原著論文

ヒューマノイドロボットに対する人間の多感覚的な感情認知

河原 美彩子, 澤田 佳子, 田中 章浩

東京女子大学

Multisensory Emotion Perception to Humanoid Robot

Misako KAWAHARA, Yoshiko SAWADA and Akihiro TANAKA

Tokyo Woman's Christian University, 2-6-1 Zempukuji, Suginami-ku, Tokyo 167-8585, Japan

Abstract : We investigated the multisensory emotion perception from humanoid-robot. In the experiment, participants were presented with video clips containing emotional colored eyes and voice (Task 1) or body gesture and voice (Task 2) of the robot, which were either congruent or incongruent in terms of emotional content (e.g., a happy body gesture paired with a sad voice on an incongruent trial). Participants were instructed to judge the emotion of the robot as either happiness or sadness. We examined the proportion of responses based on visual or auditory cues for the robot's expression. Results showed that participants relied more on auditory cues than on the visual cues in Task 1. However, this vocal superiority was not observed in Task 2. These results suggest that the multisensory emotion perception from the robot is different whether the cues are natural or artificial. We proposed a model for multisensory emotion perception from a robot.

Keywords : *Humanoid robot, Emotion perception*

1. はじめに

現代社会において、人間とコミュニケーションをとることのできるロボットが身近な存在になってきている。これまでロボットは主に、本来人間が行っていた作業を代わりに行う用途として活躍してきた。しかしながら、近年Pepper (Softbank) やaibo (SONY) といった、人間の日常生活の中で直接コミュニケーションをとることができるロボットが増えてきている。Pepperなどのヒューマノイドロボットの形状は人間と類似しており、発声機能や身体の動きを用いた感情表出機能が組み込まれている。こうした機能を利用することによって、人間とヒューマノイドロボットが人間同士のやりとりと同様にコミュニケーションをとり、社会関係を構築できるよう日々開発が進められている。コミュニケーションをとる上で相手の感情を理解することは大変重要であるが、果たしてロボットが表出した感情を人間は正しく認知することが可能なのだろうか。

ヒューマノイドロボットによる感情表出とそれに対する人間の感情認知については、すでにロボットの音声や顔の表情 [1], 身体の動き [1, 2], ロボットの身体のパーツから発光された色 [3, 4] などを用いて検討されてきており、これらの手がかりから人間はロボットの表出した感情を一定の精度で知覚できることが示唆されている。たとえば、Tsiourtiら [1] は、ロボットが顔の表情、頭部の動き、身体 (全身) の動き、移動 (移動速度と進行方向)、音声によってそれぞれ表出した感情を、人間が読みとることができるか検討した。その結果、人間はロボットの感情を表情から最もよく読みとることができるが、音声等の他の手がかりからもチャンスレベル以上の精度で、ロボットの意図した感情を読みとることができるこ

とを報告している。また、ロボットの身体の動き (以下、身体言語とする) に着目した研究では、ヒューマノイドロボット (The Social Robot Brian 2.0) が身体言語 (body language) によって表出した8つの感情 (sadness, elated joy, anger, interest, fear, surprise, boredom, happiness) を、人間がロボットの表出した通りの感情として正しく認知できるか検討している [2]。実験の結果、interestとfearを除く感情において正答率がチャンスレベルを上回り、全体的に人間はヒューマノイドロボットの身体言語から感情を認知できることが明らかとなった。また、色と特定の感情との結びつきを用いた研究では、ロボットの頭部や目から発光された色 (以下、目の色とする) を手がかりにした場合も、人間がロボットの感情を読みとれることが示されている [3, 4]。

このように、人間は目の色の変化や身体言語、音声といった、単一の情報を手がかりにした場合に、ヒューマノイドロボットの表した感情を認知できることが先行研究によって示されている。しかしながら、人間のコミュニケーション場面においては、顔や身体の動きによって同時に感情が表出され、複数の情報を手がかりにする場面も多く、単一の情報のみを手がかりに他者の感情を判断する機会は少ないだろう。また、顔や身体といった視覚的な手がかりに加え、音声という聴覚的な手がかりを同時に受けとる場合もある。したがって、人間は顔の表情や音声といった複数の情報を同時に知覚し、多感覚的に他者の感情を読みとっていると言える。実際に、人間の顔の表情と音声を用いた多感覚的な感情認知では、顔の表情という視覚情報と音声という聴覚情報がそれぞれ単一で呈示された場合と比較して、視覚と聴覚の両方の情報が同時に呈示されたほうが感情認知の精度が高く [5], 視覚情報と聴覚情報は相互に影響を及ぼすことが示されている [6]。ロボットからの感情認知においては、上記の先行研究によっ

