

# 遠隔地の複数鑑賞者の存在感を表現するテレプレゼンスロボット

米澤謙<sup>†1</sup> 上田博唯<sup>†1</sup>

テレプレゼンスシステムのエンターテインメント的応用の1つとしてロボットが遠隔地の友人らの感情やリアクションを表現することによって、同じ部屋で友人と一緒に映画を楽しんでいるかのような臨場感を与えるシステムを提案する。ロボットが遠隔地の鑑賞者の様子を表現することで臨場感を生み出すことができるという可能性を検証するための基礎的な実験を行ったので報告する。

## Telepresence Robots representing presence of multiple remote people

KEN YONEZAWA<sup>†1</sup> HIROTADA UEDA<sup>†1</sup>

We propose a telepresence system that robots beside a user properly represents an emotion or a behavior of remote friends so that those robots can raise a sense as if enjoying watching a movie with those friends in the same room. In this paper, we report about fundamental experiment to prove the potentiality that our proposed approach has an effect on raising the sense by representing the presence of multiple remote people.

### 1. はじめに

近年、遠隔コミュニケーションをサポートするためのシステムとしてビデオ会議システムやテレプレゼンスシステムが多く開発されてきた。大塚らは遠隔地にいる複数の人々の会話をあたかもその場で聴いているかのように再現することを目指し、MM-Spaceを開発し[1]、実際の対面対話において重要な視線や首振り、頷きなどの非言語行動の再現を可能にした。実体性や3次元表現度を強調して、遠隔地の人の存在感を高めるテレプレゼンスロボットとして、柏原らは TEROOS というウェアラブルのテレプレゼンスロボットを開発した[2]。遠隔地の操作者がロボットの視線方向を制御することで、装着者が自由に移動をしながらも、視線の共有を可能にし、円滑なコミュニケーションを実現した。また長谷川らはユーザの無意識的身振りをそのまま表出するテレプレゼンスロボットを開発した[3]。それにより、無意識的身振り情報が欠落することで生じていた発話衝突という問題を解決し、話者交替の円滑化を実現した。

我々の関心はテレプレゼンスシステムのエンターテインメント(エンタメ)的応用である。動画共有(ニコニコ動画等)、オンラインゲーム、ネットコミュニティ(facebookやmixiなど)といったソーシャルメディアの普及に伴い、遠隔地の複数の人々との娯楽的交流が多様化している。そこで我々はそのような遠隔地の複数人とのエンタメ的コミュニケーションを複数台のテレプレゼンスロボットを用いて人々の存在感を演出し、あたかも同じ部屋にいるかのような臨場感を醸し出すシステムの開発を目指した。

会議に限らず、動画鑑賞時においても、視線の共有や一致、非言語行動の表出はロボットの存在感を高める効果を

持つため、再現されなければならない。大塚らの手法[1]を用いることで、それらを実現することは可能ではあるが、ユーザに対して絶対的な鑑賞場所をシステムが指定することは快適な場所での鑑賞を阻む恐れがある。そこで、ユーザに自らの鑑賞場所とロボットの配置場所を任意で指定させ、それぞれの部屋において、異なる位置関係にも関わらず、視線や非言語行動の整合性を保たせる必要がある。また、柏原ら[2]や長谷川ら[3]が開発したテレプレゼンスロボットは1対1のコミュニケーションを主な対象としている。これに対し、映画を鑑賞するという状況において、対面状態よりTV画面を共同注視した状態が長いことから、長谷川らが開発したユーザの身振りをそのまま再現する手法ではなく、ロボットを通して遠隔地のユーザの存在感を適切に伝達するのに適した新しい手法を本研究では探ろうとした。

本稿では、ロボットが遠隔地の鑑賞者の様子を表現し臨場感を生み出すために必要な情報である感情に焦点を当て、ロボットによる感情表現動作をどのようにすれば適切に感情を伝達できるかを論じ、評価ビデオを用いた検証実験についても報告する。

