

親和的インタラクションにより引き出す表情の観察によるユーザ状態把握

Interaction-based user monitoring by the observation of nonverbal expression

今井 岳*¹ 金岡 利知*¹ 立田 隼人*¹ 渡辺 一郎*¹ 安川 裕介*¹
 Takashi Imai Toshikazu Kanaoka Hayato Tateda Ichiro Watanabe Yusuke Yasukawa

*¹ 株式会社富士通研究所
 Fujitsu Laboratories Ltd.

We have been developing the bear cub social robot, with the appearance of a stuffed bear cub and a child-like personality, which aims to be a friendly companion in daily life. In this paper, we propose an interaction-based user monitoring method by the face-to-face observation of user's nonverbal expression and show some experimental results.

1. はじめに

近年、ロボットは、実世界の情報を取り込み、その情報を処理することで、新たなサービスが提供できるインタラクションデバイスとしても期待されており、人とのインタラクションを前提としたサービスを目的とした多くのロボットが研究されている。

筆者らは、従来 IT があまり浸透していない介護、教育、一般家庭などでの日常生活シーンに溶け込み、生活支援などのサービスを提供する「人に優しい端末」の実現を目指して、「子ぐま型ソーシャルロボット」(図 1)の研究開発を行っている[富士通 2010]。従来の PC やスマートフォンのような、単に情報伝達をするだけの無機質な端末ではなく、ユーザの感情に訴えかけユーザとの愛着・絆を形成できる新しい端末の実現を目指している。

そのために、端末自体をあたかも意思を持っているかのように自律的・生物的に動作させながら、ユーザとのアイコンタクトや同調動作などの社会的振る舞いを可能とする親和的インタラクション技術を開発してきた[渡辺 2007, 今井 2011]。これまでの研究で、高齢者との触れ合いによる生活活性化など、親和的インタラクション技術の有用性が示されてきており [渡辺 2007, 山岡 2010, 雨宮 2011], 「子ぐま型ソーシャルロボット」がユーザとの間に親和的関係を築くことができる可能性を示唆している。

このような「人に優しい端末」を実現するために、利用者の心的状態を理解し、それにより適切な態度や振る舞いを行えることが重要であると筆者らは考えている。「子ぐま型ソーシャルロボット」は、生物的、有意識的、社会的な振る舞いの特徴により、人とのインタラクションで表出するような表情や反応をユーザから引き出すことが期待できる。その為、対人コミュニケーションで表出される非言語の振る舞いや反応を観察する事により、ユーザの様々な心的状態を推定できると考えている。



図 1 子ぐま型ソーシャルロボット

しかしながら、人とロボットの自然なインタラクションのなかでは、不確かさが多く、いつどのようにユーザの表情や反応をセンシングするのが良いのかはわかっていない。本稿では、人とロボットの自然なインタラクションのなかでユーザの表情や振る舞いを効率的にセンシングする方法を提案し、応用例の一つである、高齢者見守りの場面における本センシング手法の有用性について検証する。

2. 子ぐま型ソーシャルロボットによるユーザ観察

2.1 従来の感情推定技術の課題

表情から人の感情を推定する技術の研究はこれまでも盛んに行われている[Zeng 2009]が、多くの研究は意図的に作られた表情の画像や動画データに対してのクラスタリングを行うものであり、これらの研究ではカメラが表情や顔を安定的に捉えている事が前提となっているものが殆どである。

実際に人とロボットが触れ合う場面では、人とロボットの位置関係や、シチュエーションも多様であり、表情を安定的に観測して識別する事は難しい。我々は実際のインタラクション場面におけるユーザの心的状態の観測の課題を以下のように整理した。

1. センシングの不確かさ
2. 心的状態の表出に関する不確かさ
3. 観測時のユーザ状況の多様性

センシングの不確かさとは、人とカメラの距離や角度、ユーザの振る舞いや環境の影響によってセンシング結果の信頼性が下がってしまう問題である。心的状態の表出に関する不確かさとは、例えば喜びなど、特定の感情を持っていたとしても、その間常に喜びを表す表情や振る舞いを表出し続けているわけではなく、何かの折に喜びの感情が垣間見える振る舞いを表出すると考えられる。すなわち、ユーザの心的状態を表す表情や振る舞いが表出されるタイミングが不確かである問題である。観測時のユーザ状況の多様性については、表出される振る舞いは心的状態の他に、ユーザが置かれている状況に大きく影響を受けると考えられるが、多くの場合はユーザ状況を厳密に把握する事は難しく、得られたセンシング結果を単純に比較する事はできないという問題である。