て示されているように、単一の視覚的な手がかりからある程度正しく感情を読みとることができるが、それでは不十分なケースも存在するだろう。たとえば、表情の動きを実装したヒト型ロボットを用いた研究は、ロボットが笑顔を見せようとしてもその動作速度が遅い場合に、受け手に不気味さを与えてしまう可能性を示唆しており [7]、人間の表情と類似した動きであっても、それが意図された感情として人間に受けとられない可能性もある。また、人間の表情の動きを実装したロボットの制作にはコストがかかるため、一般家庭等に導入されるロボットには、Pepper (Softbank) のように顔の動きがないものや、顔のパーツのみがモニタに表示されることで表情を表現するものもある。表現を簡略化することにより、視覚的な手がかりが曖昧になる可能性があり、このように視覚的な手がかりが不十分な場合は、聴覚的な手がかりによって補う必要があるかもしれない。したがって、ロボットと人間の感情コミュニケーションにおいても、ロボットによる多感覚的な感情表出とそれに対する人間の認知に焦点を当てる必要がある。その際、人間同士のコミュニケーションと異なる点として、ロボットは人間が用いることができない方法で感情を表出できることにも注目すべきである。たとえば、人間はロボットの目の色から感情を読みとることができるが [3, 4]、目の色を変化させることは人間には不可能である。そのため、顔の表情や身体言語、音声といった、人間にとって親しみのある手がかりのほうが、感情の手がかりとしてより分かりやすいかもしれない。人間のロボットに対する感情認知精度は、その手がかりが人間にとって親しみのあるものであるか否かの影響も受ける可能性があるため、ロボットの感情表出手段の種類も考慮する必要があるだろう。

Tsiourti らは、ロボットの音声と身体言語からの多感覚的な感情認知について検討している [8]。実験では、喜び、悲しみ、驚きを喚起する動画を実験参加者に視聴させた後、その動画に対する反応として、ロボットが喜び、悲しみ、驚きを表出している様子を呈示し、ロボットの表した感情を回答させた。実験には、動画の内容とロボットの表出した感情が一致している条件 (たとえば、悲しみの動画に対してロボットが悲しみの感情を音声と身体言語で表出する) に加え、動画の内容とロボットの表出した感情が一致していない条件 (たとえば、悲しみの動画に対してロボットが喜びの感情を音声と身体言語で表出する) があつた。また、ロボットの音声と身体言語の表す感情が矛盾しており、動画の内容とロボットの表出した感情が部分的に一致しない条件 (たとえば、悲しみの動画に対してロボットが音声では喜びを、身体言語では悲しみを表出する) もあつた。実験参加者のロボットに対する感情判断を条件間で比較した結果、動画の内容とロボットの表出した感情が一致している条件に比べて、動画の内容とロボットの表出した感情が一致していない条件と、ロボットの音声と身体言語の表す感情が矛盾している条件では、人間のロボットに対する感情の判断が混乱することが示された。しかしながら、この研究ではロボットの感情を人間が判断する際に、ロボットによって表出された感情が動画の内容という文脈と

合致しているか否かという要因が必ず含まれている。そのため、ロボットの身体言語と音声を手がかりとして、人間がどのようにロボットの感情を多感覚的に認知するのかについて、文脈の要素を排除した状態では検討されていない。また、人間同士の感情認知においては、視覚情報 (顔) が明瞭である場合は、聴覚情報 (音声) と比較して視覚情報が重視される傾向が報告されている [9, 10]。しかし、ロボットからの多感覚的な感情認知において、視覚情報と聴覚情報のどちらを重視するのか、相対的なモダリティの優位性についても検討されていない。そこで本研究では、Pepper (以下、ロボットとする) を用いた実験を実施し、ロボットが視覚と聴覚によって多感覚的に表出した感情を人間がどのように認知するのか、文脈を統制したうえで検討することを目的とした。

実験ではロボットが視覚情報 (目の色または身体言語) と聴覚情報 (音声) によって多感覚的に感情を表出する様子を実験参加者に視聴させ、ロボットがどのような感情を表しているか判断させた。このとき、視覚と聴覚がそれぞれ表出する感情が矛盾する組み合わせ (たとえば、喜びの目の色 + 悲しみの音声) も作成し、2つの感覚情報が異なる感情を表している場合に人間がどちらの情報を重視するのかを調べることで、視覚と聴覚の相対的な優位性を検討した。

