

会話時の自然な視線インタラクションを再現する 会話エージェント

葛島 健人^{1,a)} 三武 裕玄^{1,b)} 長谷川 晶一¹

概要: 会話エージェントが生活を楽しむような対話相手となるには、会話エージェントが社会性を感じさせる自然な会話動作・視線を行って機械的でない会話への関心を対話相手に示す必要がある。本研究では会話において話者の動作・視線に応じてこれら自然な反応動作・視線を実現する手法を提案する。実会話における動作・視線・音声を記録し、記録を再現するような行動決定モデルを機械学習により獲得することで、実会話事例が含む細やかな会話動作の特徴を持つ会話インタラクションを再現する。

Conversational Agent with Natural Eye Gaze Behaviour

KENTO KUZUSHIMA^{1,a)} HIRONORI MITAKE^{1,b)} SHOICHI HASEGAWA¹

Abstract: Conversation agents are expected to be a life partners with social behaviours. Agents are required to give social and natural action, motion and gaze to express interest of the agent to ongoing conversation. We propose a method to generate reaction motion and gaze according to users' motion and gaze during conversation. The method includes to capture motion, gaze and voice in conversation, and learning HMM with the data. Using the HMM as action determination model, our conversation agent achieve naturalness based on real example of conversation.

1. はじめに

近年、キャラクターを持ったロボットの登場や、没入 VR の発達によって人とキャラクタを同じ世界に立たせることが可能になってきた。人はこれまで、ゲーム等の仮想世界においてキャラクタに様々な役割を与え、コントローラを通してインタラクションを行ってきたが、キャラクタと同じ世界に立つことが可能になった今、キャラクタに会話等の直接的な社会的コミュニケーションを含むインタラクションを求めている。中でも会話は、社会的コミュニケーションの重要な要素であり、人はキャラクタと自然な会話を行えないと相手に社会性や知性を感じられずコミュニケーションへの期待を削がれてしまう。そのため、現在、自然な会話を実現するための会話エージェントの研究が盛

んに行われている。

会話エージェントとの自然な会話を実現するためには、話し言葉や文字といった言語インタラクションだけでなく視線やジェスチャーといった非言語インタラクションを実現することが必要である。

また、自然な会話の実現に必要なこととして会話のインタラクティブ性の再現が挙げられる。会話エージェントは話者の発話・視線・動作といった会話行動に対してリアルタイムに反応動作を決定、行動しなければならない。従来の会話エージェントは人手で作成した行動決定モデルに従って動作するものが多いが、人手で作成するためには実会話を注意深く観察し、法則性を発見、アルゴリズム化する必要がある、容易な作業ではない。また「落ち着いている」や「はつらつとしている」等、特徴の違うエージェントを作る際には、新しく行動決定モデルを作成せねばならず、行動決定モデルを全て人手で作成することは現実的ではない。

そこで本研究では、会話エージェントが話者の会話行動

¹ 東京工業大学
Tokyo Institute of Technology, Yokohama, Kanagawa 226-8503, Japan

a) kento.kuzushima@haselab.net

b) mitake@haselab.net

に応じてリアルタイムに反応動作を決定するための行動決定モデルを，実会話事例から HMM を用いた機械学習によって獲得することを目的とする．実会話における会話行動を記録し，それを再現するような行動決定モデルを機械学習により獲得することで，従来の人手で行動決定モデルを作成する方法よりも自然な会話を実現することが期待できる．また，学習に用いるデータを変更することで特徴の違うエージェントを簡単に作成できることが期待される．

2. 関連研究

非言語情報も用いてコミュニケーションを行う会話エージェントの研究は数多くなされている．渡辺らの研究 [1] では話者の発話の切れ目を検知するモデルを作成し，それを用いることでエージェントが話者に頷きを返すタイミングを適切に取得できることを示した．ただし，このエージェントは発話にのみ反応し，話者の視線や動作には反応しないため，不自然さが生じる恐れがある．DeVault らの研究 [2] では話者の発話や視線，表情や動作といった多くの会話行動に応じて適切な動作やタイミングを返すエージェントを作成した．この研究では，行動データや行動決定モデルを人手で精巧に作成することでジェスチャーの選択やそれを話者に返すタイミングについては適切である一方，視線についてはモデルの作成が十分ではなく，エージェントが話者から視線を動かさない，といった不自然さを感じることがある．

