

# 情動的共感表現に向けた 多層模倣システムの開発とアンドロイド実装

## Development of a Multi-layer Mimicry System and Its Implementation on an Android for Emotional Empathy Expression

木下 邑輝<sup>1</sup> 平山 聖輝<sup>1</sup> 大原 颯太<sup>1</sup>

林里奈<sup>2</sup> 石原尚<sup>1\*</sup>

Yuki Kinoshita<sup>1</sup> Ibuki Hirayama<sup>1</sup> Sota Ohara<sup>1</sup>

Rina Hyashi<sup>2</sup> Hisashi Ishihara<sup>1</sup>

<sup>1</sup> 大阪大学大学院 工学研究科

<sup>1</sup> Graduate School of Engineering, The University of Osaka

<sup>2</sup> ほんわか研究所

<sup>2</sup> Honwaka Laboratory

**Abstract:** Facial mimicry by robots has an important role on conveying emotional empathy in human-robot interaction. However, continuous long-term postural mimicry often provides awkward impressions due to its conspicuousness. This study proposes a multi-layer mimicry system that integrates Superficial Postural Mimicry and Deep Mood Mimicry. Superficial Postural Mimicry serves as apparent postural mimicry only in response to salient human facial actions, while Deep Mood Mimicry serves as inconspicuous but continuous empathy expression in which robots behaves based on the mood copied from the estimated mood state of humans. These two mimicries were dynamically generated by individual system modules and integrated effectively. The proposed system was installed on an android robot face to test its performance.

### 1 緒言

人間同士のインタラクションにおいて、相手の仕草や表情を模倣することは、円滑なコミュニケーションや、親近感・信頼感の構築に寄与する。この模倣行動は情動的共感の伝達手段の一つであり、相手の感情状態を理解および共有していることを相手に示す重要な機構である [1]。

この観点から、ロボットの模倣に関する研究は数多く行われてきた。その多くは人間の観測された動作や感情状態に対して、あらかじめ定義された一時的な目立つ反応を対応づけてその状態を再現する枠組みに基づいている [2, 3, 4]。たとえば、笑顔が検出された場合にロボットも笑顔を提示する模倣や、喜びと推定された感情状態に対して対応する喜びの表情を生成する模倣などが挙げられる。

しかし、このような模倣では持続的に情動的共感を伝達することは容易ではない。この模倣は顕著な反応



図 1: Appearance of the female-type android robot “A-Lab Android ST”

により共感を明示的に示しやすい一方で、長時間継続させると意図的であると知覚され、わざとらしさを感じさせて居心地の悪さの原因になる恐れがある。ゆえに、短時間において共感を強調する手段として有効であるものの、持続的な共感表現としては限界を有する。本研究では、相手の形態を再現する模倣を表層的模倣

\*連絡先： 大阪大学工学研究科  
〒 565-0871 大阪府吹田市山田丘 2 - 1  
E-mail: ishihara@mech.eng.osaka-u.ac.jp

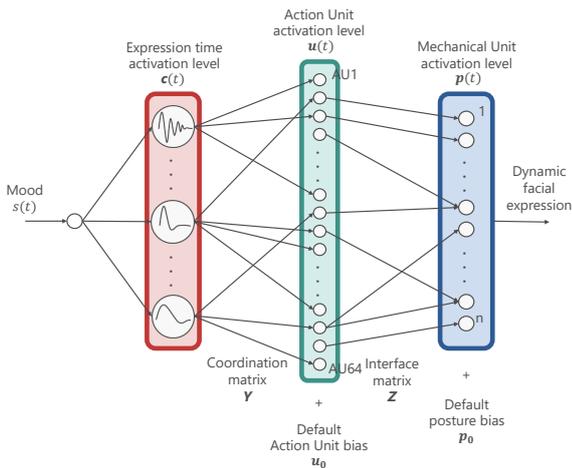


図 2: Overview of AMESS.