上述のように、人間同士の感情認知においては、顔の表情を視覚情報、音声を聴覚情報としたとき、視覚情報である顔が明瞭である場合は、聴覚情報と比較して視覚情報が重視される傾向が報告されている [9, 10]。また、身体言語と音声を手がかりとした視聴覚的な感情認知においても、顔の表情と音声による感情認知と同様に、感情の判断に互いの情報が影響を及ぼし合うことが報告されている [11]。身体言語と音声の組み合わせにおいては、どちらの情報がより優位かという直接的検討はなされていないが、身体言語を顔の表情と同様に視覚情報と捉えれば、聴覚情報に比べて視覚情報が優位に重視される可能性がある。このことから、ロボットの感情を読みとる際にも聴覚情報よりも視覚情報を手がかりとして重視する傾向がみられると仮説を立てた。また、ロボットの感情の表出方法として、身体言語と音声は人間の感情表出においても用いられるものであるが (以下、人間との共通表現とする)、目の色は人間には変化させることができないロボット特有の表現方法 (以下、人工表現とする) である。そのため、人間にとって人工表現よりも人間との共通表現のほうが、感情判断の手がかりとしての優位性が高くなると予測した。

2. 方法

2.1 実験計画

実験は2つの課題から構成された。課題1では、目の色と音声を組み合わせた刺激を呈示し、課題2では、身体言語と音声を組み合わせた刺激を呈示した。課題1は2 (目の色: 喜び, 悲しみ) $\times$ 2 (音声: 喜び, 悲しみ) の被験者内計画、課題2は2 (身体言語: 喜び, 悲しみ) $\times$ 2 (音声: 喜び, 悲しみ) の被験者内計画だった。従属変数はいずれの課題においても、

ヒューマノイドロボットに対する人間の多感覚的な感情認知

感情判断の手がかりとして視覚と聴覚のどちらの情報を選択したか、その割合とした。

2.2 実験参加者

実験参加者は、正常な視覚及び聴覚を有する大学生47名(全員女性)であった。平均年齢は18.60歳($SD=0.96$)であった。なお、本研究は、東京女子大学人を対象とする研究に関する倫理審査委員会により、倫理審査は不要であると判断された(受付番号: R2020-101)。人を対象とする医学系研究に関する倫理指針(文部科学省, 厚生労働省)に基づき、インフォームドコンセントは口頭にて取得した。

2.3 実験刺激

ロボットが目の色(静止画)、身体言語(動画)および音声によって感情(喜び、悲しみ)を表す様子を記録し、静止画または動画を音声と組み合わせたものを刺激として使用した。目の色は人間が表現できない人工表現であり、それによって生じた不自然さが感情認知に影響する可能性がある。また、点灯の仕方や速度が感情認知に及ぼす影響も避けるため、静止画の刺激を使用した。なお、ロボットの感情表出はPepper開発用ソフトウェア(Choregraphe)を用いて操作した[注1]。刺激を視覚あるいは聴覚の単一の感覚モダリティによって呈示した場合の感情判断の正答率を統制するため、予備実験を実施し、静止画、動画、音声の正答率が約80%で揃うよう選定したうえで、それらを組み合わせて刺激を作成した。人間の視聴覚による多感覚的な感情認知を検討した先行研究[12, 13]では、天井効果避け、且つ、視覚と聴覚の手がかりの認識容易レベルを揃えるために、各モダリティの刺激正答率が80%程度以下になるよう統制されていた。そのため、本研究でも視覚および聴覚刺激を単独で呈示した際の正答率が80%程度に揃うよう、刺激の選定を行った。

目の色は、先行研究[3, 4]と同様にPulchikの感情モデル[14]に結果が合致することを予備実験にて確認した上で、喜び刺激として彩度の異なる2種類の黄色を、悲しみ刺激として彩度の異なる2種類の青色を使用した。身体言語は、各感情につき2種類の身体の動き(喜び: 右手でガッツポーズをする動き、上を向いて両手を広げる動き; 悲しみ: 両手で頭を抱える動き、うつむいて左手を額に手を当てる動き)を選定した。音声は、「ぼくの取り分はそれですね」というセリフを喜びと悲しみの感情を表す言い方で発話させたものを使用した。各感情につき2種類の音声を選定し、喜びの音声はChoregraphe上のパラメータにおいてピッチが180、話速が155の音声と、ピッチが180、話速が135の音声を使用した。悲しみの音声は、ピッチが140、話速が75の音声と、ピッチが110、話速が155の音声を使用した。なお、セリフが意味的に中立であることを確認するため、文字で呈示された当該セリフから感じられる感情を7件法(「1. 悲しみ」～「4. どちらでもない」～「7. 喜び」)で評定させる調査を実施した結果、「4. どちらでもない」が最頻値となり、意味的に中立であることが確認された。