会話における視線の機能の研究として Kendon らの研究 [3] が挙げられる．この研究では，二人会話の視線を観察することによって，視線が聞き手によって「次話者になるための発話準備中であること」を伝えるために使用されたり，話者によって「発話権を譲ろうとしていること」を伝えるために使用されたりと，様々な用途で視線が会話に利用されていることが分かった．

会話時の視線の重要性に着目し，それを再現する研究として益子らの研究 [4] が挙げられる．この研究では，話者の発話，視線，動作と，エージェントの発話情報から，エージェントの会話状態を同定し，その状態に適切な視線と頭部の動作を決定，再現した．ただし，この研究ではエージェントの行動決定モデルは人手で用意したものであるため，自然な会話を十分に再現できない可能性がある．

3. 提案手法

人間同士が会話で行う自然な視線移動の振る舞いを再現するため，実会話事例における視線移動の時系列記録に基づき隠れマルコフモデル (HMM) を学習し，獲得した HMM を行動決定モデルとして用いることでインタラクティブに会話行動を決定する．決定に基づきキャラクタ身体物理モデルを駆動することで会話動作を実現する．全体像を図 1，図 2 に示す．

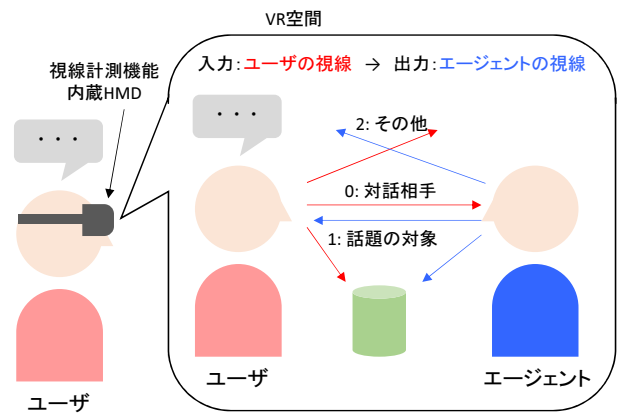


図 1 提案システムによるインタラクション

Fig. 1 Interaction System

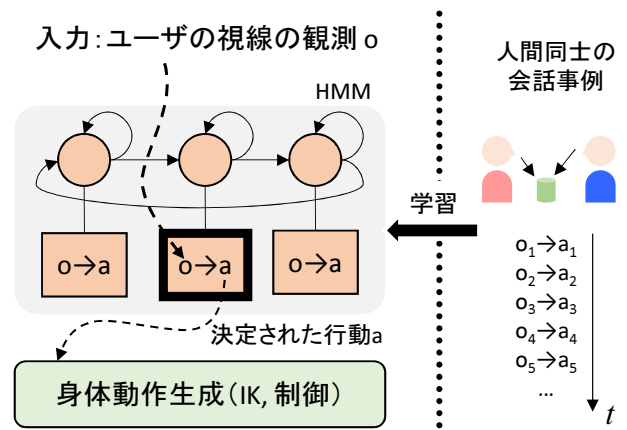


図 2 会話動作生成システム

Fig. 2 Action Generation Method

3.1 実会話事例に基づく HMM の学習

会話エージェントの行動決定の規範となる，人間同士の会話の様子を記録する．今回は 2 名の話者が話題の対象となる物体を挟んで対面会話を行う様子を記録した．次に注視状態の時系列データを作成する．今回は，話者それぞれについて注視対象を 0:対話相手，1:話題の対象となっている物体，2:それ以外の 3 種類に分類した．話者 1 の注視状態を o ，話者 2 の注視状態を a として，両者の組 $(o, a) (o, a \in \{0, 1, 2\})$ を 0.1 秒刻みで並べた時系列データを用意した．得られた時系列データにより HMM の学習を行った．学習には機械学習ライブラリ Accord.NET [5] を用いた．図 3 に得られた HMM の例を示す．

3.2 HMM によるインタラクティブ行動決定

HMM の遷移行列を T ，状態 i の平均出力を μ_i ，時刻 n に状態 i にいる確率を $p_i(t)$ とする．前節で獲得した HMM の出力記号は注視状態の組 (o, a) であるが，行動決定手法として用いる際は o を対話相手の注視状態 (=観測)， a を対応するエージェントの注視状態 (=行動) と見なす．すなわち，本手法は観測 o を入力として行動 a を決定する．まず，HMM の遷移確率行列 T に基づき状態確率分布 $p(t)$ を

