3O1-OS-3a-6

対話シーンによる違いを反映可能な 擬人化エージェントのノンバーバル表現の表出モデルの検討

Modeling tendencies of nonverbal expressions in different situations of human dialogues using a pair of embodied agents

奥内 啓太[*1] Keita Okuuchi　角所 考[*1] Koh Kakusho　小島 隆次[*2] Takatsugu Kojima　片上 大輔[*3] Daisuke Katagami

[*1] 関西学院大学 Kwansei Gakuin University　[*2] 滋賀医科大学 Shiga University of Medical Science　[*3] 東京工芸大学 Tokyo Polytechnic University

In this paper, we discuss a mathematical model to reproduce nonverbal expressions appearing in dialogues performed in informative TV programs, which include news, interviews, educational programs etc., in order to present various kinds of information to users in the form of dialogues by a pair of embodied agents. Although some knowledge about qualitative tendencies of nonverbal expressions in human conversations have already been obtained in the field of social science, quantitative descriptions of those tendencies are required to calculate the actual amount of nonverbal expressions worn by embodied agents. It is also important to describe differences in the appearance of each nonverbal expression in various situations of dialogues. We analyze nonverbal expressions appeared in various situations of informative TV programs by considering the knowledge of social science, and describe quantitative tendencies of their appearance in different situations in the form of a mathematical model .

1. はじめに

視線やしぐさ等のノンバーバル表現はコミュニケーションに重要な役割を果たすため，HAI (Human Agent Interaction)の分野では，人と擬人化エージェント間のコミュニケーションの円滑化を目指し，擬人化エージェントのノンバーバル表現の自動生成に関する研究が行われている．このような研究は，擬人化エージェントが人と直接対話することで様々な情報を提供するような用途を想定した議論[Cassell 2000]が多いが，人間同士の間の情報提供では，このような直接対話によるものだけでなく，TV のニュースや教育番組等に多く見られるように，複数人同士で意図的に演じられる対話を第三者に視聴させる形式も多用されることから，擬人化エージェントによる情報提供にエージェント間の対話を用いる試みもある[久保田 2002] [高橋 2011] .

人間同士の対話時に表出されるノンバーバル表現は当然，発話の内容に依存するが，理想的な対話では，片方の対話者が発話すると他方は相手への視線や頷きを表出する，といったように，発話内容に依存しない特徴的な表出傾向も見られるため，上述のような TV 番組の対話場面でも，出演者間に対話が成立しているように見せるための演出として，同様の表出傾向を伴う対話が演じられる．このような理想的な対話におけるノンバーバル表現の表出傾向には，定性的なレベルでは，社会心理学の分野で様々な知見が得られていることから，筆者らはこれを定量的な数理モデルの形で表現することにより，擬人化エージェント間対話において，同様の表出傾向を持つノンバーバル表現を生成することを試みてきた[伊藤 2003].

しかしながら，TV 番組の対話場面に限っても，定量的なレベルで考えれば，上のような表出傾向は必ずしも一律普遍ではなく，例えば真剣な対話シーンと和やかな対話シーンでは笑顔の表出傾向が異なるなど，対話シーンに依存した表出傾向の違いが存在すると考えられる．そこで本稿ではこのような違いを反映可能なノンバーバル表現の表出モデルについて検討する．

連絡先：奥内啓太，関西学院大学理工学部角所研究室，三田市学園 2-1 (Email: ccd77074@kwansei.ac.jp)

2. 対話シーン毎のノンバーバル表現の表出傾向

社会心理学の分野では，人間同士の対話における，表1①~⑯の発話量や視線，笑顔，頷き，身体動作等のノンバーバル表現間に，片方が増加すればもう一方も増加するという正の相関性が見られることが知られている[工藤 1999][Beattie 1978] .