### 2. ロボットによる鑑賞者の存在感の表現

予備実験の結果、ロボットが鑑賞者の動作をあまりに忠実に再現すると、鑑賞時の鑑賞者の動作は控え目であることが多いため、効果的にその鑑賞者の存在感を他者へ伝えることは難しいことが分かった。そこで、我々は鑑賞者から必要な情報を検出し、やや誇張的にその鑑賞者の存在感をロボット動作で表現する手法を用いることとした。

まず、最初に我々は3人の友人同士がリビングルームで映画(スリラーとコメディ要素のある「エクスクロス」)を見る様子をビデオ記録した。そしてその記録映像を分析す

<sup>†1</sup> 京都産業大学  
Kyoto Sangyo University

ることによって、検出されるべき鑑賞者の情報（以下ファクター）を鑑賞者の表情、動作、声とし、そしてこれらのファクター(F)は検出されるフェーズ(P)によって表現されるべき重要性に応じて表1のように分類した。

表1. ファクターとフェーズの組み合わせ

F	P	行動の細目	代表例	ロボットの対応			
表情	情動		大きく感情表出	表情から推定された情動を表現するロボット動作に翻訳			
	ムード		小さく感情表出	動作がなかった場合に情動表現動作に翻訳			
動作	情動	会話	意図的	頷き	頭部運動を忠実に再現		
			自己親密	手で口を隠す動作	両手、胴体だけで情動表現		
			無意識的	驚いた時のとっさの動作	両手、胴体だけで情動表現		
		非会話	無意識的	驚いた時のとっさの動作	情動表現動作に抑揚としてインクリメント		
			ムード	会話	意図的	頷き	頭部運動を忠実に再現
					自己親密	髪を触る	両手、胴体だけでムード表現
	無意識的	背伸び			両手、胴体だけでムード表現		
	非会話	姿勢変更	上体を前に傾げる	両手、胴体だけでムード表現			
		自己親密	髪を触る	行動から推定されたムードを表現する動作に翻訳			
		無意識的	背伸び	行動から推定されたムードを表現する動作に翻訳			
	声	情動	会話	高揚気味で他者へ話しかけ	情動表現動作に抑揚としてインクリメント		
			非会話	叫び声のような反応	情動表現動作に抑揚としてインクリメント		
ムード		会話	他者への話しかけ	声をそのまま伝える			
		非会話	独り言	声をそのまま伝える			

(1)突発的な情動が生じたフェーズ

映画の中の刺激的なシーンや意外な会話によって、鑑賞者の感情が高ぶった状態において、他者に伝えられるべきその鑑賞者の情報は「情動」である。人間が情動を最も表出する部位は顔の表情である。従って、突発的な情動が生じたフェーズにおいて、表情から得られる情動情報をロボットが表現をすることが、鑑賞者の存在感を表現することにおいて最も重要と考える。表情から推定された情動をロボット動作で表現する手法には、エクマン6つの表情(怒り、嫌悪、恐れ、喜び、悲しみ、驚き)をラバン特徴量から基本的心理尺度を求め、ラッセルの円環モデルを用いる手法[4]を応用することで、ロボット動作に「翻訳」する。各情動とロボット動作の対応を表2に示す。なお、この情動表現動作の程度をフレキシブルに表現させるために、動作や声の情報をを用いる。動作や声は感情表出度に個人差があるが、それらの程度を検出し、抑揚として情動表現動作にインクリメントする。

(2)平常的な心理状態のフェーズ

穏やかな心理状態時においては、情動ほど強い感情ではないが長期的な感情である「ムード」が鑑賞者のさりげないしぐさや姿勢変更から表出される。しぐさにおいては、自らの髪を触るといった行為や鼻をつまむといった行為によって、緊張や不安を鎮めようとする「自己親密行動」[5]や、背伸びやあくびといった疲労が起因した退

