Vol.2018-CVIM-212 No.33
2018/5/10

瞬きの引き込み現象を援用した対話エージェント

李 明輝[1,a)]　福田 悠人[1,b)]　小林 貴訓[1,c)]　久野 義徳[1,d)]

概要：人間同士が会話をするとき，聞き手が話し手の瞬きに合わせて瞬きをする現象，いわゆる瞬きの引き込み現象が起きる．この瞬きは，聞き手が無意識的に行うものであるが，ジェスチャや表情変化などと同様に，コミュニケーションの一部であると考えられる．そこで，本稿では，この瞬きの引き込み現象を援用したエージェントを提案し，人間とエージェントの会話においても，同様の現象が起きるかを実験により確認する．また，この瞬きの引き込み現象をエージェントに真似させることで，エージェントのイメージアップが図れるかを調査した．

Communication Agent Incorporating the Effects of Eyeblink Entrainment

Minghui Li[1,a)]　Hisato Fukuda[1,b)]　Yoshinori Kobayashi[1,c)]　Yoshinori Kuno[1,d)]

Abstract: Spontaneous eye blinks are highly synchronized between humans when they speak face to face. These kinds of eyeblinks can be considered a type of communication modality in the same sense as gestures or facial expressions. These eye blinking can be used as communication signals that indicate the listener's attention towards the speaker. In this paper, we propose a novel communication agent that incorporates eyeblink entrainment effects to improve its communication abilities. We confirm that the eyeblink entrainment occurs even when a computer graphics agent plays the role of communication partner (speaker or listener) through the experiments.

1. はじめに

高齢化社会の進行に伴い，今日では，現役世代 2.3 人で 1 人の高齢者を支える社会となっている [1]．平成 28 年に行われた「介護労働実態調査」によると，「従業員不足」と感じる介護サービス事業所が全体の 62.6%を占め，介護サービスの一番の問題として「人手不足」が取り上げられている [2]．このような介護者不足の問題を解決するために，人間の代わりに高齢者の介護をするロボットの研究が盛んに行われている．実際に，コミュニケーションロボットの利用により，被介護者の自立性の向上や生活の活発化の効果も認められ，コミュニケーションを「目的」とするだけでなく，介護手段としてうまく利用することによる「促し」の効果も大きいことが報告されている [3]．また，武川は，エージェント・ロボットは「ユーザが使うときに人と接するときと同じ手続きを使うことができるような設計」にすることが重要であり，良いコミュニケーションができるエージェント・ロボットは，会話の内容だけでなく，人間が話をしているときの振る舞いも表現できることが大切であると述べている [4]．

より良いエージェント・ロボットとユーザのコミュニケーションのために，我々のグループでも，これまでに，音声認識や表情認識を用いて対話相手の反応を測定し，それに合わせて話題を変更したり，ロボットに特定の動きをさせたりする研究を行ってきた [5],[6]．本稿では，人間が話をしているときに無意識的に行う行動である瞬きに注目して，人間のような瞬きができるエージェントの開発をめざす．また，会話エージェントに人間のような瞬きをさせることで，ユーザのイメージがどのように変化するか．会話全体の流れの改善につなげられるかどうかを調査する．

1 埼玉大学
Saitama University
a) zhang4bao@hci.ics.saitama-u.ac.jp
b) fukuda@cv.ics.saitama-u.ac.jp
c) yosinori@hci.ics.saitama-u.ac.jp
d) kuno@cv.ics.saitama-u.ac.jp

© 2018 Information Processing Society of Japan

中野らは，テレビ演説を視聴したときの 14 名の被験者たちの瞬きを調べ，話し手 (演説者) の瞬き開始時刻から 250〜500 ミリ秒遅れたところで，聞き手 (被験者) の瞬き生起確率が有意に増加していることを明らかにした．また，このような現象は，音声か画像のどちらが一方のみ提示した場合は観測されず，「対面して話を聞く」という状況に対して特異的に生じた現象であった．そして，演説者の瞬きのタイミングを調べたところ，約 7 割の瞬きは発話の合間 (OFF タイプ)，あるいは発話の終わり (OFFSET タイプ) で生じており，発話開始時 (ONSET タイプ) や発話中 (ON タイプ) で生じた瞬きは 3 割程度であった．瞬きの引き込み現象は OFF タイプか OFFSET タイプの瞬きに対して起きていた [7]．これらの結果から，瞬きが個人内の情報処理だけでなく，個人間の相互作用という社会性認知においても重要な役割を果たしていることを示した [8]．

