，対話の自然さについても評価した．各キャラクタ特性の評定点は，対応する項目の平均値とした．

3.2 実験結果

音声特徴量ごとに基準対話を含めた 3 群間で分散分析を行った (表 2)．発話量は，2 条件のため t 検定を行った (表 3)．自然さの評定点については，交替潜時の長さを除いた条件に関しては基準対話と同等以上であり，交替潜時の長さ以外の音声特徴量においては本実験の調整において，問題となる不自然さは生じていないと考えられる．相槌の頻度は，外向性と丁寧さのキャラクタ印象に影響する．相槌の種類は，2 シナリオともに印象評定の有意差が見られた特性はほとんどなかった．1 分間の対話では種類の差異に気づきにくかったことが考えられる．フィラーの頻度は，情緒不安定性と丁寧さのキャラクタ印象に影響する．交替潜時の長さは，印象への影響が特に大きく，すべてのキャラクタ特性に影響していた．しかし，自然さの評価は低くなった．本実験では，発話量の違いは，外向性，丁寧さへの影響が見られたが，対話の文脈に依存している可能性も考慮しなければならない．これらより，発話量，相槌の頻度，フィラーの頻度，交替潜時の長さは，いずれかのキャラクタ特性の印象に影響していることがわかったため，振る舞い制御モデルに採用する．

4. 振る舞い制御モデル

ロジスティック回帰を用いて，与えられたキャラクタから音声特徴量の制御量を決定するモデルを構築する (図

Controlling dialogue behavior for character expression of an autonomous android : *Kenta Yamamoto, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, Tatsuya Kawahara (Kyoto Univ.)*

表 2: 各キャラクタ特性の分散分析結果

外向性

		高/多/長		基準		低/少/短			
	音声特徴量	平均	SD	平均	SD	平均	SD	F 値	多重比較
S1	†相槌の頻度	5.44	1.08	3.99	1.16	4.08	1.07	27.174*	**高 > 低**, 高 > 基準
	相槌の種類	4.64	1.22	3.99	1.16	4.76	0.90	9.342**	多 > 基準, 少 > 基準
	†フィラーの頻度	3.51	1.02	3.99	1.16	4.84	1.07	18.963**	**低 > 基準 > 高**
	†交替潜時の長さ	2.56	0.91	3.99	1.16	5.20	1.06	70.495**	**短 > 基準 > 長**
S2	†相槌の頻度	5.35	0.89	4.67	1.17	4.73	0.83	12.020**	**高 > 低**, 高 > 基準
	相槌の種類	4.88	0.94	4.67	1.17	4.80	0.88	0.813	
	†フィラーの頻度	3.53	1.06	4.67	1.17	5.17	0.89	35.103**	**低 > 基準 > 高**
	†交替潜時の長さ	2.70	0.97	4.67	1.17	4.64	1.19	64.468**	**短 > 長**, 基準 > 長

($^*p < 0.05$, $^{**}p < 0.01$)

情緒不安定性

		高/多/長		基準		低/少/短			
	音声特徴量	平均	SD	平均	SD	平均	SD	F 値	多重比較
S1	相槌の頻度	2.55	1.05	3.38	1.17	3.10	1.10	9.149**	**低 > 高**, 基準 > 高
	相槌の種類	3.08	1.15	3.38	1.17	2.73	0.94	7.060**	基準 > 少
	†フィラーの頻度	4.71	1.42	3.38	1.17	2.42	1.09	56.378**	**高 > 基準 > 低**
	†交替潜時の長さ	4.76	1.16	3.38	1.17	2.38	1.03	62.988**	**長 > 基準 > 短**
S2	相槌の頻度	2.91	1.21	3.92	1.45	3.22	1.16	15.119**	基準 > 高, 基準 > 低
	相槌の種類	3.05	1.24	3.92	1.45	3.16	1.11	13.799**	基準 > 多, 基準 > 少
	†フィラーの頻度	5.07	1.50	3.92	1.45	2.37	1.03	67.347**	**高 > 基準 > 低**
	†交替潜時の長さ	4.84	1.26	3.92	1.45	3.09	1.31	32.082**	**長 > 基準 > 短**

($^*p < 0.05$, $^{**}p < 0.01$)

丁寧さ

		高/多/長		基準		低/少/短			
	音声特徴量	平均	SD	平均	SD	平均	SD	F 値	多重比較
S1	†相槌の頻度	4.37	1.62	5.38	1.19	5.21	1.05	9.512**	**低 > 高**, 基準 > 高
	相槌の種類	5.10	1.17	5.38	1.19	5.46	1.05	1.508	
	フィラーの頻度	4.77	1.33	5.38	1.19	5.21	1.00	4.078*	基準 > 高
	†交替潜時の長さ	4.60	1.18	5.38	1.19	2.84	1.09	63.181**	基準 > **長 > 短**
S2	†相槌の頻度	4.12	1.62	4.05	1.41	4.85	0.96	7.382**	**低 > 高**, 低 > 基準
	相槌の種類	5.01	0.96	4.05	1.41	4.88	0.94	15.347**	多 > 基準, 少 > 基準
	フィラーの頻度	4.54	1.02	4.05	1.41	4.75	1.21	5.887**	高 > 基準, 低 > 基準
	†交替潜時の長さ	4.41	1.12	4.05	1.41	3.10	1.59	18.001**	**長 > 短**, 基準 > 短

($^*p < 0.05$, $^{**}p < 0.01$)