屈によって、無意識のうちに表出される「無意識的行動」がある。姿勢においては、上体を前に傾けたり、両足を後ろに引いたりすることは興味を示し、頭を下げたり、頭を片方に傾けたり、TV モニタから顔をそむけたり、片手で頭を支えたり、後ろにもたれかかって両足を伸ばしたりすることは退屈を示すとされている[6]。そのようなムードをロボット動作で表現するために、鑑賞者のしぐさや姿勢を忠実に再現することが考えられるが、ロボットの持つ動作自由度に制約があるため難しい。そこで、あらかじめそのようなしぐさや姿勢をテンプレートとしてシステムが記録しておき、それぞれのしぐさや姿勢から推定されるムードを表現するロボット動作を記録したルックアップテーブルを手法[4]を応用し、作成することで、鑑賞者とロボットの動作は異なるが、それぞれの動作から同じ感情が表現され、鑑賞者のムードを表現したロボット動作が可能になる。

(3)会話状態のフェーズ

会話をしている状態において、同意、否定、納得といった鑑賞者の「意図」が頷きや首振りといった頭部運動によって表現されている。また、会話状態は(1)情動フェーズ、(2)ムードフェーズの一方と同時に生じるフェーズでもある。鑑賞者の頭部運動はロボットの動作自由度で十分に忠実な再現が可能であり、通常は頭部運動をロボット動作に対してさらに翻訳を加える必要はない。しかし、「意図」の方向は特定の人物、或いは不特定多数の人物のいずれかであると考えられる。そのため、それぞれの意図の方向についてロボットの動作を考える必要がある。特定の人物に向けられた意図については、発話者の視線をシステムが検出し、ロボットの視線をその視線方向に向けることで意図の方向を再現する。本システムではロボットの配置をユーザの自由にするので配置について精密な計測と計算を行うようにする。不特定多数の人物に向けられた意図については、TV 画面への注視時や不明確な視線方向で発話した場合に不特定多数の人物に向けた発話と判定し、鑑賞者全員に対して話しかけるようなロボット動作で再現する。

表2. 情動とロボット動作の対応

感情	感情表現動作			ロボット動作		
	領域面積	重心	移動速度	首上下	上体左右	両手上下
喜び	大	高	中	大きく上	不動	大きく上下
驚き	中	中	速	小さく上	小さく左右	小さく上下
嫌悪	小	低	速	小さく下	大きく左右	小さく下に
怒り	中	中	中	小さく下	小さく左右	小さく上下
悲しみ	中	低	遅	小さく下	小さく左右	小さく下に
恐怖	小	中	中	小さく下	大きく左右	小さく上下

### 3. システム構成

本システムでは株式会社ユウビ造形のコミュニケーションロボット Phyno を使用し、複数のロボットがネットワークで繋がり、遠隔地の鑑賞者の様子を表現する。図1に提案システムの外観を示す。



(a) 実風景



(b) システム使用風景  
図1. システムの外観

本システムではユーザである鑑賞者と遠隔地にいる友人などの代理人であるロボットの部屋の中での配置を自由にできるようにするため、それぞれの部屋における鑑賞者の位置関係が異なる。そのような位置関係においても、全ての部屋で鑑賞者（の代理人であるロボット）の視線を再現するために、システムは鑑賞者が視線向けている物体がTV画面であるか、或いは他のロボットであるかを常に特定する。視線方向の検出にユーザのヘッドフォンに付けた赤外線カメラと天井部の不可視マーカを使っている[7][8]。

図2に制御情報の流れを示す。まず、各部屋のサーバはロボットの位置・姿勢情報を検出し、それぞれの部屋のロボットの位置関係を理解する。そして、ユーザの位置と姿勢情報を常に検出し、それらのサーバはその情報からユーザがロボットとTV画面のどちらを見ているかを判定する。その後、その判定結果は他のユーザの部屋のサーバへ送信される。各部屋のサーバは判定結果を受け取り、判定対象へのロボットの視線方向を算出し、ロボットを制御する。一方、ユーザの視線方向が判定されなかった場合は、その

ユーザの位置・姿勢情報から頭部運動の方向と大きさを算出し、傾きや首振りといった意図を示すものであったかを判定する。意図を示すと判定された場合は、その情報からユーザの頭部運動を忠実に再現する。一方、意図を示すと判定されなかった場合は、部屋の中の掛け時計のような他の部屋では表現できない対象を見たと判断し、その頭部運動は再現しない。

