TV 番組の対話シーンに基づく 対話エージェントのノンバーバル表現のモデル化

Modeling nonverbal expressions for conversations between embodied agents based on dialogue scenes in TV programs

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

Abstract: A mathematical model to reproduce nonverbal expressions appearing in dialogues of TV programs such as news, interviews, educational program is proposed. The model is used to control nonverbal expressions of the embodied agents that present various information to users indirectly through conversations between them. In the field of social science, it is known that some tendencies are observed in the relationship between nonverbal expressions in human conversations. These tendencies are often emphasized especially in the case of conversations acted in TV programs so that those conversations look realistic to viewers. We analyze these tendencies to describe them in the form of a mathematical model for controlling nonverbal expressions of embodied agents in conversations with various situations for information presentation.

1. はじめに

人のコミュニケーションにおいて，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．このため，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きた．このときの擬人化エージェントには，人と直接対話して様々な情報を提供すること目指したもの[1]が多いが，人間同士の情報提供では，情報を提供する側の人間が，情報を受け取る側の人間と直接対話するような形式に加えて，TV のニュースや教育番組等に多く見られるように，情報提供側で複数人が対話を演じ，それを視聴者に対して一方的に提示するような形式も存在する．このため，この形式にならって，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[2][3]．本稿では，このような場合の擬人化エージェントのノンバーバル表現の自動生成について議論する．

人が発話する際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，対話の場合には，これに加えて，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，対話者間のノンバーバル表現間にも関係性が生じる．このような関係性については，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けるノンバーバル表現の自動生を試みてきた[4]．しかし，上の社会心理学分野の知見は，普遍的に見られるノンバーバル表現間の定性的な関係性を述べたものであり，実際の対話の各場面における個々のノンバーバル表現の表出量は，対話の状況によって様々に異なる．

一方，前述のような TV 番組の対話場面では，出演者間に対話が成立しているように見せるために，対話相手の発話に対して大きく頷くなど，このような関係性を明示するような形でノンバーバル表現が表出されると共に，対話状況の違いをわかりやすくするために，ユーモラスなシーンでははっきりした笑顔を表出するなど，状況に応じた表出量の強弱の違いも明確である．

そこで本稿では，上のような TV 番組の対話における各状況毎のノンバーバル表現の表出傾向をモデル化し，擬人化エージェントの対話を用いた情報提供におけるノンバーバル表現の自動生成に利用することを試みる．

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: d77074@kwansei.ac.jp)

2. ノンバーバル表現の表出傾向

2.1 ノンバーバル表現間の関係性

社会心理学の分野では，人間同士の対話において，発話量と視線，笑顔，頷き，身体動作等のノンバーバル表現の間に，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[5][6][7][8]．これらを異なる人物間および同一人物内についてそれぞれまとめたものが表 1 の(a),(b)であり，それぞれの表で

番号の付いているノンバーバル表現の組合せの間に関係性が見られることを示している.

表 1. ノンバーバル表現間の関係性

(a) 対話者 A,B 間のノンバーバル表現間の関係性

		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			

(b) 対話者 A,B 毎のノンバーバル表現間の関係性

		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

2.2 対話状況の分類

TV 番組では, 視聴者に対して出演者の対話がかみ合っているように見せるため, 上のような相関性がより明確に現れると考えられるが, その量的な表出傾向には対話状況に応じた違いがあると考えられる. これについて議論するには, まずそのような表出傾向に違いが生じる対話状況の種類を明らかにする必要がある.

この際に参考となるものとして, 会話分析の従来研究[9][10]がある. これらの研究では, 会話状況を分析する際に, 主導権や参与構造の違いに着目している. このうち, まず主導権の違いについては, グループ会話時における会話進行を主導する役割を表すものとして“フロア”という概念が導入されている. すなわち, グループ会話において, 会話の進行を主導する役割を持つ話し手は, フロア支配する権限を有することになる. 会話の各参与者のノンバーバル表現の表出傾向は, 誰がフロアを支配しているの違いによる影響を受けると考えられる.