つまり、実際のインタラクションでのユーザセンシングは、様々なユーザ状況で、いつ表出されるかわからない情報を不確実なセンサでセンシングしている事になり、この情報のみにより何らかの意味づけを行う事は非常に困難である。

これらの課題に対し、インタラクションの状況に基づいてユーザ観察を実行する事により、安定したユーザ観測を実現する方

法を提案する。インタラクションでは、人とロボットが相互に影響を与え合っており、インタラクションの状況を判断する事により、情報を取得しやすいタイミングの把握や、観測時のユーザ状況の統制を実現する事ができると考えた。次節で提案するユーザ観察手法の詳細を説明する。

2.2 自然なインタラクションにおけるユーザ観察手法

自然なインタラクションにおけるユーザ観測のしやすさについて、以下のような仮説を立てた。

1. アイコンタクトにより安定して正面から表情を観測できる
2. ユーザがロボットに働きかけたタイミングで表情や振る舞いが多く表出される
3. ロボットがユーザに働きかけたタイミングで表情や振る舞いが多く表出される(引き出すことができる)

つまり、インタラクション中にどちらかが何らかの働きかけを開始するタイミングでアイコンタクトを行い、その間の表情や振る舞いを観察する事により、安定して観察ができると考えた。以下に、本手法によるユーザ観察の詳細と実装方法について述べる。

(1) 非言語の振る舞い検出

子ぐま型ソーシャルロボットでは、鼻のカメラを使って、ユーザの表情や仕草を観察することができる。また、親和的振る舞いの特徴により、ユーザが人とコミュニケーションと同様な非言語の振る舞い表出する事が期待できる。そこで、カメラ情報から表 1 に示すユーザの振る舞いを検出する機能を実装した。

表 1 検出する非言語動作

視線	ロボットのカメラから見た顔の向きで判断し、視線を向ける、顔をそむけるなどの動作や、ロボットの視線と一致するアイコンタクトを検出
位置関係	ロボットから見たユーザの相対位置や、ユーザの顔の大きさによる大まかな距離および、接近、後退などの移動を検出
身振り	肌色とオプティカルフローによる手の動作の検出、顔特徴点の動きによる頷きの検出、検出した顔の傾き変化による首の傾げの検出
表情	顔特徴点による笑顔度および口の開閉の検出
接触行動	接触センサによるロボットへの接触の検出

(2) アイコンタクト時の振る舞いによるユーザ観察

センシングの不確実性と観測時のユーザ状況の多様性を軽減するため、アイコンタクト時にユーザの振る舞いを観察するようにした。アイコンタクトにより、人に対してロボットへの注目を促し、比較的安定して表情を正面から観測できる状況を作り出すと同時に、ロボットとのインタラクション中で、お互いを見ている場面という状況の限定を行うことになり、観測状況の緩い統制を実現している。

一回のアイコンタクト時に一定時間ユーザを観測し、これを一回のユーザ観察の単位とし、この期間に検出した非言語の振る舞いの頻度などの統計量をユーザ観察結果とする。また、アイコンタクトの瞬間前後では多くの振る舞いが検出されると予想される為、アイコンタクトの瞬間から少し遡ってセンシング期間をとる。例えば、アイコンタクトの 1 秒前から 4 秒後までの 5 秒間の観測で振る舞いを評価する。

(3) ユーザからの働きかけをトリガとしたセンシング

心的状態の表出に関する不確実性を考慮すると、表出されるユーザの顕著な振る舞いをしてできるだけ逃さず取得するため、センシングを行うタイミングが重要である。ユーザがロボットに何らかの働きかけを行う際に、心的状態の表出が行われる可能性が高

いと考え、ユーザからの働きかけをトリガにセンシングを開始する機能を開発した。

ユーザからの働きかけの検出は、非言語の振る舞いの検出を利用して判断している。ユーザからの働きかけがあったと判断しているのは以下の 6 つのケースである。

1. ユーザの接近を検出した場合
2. ユーザがロボットに視線を向ける動作を検出した場合
3. ユーザの笑顔度が急上昇した場合
4. ユーザの手の動作の頻度が急上昇した場合
5. ユーザの頷きの頻度が急上昇した場合
6. ユーザのロボットへの接触頻度が急上昇した場合