2. 対話エージェント

本稿では，MMDAgent[9] を利用して対話エージェントを作成する．瞬きの引き込み現象を再現するために，ユーザの瞬きなどを映像から検出し，MMD モデルに対して，動作指示を行う．

2.1 瞬き検出モジュール

OpenCV に実装されている顔検出器を用いて，入力映像の各フレームで顔領域の検出を行ってから，Chehra[11] を用いて顔の特徴点を検出し，目の辺りの特徴点の座標を用いて瞬きを検出する．

- 瞬きの検出
  Chehra によって検出された両目の上下と左右の特徴点の座標から両目それぞれの上下と左右の距離 lh，lw，rh，rw を求める．次に，両目それぞれの横 (左右) と縦 (上下) の距離比 lq，rq を求める．最後に実験的に適切な閾値 t を決め，求めた距離比 lq か rq が t 以下であれば瞬きが発生したと判断する．
- 眼球の動きの計測
  動画の最初のフレームから，Chehra によって検出された瞳孔位置の x 座標を求め，lx_0，rx_0 として記録する．次に，各フレームにおいて瞳孔の位置の x 座標 lx，rx を求め，現在位置の x 座標と最初の x 座標の距離 ld，rd を求める．

  $$ld = lx - lx_0, rd = rx - rx_0$$

  実験的に適切な閾値 $t1$ を決め，距離 ld か rd が $t1$ を超えたときに眼球が動いたと判断し，その時刻の両目の瞳孔の x 座標を記録する．また，その時刻の両目の瞳孔の x 座標 lx，rx を用いて lx_0，rx_0 を更新する．最後に両目の中心点の x 座標 lxc，rxc と瞳孔の x 座標の差 ldc，rdc を求め，目の横幅に基づいて瞳孔の相対位置 lqc，rqc を求める．

  $$ldc = lx - lxc, rdc = rx - rxc$$

  $$lqc = ldc/lw, rqc = rdc/rw$$

- MMDAgent との通信
  瞬きと眼球の動きの検出結果に基づいて MMD モデルを動かすため，瞬き検出モジュールと MMDAgent の間で通信を行う必要がある．具体的には，瞬き検出モジュールはクライアントとして動作させ，MMDAgent に動作生成モジュールをサーバとして組み込んで，ソケット通信を行う．瞬きや眼球の動きを検出するたびに，その時刻のフレーム ID と瞳孔の相対位置 lqc，rqc を送信する．

2.2 動作生成モジュール

カメラやビデオなどの映像中のユーザの瞬きと眼球の動き，及び発話中の口の動きを対話エージェントで再現するために，MMDAgent に動作生成モジュールを組み込む．

- 映像の音声に合わせて口パクする方法
  MMDAgent の音声認識用のツール Julius[10] を有効にし，パソコンの音声入力をステレオミキサーに設定してから映像を再生すると，Julius によって映像中の音量を検出できる．Julius が認識できる最小音量を超えるときに音声があると判定し，そうでないときは音声がないものとして判定する．音声が「有から無」と「無から有」になるタイミング (フレーム ID) を記録し，MMDAgent のシナリオに「無から有」のタイミングが来たら MMD モデルに口の開閉を繰り返させ，「有から無」のタイミングが来たら動きを止めるように設定する．
- 瞬き検出モジュールとの通信
  瞬きや眼球運動を動作生成モジュールで扱うために，瞬き検出モジュールが検出した瞬きのタイミングや眼球が動いたフレーム ID と瞳孔の相対位置をソケット通信を介して受信する．
- 再現モーションの生成
  動画の中の人の瞬きと眼球の動きを再現するために，検出プログラムによって検出した瞬きと眼球運動が発生したフレームと眼球の相対位置を利用して，MMD モデルを同じ動きをさせるための再現モーションを生成する．再現モーションの開始と同時に口パクのシナリオを開始することで，MMD モデルによって映像中のユーザの瞬きと眼球運動，口の動きを再現できる．

3. 実験

先行研究により，人間の演説者と視聴者同士の間に「瞬きの引き込み現象」が存在することが分かった．また，人間同士の会話においては，瞬きによって無意識的にコミュ

ニケーションを取っているとも考えられる．我々はこのような現象が人間とエージェントの会話にも存在するのではないか．また，このような無意識的なコミュニケーションの真似をすることで，エージェントが利用者に与える印象が改善できるではないかと考えた．この二つの仮説を確かめるために，「被験者にエージェントによる演説 (の再現動画) を聞いてもらう」実験と「被験者にエージェントに話してもらう」実験の二通りの実験を行った．すなわち，会話においてエージェントが「話をする側」と「話を聞く側」の 2 パターンで被験者の反応を調べる実験である．実験では前節で述べた対話エージェントを用いて，眼球運動や瞬きなどの動き再現し，対話エージェントが表示された画面を被験者に見せて，被験者の反応を分析する．