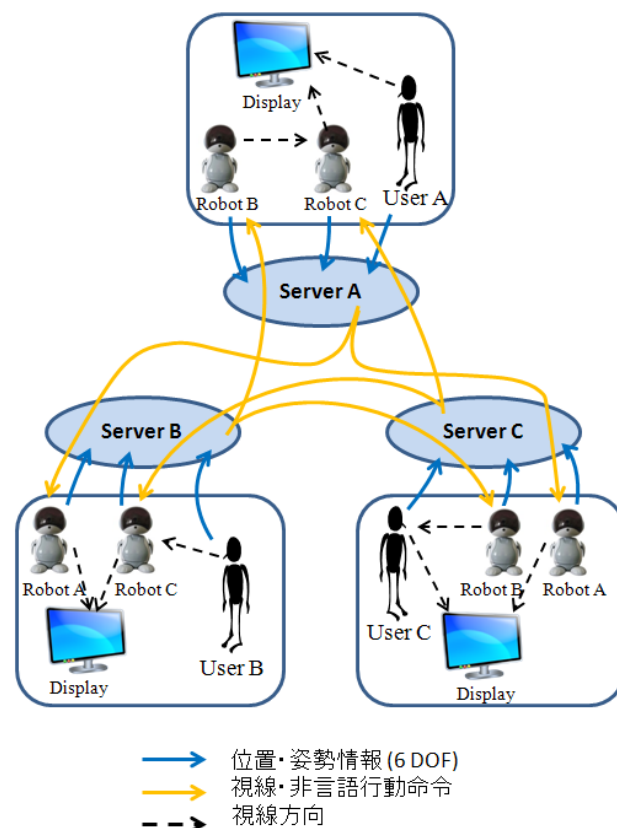


図2. ロボット制御処理

### 4. 印象評価実験

本実験では、実際に複数の鑑賞者が映画を鑑賞している様子を撮影したビデオ（以下オリジナルビデオ）と人間をロボットに置き換えたビデオ（以下ロボットビデオ）を用いた。ロボットビデオを制作するために、オリジナルビデオを分析し、前述の提案手法によって鑑賞者の様子をロボット動作に翻訳した。そして、鑑賞者の位置にロボットを設置し、ある鑑賞者がロボットと一緒に映画を鑑賞しているかのように見えるビデオに合成し、ロボットビデオを作成した。評価用ビデオの長さは1分27秒である。それぞれのビデオの画面例を図3に示す。

我々は被験者に対し、ビデオを見ながら指定したシーンにおけるある鑑賞者、或いはロボットの感情を推定し、記録することを求めた。評価を求めたシーン数は11であり、被評価者の表情が高いレベルで感情を表出しているシーンではロボットは情動表現動作を行い、傾きや視線を合わせ



るといった動作をしているシーンではロボットはやや誇張的にその動作を再現している。順序効果を考慮し、被験者によって評価するビデオの順番を交互にしている。今回の実験で被評価者となる鑑賞者にオリジナルビデオを見せ、自らの感情を評価させ、それを真値とし、ロボットビデオで評価された感情との正解率を求めた。これにより、ロボットによってどの程度正しく鑑賞者が表出した感情が伝えられたかということが分かる。また、鑑賞者本人が申告した真値とオリジナルビデオで被験者が評価した感情との正解率も求め、従来の人同士の場面から実際人はどれだけ正しく感情を読み取っていたかを調査することに加え、ロボットビデオでの正解率と比較することで、どちらがより正しく被評価者の感情を伝達できているかを調査する。さらに、被験者が推定した感情評価がオリジナルビデオとロボットビデオ間において一致しているかを $\kappa$ 検定により求める。



(a)オリジナルビデオ



(b)ロボットビデオ

図3. 印象評価に使用したビデオ画面例

## 5. 実験結果

ロボットビデオに対する評価結果を表3に示す。表3では鑑賞者がビデオに記録された自らの様子を見て付けた感情評価を真値と表しているが、この真値を被験者が評価した感情の正解率は100%や80%という高いものもあるが0%というものもあり、全体の平均では56.3%であった。表4はオリジナルビデオにおいて被験者が評価した感情の正解率を示す。表3と同じ傾向で平均は54.5%であった。