目の色によって感情を表出しているロボットの顔を記録した静止画と、身体言語によって感情を表出しているロボットの全身の様子を記録した動画は、解像度を 640×480 pxに設定し、音声はサンプリングレートを48 kHz、量子化を16 bitに設定した。感情を表す音声と静止画または動画を組み合わせた刺激ファイルの長さは、すべて3500 msに統制した。

2.4 実験装置

視覚刺激を呈示するため、パーソナルコンピュータ(Lenovo, G580)を1台使用した。なお、画面解像度は 800×600 pxで統一した。また、音声呈示のためにヘッドホン(SENNHEISER, HDA200)を使用した。なお、音声刺激呈示時の音圧は58 dBとした。

2.5 手続き

防音室にて個別に実験を実施した。実験参加者は、課題1(音声+目の色)と課題2(音声+身体言語)の両方に参加し、実施順序は実験参加者間でカウンターバランスをとった。課題1では、音声(喜び2種類、悲しみ2種類)と目の色(喜び2種類、悲しみ2種類)を組み合わせた計16刺激を3回ずつ呈示し、計48試行から構成された。課題2は、音声(喜び2種類、悲しみ2種類)と身体言語(喜び2種類、悲しみ2種類)を組み合わせた計16刺激を3回ずつ呈示し、計48試行から構成された。なお、刺激はすべてランダムな順で呈示した。

実験手順を図1に示す。実験では、ロボットの静止画または動画と同時に「ぼくの取り分はそれですね」という音声をパーソナルコンピュータおよびヘッドホンを通してそれぞれ呈示した。その後、ロボットの表す感情が喜びと悲しみのどちらであったか、キー押しによって二肢強制選択で実験参加者に回答させた。実験課題に正解はなく、思ったように回答するよう教示を行った。なお本研究では、人間がロボットの感情を判断する際に、視覚と聴覚のどちらの情報に自動的に注意を向けるかを検討することを目的としていたため、参加者への事前の説明として、特定の感覚モダリティに注意を向ける教示は行わなかった。課題1および課題2の終了後、Pepperとの触れ合いの経験を問う3項目について質問紙で回答させた。また、「プログラミングをしてPepperを動かしたことがあるか」「今までに学内でPepperに関する実験に参加したことがあるか」の2項目について、その有無を回答させた。

図1 実験の手順

3. 結果

質問紙調査の結果、全実験参加者のうち、実験参加以前に Pepperを知っていた人は86.96%で、そのうちPepperのアプリ開発ソフトウェア(Choregraphe)を使用したことがある人はいなかった。

ロボットから多感覚的に表出された感情を人間がどのように認知するのかを検討するため、課題において視覚と聴覚の表す感情が矛盾する刺激を視聴した試行の結果に着目した。まず実験参加者ごとに、感情が矛盾する刺激に対して視覚の表す感情を選択した割合(以下、視覚選択率とする)と聴覚の表す感情を選択した割合(以下、聴覚選択率とする)を算出した。たとえば課題1において、喜びの目と悲しみの音声を組み合わせた刺激に対して、喜びと回答した場合は、その回答を視覚選択率として集計した。視覚情報と聴覚情報のどちらが優位に選択されるかを検討するため、各課題において、刺激の種類(課題1: 悲しみの音声+喜びの目、喜びの音声+悲しみの目; 課題2: 悲しみの音声+喜びの身体、喜びの音声+悲しみの身体)ごとに聴覚選択率から視覚選択率を差し引き、この差分値について1サンプルの t 検定(検定値=0)を行った。差分値が0の場合、視覚情報と聴覚情報を同程度に選択したことを示す。差分値が正の値である場合は聴覚情報を、負の値である場合は視覚情報を優位に選択したことを表す。

結果を図2に示す。課題1では、感情の組み合わせに関わらず、目の色よりも音声の表す感情のほうが優位に選択された(悲しみの音声+喜びの目: $t(46)=7.29, p<.001$; 喜びの音声+悲しみの目: $t(46)=8.31, p<.001$)。一方、課題2では、感情の組み合わせによって結果が異なり、喜びの音声+悲しみの身体言語と組み合わせられた場合においてのみ、音声よりも身体言語のほうが選択されやすく($t(46)=-3.18, p<.01$)、悲しみの音声+喜びの身体言語と組み合わせられた場合には感情の選択率に有意な差はみられなかった($t(46)=0.01, n.s.$)。