と呼ぶ。

そこで本研究では、相手の気分をロボット自身の気分として反映し、その気分をロボット独自の動きで表現する深層的模倣を提案する。本システムは、相手の感情状態との一致が一連の動作の背後に透けて見える、すなわち、気分が同調しているかのように知覚されることを特徴とする。この反応は目立ちにくく、継続的に提示されてもわざとらしさを生じにくい表現となることが期待される。一方で、顕著な反応を伴わないため、共感を示されていることが必ずしも明示的に知覚されない恐れがある。

以上の背景を踏まえ、本研究では、一長一短のある表層的模倣および深層的模倣を相補的に組み合わせた多層模倣システムの構築を目的とする。本システムにより、わざとらしさを感じさせることなく、共感を示していることが明示的に、また持続的に伝えることを目指す。このシステムは、感情の幅広い表現が可能なアンドロイドロボット（以下、アンドロイド）に実装し、その有効性を検証する。

## 2 模倣システムの構築

### 2.1 採用するアンドロイド

多層模倣では、顕著な形態変化による明示的な表層的模倣から、相手の気分をロボットが動作全体の傾向として表現する深層的模倣までを同一媒体上で実現する必要がある。そこで本研究では、高い自由度を有する顔アクチュエータを備えた「Android ST（株式会社 A-Lab 製）」を用いた。図 1 に、そのアンドロイドを示す。本アンドロイドは顔に 13、首に 3 の動作箇所を有し、顔および頭部の複数の動作部位の配置を制御可

能である。微細な運動変化から大きな表情変化まで表現できるため、多層模倣の実装に適している。

### 2.2 アンドロイドのための表情表出システム

各模倣システムでは、アンドロイドにおける形態レベルあるいは感情レベルでの運動表現の生成ができる表情表出システムが必要である。そこで本研究では、石原らによって提案されたシステム「Automatic Mood Expression Synthesis System（以下、AMESS）」を採用する [5]。AMESS のシステム概略を図 2 に示す。AMESS は、連続値として表現される気分から時間的および空間的な特徴を有した顔運動で気分表現するシステムであり、感情状態を一時的な反応としてではなく、連続的な顔運動の中に内在化された形で表出することが可能なシステムである。たとえば、欠伸や呼吸、瞬きなど、生理状態を反映しつつ生じる無意識的動作の現れ方が、気分に応じてどう変わるかがあらかじめ定義されており、眠たい気分を設定すると、大きな欠伸が頻繁になり、呼吸が浅くゆっくりになり、頭が揺れたりし、その動きの中で眠気が表現される。この特性は、本研究で導入する、気分が相手と同調しているように知覚される深層的模倣の実現に適している。さらに、AMESS はアクチュエータ指令までを含む枠組みであるため、アクチュエータ指令を直接制御することで表層的模倣にも活用できる。

### 2.3 人顔運動の取得と前処理

模倣システム全体で用いる人顔運動データを取得するため、本研究では人とロボットのインタラクションを想定した人表情動画を撮影した。本研究の模倣システムは、情動的共感における模倣を動的な過程として捉えることを目的としているため、動画には豊富な種類の人顔運動による形態変化が十分に含まれているとともに、気分の切り替わる過程が明確に観測される必要がある。そこで本研究では、眠気を帯びた低覚醒状態から、興奮傾向の高覚醒状態へと移行するシチュエーションを設定し、気分の連続的な変化と、それに対応して現れる多様な顔運動が観測されるようにした。実際に撮影した動画のスナップショットを図 3 に示す。

人顔運動の計測には OpenFace [6] を用い、FACS [7] に基づく人顔運動の動作単位である Action Unit (AU) の活動レベル 17 種類、および頭部回転運動 (Yaw, Pitch, Roll) の回転量 3 種類を、それぞれ時系列データとして取得した。AU については、OpenFace が出力する活性判定情報 (0 または 1) も併用し、AU の活動レベルベクトルとのアダマール積による統合処理を行うことで、AU 値の誤検出を抑制した。さらに、各 AU 値お

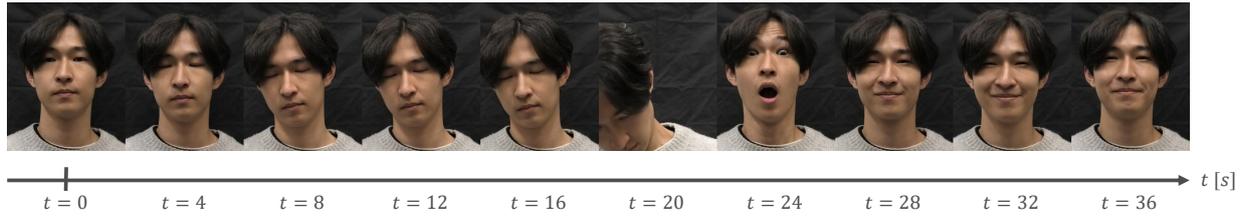


図 3: Snapshots of the human video used for facial motion acquisition, showing transitions between low-arousal and high-arousal states.