対話者A,B間のノンバーバル表現の相互依存性					
		対話者A			
		発話量	視線	笑顔	頷き
対話者B	発話量		④	⑥	⑧
	視線	①	⑤		
	笑顔	②		⑦	
	頷き	③			

同一対話者A,Bのノンバーバル表現の相互依存性					
		視線	頷き	笑顔	身体動作
対話者A	発話量	⑨	⑩	⑪	⑫
対話者B	発話量	⑬	⑭	⑮	⑯

表 1. ノンバーバル表現間の関係性

TV 番組では理想的な対話が演じられるため，上のような相関性が多く現れると考えられるが，定量的には前述のような対話シーンによる違いも生じると考えられる．このような違いを生む対話シーンの分類基準として，本稿では，会話分析の従来研究[Clark 1996][Gatica-Perez 2009]等も参考に，(a)対話の主導権に関する役割関係，(b)対話者と視聴者の参与構造，(c)対話自体の雰囲気，の三条件に注目し，これらが異なる TV 番組映像における上記①~⑯の相関性を分析した．(a)の条件が異なる例として(A)ニュース番組，(B)バラエティ番組，(C)教育番組，(D)対談番組，(E)ショッピング番組，を選び，この中の対話場面を，(b)の条件が異なる例として，対話者同士の直接対話と視聴者に対する語りかけ（以後“対話”と“語り掛け”と呼称），(c)の条件が異なる例として，和やかと真剣，にそれぞれ分類し，これら

の組合せ毎にノンバーバル表現の表出量を単位区間毎に目視で判定した(最大値:1～最小値:0). その結果, 図 1～2 に示すように正の相関性が成り立つシーンと成り立たないシーンが存在し, 成り立つ場合も定量的な関係性は様々であった. 表 2 はこのときの相関性の有無をまとめたものである.

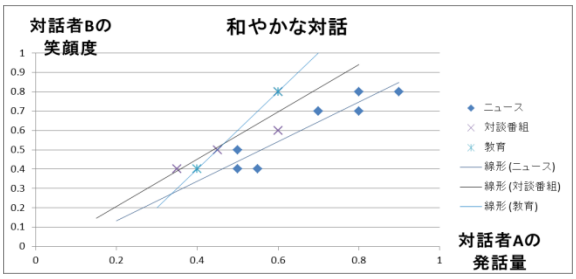


図 1.正の相関性に定量的な違いが見られるシーンの例

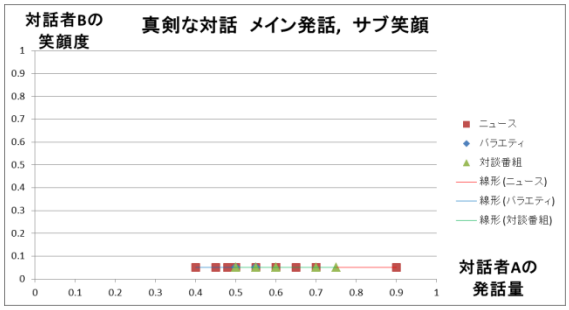


図 2.正の相関性が見られないシーンの例

	和やかな対話	真剣な対話	和やかな語りかけ	真剣な語りかけ
A)発話－B)視線	○	○	×	×
A)発話－B)笑顔	○	×	○	×
A)発話－B)頷き	○	○	○	○
B)発話－A)視線	○	○	×	×
B)発話－A)笑顔	○	×	○	×
B)発話－A)頷き	○	○	○	○
A)発話－A)視線	○	○	×	×
A)発話－A)笑顔	○	×	○	×
A)発話－A)頷き	○	○	○	○
A)発話－A)身体動作	○	○	○	○
B)発話－B)視線	○	○	×	×
B)発話－B)笑顔	○	×	○	×
B)発話－B)頷き	○	○	○	○
B)発話－B)身体動作	○	○	○	○
A)視線－B)視線	○	○	○	○
A)笑顔－B)笑顔	○	○	○	○

表 2. 相関性の有無

3. ノンバーバル表現の表出モデル

上述の議論を基に, ノンバーバル表現間の関係性を数理モデルとして記述することを試みる. まず, 従来の社会心理学で報告されている定性的なレベルでのノンバーバル表現間の正の相関性を表現するために, 以下のような制約条件を定義する.

$$E \equiv \sum_{i=1}^{16} E_i = \sum_{i=1}^{16} (x_i^X - y_i^Y) = 0 \qquad (1)$$

ここで, x_i^X, y_i^Y (i=1,･･･16; X, Y∈{A, B})は, 表 1 中の①～⑯で表したノンバーバル表現の組合せにおける対話者 X, Y のノンバーバル表現の単位区間毎の表出量を表す.