課題1と課題2では感情判断のパターンが異なることが示されたが、視覚刺激の種類と感情の組み合わせの2つの要因の

うち、どちらがよりロボットからの感情認知に影響を及ぼすのだろうか。これを検討するため、課題1および課題2における聴覚選択率と視覚選択率の差分値について、視覚刺激の種類(目の色、身体言語)×感情の組み合わせ(悲しみの音声+喜びの視覚、喜びの音声+悲しみの視覚)の被験者内分散分析を行った。その結果、視覚刺激の種類の主効果が有意であり($F(1,46)=61.65, p<.001$)、視覚情報が身体言語であるときよりも目の色であるときのほうが聴覚選択率が有意に高かった。一方、感情の組み合わせの主効果はみられなかった($F(1,46)=0.24, n.s.$)。この結果は、ロボットからの感情認知が、視覚と聴覚の表す感情の組み合わせよりも、手がかかりとなる要素が人工表現であるか人間との共通表現であるかの影響を強く受けることを示している。また、交互作用効果も有意だった($F(1,46)=21.16, p<.001$)。感情の組み合わせによって聴覚情報あるいは視覚情報が選択される程度に違いがみられるか検討するため、視覚刺激の種類(目の色、身体言語)において単純主効果検定を行った。その結果、どちらの条件においても感情の組み合わせの単純主効果が有意であり(目の色では $p<.05$; 身体言語では $p<.01$)、聴覚情報あるいは視覚情報が選択される程度は、視覚と聴覚の感情の組み合わせによっても異なっていた。

なお、音声、目の色、身体言語の刺激は、各感情につき2種類用意されていたが、視覚と聴覚の示す感情が不一致の刺激に対する回答においては、視聴覚刺激の組み合わせの種類の影響がみられた。悲しみの音声+喜びの目の表現では、目の色の種類によらず、ピッチが140の悲しみ音声+喜びの目の組み合わせは(63.8%)、ピッチが110の悲しみ音声+喜びの目の組み合わせ(83.3%)と比較して聴覚選択率が低かった。この傾向は、悲しみの音声+喜びの身体の表現でも同様にみられ、身体言語の種類によらず、ピッチが140の悲しみ音声+喜びの身体の組み合わせは(35.3%)、ピッチが110の悲しみ音声+喜びの身体の組み合わせ(65.2%)と比較して聴覚選択率が低かった。上記の結果は、視覚あるいは聴覚の単一感覚モダリティにおける感情判断正答率を統制した刺激を用いても、それらの組み合わせの種類が多感覚的な感情判断に影響する可能性を示しており、この組み合わせの影響については今後検討が必要である。

4. 考察

4.1 結果の考察

本研究では、ロボットが視覚と聴覚によって多感覚的に表出した感情を、人間がどのように認知するのかを検討することを目的とした。実験の結果、課題1では目の色(つまり、視覚)と音声(つまり、聴覚)の表す感情の組み合わせによらず、聴覚情報を重視してロボットの感情を判断することが明らかとなった。一方、課題2では、音声(つまり、聴覚)よりも身体言語(つまり、視覚)が表す感情が重視される傾向がみられ、感情の組み合わせによっても結果が異なっていた。これらは、人間がロボットから感情を認知する場合、聴覚情報よりも

図2 ロボットが視覚(目の色または身体言語)と聴覚(音声)で矛盾する感情を示したときの視覚選択率と聴覚選択率の差分

ヒューマノイドロボットに対する人間の多感覚的な感情認知

視覚情報を、人間の感情表出には用いることができない人工表現よりも人間との共通表現を感情認知の手がかりとしてより重視する、という仮説をおおむね支持する結果である。

実験で用いた目の色、身体言語および音声刺激は、予備実験を行うことによって、正答率を刺激の種類間で同程度に揃えられていた。そのため、課題間の聴覚選択率の差異は、視覚刺激の種類による感情認知の難易度によってもたらされたものではないと言える。課題1で用いられた目の色はロボット特有の人工表現であり、課題2で用いられた身体言語は人間との共通表現であった。本研究では視覚的手がかりが身体言語であるとき（課題2）よりも目の色であるとき（課題1）のほうが、聴覚を重視した感情判断がなされていた。このことは、ロボットからの感情認知において、視覚情報が人間との共通表現によるものであるか人工表現によるものであるかの影響が大きいことを示している。