3.1 実験 1(エージェントが話をする側)

3.1.1 実験方法

まず，前節で述べた対話エージェントを用いて，先行研究 [7] でも利用された木村拓哉がドラマの中で行った演説シーンの再現動画 1 を作成した．本稿では，木村拓哉にイメージの近い MMD モデル (図 1) を利用した [12]．また，比較のために，口と眼球の動きが同じで瞬きだけを等間隔にする再現動画 2 も作成した．

図 1 MMD モデルによる再現動画 (左) 瞬きをする瞬間 (右) [12]

次に，6 人の被験者に，順に元の動画、再現動画 1 と再現動画 2 を見てもらい，被験者が動画を見ているときの様子をビデオで記録した (図 2)．見終わったら動画に関するアンケートを答えてもらった．実験後，撮影した被験者のビデオ映像を解析して，瞬きの引き込み現象が発生したかどうかを調べ，各動画のアンケート得点を分析した．

図 2 実験 1 のイメージ図

3.1.2 実験結果

実験で記録した被験者たちの瞬きタイミングと演説者の瞬きタイミングの関係を，先行研究と同様にサロゲート法を用いて分析した結果を図 3-5 に示す．

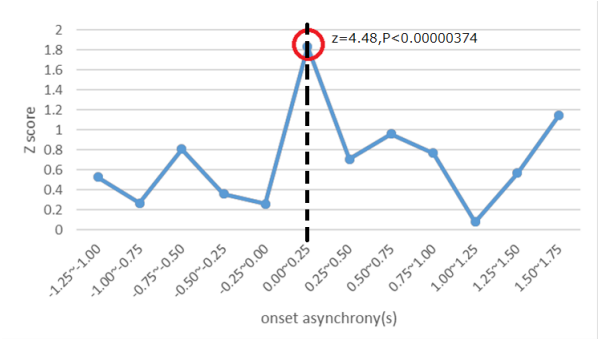

図 3 元動画

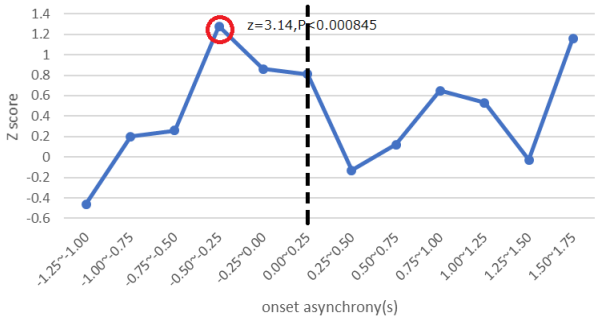

図 4 再現動画 1(元動画と同じタイミング)

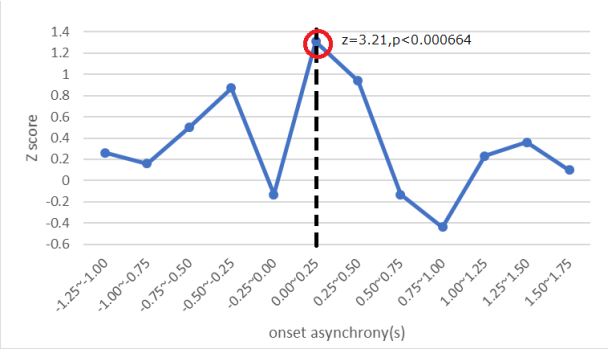

図 5 再現動画 2(等間隔)

3.1.3 考察

実験 1 では被験者に順に 3 本の動画を見せた．最初の 1 本目に本物の演説動画を見せるのは先行研究 [7] の結論の確認のためである．そして，2，3 本目の再現動画がエージェントによる瞬きの引き込み現象が存在するかどうかを調べる実験である．その結果，1 本目の本物の演説動画を見る時に被験者たちの瞬きのタイミングを示すそれぞれのビンの Z スコアのグラフは，3 本のうち一番高いピーク値

© 2018 Information Processing Society of Japan

(図 3 の赤い丸) を持ち，グラフの形状も先行研究と同様の結果となった．ただし，ピークのビンが先行研究と 1 個ずれている．これは，瞬きの計測手段 (先行研究では垂直 EOG 法を利用) による誤差だと考えられる．

