H-018

ロボットとの対話における人間の「退屈」状態の解析
Analysis of human behavior related to boredom in multiparty conversation including robot agents

芝﨑 泰弘 † Yasuhiro Shibasaki
船越 孝太郎 ‡ Kotaro Funakoshi
篠田 浩一 † Koichi Shinoda

1. はじめに

情報システムの活用される場面はユーザが 1 名のみ存在する場合および複数存在する場合が考えられ，教育システムなど高度に知的なシステムの場合は，特に後者で用いられる場合が多い．ただ，よく知られているように 3 者以上のインタラクションを扱う場合，複数の聞き手が存在するため話者交替が複雑化し，またジェスチャーなどの非言語情報がより重要な役割を果たす．よって，2 者対話の状況をそのまま適用することは困難である [1].

多人数環境下における情報システムとユーザとのインタラクションを実現するためには，ユーザと音声を通じやりとりできるのみならず，非言語情報を通して対話できることが望ましい [2]．また，画像データなどの客観的情報を処理するだけでなく，ユーザの内部心理状態などのより主観的情報を解釈することが求められる．中でも，情報システムを利用しているユーザにおける，遂行中タスクに対する満足状態と退屈状態を機械的に理解することによって，よりユーザとの親和性の高いシステムを実現できると考えられる．

我々は，多人数環境下における情報システムユーザの退屈状態検出に着目する．従来研究は，1 対 1 の対話を通じたユーザインタラクションしか考慮できていない点や自発的動作を十分に検出できていない点，また退屈状態を直接研究対象としていない点など，課題も多い．また，ユーザの退屈状態などを推定するために，従来手法では参加者へのアンケート調査などの主観的指標を用いている [3]．情報システムがユーザの退屈状態を機械的に推定するためには，上記に加えて，Kinect[4] などのより情報量の多い機器に基づいた，自発的動作の有無などのより客観的指標を判断材料とすることが効果的である．

そこで，本研究は，ロボットを含む多人数環境下における人間の退屈状態の解析を目的とする．我々は男女 3 名の参加者が対話ゲームを進める場面を収録し，参加者の退屈状態および自発的動作の 2 つの観点に基づくタグ付けを実施した．その後タグ付けデータの定量的解析と自発的動作の分類についての認識実験を行った．

2. 画像情報を用いた内部心理状態推定と動作解析

2.1. 内部心理状態一般の推定

Castellano らは，ユーザの内部心理状態を把握することで長時間の連携が可能な情報システムを構築するため，対話的ゲーム参加者のゲームへの集中度合いを機械的に検出する研究を行っている [5]．この研究では，電子チェスボード上でロボットと 1 対 1 の環境下でチェスを進める子供の注視状況や，ゲームの進捗状況を分析している．

また，Gunes らは，マルチモーダル情報を用いる感情認識システムを構築するため，被写体の非言語的動作と心理状態との間の関連性を分析する研究を行っている [6]．被写体は指示に従って身体動作を行っており，この研究では自発的動作を扱っていない．

そして，岡村らは，製造業のライン生産現場におけるヒューマンエラーを軽減する情報システムを実現するため，作業者の各身体部位の動作特徴と集中状態との間の関連性を分析する研究を行っている [7]．ここでは，単調作業中の作業者の頭部と肘部また掌部の 3 カ所の動きのみを検出対象としている．

2.2. 退屈状態の推定

Jacobs らは，対話相手の満足状態また退屈状態を検出し文脈解釈可能なロボットシステムを構築するために，ビデオ映像視聴者の退屈状態を機械的に推定する研究を行っている [8]．この研究は，被写体頭部の位置情報のみを解析対象としている．

また，Nataliia らは，参加者の非言語的情報を含むマルチモーダル情報を用い退屈状態を解析するより知的な情報システムを実現するために，参加者の静止状態と退屈また非退屈状態との間の関連性を分析する研究を行っている [9]．解析対象の対話的ゲーム中に，参加者が他者の指示に従って受動的に動作を行う場面が含まれている．

2.3.Kinect を用いた多人数環境下での動作の解析

Wang らは，情報システムへの動作による非接触的指示入力を実現するために，ハンドジェスチャーを実時間的に分類する研究を行っている [10]．著者らは，Kinect から得られる深度情報を活用することによって，10 通りのハンドジェスチャーを 75%以上の精度で分類することが可能であると述べている．

一方，石川らは，多人数環境下で動作し参加者間の対話状況を認識できる情報システムを実現するために，多人数環境下における参加者間の会話の受話者を推定する研究を行っている [11]．ここでは，参加者頭部の位置情報のみを推定に用いている．