表 1: Modified correspondence between Action Units obtained from OpenFace and those used in AMESS for Superficial Mimicry. Only AUs whose correspondence was adjusted are listed.

OpenFace AU	AMESS AU	Description
AU7	AU44	Lid tightener mapped to blink-related eyelid motion
AU26	AU27	Jaw opening motion aligned with AMESS mouth opening definition
AU25	–	Excluded due to redundancy with jaw-related AUs
AU51	AU52	Head rotation mapped with left-right mirroring
AU52	AU51	Head rotation mapped with left-right mirroring
AU53	AU54	Head tilt mapped with up-down mirroring
AU54	AU53	Head tilt mapped with up-down mirroring
AU55	AU56	Head roll mapped with left-right mirroring
AU56	AU55	Head roll mapped with left-right mirroring

よび頭部回転量に対して、事前に計測しておいた操作者ごとに決定した最大値に基づく正規化を行い、0~1のスケールに変換した。この処理により、頭部回転量を AU51~56 の活動レベルとして扱うことも可能となり、OpenFace によって得られる 17 種類の AU と合わせて、計 23 種類の AU 値を同一の枠組みで入力特徴量として構成した。また、本正規化は、顔や頭部の可動域の違いといった操作者間の個人差の影響を抑制することも目的としている。最後に、これらの時系列データに含まれる短周期的ノイズを低減するため、指数移動平均による平滑化を施し、模倣システム全体の入力として用いる安定した人顔運動特徴量を構成した。

## 2.4 表層的模倣システムの構築

2.3 節で得られた人顔運動特徴量である AU 値を、AMESS に含まれるアンドロイドのための制御用 AU ベクトルへ対応づけた。ここで、相手の顔動作を模倣しているという印象がより効果的になるような対応の調整を実施した。たとえば、頭部回転運動に関する AU は、人間と対面するインタラクションを想定しているため、相手と鏡像関係になるように対応付けた。調整を行った AU を表 1 に示す。

## 2.5 深層的模倣システムの構築

深層的模倣では、相手の人顔運動から推定された気分をアンドロイドの気分として反映し、AMESS を用いて気分に基づく表情生成を行った。AMESS の入力となる気分は覚醒度の 1 次元モデルとして定義されているため、本研究においても気分は覚醒度の 1 次元で連続的に変化するものとして扱う。人顔運動から覚醒度を推定するため、先行研究において覚醒度との関連が指摘されている特徴 [8, 9, 10, 11] に基づき、2.3 節で得られた AU 情報から構成可能な 4 つの特徴量を設計した。これらの特徴量を表 2 に示す。

この 4 つの特徴量を線形結合したものを、相手の覚醒度の推定値とした。各重みは、先述した人の動画とは別の、人が意図的に演技した高覚醒あるいは低覚醒状態を維持している事前動画（各約 30 秒）それぞれに対して、2.3 節の処理を施した人顔運動データに基づきあらかじめ決定した。具体的には、各特徴量を単独で用いた場合に算出される覚醒度の平均値が、高覚醒状態では高覚醒に寄与する特徴量で正の値（約 0.5）、低覚醒状態では低覚醒に寄与する特徴量で負の値（約 -0.5）となるよう、小数第 1 位のオーダーで重みを調整した。

さらに、推定された覚醒度とアンドロイドの現在の覚醒度との差に基づいて気分を更新する動的モデルを

表 2: Features used for arousal estimation based on facial and head movements. Each feature is derived from OpenFace outputs and reflects activity patterns associated with high or low arousal states.

