

擬人化エージェント同士の 対話における視覚的ノンバーバル表現の制御

Adjustment of Visual Nonverbal Conversation Signals
Displayed by Embodied Agents during Dialogue

伊藤 淳子*1
Junko ITOU

角所 考*2
Koh KAKUSHO

美濃 導彦*2
Michihiko MINOH

*1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

*2 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University

In this article, we discuss how to control nonverbal conversational signals (NCS) including eye gazes, noddings and facial expressions displayed by a pair of embodied agents that present various information to their users through their dialogues. In the field of social psychology, it is reported that there are interdependences among NCS displayed by persons during dialogues. We adjust NCS displayed by the agents for information presentation so that the interdependences among the NCS are properly maintained. By comparing cartoon of embodied agents produced by our approach to those without considering interdependences, we verified the validity of our approach.

1. はじめに

web などを通じた情報提供を擬人化エージェントの発話と連携させた形で実現する試みが盛んとなっている。擬人化エージェントとの連携により、情報取得者の情報理解にかかる負荷を低減することが期待されている。しかし、擬人化によって、エージェントは発話によるバーバル情報だけでなく、その人間的な外見から視線や表情といった様々な視覚的ノンバーバル表現を情報提供側の意図に関わらずユーザに提供してしまうという問題も持っている。このため、このようなノンバーバル表現をタスクや情報の内容に応じて適切にエージェントに表出させる研究が行われている [Cassell 99]。

一方、エージェントによる情報提供の方式として、単一のエージェントではなく複数のエージェント同士の対話を利用することも試みられている [久保田 02]。ところが、人間同士の対話においては、対話者らが表出する視覚的ノンバーバル表現の間に様々な相互依存関係が存在することが社会心理学の分野で指摘されているため [Argyle 65]-[Dimberg 82]、上記のように複数のエージェントの対話を利用する場合、エージェントの発話内容やタスクとの整合性だけでなくエージェントの表出する視覚的ノンバーバル表現についても、その間に存在する相互依存性を考慮する必要がある。そこで本研究では、対話形式で情報提供を行う擬人化エージェントの視覚的ノンバーバル表現間に上記のような相互依存性を実現することを試みる。

2. 対話におけるノンバーバル表現間の関連性

2.1 対話者間の相互依存性

社会心理学の分野では、人間同士の対話において対話者が表出するノンバーバル表現の間には相互依存性があることが報告されている。例えば、Kendon [Kendon 67] や Beattie [Beattie 78] は、話し手の発話量と聞き手の視線量との間に正の相関関係があることを実験により確かめている。また、Matarazzo [Matarazzo 64] は聞き手の頷き行為と話し手の発言時間との間に、更に Dimberg [Dimberg 82] は対話者同士の顔表情間にそれぞれ正の相関が見られることを報告している。このような研究結果を統合すると、対話を行う二者のノンバーバル表現の間

には表 1 のような相互依存関係が存在すると考えられる。表中の*印は正の相関関係を表す。

表 1: ノンバーバル表現間の相互依存性

		人物 A			
		視線	笑顔	頷き	発話
人物 B	視線	*			*
	笑顔		*		*
	頷き				*
	発話	*	*	*	*

2.2 同一人物によるノンバーバル表現間の相互依存性

前節のような対話者間の相互依存性に加え、更に各対話者が表出する個々のノンバーバル表現間にも相互依存性が存在することが報告されている。その一つとして、Argyle らは、視線、対人距離などのノンバーバル表現の総量は対話者間の人間関係に基づく親密性の度合いによって決まり、あるノンバーバル表現が制約を受けた場合に別のノンバーバル表現を変化させることによってその平衡を補償するという親密性平衡モデルを提案した [Argyle 65][Argyle 76]。この具体例として、親密性の低い第三者同士が偶然近接して座ることになった場合、互いに相手の視線を避ける行動をとるということが挙げられる。人間の心理状態がお互いの距離によって大きく変化するとは考えにくいことから、距離によってこのようにノンバーバル表現の表出の程度が変化するのは、親密性平衡モデルの平衡状態の維持が、各人が自分の心理状態をどの程度表出するかについての意図を変化させることで実現されているものと推察される。

3. ノンバーバル表現の制御モデル

3.1 ノンバーバル表現の定量化

前節で述べた知見をもとに、情報提供側が与えた発話に従って対話する二名の擬人化エージェントのノンバーバル表現を自動的に制御することを考える。本稿では擬人化エージェントに表出させるノンバーバル表現として視線、笑顔、頷きの三種類を考え、エージェント $Y (Y \in \{A, B\})$ が表出するこれら三種類のノンバーバル表現および発話の度合い (これを“表出度”と呼ぶ) を次のように定量化する。