3. 退屈状態解析にむけたデータベースの構築

我々は Aldebaran 社製の人型ロボット「NAO」[12]1 台と成人 3 名の参加者とが，会話やジェスチャーなどの対話を通じてゲームを進める場面をマイクロソフト社製の RGB-D カメラである Kinect とマイクロフォンを用いて収録した [11]．収録にあたっては Wizard of OZ 方式を採用しロボットとして NAO1 台を配備した．収

†東京工業大学 大学院情報理工学研究科
‡(株) ホンダ・リサーチ・インスティチュート・ジャパン

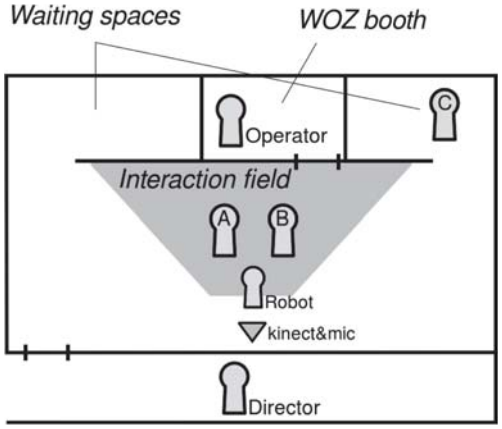

図 1. 多人数環境下における人間の退屈状態解析のためのデータ収録環境見取り図 [11]

録環境の見取り図を図 1 に示す.

収録場 (Interation field) には NAO が待機しており，3 名の参加者は各々収録場に出入りしつつ NAO および他の参加者と対話を進めながら NAO とのゲームを約 25 分間行った.

監督者 (Director) は，収録場に存在する参加者数や各参加者のゲーム参加累積時間を考慮しながら各参加者にゲームへの参加および離脱を指示した．また，ゲームの進行状況を考慮しながら，適宜ゲームのヒントを提示するよう操作者 (Operator) に指示を出した．操作者は監督者の指示に従い用意された API を通じて NAO を操作した.

なお，3 名の参加者は NAO が Wizard of OZ 方式によって遠隔操作されていること，またデータ収録環境に操作者が存在していることを事前に知らされていない.

ここで，対話的ゲームは「20 のとびら」と呼ばれるもので，出題者が想定しているコンセプト (例：ペンギン，バナナなど) に対して参加者が YES/NO 形式の質問を繰り返しそれに対する出題者の正解また不正解の応答をふまえて，それが何かを当てるゲームである.

上記のゲームを 25 分間を 1 セッションとし 30 セッション，合計約 15 時間分収録した．参加者は，各セッション毎に別々の主に 20 歳代から 60 歳代の日本人の男女 3 名であった.

その後，収録済 RGB 動画を参照することでゲーム中の参加者が見せる自発的動作をリストアップした．その結果を表 1 に示す．なお，退屈動作の分類については 4 節で述べる.

リストアップした自発的動作のうち，歓声と共に手を挙げている場合など明らかに満足している場面での動作を除いたものを退屈動作と定義した．参加者が，「腰に手を当てたり腕組みしたりしている」に該当する退屈動作を見せているシーンの一例を図 2 に示す.

次に，収録された RGB 動画に対して，タグ付け用ツール ELAN[13] を用いて 5 秒間の固定長区間 (以下，サンプルと呼称) 毎に 2 名による人手によるタグ付けを実施した．我々は下記の通り，退屈状態と自発的動作について独立した 2 つのタグ付け観点を設定する方針を採用した.

表 1. 自発的動作のリストアップ (3 節) および退屈動作の分類 (4 節)

自発的動作	胴体	足	手	頭
胴体を前後左右に揺すっている	○			
屈めていた体を伸ばしている	○			
胴体を左右に回転させている	○			
後傾に座っている	○			
体の重心を左右の足に載せ替えている		○		
片足重心の姿勢である		○		
足を組んだり組み替えたりしている		○		
足踏みしている		○		
行ったり来たり左右前後をうろうろする		○		
しゃがんでいて手を膝にあてている		○	○	
前屈し両手を両膝にあてている		○	○	
髪や服を直している			○	
顔や耳を触っている			○	
頬杖をついている			○	
指先をいじっている			○	
腰に手を当てたり腕組みしたりしている			○	
床に手をついている			○	
手をすり合わせる			○	
NAO や参加者以外 (床や天井) を見ている				○
ゲームに参加していない人間を見ている				○
目を閉じている				○
首を傾げている				○
あくびをしている				○
歓声と共に手を挙げている	退屈動作でない			
他の参加者の体を触っている	退屈動作でない			

- 退屈状態：ゲーム進行中の動画を見たタグ付け者から主観的にみて，参加者がゲームを楽しんでいる (Bored) か，退屈している (Not bored) か，もしくはどちらとも言えない不明な状態 (Cannot say) か
- 自発的動作：参加者が退屈動作を行っている (Gesture 有) か，行っていない (Gesture 無) か