Feature	Description	Related AU / Quantity	Arousal tendency
F1	Temporal change of facial AU activity	AU2, AU6, AU7, AU9, AU10, AU12, AU17	High
F2	Temporal change of head pitch rotation	AU53 / AU54 (Head pitch)	High
F3	Ratio of eye-blink activity time	AU45 (Blink)	Low
F4	Ratio of head roll exceeding threshold	AU55 / AU56 (Head roll)	Low

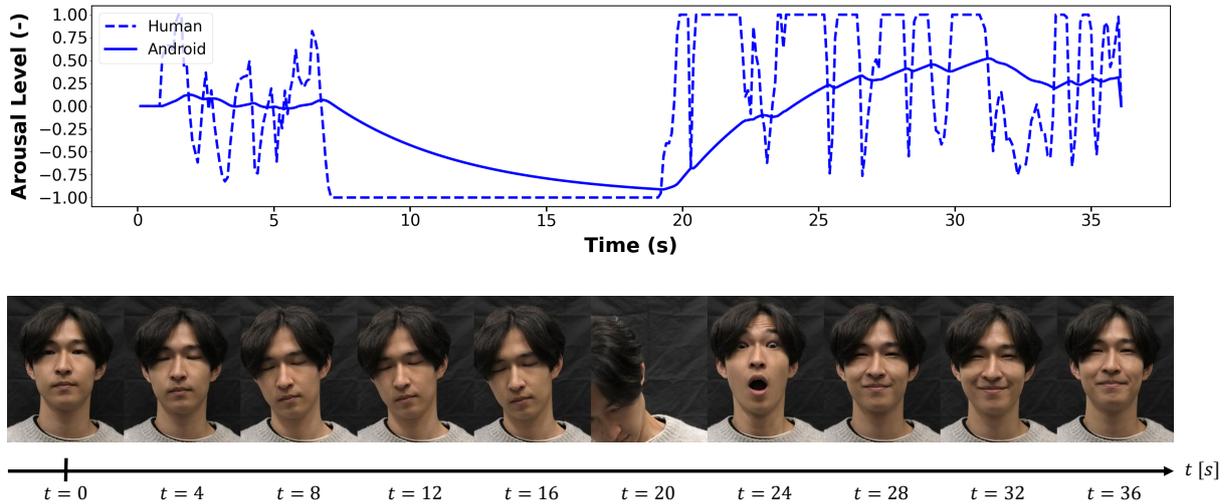


図 4: Example of arousal estimation and mood update in Deep Mood Mimicry.

構築した。その更新則を式 (1) に示す。

$$s[j] = s[j - 1] + K_p(A[j] - s[j - 1]) \quad (1)$$

ここで、 $s[j]$  は  $j$  フレーム目におけるアンドロイドの覚醒度、 $A[j]$  は人顔運動から推定された相手の覚醒度、 $K_p$  は比例ゲインである。 $K_p$  が過度に大きい場合には気分変化が唐突となり違和感やわざとらしさを生じさせる一方、小さすぎる場合には気分同調が知覚されにくくなる恐れがある。 $K_p$  を決定するにあたって、本研究では、2.3 節で撮影した人表情動画に対して覚醒度推定を行い、その推定値を式 (1) に基づいてアンドロイドの気分更新に用いた。このとき、アンドロイドに実装した際の人表情動画に対する挙動を観察しながら、気分変化が唐突と知覚されず、かつ気分同調が認識可能となるよう、試行錯誤的に調整を行った。その結果に基づいて、 $K_p = 0.02$  を採用した。この更新により、推定値に含まれるノイズの影響を抑制しつつ、過度な即時追従を避けた緩やかな気分同調が良好に実現された。人表情動画に対する覚醒度推定結果およびアンドロイドの気分変化を図 4 に示す。

## 2.6 多層模倣システムの構築

多層模倣では、表層的模倣による明示的な反応と、深層的模倣によるわざとらしくない気分同調の反応を両立する表情生成を目指す。両者を静的に重ね合わせた場合、目立つ表層的模倣が知覚的に優勢となり、真似されているというわざとらしい印象を与える恐れがある。一方で、表層的模倣の表出強度を抑制すると、今度はその明示性が低下し、共感的応答として十分に知覚されなくなる。このように、ある時間フレームにおいて空間的特徴を混合する静的な統合では、表層的模倣の明示性と深層的模倣のわざとらしさを同時に満たすことが困難である。