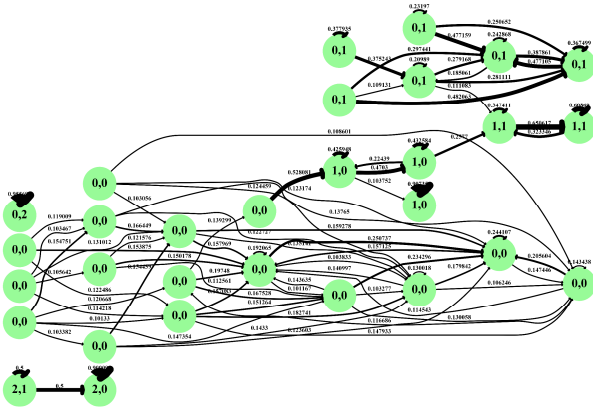


図 3 会話動作を学習した HMM の例

Fig. 3 Example HMM from Conversation Data

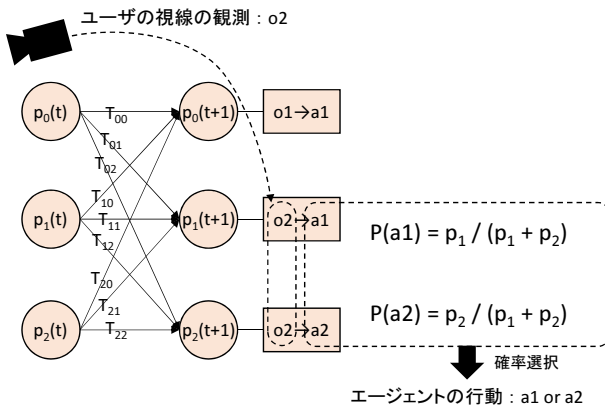


図 4 HMM による行動決定

Fig. 4 HMM Based Action Determination

更新する。

$$p(t+1) = T^t p(t) \quad (1)$$

次に観測 $o(t+1)$ を出力に含む状態の集合を求める。これらが行動の候補状態となる。

$$S(t+1) = \{i \in N | \exists a; \mu_i = (o(t+1), a)\} \quad (2)$$

行動候補状態集合において行動 a_n を出力する状態にいる確率を足し合わせ、行動 a_n が選ばれる確率 $P(a_n)$ を得る。

$$P(a_n) = \frac{\sum_{i \in \{ \exists o; \mu_i = (o, a_n) \}} p_i(t+1)}{\sum_{i \in S(t+1)} p_i(t+1)} \quad (3)$$

最後に、確率 $P(a_n)$ に基づき行動を選択する。以上の処理を学習に用いた時系列データと同じ時間刻み (今回は 0.1 秒) ごとに実行することで、インタラクティブに行動を決定する。本手法の概要を図 4 に示す。

3.3 身体動作生成

行動決定結果に基づき、キャラクタ身体物理モデルの眼球・頭部・胴体を動作させ注視動作を実現する。眼球・頭部が注視対象の方を向くような姿勢を目標姿勢とし、各関節角度を逆運動学により決定する。目標関節角度を実現す

るよう関節トルクを PD 制御により与える。注視対象が変化した際は目標姿勢間を滑らかな加減速で補間する。逆運動学における関節の重み付け、PD 制御の係数を調整し自然な身体動作を得る。

4. 結果と考察

実会話から学習で得られた HMM を用いて会話エージェントと視線インタラクションを行った。インタラクション中の様子を図 4-1 から図 4-5 に示す。図の右側のキャラクタが機械学習で得られたエージェントであり、ユーザが操作するのは左側のキャラクタである。また中央の白い板が話題の対象となっている物体を表す。

インタラクション中、ふたりとも向き合っている状態 (図 5-1) からユーザが物体を注視する (図 5-2) と、エージェントもそれに応じて物体を注視する状況 (図 5-3) がしばしば見ることができた。これは学習データに使った実会話事例にも見られた共同注視を、学習によってエージェントで再現することができたと考えられる。また、両者が物体を注視している状況 (図 5-3) や、エージェントが物体を注視している状況 (図 5-4) からユーザがその他の場所を見ると、一瞬でエージェントがユーザの方に視線を戻す状況 (図 5-5) をしばしば見ることができた。これも、実会話事例内に存在した「話者が視線を逸らしながら喋りかけはじめ、それを受けて聞き手が相手の顔を見て傾聴する」という現象を再現できたと考えられる。これらの結果から、実会話事例を学習し、会話エージェントを構築することで、実会話事例内に含まれている特徴を再現できることが分かった。