上記方針に従ってタグ付けを行った後，2 名のタグ付け者間でのタグ付け結果の統合を行った．退屈状態に関しては，2 名のタグ付け者が共に Bored とタグ付けした場合のみ Bored，また共に Not bored とタグ付けした場合のみ Not bored とし，他の場合は Maybe bored

図 2. 対話的ゲーム参加者 (左女性) が退屈動作を見せている状況図

とした．自発的動作に関しては，2 名のタグ付け者が共に Gesture 有とタグ付けした場合のみ Gesture 有，他の場合は Gesture 無とした．

タグ付け結果の統合の結果，データベースに含まれる全 6,105 サンプルの内，退屈状態が Bored に該当するサンプルは 438 個，Maybe bored に該当するサンプルは 1,391 個，Not bored に該当するサンプルは 4,276 個存在した．また，自発的動作が Gesture 有に該当するサンプルは 4,808 個，Gesture 無に該当するサンプルは 1,297 個存在した．

4. リストアップした退屈動作の分類

前節でリストアップした退屈動作の分布についてより詳しく調べるため，人体を構成する 4 つの部位であるところの，胴体と足，手，頭に着目しどの人体部位に特徴が見られるかの観点に基づいて，構築済データベースにおいて観察された退屈動作の分類を行った．結果を表 1 に示す．

胴体と足，手，頭のいずれの箇所においてもそれらの部位に特徴を持つ退屈動作が，データベース内に存在することがわかった．以下，胴体と足，手，頭に特徴的部位を持つ退屈動作を各々胴部動作，足部動作，手部動作，頭部動作と呼称する．なお，「前屈し，両手を両膝にあてている」場合など，特徴的部位が複数にわたる場合を考慮しマルチラベルによる分類方式を採用した．

その後，上記の特徴的部位に基づいた分類をふまえ，前節までで構築したデータベースに対して改めてタグ付けを行った．結果，全 6,105 サンプルの内，胴部動作がみられたサンプルは 20 個，足部動作がみられたサンプルは 3,174 個，手部動作がみられたサンプルは 3,004 個，頭部動作がみられたサンプルは 1,073 個存在した．

5. 退屈動作についての解析

退屈動作の 4 分類毎に，各退屈状態下における出現率を比較した結果を図 3 に示す．

全 3 退屈状態における全 4 退屈動作分類のうち，退屈状態が Maybe bored の状況下で足部動作に該当する退屈動作を行う頻度が最も高い (53%) ことがわかった．退屈状態が Maybe bored の状況下で胴部動作に該当する退屈動作を行う頻度が最も低い (0.2%) ことがわかった．

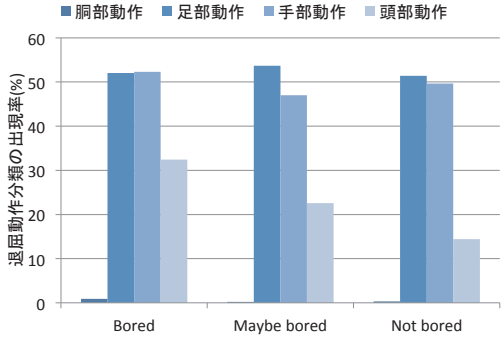

図 3. 各退屈状態における各退屈動作分類の出現率分布図

また，頭部動作に着目した場合，退屈状態が Bored に該当するサンプルにおいて最も出現率が高く (32%)，また退屈状態が Not bored に該当するサンプルにおいて最も出現率が低い (14%) ことがわかった．よって，頭部動作の有無を退屈状態分類の判定基準とできる可能性がある．

一方で，胴部動作および足部動作，手部動作に着目した場合，各退屈動作は，3 通りの退屈状態いずれの下にあってもほぼ等しい確率で出現しうることがわかった．これらの退屈動作は退屈状態の判定基準として用いることができないと想定される．

6. 退屈状態の推定に関する予備実験

参加者の退屈状態を機械的に推定するためには，認識器を用いた実験によって，退屈動作の有無の認識可能性を検証する必要がある．そのための予備実験として，特徴的部位に着目した場合の退屈動作分類についての認識可能性を検証するために，前節で構築したデータベースのうち，4 セッション分の約 100 分のタグ付き動画データを用いて退屈動作の分類についての認識実験を行った．また，対象 4 セッションの選定にあたっては，認識対象のラベルが可能なかぎり均一に分布するように配慮した．

認識対象サンプルおよび分類方式としては，足部動作と手部動作，頭部動作の 3 通りとした．なお，胴部動作についてはサンプル数が限られていたため実験対象から除外した．認識実験にあたっては隠れマルコフモデル (HMM) を認識器として用い，Kinect v1 によって収録された参加者の全身骨格 20 カ所についての 3 次元時系列データを特徴量とした．骨格データの取得頻度は 30fps であった．