次に, 上の相関性には, 定量的なレベルでは図 1 に示したような関係性の違いが存在することを考慮し, この違いをそれぞれのノンバーバル表現の組合せ毎に係数 α_i で表すことにより, 次のように制約条件を修正する.

$$E' \equiv \sum_{i=1}^{16} E_i' = \sum_{i=1}^{16} (x_i^X - \alpha_i y_i^Y) = 0 \qquad (2)$$

さらに, 対話シーン毎の表 2 のような相関性の有無を表現するために, どの E_i' の充足が要求されるかを表現するフラグ l_i を導入し, 次のように制約条件を修正する.

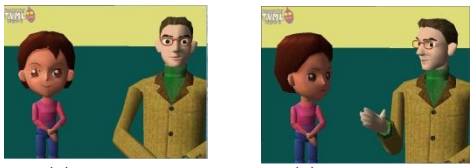
$$E'' \equiv \sum_{i=1}^{16} l_i E_i' = \sum_{i=1}^{16} l_i (x_i^X - \alpha_i y_i^Y) = 0 \qquad (3)$$

4. 実験結果

3.の式(3)をモデルに用いた場合の表出量の再現性を調べるため, まず x_i^X, y_i^Y と l_i に 2.で求めた TV 映像の各区間の表出量と相関性の有無を与えて E'' を α_i について最小化することにより, TV 映像のノンバーバル表現間の相関性を比例関係で近似する a_i の値を求め, 次にこの a_i を用いて、l_i と各区間の発話量を与えたときの他の表出量を, E'' を発話量以外の x_i^X, y_i^Y について最小化することで求めたところ, 実際の表出量との平均誤差は 0.2 程度であった. 図 3～4 は得られた表出量を TVML(TV program Making Language)を用いて擬人化エージェントで表現した結果であり, 図 3 では対話の主導権に関する役割関係が同じ場合の参与構造や雰囲気による違いが, 図 2 では雰囲気と参与構造が同じ場合の役割関係による違いが表れている.

(a)和やかな対話　(b)真剣な対話　(c)真剣な語りかけ

図 3. ニュース番組に対する異なるシーンの再現例



(a)ニュース　(b)バラエティ

図 4. 和やかな対話に対する異なるシーンの再現例

5. まとめ

TV 番組のような対話による情報提供を擬人化エージェント間の対話を用いて実現することを目標に, 実際の TV 番組におけるノンバーバル表現の表出傾向の対話シーン毎の違いを反映可能な表出モデルについて検討した. 今後は, ノンバーバル表現の認識処理の自動化により, より多量の TV 映像データを用いた表出傾向の分析とモデルの改良を進めていく予定である.

参考文献

[Cassell 2000] J.Cassell: Human Conversation as a System Framework: Designing Embodied Conversation Agent, MIT Press, 2000.

[久保田 2002] 久保田秀和他: 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, 2002

[高橋 2011] 高橋朋裕他:常時稼働を想定した情報インタフェースとしてのエージェント設計, HAI シンポジウム, 2011.

[伊藤 2003] 伊藤淳子他: 会話エージェントのためのノンバーバル表現間の相互依存性のモデル化, 情処研報, 2003

[工藤 1999] 工藤力:しぐさと表情の心理分析, 福村出版, 1999.

[Beattie 1978] G.W.Beattie: Sequential Temporal Patterns of Speech and Gaze in Dialogue, Semiotica, 1978.

[Clark 96] H.H.Clark: Using Language, Cambridge University Press, 1996.

[Gatica-Perez 09] D.Gatica-Perez: Automatic nonverbal analysis of social interaction in small groups, J. Image & Vision Computing, 2009