そこで本研究では、両者を同一時間フレームで統合するのではなく、時間的に役割分担させる動的な組み合わせを採用した。具体的には、深層的模倣は常時動作させる一方、表層的模倣は顕著な人顔運動が観測された場合にのみ活性化させ、深層的模倣の出力に加算するものとした。アンドロイドの表情は、この加算結果を目標値として更新される。このとき、表層的模倣が一連の動作に滑らかに重畳されるように、出力は目標値に対して唐突に切り替えるのではなく、AU 値の最大変化速度を制限した遷移過程を経て更新する。表層的模倣の活性化が終了した場合にも、同様の速度制

表 3: Final design parameters for Multi Layer Mimicry and their functional roles.

Parameter	Symbol	Value	Primary role
Activation threshold based on human AU change rate	$\theta$	80 %	Controlling sensitivity of Superficial Mimicry activation
Minimum duration of Superficial Mimicry	$T_{min}$	0.5 s	Ensuring perceptual coherence of Superficial Mimicry as a discrete response
Maximum AU change rate at layer switching	$\epsilon$	1.0 s <sup>-1</sup>	Maintaining continuity of Deep Mood Mimicry while enabling salient Superficial Mimicry

限の下で深層的模倣の出力へと遷移する。この仕組みにより、深層的模倣によるわざとらしくない共感表現と、表層的模倣による明示的な共感表現のなめらかな相互遷移の実現を図る。

深層的模倣に対して表層的模倣を知覚可能かつ滑らかな形で重畳するため、本研究では、多層模倣の振る舞いを規定する主要な制御要因として「AU 変化率に基づく表層的模倣の活性化感度」、「表層的模倣が活性化する最低継続時間」、「模倣出力の切り替え時における AU 値の最大変化速度」という 3 つのパラメータを導入する。

AU 変化率に基づく表層的模倣の活性化感度は、人顔運動に対して表層的模倣が活性化するか否かを決定するパラメータであり、表層的模倣の発生頻度を規定する。感度を大きく設定した場合、微小な AU 変動に対しても表層的模倣が活性化しやすくなり、結果として落ち着きのない、わざとらしい印象を与える可能性がある。一方で、小さく設定した場合には、顕著な人顔運動に対しても表層的模倣が活性化されにくくなり、明示的な共感表現として十分に知覚されない恐れがある。

表層的模倣が活性化する最低継続時間は、一度活性化した表層的模倣が、どの程度の時間継続されるかを決定するパラメータであり、表層的模倣のわざとらしさや知覚性に影響を与える。最低継続時間を長く設定した場合、真似されているというわざとらしい印象を抱かせる可能性がある。一方で、短く設定した場合には、表層的模倣が短時間で消失するため、共感的反応として十分に認識されない恐れがある。

模倣出力の切り替え時における AU 値の最大変化速度は、深層的模倣の出力と表層的模倣活性時の加算出力間の遷移の最大速度を決定するパラメータであり、模倣挙動全体の滑らかさを規定する。最大変化速度を大きく設定した場合、模倣出力が急激に変化し、表層的模倣が深層的模倣から不連続に切り替わったように知覚され、違和感を生じさせる可能性がある。一方で、小さく設定した場合には、遷移が過度に緩やかとなり、表層的模倣が顕在化せず、明示的な模倣表現として機能しなくなる恐れがある。

これらの印象特性を考慮し、アンドロイドによる多

層模倣の挙動を確認した上で、深層的模倣に重畳される表層的模倣の知覚性および表出の滑らかさを観察しながら、各パラメータを試行錯誤的に決定した。ここで、感度は、AU 変化率に基づく閾値に基づいて定義した値である。AMESS における AU 値は [0, 1] の範囲で正規化されており、本研究の制御周期は 0.1 s と設定しているため、AU 変化率の理論上の最大値は 10 s<sup>-1</sup> となる。本研究では、この最大値に対する変化率の閾値の比を感度として扱った。また、模倣出力の切り替え時における AU 値の最大変化速度は、本研究で用いる AU 値の出力レンジ [0, 1] での速度として設定している。表 3 に、最終的に採用した多層模倣システムの設計パラメータを示す。