各センシング値の急上昇は、最新のセンシング値とそれより前の一定時間のセンシング値の平均値の差が閾値を上回る事により、検出している。例えば一定時間を 4 秒として、過去 4 秒間のセンシング値の平均を大きく上回る値を検出した場合に急上昇したとしている。ここでの各センシング値は、ノイズによる誤検出を避ける為、移動平均によるフィルタ後の値を用いている。

(4) ロボットからの働きかけをトリガとしたアクティブセンシング

心的状態の表出を引き出し、また、観測時のユーザ状況を統制する、アクティブセンシングの機能を開発した。これはロボットからの働きかけをトリガにセンシングを開始する方法である。具体的には、ロボットの動作により意図的にユーザに刺激を与えてからその反応をセンシングする。これにより、ユーザ反応を意図的に引き出し、表情や振る舞いを観測する事ができると考えた。また、与えた刺激動作によって、観測時のユーザ状況を統制する事もできる。

刺激として与える動作は、次のような情報と共に複数定義する。

1. 動作情報
2. 動作条件
3. アイコンタクトのタイミング

動作情報は、実際にその動作を動かすためのモータコマンドなどの情報である。動作条件は、この動作が発動してよい条件であり、ロボットの現在の動作と矛盾する動作を実行しない、また、ユーザの状態により動作を変えるために定義する。アイコンタクトのタイミングは、動作を開始してからどの程度の時間が経過してからアイコンタクトをするかを定義する。例えば、声をかけるような動作であれば、動作の開始と同時にアイコンタクトを試みるし、くしゃみやあくびのような生理的動作であれば、刺激動作が終わった後にアイコンタクトを試みるのが適している。刺激動作は動作条件に一致するものからランダムに選択されるようにした。

刺激動作として、挨拶、あくびやくしゃみ等の生理的動作、驚かせる動作、構って欲しい動作、つまらなそうな動作、派手に喜ぶ動作、少し怒った動作をそれぞれ数種類用意した。

(5) ユーザ観察の実装

これらの機能を使って実際にユーザ観察を実行する手続きについて述べる。

まず、ユーザ観察の開始のタイミングを以下のように決定する。

1. ユーザとのインタラクションが新規に開始された場合
2. 既にインタラクション中であり、ユーザからの働きかけを検出した場合
3. 既にインタラクション中であり、一定時間センシングを実行しなかった場合(アクティブセンシング)

インタラクションの開始は、ロボットの動作モードと関係する。本システムは、ロボットが人とは無関係に振る舞うか、人がいる事を意識した振る舞いをするかを切り替えて動作しており、人を意識した振る舞いに切り替えたタイミングでインタラクションを開始したとみなす。この切り替えはアプリケーション依存であるが、殆どの

場合、人を見つけたらインタラクションを開始するように動作させており、またその後アイコンタクトが成立しない場合や、あまりユーザから反応がない場合にはインタラクションを終了し、人とは無関係な振る舞いに切り替える。

ユーザ観察が開始されると次のように動作する。

1. アクティブセンシングの場合刺激動作の提示を開始し、アイコンタクトのタイミングまで待つ
2. 対象のユーザの顔追従を開始
3. 対象ユーザの顔をカメラ中央で捉え、且つ対象ユーザの視線(顔の向き)が正面(ロボットの方向)を向いていた場合アイコンタクトの成立と判断
4. センシング期間が完了するまで顔追従を継続
5. 顔追従を終了し、センシング期間の非言語の振る舞い検出結果を集計する

顔追従を行う際には、ユーザに注目を促すため、身体の動きを一瞬止めるなどの視線強調動作[金岡 2012]を行う。

アプリケーションの動作の都合により、アクティブセンシングの刺激動作の実行を禁止している場合がある。その場合には刺激動作は実行せず、アイコンタクトとセンシングのみを行う。この場合をアクティブセンシングと区別するため、定時センシングという事にする。

3. 高齢者見守りへの適用と評価

提案した観察手法の有用性を示すため、「子ぐま型ソーシャルロボット」の応用例の一つである、高齢者見守りシーンに適用し、自然なインタラクションから意味のあるセンシングが可能であるかを評価する。

3.1 実験条件

認知症を発症している 90 歳女性の 1 名の被験者に対し 3 日間実験を行った。13:30~16:30 の約 3 時間ロボットと自由に触れ合ってもらい、その間に本手法により、ユーザ観察結果と、各ユーザ観察におけるユーザの振る舞いの動画を収集した。実験は被験者が主に生活する被験者宅のリビングで行い、同じ目線の高さで自由に触れることができる位置に子ぐま型ソーシャルロボットを設置した。