課題2では、感情の組み合わせによっても手がかりに対する重みづけが異なることが示された。先行研究では、表情と音声のどちらにおいてもネガティブ感情のほうがポジティブ感情よりも知覚されやすい傾向が示唆されている [10, 12]。したがって本研究の結果から、視覚と聴覚の2つの感情の手がかりがどちらも人間との共通表現である場合には、ロボットの感情を認知する場合も人間の感情を認知するときと同様に、聴覚情報よりも視覚情報が、ポジティブ感情よりもネガティブ感情が優位に重視されやすいと考えることができる。

以上の結果を踏まえ、ロボットに対する感情認知のフローチャートを提案する（図3）。人間はロボットによる感情表出を知覚した後、まずその感情の手がかりが人間との共通表現と人工表現のどちらを用いたものであるかを判断する。ここで、同時に呈示された2つの手がかりのどちらか一方のみが人間との共通表現であれば、それが示す感情がロボットの感情として優先される（ステージ1）。2つの手がかりがどち

らも人間との共通表現である場合は、次の段階に判断が持ち越され、感情情報の重みづけ（視覚情報か聴覚情報か）と、感情価の判断（ポジティブ感情かネガティブ感情か）がそれぞれなされ、感情が決定する（ステージ2）。このフローチャートを本研究の結果に適用すると、以下のような説明が可能である。

まず、目の色と音声が組み合わせられた場合には、ステージ1で人工表現である目の色よりも、人間との共通表現である音声の感情が優先されるため、悲しみの目と喜びの音声の組み合わせでは喜びが、喜びの目と悲しみの音声の組み合わせでは悲しみが、それぞれロボットの感情として決定する。一方、悲しみの身体言語と喜びの音声の組み合わせられた場合には、どちらの手がかりも人間との共通表現であるためにステージ1では判断がつかず、ステージ2に持ち越される。ここで、視聴覚情報に対する重みづけが適用され、視覚情報である身体言語の表す感情（つまり、悲しみ）が優先される。さらに、感情価の判断でもネガティブ感情を示す身体言語が優先されるため、ロボットの感情は悲しみと判断される。喜びの身体言語と悲しみの音声の組み合わせられた場合には、もう一方の感情の組み合わせと同様に判断がステージ2に持ち越され、視聴覚情報に対する重みづけの決定において、視覚情報である身体言語の表す感情（つまり、喜び）が優先される。しかしながら、感情価は音声を表すネガティブ感情（つまり、悲しみ）が優先される。これにより、喜びと悲しみの情報が拮抗し、結果的に視覚選択率と聴覚選択率に有意差がみられなかったと考えられる。

ただし、課題2では身体言語の動画を視覚刺激として使用したのに対し、課題1では目の色の変化の不自然さが感情認知に影響を与える可能性を考慮して静止画を使用した。そのため、視覚刺激が動的であるか否かによって、視覚情報の印象が異なっていた可能性がある。特に、目の色に関しては、色の変化に気付いていない実験参加者がいた可能性も否定できない。今後は、目の色を消灯させた状態から特定の色へと変化する動画刺激を使用し、検討することが必要である。また、実験の結果、聴覚選択率は視覚刺激と聴覚刺激の組み合わせの種類の影響を受ける可能性が示された。この結果は、単感覚的に刺激を呈示した場合の正答率が同程度に揃えられていても、その視覚刺激と聴覚刺激が組み合わせられた場合、多感覚的な感情認知における視覚あるいは聴覚情報への重みづけが一定にならず、組み合わせによって異なることを示唆している。今後は、刺激の組み合わせの影響について検討するとともに、多感覚刺激を作成する際には、音声やジェスチャーといった個々の手がかりの認識容易性だけでなく、それらの組み合わせを考慮したうえで刺激を選定する必要がある。

図3 ロボットに対する感情認知のフローチャート

ステージ1では、ロボットが表出した多感覚的な感情の手がかりが人工表現と人間との共通表現のどちらによるものが判断する。人工表現と人間との共通表現が組み合わせられている場合、人間との共通表現が優先される。人間との共通表現同士の組み合わせの場合、感情認知はステージ2に持ち越される。ステージ2では、手がかりが視覚情報か聴覚情報か、また表している感情がポジティブかネガティブかの判断が行われ、視覚情報とネガティブ感情がそれぞれ優先される。

4.2 ロボットデザインへの応用的示唆

本研究では、ロボットからの感情認知は、その手がかりが人工表現によるものであるか、人間との共通表現によるものであるかの影響を受けることが明らかとなった。この結果は、これまでの研究で報告されてきた、聴覚よりも視覚の情報を重視して