また，認識対象データ数は 737 サンプルであった．実験にあたってはラベル分布の偏りを除去した上で 5-fold cross validation を実行した．HMM 状態数は 50，Gauss 混合数は 4 とした．

退屈動作の分類についての認識実験の結果を表 2 に示す．

認識対象データ中に 290 サンプル存在した足部動作に関しては，90.3%に該当する 262 サンプルで分類の認識に成功した．また，241 サンプル存在した手部動

表 2. 退屈動作の分類についての認識実験結果 (認識誤り率：12.3%)

正解/認識結果	足部動作	手部動作	頭部動作	合計
足部動作	262	12	16	290
手部動作	7	219	15	241
頭部動作	28	12	166	206
合計	297	243	197	737

作および，206 サンプル存在した頭部動作に関しては，各々 90.9%に該当する 219 サンプル，80.6%に該当する 166 サンプルにおいて認識に成功した．

今回の退屈動作分類についての認識実験において認識誤りが最も多かったケースは，頭部動作を足部動作と誤認識するもので全 737 サンプルの中 28 サンプルでこの傾向が見られた．逆に，認識誤りが最も少なかったケースは，手部動作を足部動作と誤認識するもので 7 サンプルがこのケースに該当した．そして，足部動作と頭部動作との間の分類は，足部動作と手部動作の分類および手部動作と頭部動作の分類に比べてより困難であることがわかった．

さらに，認識失敗サンプルを定性的に分析した結果，「NAO や参加者以外 (床や天井) を見ている」場面 (以下，場面 A と呼称) を「片足重心の姿勢である 」場面 (以下，場面 B と呼称) と取り違えているケースが多く見られた．逆に場面 B を場面 A と誤認識するケースも観察された．本実験条件の場合，場面 A と場面 B を分類することは困難であると考えられる．

7. おわりに

多人数環境下で対話的ゲームを進める参加者が退屈時に見せる自発的動作の分類を行った．また，退屈状態と自発的動作との関係を分析することで，頭部動作の有無を退屈状態の判定基準とできる可能性があることがわかった．

さらに，退屈動作の分類についての認識実験を行った結果，認識誤り率は 12.3%となり退屈動作を機械的に分類できる可能性があることがわかった．

今後は，隠れマルコフモデルやサポートベクターマシン (SVM)，隠れ条件付き確率場 (HCRF) などの認識器を用いて参加者の退屈状態を機械的に推定する手法の実装と検証を進める予定である．

参考文献

[1] 坊農真弓, 高梨克也, 人工知能学会. 多人数インタラクションの分析手法. オーム社, 2009.

[2] Julie A Jacko. *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. CRC press, 2012.

[3] Thomas Goetz, Anne C Frenzel, Nathan C Hall, Ulrike E Nett, Reinhard Pekrun, and Anastasiya A Lipnevich. Types of boredom: An experience sampling approach. *Motivation and Emotion*, Vol. 38, No. 3, pp. 401–419, 2014.

[4] Kinect. `https://www.microsoft.com/en-us/kinectforwindows/`.

[5] Ginevra Castellano, André Pereira, Iolanda Leite, Ana Paiva, and Peter W McOwan. Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pp. 119–126. ACM, 2009.

[6] Hatice Gunes and Massimo Piccardi. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Vol. 1, pp. 1148–1153. IEEE, 2006.

[7] 岡村瞬, 梶原祐輔, 原田史子, 島川博光. J-031 kinect を用いた単調作業における特徴部位の推定 (j 分野: ヒューマンコミュニケーション & インタラクション, 一般論文). 情報科学技術フォーラム講演論文集, Vol. 12, No. 3, pp. 447–448, 2013.

[8] Allison M Jacobs, Benjamin Fransen, J Malcolm McCurry, Frederick WP Heckel, Alan R Wagner, and J Gregory Trafton. A preliminary system for recognizing boredom. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 299–300. ACM, 2009.

[9] Nataliia Biriukova, Kotaro Funakoshi, and Koichi Shinoda. Collection and analysis of multiparty interaction data for automatic boredom recognition. In *Proc. The 28th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI) 2014*, 2014.

[10] Chong Wang, Zhong Liu, and Shing-Chow Chan. Superpixel-based hand gesture recognition with kinect depth camera. *Multimedia, IEEE Transactions on*, Vol. 17, No. 1, pp. 29–39, 2015.

[11] 石川真也, 船越孝太郎, 篠田浩一, 中野幹生. 多人数対話ロボットの実現にむけたマルチモーダル対話データの収集と分析. 人工知能学会第 27 回全国大会論文集 1K3-OS-17a-5, 2013.

[12] Nao aldebaran robotics. `http://www.aldebaran-robotics.com/`.

[13] Elan. `https://tla.mpi.nl/tools/tla-tools/elan/`.