2 本目の本物と同じタイミングで瞬きをする動画を見るときの被験者たちの Z スコアグラフにもピークが存在するが，タイミングが再現動画の瞬きが発生した後ではなく，その前の-0.5s〜-0.25 のタイミングとなった (図 4 の赤い丸に対するビン)．これは，最初に一度同じ内容の演説 (本物動画) を見たことで話の切り目や演説者の瞬きを一部覚えたので，もう一度見た際に話の切れ目を予測して少し早いタイミングで瞬きをしたと考えられる．

3 本目の動画の Z スコアグラフにも唯一のピークが存在する (図 5 の赤い丸)．これは，我々の予想と異なり，3 本目の動画を見る時もエージェントによる瞬きの引き込み現象が起きていることを示している．先行研究では，音声と画像を一方だけ見せたときには瞬きの引き込み現象が確認されなかったが，本稿のように，音声は同じで瞬きだけ本物と異なる画像を見せる実験は行っていない．本実験の結果では，このような組み合わせでも瞬きの引き込み現象が観察されたが，偶然に起きた可能性も否定できない．その理由として，まず等間隔の瞬きをしたとはいえ，本物と同じタイミングで発生した瞬きが何回か存在すると考えられる．そのため，果たして「瞬きの引き込み」に貢献したのはその同じタイミングの瞬きか，それとも他の瞬きかを調べることができない．また，アンケートの感想によると 3 本目の動画を見るときに「眠くなった」などの意見もあり，途中で休憩を入れるなど，実験方法をより注意深く検討する必要があったと考えられる．

3.2 実験 2(エージェントが話を聞く側)

3.2.1 実験方法

まず，実験用プログラムとして前節で述べたエージェントを使って，ウェブカメラに映る被験者と同じタイミングで瞬きをするエージェントを作成した．また，比較のため，一定間隔 (本実験では 3 秒) で瞬きをするエージェントも作成した．この実験では，エージェントは聞き役であるが，被験者の話しやすさのため，挨拶や実験の簡単な説明，最初の被験者の発話を促すセリフなどは話すようにした．

次に，14 人の被験者にエージェントの画面に向かって，2 つの話題について，一分以内で話をしてもらった．話の内容はそれぞれ，事前に配ったレシピに基づき「エージェントにおいしいカレーの作り方を教える」ことと「大学の正門から実験を行う場所までの道案内」である．

エージェントの瞬きは交代で「被験者と同じタイミングで瞬きをする」と「等間隔で瞬きをする」ように設定した．例えば，1 番目の被験者がレシピの説明をするときにエージェントは被験者と同じタイミングで瞬きをし，道案内の話をするときにエージェントが等間隔で瞬きをした場合，2 番目の被験者がレシピをするときはエージェントは等間隔で，道案内をするときは被験者と同じタイミングで瞬きをする．

実験終了後，被験者にエージェントに関する以下の質問に答えてもらった．

(1) より自然と思う動画はどちらですか？
(2) より聞いているように見える動画はどちらですか？
(3) より知的に見えると思う動画はどちらですか？
(4) 瞬きのタイミングの違いに気づきましたか?(y/n)

(1)-(3) の質問には，どちらの話題の時の方がよかったかを 6 段階で答えてもらった (同じタイミングの瞬きの方が良いを 6 として，等間隔の瞬きの方が良いを 1 とした)．

3.2.2 実験結果

「被験者と同じタイミングで瞬きをする」場合と「等間隔で瞬きをする」場合で分けて集計した各質問の得点を図 6 に示す．また，エージェントへのイメージに関する (1)-(3) の質問の得点を話の内容別に集計した結果を図 7 に示す．なお，瞬きのタイミングの違いに気づいたか?という質問には，全員が気づかなかったという回答であった．

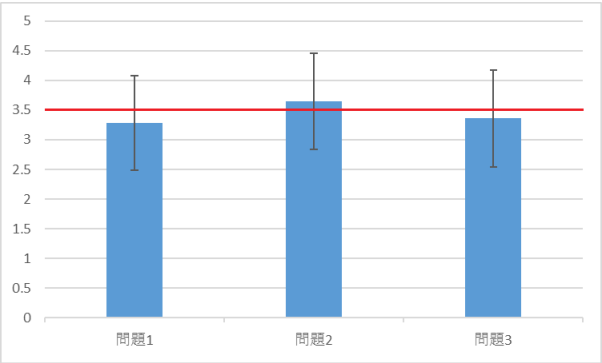

図 6 実験 2 アンケート結果

図 7 実験 2 アンケート結果 (内容別)