視線: 単位時間にエージェント Y が相手エージェントに視線を向ける時間の割合 (e_Y^V)

笑顔: エージェント Y の笑いの程度 (e_Y^F)

連絡先: 伊藤 淳子, 京都大学大学院情報学研究科, 〒606-8501
京都市左京区吉田本町, tel:075-753-9060, fax:075-753-9056, itoh@mm.media.kyoto-u.ac.jp

領き: 単位時間におけるエージェント Y の領きの回数 (e_n^Y)
 発話: 単位時間におけるエージェント Y の発話語数 (e_s^Y)

人間のノンバーバル表現の表出度は、心理状態とそれをどの程度表出するかという意図、これらに関する個人差の影響を受けると考えられるため、上記の各表出度 $e_z^Y (Y \in \{A, B\}, z \in \{e, f, n, s\})$ を更に式 (1) で表す。ただし、 c_z^Y は意図、 x_z^Y はノンバーバル表現によって表出されるべき心理状態の度合い、 i_z^Y は表出の際の個人差、 u_z^Y は左辺と右辺の単位を合わせるための定数を表す。

$$e_z^Y = c_z^Y \cdot i_z^Y \cdot x_z^Y \cdot u_z^Y \quad (1)$$

3.2 相互依存性の表現

2.1 節で述べた相互依存性は、対話者が表出したノンバーバル表現の意味を観察者が解釈した結果に基づくものであることから、3.1 節で挙げた変数のうち、心理状態に関する関係として表現するのが適切であると考えられる。このとき、式 (1) の関係を用いると、表 1 に示した各ノンバーバル表現間の相互依存性の非充足度を評価する関数 $E_1 \sim E_8$ は次式で定義できる。

$$\begin{aligned} E_1 &= \left(\frac{e_e^A}{i_e^A c_e^A u_e^A} - \frac{e_e^B}{i_e^B c_e^B u_e^B} \right)^2 & E_5 &= \left(\frac{e_s^A}{i_s^A c_s^A u_s^A} - \frac{e_s^B}{i_s^B c_s^B u_s^B} \right)^2 \\ E_2 &= \left(\frac{e_f^A}{i_f^A c_f^A u_f^A} - \frac{e_f^B}{i_f^B c_f^B u_f^B} \right)^2 & E_6 &= \left(\frac{e_f^A}{i_f^A c_f^A u_f^A} - \frac{e_s^B}{i_s^B c_s^B u_s^B} \right)^2 \\ E_3 &= \left(\frac{e_n^A}{i_n^A c_n^A u_n^A} - \frac{e_n^B}{i_n^B c_n^B u_n^B} \right)^2 & E_7 &= \left(\frac{e_n^A}{i_n^A c_n^A u_n^A} - \frac{e_s^B}{i_s^B c_s^B u_s^B} \right)^2 \\ E_4 &= \left(\frac{e_e^A}{i_e^A c_e^A u_e^A} - \frac{e_s^B}{i_s^B c_s^B u_s^B} \right)^2 & E_8 &= \left(\frac{e_n^A}{i_n^A c_n^A u_n^A} - \frac{e_s^B}{i_s^B c_s^B u_s^B} \right)^2 \end{aligned} \quad (2)$$

一方、2.2 節で述べた親密性平衡モデルは、ノンバーバル表現の表出の意図に関するものであるため、 c_z^Y によって表現するのが適切と考えられる。そこで、親密度に対する、各ノンバーバル表現間に存在する相互依存性の非充足度を評価する関数 E_9 を式 (3) のように定義する。ただし、 I は二人のエージェント間に定まる親密度、 w_z^Y はどの意図をどの程度重視するかを表すための係数である。

$$E_9 = \left(I - \sum_{Y \in \{A, B\}, z \in \{e, f, n, s\}} w_z^Y c_z^Y \right)^2 \quad (3)$$