## 2.7 実装結果と考察

アンドロイドに実装した人表情動画に対する各模倣システムの挙動を図 5 に示す。図 5 は、人表情動画と、それに対する表層的模倣、深層的模倣および多層模倣それぞれのスナップショットである。それぞれの模倣挙動における表出の違いと、多層模倣において表層的模倣と深層的模倣がどのように統合されているかを視覚的に確認することができる。

多層模倣では、表層的模倣による明示的な反応と、深層的模倣による緩やかな気分同調とが、相補に統合することができた。表層的模倣は、人顔運動が顕著に変化したタイミングにおいてのみ一時的に活性化し、笑顔や驚きといった大きな顔運動変化に対して明確な共感的反応を提示する。一方で、深層的模倣は、表層的模倣が活性化していない区間においても継続的に動作し、人の気分変化に対応した緩やかな気分同調表現を生成している。この結果、多層模倣では、真似されているというわざとらしい印象を抑制しつつ、共感的応答としての明示性を保持した表情生成が実現されたと考えられる。

一方で、本研究の多層模倣システムには課題も残されている。まず、本研究では導入したパラメータを人表情動画およびアンドロイド実装時の挙動観察に基づ

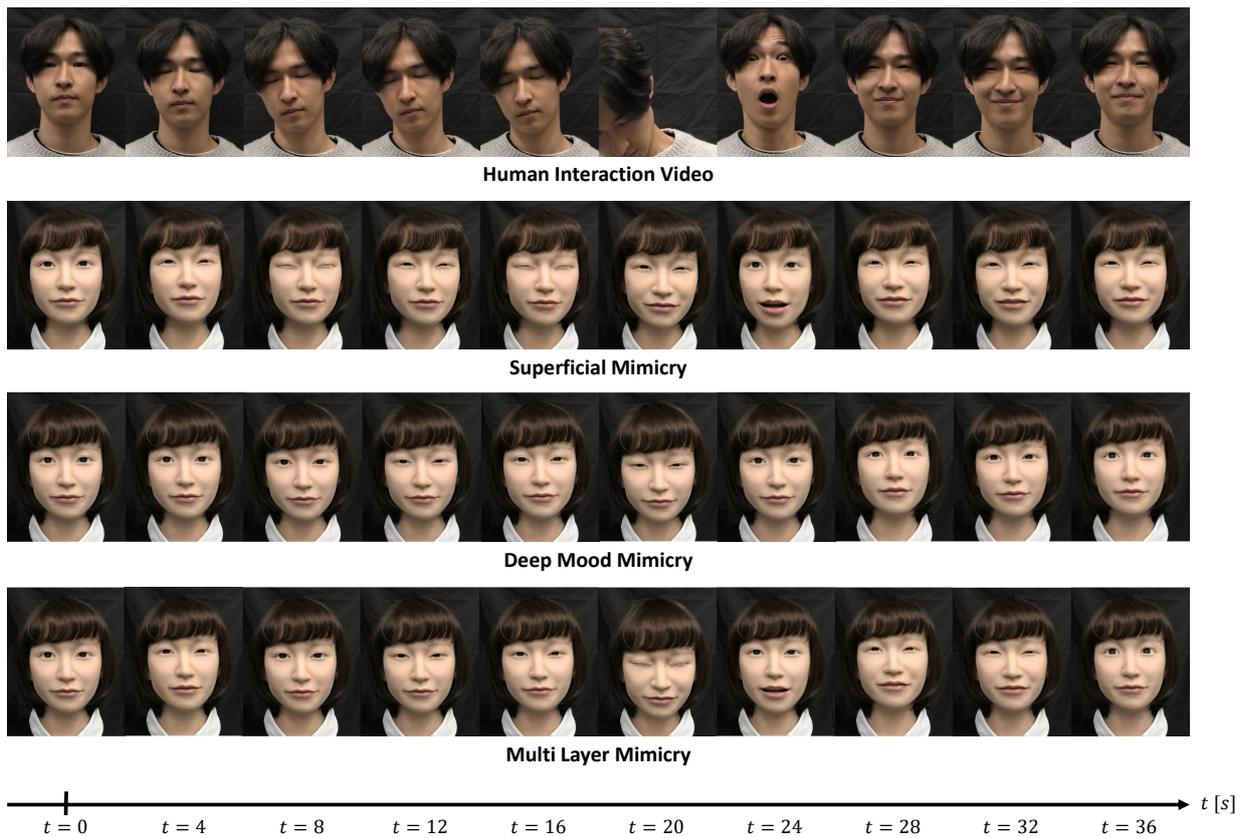


図 5: Implemented Mimicry on Android ( $t = 0 \sim 36$  [s]).