また, 参与構造に関する議論では, 会話の参与者は“話し手”, “受け手”, “傍参与者”に分類される. このうち話し手は文字通り参与者の中の発話者である. 受け手は, 話し手の発話が直接向けられている相手のことで, 話し手からは何らかの直接的な応答を期待されている. 傍参与者は, 会話には参加しており, 他の参与者から同じ会話の参与者として認識されてはいるが, 話し手と受け手の会話を傍らで一方的に聞く立ち場にある人のことで, 話し手からは必ずしも直接的な応答を期待されていない. グループ会話では, 参与者間で誰が話し手, 受け手, 傍参与者となるのかは時々刻々変化し, それが参与構造の違いとなる.

上のような主導権や参与構造の違いは, TV 番組においても見られる. 例えば, 主導権の違いは, ニュース番組のようにメインキャスターが継続的に主導権を握り続けるような場合や, バラエティ番組での掛け合いのように主導権が交互に移り替わる場合などに見られる. また参与構造の違いは, 出演者同士が話し手と受け手の関係となって会話を進行し, 視聴者はそれを傍参与者として視聴するような場合や, 出演者のうちの片方が視聴者を受け手とみなして画面から直接語りかけ, もう一方の出演者はそれを傍参与者として聞いているような場合などに見られる. また, TV 番組の対話状況には, 上の他にも, 対話の雰囲気が真剣なものであるか, 和やかなものかといった雰囲気による違い考えられる.

以上の考察から, TV 番組の対話状況には, (A)対話の主導権に関する出演者間の役割関係, (B)出演者と者と視聴者の間の参与構造, (C)対話自体の雰囲気, の3つによる違いが存在することが予想される.

2.3 被験者による対話状況変化の検出実験

上のような分類による対話状況の違いが, 実際に視聴者にとって意識されているのかを調べるために, 被験者による実験を行った. 被験者は 18 人の大学生・大学院生で, 様々な TV 番組の対話シーンを各数分程度視聴してもらい, 映像の途中で対話状況に変化があったと思った瞬間を指摘してもらうと共に, 何故そのように思ったかの理由を尋ねた. TV 番組としては, 対話による情報提示を意図していると思われる TV 番組として, (ア)ニュース番組, (イ)バラエティ番組, (ウ)教育番組, (エ)対談番組, (オ)ショッピング番組の5種類に注目し, それぞれの例として, (ア)にはテレビ朝日の“報道ステーション”, (イ)にはテレビ大阪の“きらきらアフロ”, (ウ)には.NHK の“高校講座”, (エ)にはテレビ朝日の“徹子の部屋”, (オ)にはサンテレビの“ジャパネットたかた”を用いた.

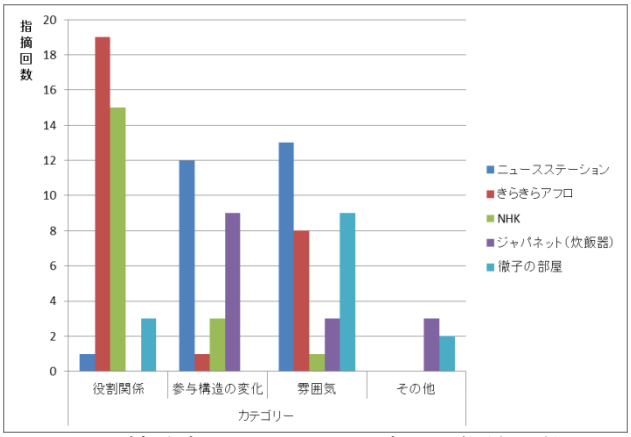


図 1. 被験者による対話状況変化の指摘回数

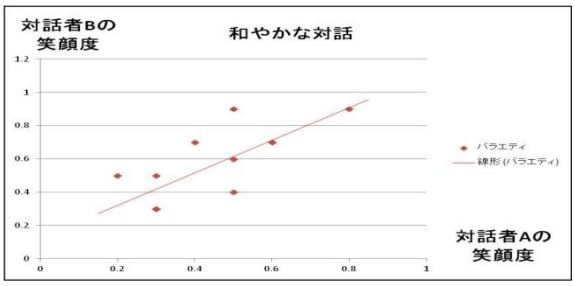
これらの映像に対して被験者が対話状況の違いとして指摘した理由には,「二人の雑談から指示者と従事者の関係になった」,「アシスタントが主導的に会話を進めるようになった」など, 2.2 の(A)に関するもの,「二人の対話から, 視聴者向けの話に変わった」,「二人が話している状態から, 視聴者に問いかけるような形になった」といった(B)に関するもの,「雰囲気が真面目な感じから, 和やかな感じになった」,「真剣な感じから笑顔な感じに」といった(C)に関するものが実際に見られた. また, それ以外の指摘理由としては,「話題が転換した」など, 発話内容に関するものが多くみられた. 本研究では, 擬人化エージェントのノンバーバル表現の生成手法に焦点を当てていることから, 上の指摘理由の中から発話の内容に関するものを除き, 残りについての指摘回数を(A)～(C)およびそれ以外について集計してみた. 結果を図 1 に示す. この結果より, 上述した(A)～(C)に関する指摘