今回の実験では、被験者に情報提供を行うアプリケーションを動作させている[金岡 2012]。このアプリケーションはおやつの時間や現在の時刻などのお知らせを行う。事前に被験者に対して触れ合い方についての教示は行わず、普段通りの生活の中で好きな時間に自由に触れ合ってもらいようをお願いした。

3.2 ユーザ観察の結果と考察

3 日間の被験者との触れ合いで合計 829 回のユーザ観察を行い、そのうち、654 回が有効な結果であった。アイコンタクトが成立しなかった場合や、アイコンタクトが成立したものの、正面に捉

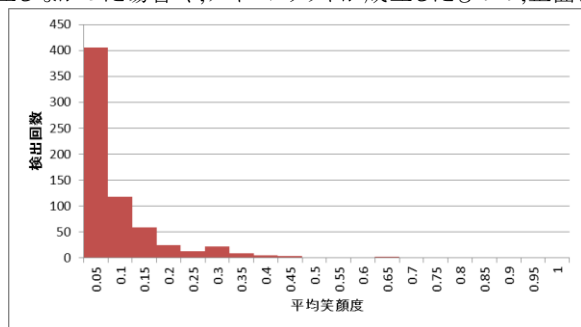


図 2 平均笑顔度の度数分布

える事が出来た顔画像が極端に少ない場合にユーザ観察を無効としている。観測した動画を確認したところ、654 回分のユーザ観察時の動画については、中央に被験者を捉えており、アイコンタクト時の振る舞い観察により、安定した観測が実現できている事が確認できた。

各ユーザ観察でのセンシング期間(一回のユーザ観察でセンシングする時間)での平均笑顔度(0.0~1.0)について、0.05 単位の度数の分布を図 2 に示す。654 回のうち 400 回以上が 0.05 以下の笑顔度であり、6 割以上のユーザ観察では笑顔がほとんど見られなかったことがわかる。観測した動画を見ても、確かに笑顔が豊富な動画は多くはなかった。2.1 で議論したように、機嫌が良い時間帯であっても笑顔などの顕著な表情が観測される頻度は少ない。また特に被験者が高齢者の為、表情の変化もあまり大きくなかったため、このような分布となっていると考えられる。

次に、ユーザ観察のタイミングによる違いについて考察する。2.2 で述べたように、本システムの実装では、ユーザ観察が実行されるタイミングは、(A)インタラクション開始時、(B)ユーザの働きかけによるセンシング、(C)アクティブセンシング、(D)定時センシングの 4 パターンがある。今回の実験では、それぞれのユーザ観察タイミングについて(A)170 回、(B)190 回、(C)62 回、(D)232 回が実行された。ユーザ観察タイミング別の、平均笑顔度の観測結果の分布を図 3 に示す。start, user trigger, active, timer はそれぞれ(A)~(D)を示している。図 3 から、(B)の場合は、平均笑顔度が 0.05 以下は約 50% であるのに対し、それ以外の場合では 65% 以上となっており、(B)のユーザがロボットに働きかけるタイミングでのユーザ観察で笑顔が良く観測されたことを示している。ユーザ観察のタイミングによってどの程度取得できる情報の質が変わるかを評価するために、(A)~(D)のそれぞれについてユーザ観察により得られる平均情報量を計算すると、(A)0.492、(B)0.658、(C)0.400、(D)0.484 となり、(B)の場合により高い情報量が得られていることがわかる。平均情報量は、センシングの効率を示しており、平均情報量を比べる事により、意味のある観測結果がどのユーザ観察方法でよく観測されたのかがわかる。

平均笑顔度以外の観測量に関してもユーザ観察のタイミング別の平均情報量を計算した結果を表 2 に示す。笑顔に関するセンシング情報に関して、(B)の場合に取得情報量が高くなっており、ユーザからの働きかけのタイミングで表情が表出されやすく、我々の手法が表情のセンシングに効果的であることが示された。

また、首振り、傾き、接近、手の動作(ジェスチャ)などの身体動作は(C)のアクティブセンシングから得られる情報量が高い事がわかる。実際に実験の動画からも、反応を示す様子が観察され、反応を引き出して観察する事には成功している。一方で、アクティブセンシングにより表情を引き出して観測するという事を期待していたが、本実験からはアクティブセンシング時には表情の表出が少ないという結果となった。今回の実験では被験者の年齢等