いて決定しており，観察者による印象評価やインタラクション実験を通じた定量的検証は行っていない．今後は，第三者評価や対話実験を通じて，共感的印象やわざとらしさの低減効果を定量的に評価する必要がある．また，本研究では覚醒度の1次元モデルに基づいて深層的模倣を構成しているが，感情価の気分次元を導入することで，より多様な気分同調表現が可能になるとも考えられる．

### 3 結言

本研究では，情動的共感表現を持続的かつわざとらしくなく実現することを目的として，表層的模倣と深層的模倣を組み合わせた多層模倣システムを提案し，アンドロイドへの実装を行った．表層的模倣では，人顔運動の顕著な変化に対する明示的な共感反応を生成し，深層的模倣では，人の気分変化に対応した緩やかな気分同調表現を生成した．これらを動的に相補統合することで，共感の明示性を保持しつつ，真似されているというわざとらしい印象を抑制した表情生成の可能性を示した．

一方で，本研究ではパラメータの決定や有効性の検証を主に挙動観察に基づいて行っており，観察者評価やインタラクション実験による定量的評価は今後の課題である．また，本研究では覚醒度の1次元モデルに基づく深層的模倣を扱ったが，感情価の気分次元を導入することで，より多様な共感表現へ拡張できるとも考えられる．

### 謝辞

本研究を進めるにあたり，JSPS 科研費 JP25H01236, JP24K00851, JP25K03425 の助成，および，立石科学技術振興財団研究助成 (A) の支援を受けました．ここに記し，深く感謝の意を表します．

### 参考文献

- [1] 福田正治: 共感と感情コミュニケーション (I): 共感の基礎, University of Toyama 博士論文 (2008)

- [2] Riek, L. D. and Robinson, P.: Real-time empathy: Facial mimicry on a robot, *Proc. of Workshop on affective interaction in natural environments (AFFINE) at ICMI 08* (2008)
- [3] Hegel, F., Spexard, T., Wrede, B., Horstmann, G., and Vogt, T.: Playing a different imitation game: Interaction with an Empathic Android Robot, *Proc. of the 6th IEEE-RAS International Conference on Humanoid Robots*, pp. 56–61 (2006)
- [4] 山崎真由香, 港隆史, 飯尾尊優: アクションユニットレベルでの表情模倣と記号レベルでの表情模倣における共感性の比較, HAI シンポジウム 2025 予稿集, P2-1 (2025)
- [5] Ishihara, H., Hayashi, R., Lavieille, F., Okamoto, K., Okuyama, T., and Osuka, K.: Automatic Generation of Dynamic Arousal Expression Based on Decaying Wave Synthesis for Robot Faces, *Journal of Robotics and Mechatronics*, Vol. 36, No. 6, pp. 1481–1494 (2024)
- [6] Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P.: OpenFace 2.0: Facial Behavior Analysis Toolkit, *Proc. of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 59–66 (2018)
- [7] Ekman, P. and Friesen, W. V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement, *Consulting Psychologists Press* (1978)
- [8] Zhang, J., Sato, W., Kawamura, N., Shimokawa, K., Tang, B., and Nakamura, Y.: Sensing emotional valence and arousal dynamics through automated facial action unit analysis, *Scientific Reports*, Vol. 14, No. 1, p. 19563 (2024)
- [9] Li, B. J., Bailenson, J. N., Pines, A., Greenleaf, W. J., and Williams, L. M.: A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures, *Frontiers in Psychology*, Vol. 8, p. 2116 (2017)
- [10] Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., and Movellan, J.: Drowsy driver detection through facial movement analysis, *Proc. of International Workshop on Human-Computer Interaction*, pp. 6–18 (2007)
- [11] Abe, T.: PERCLOS-based technologies for detecting drowsiness: current evidence and future directions, *Sleep Advances*, Vol. 4, No. 1, p. zpad006 (2023)