が殆どであることから，視聴者にとってもこの3つが主要な対話状況として意識されていることが確認できた．

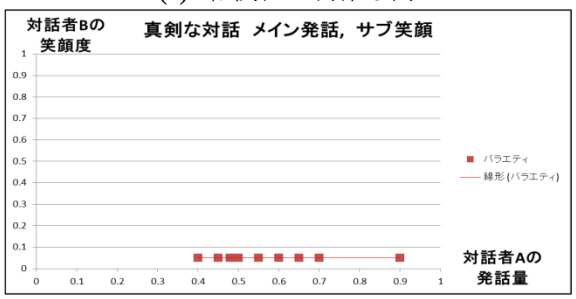
TV 番組では，通常の会話とは異なり，番組毎に出演者が対話で果たす役割が決まっていることから，(A)～(C)のうち，(A)の違いによる対話状況は番組に依存して決まると考えられる．そこで本研究では，各番組内で変化し得る対話状況のカテゴリとしては，(B)と(C)の違いに基づく，「和やかな対話」，「真剣な対話」，「和やかな語りかけ」，「真剣な語りかけ」の 4 つを考える．

2.4 対話状況毎の表出傾向の分析

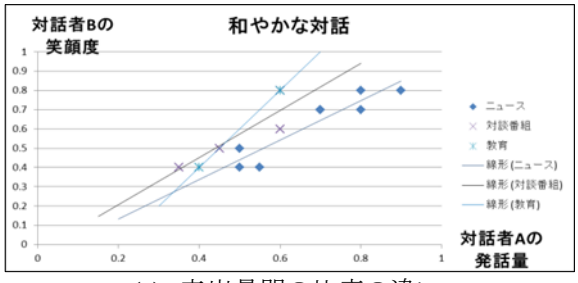
2.1～2.3 における対話状況の分類結果に基づき，各番組の各カテゴリ毎に，ノンバーバル表現の表出傾向にどのような違いがあるのかを調べた．映像を 3 秒毎に区切り，各区間毎に，表 1 と同じ発話量および相手への視線，笑顔，頷き，身体動作の表出量を最大値 1，最小値0として目視でラベル付けした．その結果，表1に示したそれぞれの関係性について，成り立つものと成り立たないものが存在することがわかった．



(a) 相関性が明確な例



(b) 相関性が見られない例



(c) 表出量間の比率の違い

図 2 ノンバーバル表現間の相関性の違いの例

図 2(a)は，(A)の主導権が主演者間で入れ替わるバラエティ番組における，(B)が出演者間の対話形式で(C)が和やかな雰囲気のときの2人の出演者 A，B の笑顔度の表出量をグラフにしたもので，両者間に正の相関性が存在することがわかる．一方図(b)は(a)と同じ番組で，(B)の参与構造が対話形式で(C)の雰囲気が真剣な場合の A の発話量と B の笑顔度の関係で，相関性は見られない．以上の結果を表にまとめたものが表 2 である．ただし，図(c)に示すように，この関係を異なる番組で比較してもると，正の相関性が見られたとしても，その量的な比率は様々となる．

表 2. 対話状況における相関性の有無

	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	○	○	○	○

3. ノンバーバル表現の表出モデル

2.4 において，相関性が明確であったものを擬人化エージェントによる対話状況毎のノンバーバル表現として再現することを目指し，これを数理モデルとして記述することを試みる．このためにまず，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，以下のような制約関数で表現する．

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

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各組合せに対する対話者 X, Y の単位区間毎の表出量を表す．

上の表出量には，図 2(c)に示したような量的な比率の違いが存在する．そこで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で表し，上の制約条件を次のように修正する．