3.2.3 考察

実験 2 では，実験 1 を通じて分かったエージェントによる瞬きの引き込み現象を利用して，エージェントが被験者の話に対して異なる瞬きを見せることで被験者に与えるイメージが変わるかどうかを調べるのが目的である．アンケートの結果 (図 6) に対して符号検定を行ったところ，予

想と異なり両者に差は見られなかった．その原因は実験の設計にあると考えられる．被験者に 2 回同じタスクを与えると，2 回目の結果が 1 回目の影響を受ける可能性があると考え，２つの異なる話題を用意したが，話題ごとに分けて集計した結果 (図 7) を見ると，道案内よりレシピ説明が 3 問ともに高い得点を得ていることが分かる．つまり，瞬きよりも，話の内容が得点に影響を与えていることになる．これは，被験者はレシピの説明をする時には手元のレシピを見ながら話すのに対して，道案内をする時には始終エージェントの画面を見て話していたので，瞬き以外のエージェントの不自然な部分が目立ってしまったと考えられる．

4. おわりに

本稿では，よりスムーズにユーザと会話するため，瞬きの引き込み現象を利用するエージェントを提案した．録画ビデオやカメラ映像に映る人間の瞬きを再現するエージェントを作成し，これを用いて「エージェントが話をする時」と「エージェントが話を聞く時」の二つの場合において，人間のような瞬きをエージェントにさせることによるユーザの反応の違いを実験を通して調べた．その結果，前者の実験では，人間の演説者と同じくエージェントが話をするときも瞬きの引き込み現象が起きることが確認された．後者の実験では，実験の設計の問題で，適切な評価を行えなかった．

今後は，まず実験 1 では，目の疲れの対策などを検討し，「本物の演説者が瞬きをしないタイミングでのみ瞬きをするエージェントとの比較」などの追加実験を行いたい．さらに，実験 2 では，より気軽に話せる話題を選定し，被験者に 2 回とも紙を見ないで話してもらうような設定で再度実験を行いたいと考えている．

本研究の一部は JSPS 科研費 JP17K18850 の助成による．

参考文献

[1] 内閣府 平成 29 年版高齢社会白書（全体版） http://www8.cao.go.jp/kourei/whitepaper/w-2017/html/zenbun/s1_1_1.html
[2] 公益財団法人介護労働安定センター, “平成 28 年度「介護労働実態調査」の結果,” 平成 29 年 8 月 4 日.
[3] 国立研究開発法人 産業技術総合研究所, “介護分野におけるコミュニケーションロボットの活用に関する大規模実証試験報告書,” 平成 29 年 5 月 31 日.
[4] 武川直樹, “社会科学のアプローチに基づくコミュニケーションロボット・擬人化エージェントの設計に向けて：人間観察によってデザインされたロボットは「不気味の谷」を渡れるか?,” 電子情報通信学会誌, vol.93, no.12, 1027-1033, 2010.
[5] 堀江直人，小林貴訓，久野義徳, “認知症介護支援のためのキーワード検出に基づく対話システム,” 第 96 回パターン計測シンポジウム, 2015.
[6] 山本祐介, 福田悠人, 小林貴訓, 久野義徳, “DNN に基づく感情推定手法の対話ロボットへの応用,” 第 100 回パターン計測部会研究会, 2017.
[7] 中野珠実, “瞬きにより明らかになったデフォルト・モード・ネットワークの新たな役割,” 生理心理学と精神心理学, vol.31, no.1, pp.19-26, 2013.
[8] T. Nakano & S. Kitazawa (2010). Eyeblink entrainment at breakpoints of speech. Experimental Brain Research, 205, 577-581.
[9] 李晃伸, 大浦圭一郎, 徳田恵一, “魅力ある音声インタラクションシステムを構築するためのオープンソースツールキット MMDAgent,” 電子情報通信学会技術研究報告, vol.111, no.364, pp.159-164, 2011.
[10] 河原達也, 李晃伸, “連続音声認識ソフトウェア Julius,” 人工知能学会誌, vol.20, pp.41-49, 2005.
[11] A. Asthana, S. Zafeiriou, S. Cheng and M. Pantic, “Incremental Face Alignment in the Wild,” in proc. CVPR2014, pp.1-8, 2014.
[12] 実験用モデルの配布ページ: https://bowlroll.net/file/6431

© 2018 Information Processing Society of Japan

正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

訂正箇所	誤	正
4 ページ 図 7		

ⓒ2018 Information Processing Society of Japan