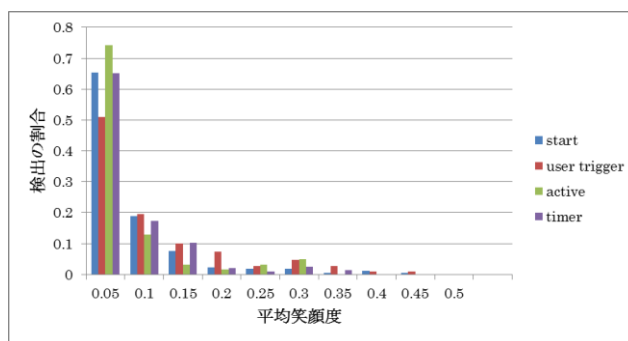


図 3 ユーザ観察タイミング別の平均笑顔度の分布

を考慮し、驚かす動作など、刺激が大きい動作を刺激動作から外していた。また、実験時に動作させていた情報提供を行うアプリケーションが、発話するなどの動作を行っている。このため、刺激動作が普段のロボットの振る舞いと大きな違いにはなっておらず、反応は引き出せたものの、刺激動作による表情変化までは引き出すことができなかった可能性がある。また、被験者により個人差もあると考えられ、刺激動作や普段の振る舞いの見直しや、他の被験者での実験などを通し、引き続き本手法の効果を検証していきたい。

表 2 各種観測量のユーザ観察タイミング別平均情報量

	start	user trigger	active	timer
平均笑顔度	0.492	0.658	0.400	0.484
最大笑顔度	0.894	1.006	0.832	0.939
笑顔変動量	0.584	0.663	0.541	0.582
首振り頻度	0.279	0.337	0.414	0.238
接近距離	0.296	0.316	0.350	0.312
頷き頻度	0.429	0.451	0.466	0.368
手の動作頻度	0.430	0.589	0.618	0.388

4. まとめ

本稿では、ソーシャルロボットが自然なインタラクションから、ユーザの非言語の振る舞いを観察する方法に関して、アイコンタクト前後の非言語の振る舞いを観測する方法を提案した。また、ユーザ観察を実行するタイミングとして、ロボットからの働きかけをトリガとするアクティブセンシングとユーザの働きかけをトリガとするセンシングの二つの方法を提案し、高齢者とのインタラクション実験により、それぞれの手法により、効果的にセンシングが行えている事を示した。ただし、現状は被験者が一名であり、他のユーザでも同様の効果があるのかなど、今後も調査を継続する必要がある。また、アクティブセンシングにより表情など心的状態を引き出すための動作刺激やインタラクションについては、引き続き検討する必要がある。

また、提案したユーザ観察方法により得られた観測結果に基づく心的状態の推定を実現するため、これらの観測結果とユーザの実際の心的状態との関係を明らかにしていくことが今後の課題である。

参考文献

- [富士通 2010] 人にやさしい端末「子ぐま型ソーシャルロボット」ユーザと親和的な関係を築くインタラクション技術, 富士通ジャーナル 2010年4月号, 富士通 2010.
- [渡辺 2007] 渡辺一郎ほか: 生き物感を有する対話ロボットによる高齢者セラピー, 日本ロボット学会講演会(RSJ2007)予稿集, 2B26, 2007.
- [山岡 2010] 山岡久俊ほか: 認知症高齢者を対象とした親和的ロボットによるロボット・セラピー, 第24回人工知能学会全国大会(JSAI2010)予稿集, 1H2-NFC3b-8, 2010.
- [雨宮 2011] 雨宮智ほか: 高齢者介護施設における子ぐま型ソーシャルロボットの影響度評価, 日本ロボット学会講演会(RSJ2010)予稿集, 1F3-1, 2011.
- [今井 2011] 今井岳ほか: 親和的なインタラクションに基づくサービスを提供する子ぐま型ソーシャルロボットの開発, 第12回計測自動制御学会システムインテグレーション部門講演会(SI2011)予稿集, 1C2-2, 2011.

- [Zeng 2009] Zhihong Zeng ほか: A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39-58, 2011.
- [金岡 2012] 金岡利知ほか: 親和的インタラクションに基づく人に優しい情報提供, 第26回人工知能学会全国大会(JSAI2012)予稿集, 3O1-OS-3a-9, 2012.