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

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，どのノンバーバル表現の各組合せに対してE_i'の充足が要求されるのかを表現するフラグ l_iを導入し，制約関数をさらに次のように修正する．

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

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た．まずx_i^X, y_i^Yとl_iに 2.3 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について最小化し，この相関性を比例関係で近似するa_iの値を求めた．これは目視で得た実際の表出量を直線で近似することに相当し，このときの実際の表出量の近似誤差の平均値は 0.2 程度であった．

上で得られたa_iの値を用いて，l_iと各区間の発話量のみを与え，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，それぞれのノンバーバル表現の表出量を求めた．図 3 は，バラエティ番組における真剣な対話，和やかな対話，和やかな語りかけの 3 つの対話状況の変化に伴って生成されたノンバーバル表現の表出量の時間変化を示したものである．この 3 つの対話状況の変化の中で，まず真剣な対話時には，真剣という雰囲気により緑線で示された笑顔の表出量が減少する一方，対話であるため，赤線で示す対話相手に視線を向ける度合いは高くなっている．続く和やかな対話には雰囲気が真剣から和やかに変化したことで笑顔の表出量が増加している．また真剣な対話同様に対話という状況は変わらないため引き続き視線の度合いが高くなっている．最後に和やかな語りかけになると，和やかな雰囲気が継続されるので笑顔の表出量が高い一方で，対話が語りかけに変化しているために対話相手に視線を向ける割合が低下している．このように本モデルでは表 2 で示した対話状況による表出傾向の違いが再現できている．図 4 は TVML(TV program Making Language)[11]を用いて図 3 における各状況の代表的な瞬間を擬人化エージェントで表現した結果である．まず真剣な対話時においては視線を相手に向ける度合いが高いため視線を相手に向け，笑顔の表出量は低いため(a)で示すようになっている．続いて和やかな対話では，笑顔の表出量が高くなったため表情に笑顔が見られるようになり，視線を相手に向ける割合は 7 割程度に減少したため，(c)が 7 割，(d)が 3 割という形で表れるようになっている．最後に和やかな語りかけでは，視線を相手に向ける割合が減少したため(d)で示すように視聴者方向に向ける割合が多くなる．

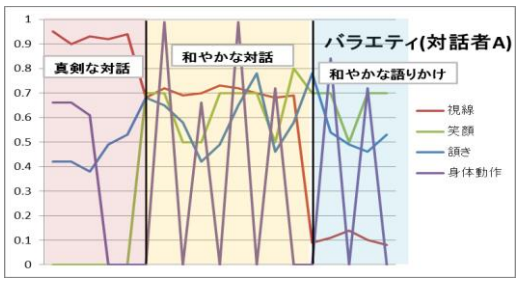


図 3. モデルで得られた表出量の時間変化


(a)真剣な対話　(b)和やかな語りかけ


(c)和やかな対話　(d)和やかな対話

図 4.擬人化エージェントによるバラエティ番組の再現例

5. まとめ

本研究では 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．被験者実験を通じてノンバーバル表現の表出傾向を左右する対話状況を明らかにすると共に，各対話状況毎の表出傾向を実際の TV 番組を基に分析し，これを再現する数理的なモデルを提案した．実験より，対話状況毎の表出傾向の違いが再現できることを確認できた．

上のモデルは，発話語数を基にノンバーバル表現の表出量を決めるものであるため，発話内容までは考慮に入れていない．しかしながらノンバーバル表現の表出量が発話内容に左右される場合も当然感がられ，これに対する対応は今後の課題の一つである．また，実際の映像の表出量を目視で獲得する場合の個人差の問題についても考慮できておらず，この影響についての分析も今後の課題と考えている．

参考文献

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

[2] 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, 2002

[3] 高橋朋裕，片上大輔: 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.

[4] 伊藤淳子，角所考，美濃導彦: 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

[5] 工藤力: しぐさと表情の心理分析，福村出版, 1999.

[6] 斉藤勇　古屋健，大坊郁夫，鈴木晶夫，白井泰子: 対人社会心理学重要研究集 3　対人コミュニケーションの心理，誠信書房, 1987.

[7] 吉川左紀子，中村真: 顔・表情の違いによる発話行動の調整，電子情報通信学会論文誌(A) Vol80, pp1324-1331, 1997.

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

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

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

[11] TVML: http://www.nhk.or.jp/strl/tvml/index.html