ロボットビデオとオリジナルビデオで推定された感情の一致度の結果を表5に示す。「喜び」が「無し」と誤判定

されたケースが4などとなっているが概ね一致している。 $\kappa$ 検定の結果でも、両者間に高い一致度が求められた( $\kappa$ 値 = 0.69)。これにより、表1の方針に従って制御されているロボットを第三者が見た場合でも実際の鑑賞者を見た場合と同様にロボットの様子から感情を読み取れたと言える。

表3. ロボットビデオに対する感情評価の結果

タイム	感情						正解率 (%)
	真値	a	b	c	d	e	
Scene1	驚	驚	喜	驚	驚	驚	80
Scene2	喜	驚	無	喜	喜	無	40
Scene3	喜	驚	驚	喜	驚	驚	80
Scene4	喜	喜	喜	喜	喜	喜	100
Scene5	無	無	無	無	恐	無	80
Scene6	喜	喜	喜	無	喜	驚	60
Scene7	恐	驚	驚	驚	驚	驚	0
Scene8	驚	無	無	恐	恐	無	0
Scene9	喜	無	無	悲	恐	恐	0
Scene10	嫌	嫌	嫌	嫌	嫌	恐	80
Scene11	嫌	嫌	嫌	嫌	嫌	嫌	100
Total							56.3

表4. オリジナルビデオに対する感情評価の結果

タイム	感情						正解率 (%)
	真値	a	b	c	d	e	
Scene1	驚	驚	驚	驚	驚	驚	100
Scene2	喜	無	無	無	驚	無	0
Scene3	喜	驚	喜	喜	喜	喜	80
Scene4	喜	喜	喜	喜	喜	喜	100
Scene5	無	無	喜	無	喜	恐	40
Scene6	喜	喜	喜	喜	喜	無	80
Scene7	恐	驚	驚	驚	驚	驚	0
Scene8	驚	驚	無	喜	無	無	20
Scene9	喜	喜	喜	喜	恐	恐	60
Scene10	嫌	嫌	嫌	恐	恐	嫌	60
Scene11	嫌	嫌	恐	嫌	恐	嫌	60
Total							54.5

表5. ロボット / オリジナルビデオ間の一致度

感情評価	オリジナルビデオへの評価							Total	Rate	
	喜び	驚き	嫌悪	恐れ	悲しみ	怒り	無し			
ロボットビデオへの評価	喜び	12	1	0	1	1	0	4	19	0.345
	驚き	1	11	0	1	0	0	0	13	0.236
	嫌悪	0	0	6	0	0	0	0	6	0.109
	恐れ	0	0	0	4	0	0	0	7	0.127
	悲しみ	0	0	0	0	0	0	0	0	0
	怒り	0	0	0	0	0	0	0	0	0
	無し	0	0	0	0	0	0	8	10	0.182
Total	15	12	9	6	1	0	12	55	1	
Rate	0.27	0.22	0.16	0.11	0.02	0	0.218	1		

## 6. 考察

ロボットビデオとオリジナルビデオで推定された感情

を $\kappa$ 検定することによって高い一致度が求められた。このことから、被験者はロボットの様子から実際の鑑賞者と同様に感情を読み取ることができたとと言える。しかし、ロボットビデオ、オリジナルビデオで推定された感情の正解率の平均値は高い結果が得られなかった。この結果はそもそも人間同士のコミュニケーションにおいて感情表出者が考える自らの感情と感情を読み取る受け手が考える感情に大きなギャップがあることによるものであると考えられる。そのような表出者と受け手の感情のギャップを埋めることが我々の提案するシステムに求められる。今回の実験において、鑑賞者の感情を表現するロボット動作は Scene3, 4, 6, 7, 10, 11 であった。これらのシーンに共通することは鑑賞者の表情が高いレベルで表出されていたことである。そこで、真値と感情を表現したロボット動作のシーンで推定された感情の正解率を求めると 70%という高い結果が得られた。一方、オリジナルビデオとの比較では 63.3%というロボットビデオより低い結果が得られた。このことから、感情を表現するロボット動作は適切に鑑賞者の感情を伝達できたとと言える。また Scene7 でロボットビデオ、オリジナルビデオ共に被験者全員が推定した感情は驚きであったのに対し、真値は恐れであった。ラッセルの円環モデルにおける驚きと恐れは、恐れが驚きより不快レベルで少し高い程度で、互いに近い座標に位置しているため区別が難しかったと考える。