感情を判断する傾向にあるという人間の視覚優位性 [9,10] が、視覚情報が人工的な表現によるものである場合には当てはまらない可能性を示唆するものである。人工表現である目の色の変化は、色と感情の結びつきによって、単感覚的な感情認知においては手がかりとして十分に有効である [3,4]。しかしながら、それに加えて人間との共通表現である音声も同時に手がかりとなる場合には、人間との共通表現を用いた手がかりが優先されてしまうことで、視覚的な手がかりとしての重みづけが小さくなり、聴覚優位な判断がなされる可能性がある。したがって、ロボットを相手とした場合の感情認知様式は、人間を相手とした場合とは異なり、手がかりが人間と共通しているか否かを考慮することも重要だといえる。それを念頭に置いた上で、たとえば身体言語においてわかりやすいジェスチャーで表現するなど、人間が感情を容易に理解できるロボットならではの感情表出の方法を考えていく必要があるだろう。

なお、上記の示唆は、視覚と聴覚によって多感覚的に感情を表現することができるコミュニケーションロボットであれば、ヒューマノイドロボットでなくても適応される可能性がある。しかし、ペット型ロボットやスマートスピーカー、CG エージェントにも応用することができるかについては、今後検討が必要である。

4.3 今後の展望

本研究では、これまで明らかにされてこなかったロボットからの多感覚的な感情認知について検討してきた。ロボットが人間との共通表現を用いて感情を示した場合には、人間を相手とした場合と類似した感情認知のプロセスがなされることが示唆された。この結果から、ロボットと人間は同じ表現方法を用いることによって双方向的なコミュニケーションをとることができると考えられる。一方、ロボットが目の色という人工表現によって感情を表出した場合、目の色のみが感情判断の手がかりとなるとときには身体言語や音声と同等に感情を伝達することができるが、音声と同時に呈示された場合には、目の色は感情判断の手がかりとしての重みづけが比較的小さいことが示唆された。これはロボットからの感情認知固有の点であり、単一の感覚モダリティによるロボットの感情研究の知見を拡張させるものであるといえるだろう。

本研究ではロボット固有の人工的な表現として、目の色に着目したが、電子音を再生することや、液晶パネルにハートや怒りを示すマーク (怒) 等の記号を呈示する方法もある。上記のような記号は通用する文化が限定される可能性があり、その点、感情と色の結びつきは通文的に利用できる可能性がある。Plutchik は、喜び-悲しみ、怒り-恐れ、嫌悪-信頼、驚き-期待の8つの感情を円環状に配置し、感情の強さを色の濃淡で表現するという、感情の立体モデルを提唱しており [14]、このモデルと類似した色と感情の連合が、日本人を対象とした研究でも示されている [3,4]。本研究ではPepperの目の色を変化させたが、頬の色を変化させることができるロボットもあるため (BOCCO emo, ユカイ工学株式会社)、目の色で

なくとも、色と感情の結びつきを利用した手がかりは汎用性が高いと考えられる。一方で、電子音も感情を表現する手段として有用であることが示唆されている。Löfflerらの研究では、「発光される色」、「ビープ音」、「回転等の動き」の3種の人工的な手段によって、人間らしさや生き物らしさをできる限り排除したロボットに感情を表現させ、人間がロボットの意図した感情を認知可能であるか検討した [15]。その結果、いずれの手段も感情の表現手段として有用であり、さらに、上記の3種類の手がかりを組み合わせた場合、感情の正答率が上昇することが示唆された。しかしながら、ビープ音のような人工表現が、ジェスチャー等の人間と共通した表現と組み合わせられた場合には、どのように感情が認知されるのかは明らかとなっておらず、また、これらの手がかりが通文的かどうかについても検討されていない。したがって、人工表現として電子音を用いた場合でも、本研究の結果が再現されるのかについては、今後検討する必要がある。

ロボット自身は感情を持たないとしても、人間はロボットの目の色や身体の動き、音声から、ロボット自身が感情を表しているように感じることができる。ロボットの感情表出様式と、それに対する人間の感情認知様式について今後さらに検討することで、より円滑な感情コミュニケーションを可能とするロボットの開発に貢献できると考えられる。