5. 今後の課題

本研究では、自然な視線移動の振る舞いを視線のみのデータから再現しようと試み、いくつかの実会話事例の特徴を再現できることが分かった。しかし、実会話での視線は「大きい声を出した相手を見る」といった様に、会話相手の視線だけでなく、発話や動作といった他の会話行動からも影響を受け生成されるものであると考えられる。また、今回は会話中の視線移動に着目し、その再現を目的としたが、今回の手法は傾き等、視線以外の非言語的な振る舞いの再現にも利用できることが期待できる。以上の点から、今後の課題として、視線だけでなく発話や動作も学習データとして利用し、また、自然な会話を再現するために複数の非言語コミュニケーションを行うことができるようなエージェントを作成することが挙げられる。

謝辞

本研究 (の一部) は国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」の支援によって行われた。

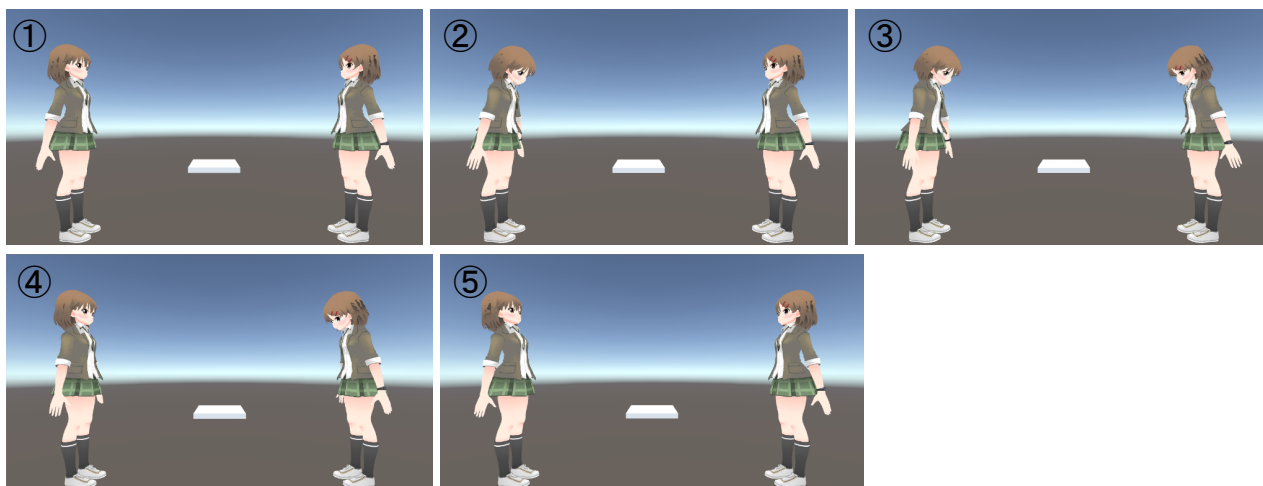


図 5 出力例

Fig. 5 Example Output

参考文献

- [1] 渡辺富夫：身体的コミュニケーションにおける引き込みと身体性-心が通う身体的コミュニケーションシステム E-COSMIC の開発を通して-, ベビーサイエンス, Vol. 2, pp. 4-12 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/10018152055/>) (2003).
- [2] DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A. and Morency, L.-P.: SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support, *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems, AAMAS '14*, Richland, SC, International Foundation for Autonomous Agents and Multiagent Systems, pp. 1061-1068 (online), available from (<http://dl.acm.org/citation.cfm?id=2617388.2617415>) (2014).
- [3] KENDON, A.: Some functions of gaze-direction in social interaction, *Acta Psychologica*, Vol. 26, pp. 22-63 (online), DOI: 10.1016/0001-6918(67)90005-4 (1967).
- [4] 益子宗：利用者注意に基づく能動的 CG, 修士論文, 筑波大学修士論文 (2004).
- [5] : Accord.NET Framework, <http://accord-framework.net/>.