2 節で述べた相互依存性をすべて満足する表出度を求めるために、 $E_1 \sim E_9$ およびそれらの重みを決定する定数 $w_1 \sim w_9$ を用いて評価関数 E を次式で定義する。

$$E = \sum_{i=1}^9 w_i E_i^2 \quad (4)$$

E は e_z^Y および c_z^Y を変数とする関数となる。よって、これらのうち、与えられた発話に基づいて各エージェントの発話量 e_s^A, e_s^B のみを固定し、残りを E に対する最急降下法を用いて更新することにより、エージェントに表出させるべき $e_z^Y (z' \in \{e, f, n\})$ の値を各時刻ごとに求める。この最小化において、意図 c_z^Y は E_9 が最小となるように親密度 I に対応した平衡を維持するためお互いに制約しあう。その結果、 c_z^Y の値が増加あるいは減少すれば、 $E_1 \sim E_8$ を最小化するために e_z^Y の値がそれに依りて増加、減少する。

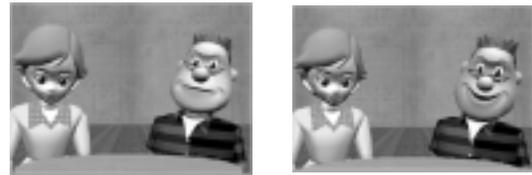
4. 対話映像の生成と評価

4.1 生成した対話映像の内容

本稿で述べた相互依存性の導入の有効性を調べるため、前節の手法で求めた表出度 e_z^Y を適用した場合とノンバーバル表現間の関連性を考慮しない場合のエージェントの対話映像を生

成、比較した。一定のノンバーバル表現を表出し続ける映像、発話時のみ発話エージェントのノンバーバル表現を制御した映像、および本手法により両方のエージェントのノンバーバル表現を制御した映像の三種類を用意し、三名の被験者に印象を聞いたところ、三名の被験者全員から本手法による映像が最も実映像から感じられる雰囲気近く、かつ自然であるとの回答が得られた。

なお、対話映像の作成には、 e_z^Y の値を直接反映させることができる、番組記述言語 TVML(TV program Making Language) を使用した [TVML]。エージェントの発話は、実世界における人間同士の対話映像から各人物の発話を書き起こして与えた。 E の係数は、映像生成により再現した部分とは別の部分の映像を用いて、エージェントの表出度が人間のそれと同等となるときに E が最小となるように逆算して事前に決定した。対話映像から表出度を求める際の単位時間は 5 秒とした。図 1 は本手法による生成映像例の一部であり、(a) が親密度を低く設定した映像、(b) が親密度を高く設定した映像である。



(a) 親密度が低い場合 (b) 親密度が高い場合

図 1: 生成映像の一部

5. おわりに

本稿では、社会心理学分野で報告されている人間同士の対話における視覚的ノンバーバル表現の研究をもとに、対話形式により情報提示を行う擬人化エージェントが表出するノンバーバル表現の程度を制御する手法を提案した。提案手法に基づいて生成した対話映像と制御を行わない映像との比較を行い、本手法の有効性を確かめた。

今後は、ジェスチャーや対人距離など、ほかの視覚的ノンバーバル表現の導入について検討する。

参考文献

- [Matarazzo 64] Matarazzo J.D., Saslow, G., Wiens, A.N., Weitman, M., and Allen, B.V.: Interviewer Head Nodding and Interviewee Speech Durations, *Psychotherapy: Theory, Research and Practice*, Vol.1, pp.54-63 (1964).
- [Argyle 65] Argyle, M., and Dean, J.: Eyecontact, distance and affiliation, *Sociometry*, 28, 289-304 (1965).
- [Kendon 67] Kendon, A.: Some functions of gaze direction in social interaction, *Acta Psychologica*, Vol.26, pp. 22-63 (1967).
- [Argyle 76] Argyle, M., and Cook, M.: *Gaze and Mutual Gaze*, Cambridge Univweaity Pewaa (1976).
- [Beattie 78] Beattie, G. W. : Sequential patterns of speech and gaze in dialogue, *Semiotica*, Vol.23, pp. 29-52 (1978).
- [Dimberg 82] Dimberg U.: Facial Reactions to Facial Expressions, *Psychophysiology*, No.6, Vol.19, pp.643-647 (1982).
- [Cassell 99] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H. and Yan, H.: Embodiment in Conversational Interfaces: Rea, CHI-99, pp.520-527 (1999).
- [久保田 02] 久保田 秀和, 山下 耕二, 福原 知宏, 西田 豊明: POC caster: インターネットコミュニティのための会話表現を用いた情報提供, エージェント, 人工知能学会論文誌, No.13 Vol.17 pp.313-321 (2002).
- [TVML] <http://www.str1.nhk.or.jp/TVML>