謝 辞

本研究は科研費新学術領域研究No.17H06345「トランスカルチャー状況下における顔身体学の構築-多文化をつなぐ顔と身体表現-」の助成を受けたものです。

また、本研究は東京女子大学の高山沙絢さん、都竹彩斗子さん、野口杏奈さん、平田美音さんの卒業研究の実験結果に基づいています。

注

[注1] Pepper開発用ソフトウェア (Choregraphe) で使われているパラメータに基づく。話速は基準 (約7 モーラ/秒) を100とした速度とおおむね対応するもので、50~400の間で調整でき、数字が大きいほど音声は速くなるものだった。ピッチは50~200の間で調整でき、数字が大きいほど音声は高くなっていた。

参 考 文 献

- [1] Tsiourti, C., Weiss, A., Wac, K., and Vincze, M.: Designing emotionally expressive robots: A comparative study on the perception of communication modalities. In Proceedings of the 5th International Conference on Human Agent Interaction pp.213-222, 2017.
- [2] McColl, D., and Nejat, G.: Recognizing emotional body language displayed by a human-like social robot, International Journal of Social Robotics, 6(2), pp.261-280, 2014.

ヒューマノイドロボットに対する人間の多感覚的な感情認知

- [3] 山内厚志, 寺田和憲, 伊藤昭: 動的な発色によるロボットの感情表出, ヒューマンインタラクション学術論文誌, 13(1), pp.41-52, 2011.
- [4] 寺田和憲, 勅使宏武, 伊藤昭: ロボットが表出する感情の社会機能的評価, 研究報告ヒューマンコンピュータインタラクション(HCI), 156(8), pp.1-8, 2014.
- [5] Kim, J., and Davis, C.: Perceiving emotion from a talker: How face and voice work together. *Visual Cognition*, 20(8), pp.902-921, 2012.
- [6] de Gelder, B., and Vroomen, J.: The perception of emotions by ear and by eye. *Cognition & Emotion*, 14(3), pp.289-311, 2000.
- [7] Mori, M., MacDorman, K. F., and Kageki, N.: The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), pp.98-100, 2012.
- [8] Tsiourti, C., Weiss, A., Wac, K., and Vincze, M.: Multimodal integration of emotional signals from voice, body, and context: Effects of (in)congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics*, 11(4), pp.555-573, 2019.
- [9] 高木幸子, 平松沙織, 田中章浩: 表情と音声と同時に感情を込めた動画刺激に対する感情知覚, 認知科学, 21(3), pp.344-362, 2014.
- [10] Takagi, S., Hiramatsu, S., Tabei, K., and Tanaka, A.: Multisensory perception of the six basic emotions is modulated by attentional instruction and unattended modality. *Frontiers in Integrative Neuroscience*, 9(1), 2015.
- [11] Eastwood, J. D., Smilek, D., and Merikle, P. M.: Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & Psychophysics*, 63(6), pp.1004-1013, 2001.
- [12] Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., and de Gelder, B.: I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychological Science*, 21(9), pp.1259-1262, 2010.
- [13] Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., and Lepore, F.: Audio-visual integration of emotion expression. *Brain Research*, 1242, pp.126-135, 2008.
- [14] Plutchik, R.: *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division, 1980.
- [15] Löffler, D., Schmidt, N., and Tscharn, R.: Multimodal expression of artificial emotion in social robots using color, motion and sound. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp.334-343, 2018.

河原 美彩子 (正会員)

2016年 東京女子大学現代教養学部人間科学学科卒業。2018年 同大学大学院人間科学研究科博士前期課程修了。現在、同大学人間科学研究科博士後期課程在学中。2018年より日本学術振興会特別研究員(DC1)。日本心理学会では、2018年度学術大会優秀発表賞受賞、2019年度学術大会特別優秀発表賞受賞。日本心理学会、日本認知科学会などの各会員。

澤田 佳子 (非会員)

2016年 東京女子大学現代教養学部人間科学学科卒業。2018年 同大学大学院人間科学研究科博士前期課程人間社会科学専攻修了。修士(人間社会科学)。2018年より東京女子大学および東京大学にて研究補助、2020年より東京大学にて学術専門職員、現在に至る。HCGシンポジウムでは、2019年度優秀インタラクティブ発表賞受賞。日本心理学会会員。

田中 章浩 (非会員)

2004年 東京大学大学院人文社会系研究科博士課程修了。東北大学電気通信研究所、東京大学大学院人文社会系研究科、オランダ・ティルブルグ大学(日本学術振興会海外特別研究員)、早稲田大学高等研究所等を経て、現在、東京女子大学現代教養学部教授。感情や感性のクロスモーダル知覚と、その文化差・個人差に関する研究に従事。博士(心理学)。