また、被評価者が喜びを表情や声で表出しているが、ロボットが情動表現動作で再現していないシーン (Scene12) も両ビデオ条件で評価させた。その結果、ロボットビデオでの感情評価は全て無感情であった。一方、オリジナルビデオでの感情評価は 2/5 人が喜びと感じていた。このことから、感情を表現したロボット動作がなければ、声による情報だけで感情を推定することが難しいことが分かる。また、今回の評価ビデオでは被評価者の顔や動作が見やすい位置での視点で感情の評価を行わせたが、実際に映画を楽しむ際は、他の鑑賞者は自らの周囲にいるため、表情や動作をより気付きにくい視点で相手の感情を読み取ることになる。その場合、今回得た結果より低い結果になることが予想される。しかし、ロボット動作においては、やや誇張的に感情を表現するので検知されやすく、実際の鑑賞者を対象としたときより高く感情を推定されることが期待される。実際の鑑賞者の視点による感情評価実験は今後検討していく。

## 7. 結言

本稿では、テレプレゼンスシステムのエンターテインメント的応用の1つとしてロボットが遠隔地の友人らの感情やリアクションを表現することによって、同じ部屋で友人と一緒に映画を楽しんでいるかのような臨場感を与えるシステムを提案した。ロボットの制御方法として鑑賞者の存

在感を表現するために重要なファクターをフェーズによってロボット動作に翻訳する手法を提案した。実際に複数の鑑賞者が映画を楽しんでいる様子を記録したビデオとその鑑賞者をロボットで表現した合成ビデオを被験者に見せ、ある鑑賞者とロボットの感情を評価させた。第三者の視点で鑑賞者の様子を見たときとロボットの様子を見たときに読み取る感情は高い値で一致していた。また、被評価者が自らに付けた感情値と両ビデオで評価された感情の比較では、被評価者の感情が表情に顕著に表れているシーンについて、70%という高い正解率が得られた。

今後、実際の鑑賞者の視点でロボットの情動表現動作から適切に感情を読み取れるかを検証する実験を行い、本稿で提案したシステムをより効果のあるものにしていく。

## 参考文献

- 1) 大塚和弘, 熊野史朗, 三上弾, 松田昌史, 大和淳司, MM-Space: 動的投影を用いた頭部運動の物理的補強表現に基づく会話場再構成, インタラクション 2012, (2012)
- 2) 柏原忠和, 大澤博隆, 篠沢一彦, 今井倫太, ウェアラブル・アパタ TEROOS を用いたフィールドテストとその分析, インタラクション 2012, (2012)
- 3) 長谷川孔明, 中内靖, テレプレゼンスロボットの身ぶりが発話交替に与える影響について, HAI2012, 2B-2, (2012)
- 4) 田辺奈々, 前田陽一郎, ペットロボットを用いた身体的特徴量に基づくフェージ情動推論, 第7回日本感性工学会大会予稿集, pp. 122, (2005)
- 5) 渋谷昌三, 『面白いほどよくわかる! 他人の心理学』, 西東社, (2012)
- 6) Peter Bull, 『Posture and Gesture』, Elsevier Science & Technology Books, (1987)
- 7) 小吹健太郎, 辻剛正, 上田博唯, スマートホームにおけるロボットの視線制御方法の提案, マルチメディア・仮想環境基礎 110(457), pp. 31-36, (2012)
- 8) 中里裕介, 榊原誠之, 横矢直和, 不可視マーカを用いた位置・姿勢認識のための環境構築とユーザ位置・姿勢認識システム, 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 2, pp. 257-266, (2008)