S1：シナリオ 1，S2：シナリオ 2

多重比較において 2 シナリオ共に両極端の条件間 (**太字**) に有意差の見られた音声特徴量 (†) を振る舞い制御モデルに採用した

表 3: 発話量に関する t 検定結果

	多い		少ない		
キャラクタ特性	平均	SD	平均	SD	t 値
外向性	5.74	0.82	5.03	0.84	4.991**
情緒不安定性	2.74	0.93	2.83	1.04	0.545
丁寧さ	4.76	1.22	5.91	1.04	5.688**

($^*p < 0.05$, $^{**}p < 0.01$)

太字は振る舞い制御モデルに採用した特性

1)．モデルの入力は，各キャラクタ特性を 1 から 7 の間で数値化したものである．モデルの出力は，各音声特徴量の制御量 [0 ～ 1] である．相槌の種類は，予備実験で各キャラクタ特性への影響がほとんど見られなかったため採用しない．学習には予備実験の実験結果で印象への影響の見られたもの (表 2-3 で † のついた特徴量) のみを使用する．各音声特徴量の低条件を 0，高条件を 1 とラベル付けする．各キャラクタ特性の評定点 [1 ～ 7] から音声特徴量の各条件 (2 値) へのマッピングを学習する．学習には，予備実験の評定点 920 サンプルを用いる．

4.1 特徴量の制御

音声特徴量は，ロジスティック回帰の出力量を用いて制御する．発話量を制御するために，あらかじめ発話量の多いパターンと少ないパターンの発話文を用意し，モデルの出力結果に応じて選択する．今回の実験では，モデルを単純化するため，ユーザ発話内のすべての節境界において相槌が生起する確率が等しいと仮定し，振る舞い制御モデルの出力を閾値とする．フィラーの頻度も相槌と同様に制御する．振る舞い制御モデルの交替潜時の長さの出力 [0 ～ 1] を [−0.5 ～ 3] に正規化し，この値を交替潜時の長さとする．負の値の場合はユーザの発話末尾にオーバラップさせる．

4.2 評価実験

3 つのキャラクタ特性の値 [1 ～ 7] の組み合わせによる 16 通りのキャラクタに対して，提案モデルにより制御した 16 対話を用意した．大学生及び大学院生の男性 11 名がこの 16 対話を印象評定した．アンケートは予備実験と同一のものを用いた．評定点を実験参加者ごとに平均と分散で正規化した値（Z スコア）と，モデルに入力したキャラクタ特性の 7 段階の点数とのピアソンの積率相関係数を表 4 に示す．外向性と丁寧さについては，p 値が有意水準 1%で有意な相関が確認された．この結果から，提案モデルにより外向性と丁寧さは表現可能であることがわかる．情緒不安定性に関して相関が小さいのは，他の 2 特性に比べてモデルの制御量に与える影響が小さいことなどが要因と考えられる．

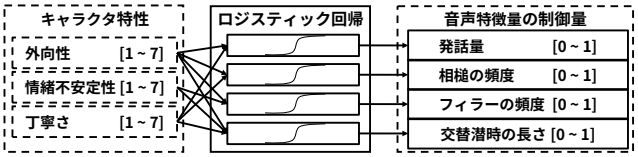

図 1: 振る舞い制御モデル

表 4: 与えたキャラクタと正規化された印象評定点のピアソンの積率相関係数

キャラクタ特性	相関係数	t 値
外向性	0.570	9.163**
情緒不安定性	−0.004	−0.056
丁寧さ	0.235	3.185**

($^*p < 0.05$, $^{**}p < 0.01$)

太字は相関の見られた特性

5. おわりに

本研究では，自律型アンドロイドのキャラクタ表現法のモデルを構築した．このモデルは，対話における音声特徴量の制御によってキャラクタに応じた振る舞いを生成する．そのため，音声特徴量を制御することによるキャラクタの印象への影響について調査した．この実験結果を用いてキャラクタに応じた振る舞い制御モデルを構築した．評価実験の結果，振る舞いの制御により，外向性と丁寧さについては表現可能であることが示された．

謝辞 本研究は，JST ERATO 石黒共生ヒューマンロボットインタラクションプロジェクト（課題番号：JPMJER1401）の支援を受けて実施された．

参考文献

[1] Koji Inoue *et al.* Talking with ERICA, an autonomous android. *SIGDIAL*, 212–215, 2016.

[2] Francois Mairesse and Marilyn A. Walker. Contorolling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37(3):455–488, 2011.

[3] 沈睿 ほか. 音声生成を前提としたテキストレベルでのキャラクタ付与. **情報処理学会論文誌**, 53(4):1269–1276, 2012.

[4] 宮崎千明 ほか. 文節機能部の確率的書き換えによるキャラクタ性変換. **言語処理学会 第 *21* 回年次大会 発表論文集**, 277–280, 2015.

[5] Rivka Levitan *et al.* Implementing acoustic-prosodic entrainment in a conversational avatar. *INTERSPEECH*, 1166–1170, 2016.

[6] 和田さゆり. 性格特性用語を用いた Big Five 尺度の作成. **心理学研究**, 67(1):61–67, 1996.

[7] Etienne D. Sevin *et al.* Influence of personality traits on backchannel selection. *IVA 2010 LNAI 6356*, 187–193, 2010.

[8] Swati Gupta *et al.* How Rude Are You?: Evaluating Politeness and Affect in Interaction. *ACII*, 203–217, 2007.

[9] 長岡千賀 ほか. 音声対話における交替潜時が対人認知に及ぼす影響. **ヒューマンインタフェースシンポジウム *2002* 論文集**, 171–174, 2002.