一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
CNR2025-2(2025-08)

表情の時間的変化に基づくオンライン会議参加者の理解度推定

佐藤　矯汰郎†　　齊藤 寛己†　　大和 淳司†

†工学院大学大学院工学研究科情報専攻 〒192-0015 東京都八王子市中野町 2665-1
E-mail: †em24025@ns.kogakuin.ac.jp,em24024@ns.kogakuin.ac.jp , yamato@cc.kogakuin.ac.jp

あらまし　新型コロナウイルスの感染拡大に伴い，オンライン会議システムの利用が急速に普及した．しかし，オフライン会議と比べて発言のタイミングや相手の意図が把握しづらく，内容理解度が低下するという課題が指摘されている．本研究では，会議参加者の理解度を非言語的な情報から可視化することを目的として，オンライン会議中の映像から顔画像データを取得し，理解度を推定する手法を検討した．従来は Action Unit を用いた推定が主に行われていたが，本研究では新たに頭部の回転運動（pitch, yaw, roll）に着目し，その時間的変化が理解度推定に有効かを検証した．オンライン会議実験により収集した映像データをもとに，Support Vector Machine および Random Forest による 2 値分類を行った結果，特に Random Forest では理解／不理解ともに高い F1 スコアを示し，頭部運動が理解度推定の有力な手がかりとなる可能性が示唆された．

キーワード　オンライン会議，顔画像，理解度，推定，時系列

Estimating Comprehension in Online Meetings Based on Temporal Changes in Facial Expressions

Kyotaro SATO†　　Hiroki SAITO†　and　Junji YAMATO†

†Information Program, Graduate School of Engineering, Kogakuin University　2665-1 Nakano-machi, Hachioji, Tokyo, 192-0015 Japan
E-mail: †em24025@ns.kogakuin.ac.jp,em24024@ns.kogakuin.ac.jp , yamato@cc.kogakuin.ac.jp

Abstract　With the spread of the novel coronavirus, the use of online meeting systems has rapidly increased. However, compared to face-to-face meetings, it is often more difficult to grasp the timing of remarks and the intentions of other participants, potentially leading to a reduced understanding of the content. This study aims to visualize participants' levels of understanding using non-verbal cues. Specifically, we extracted facial image data from recorded footage of online meetings and investigated methods for estimating comprehension levels. While prior research has primarily relied on facial Action Units for such estimation, this study focuses on head rotation movements (pitch, yaw, and roll) and evaluates the effectiveness of their temporal dynamics in estimating understanding. Using video data collected from controlled online meeting experiments, we conducted binary classification using Support Vector Machines and Random Forest classifiers. The results demonstrated that the Random Forest classifier achieved high F1 scores for both "understanding" and "non-understanding" classes, suggesting that head movements can serve as a valuable indicator for assessing participants' comprehension levels.

Keywords　Online conference, Facial image, Comprehension, estimation, Time series

1. はじめに

コロナウイルス感染拡大以前から，オンライン会議システムは一部で活用されていたが，感染拡大を契機にその利用は急速に拡大した．総務省の調査[1]によれば，テレワークを導入している企業の割合は 2020 年に 47.5%と急増し，その後 2021 年・2022 年も約 5 割の水準で維持されている．テレワークの導入形態としては，在宅勤務が 91.3%と最も一般的であり，次いでモバイルワーク（27.0%），サテライトオフィス勤務（12.9%）が続いている．

一方，全国調査の結果[2]では，64.7%の回答者が「導入しておらず，導入予定もない」と回答しており，依然としてテレワークが十分に浸透していない実態も明らかとなっている．導入企業の約 45.8%は 2020 年 4〜6 月にテレワークを導入しており，これは感染拡大が本格化したタイミングと一致している．加えて，導入企業のうち 71.9%が今後もテレワークを活用すると回答しており，一定の定着傾向も確認されている．

オンライン会議は，地理的な制約を受けずに実施できるという利点がある一方で，いくつかの課題も指摘されている．宮内ら［2］の研究では，オンライン会議においては発言のタイミングを逃しやすく，また相手の思考が伝わりにくいため，オフライン会議と比べて参加者の理解度が低下する傾向があると報告されてい

る．特に，7 人程度の中規模会議では，会議の流れを把握するために高い集中力を要したという指摘がある．これは，多人数が参加する場面では，複数の参加者の表情や身振りを同時に把握することが困難となり，結果として会議の円滑な進行を妨げる可能性があるためである．

本研究では，オンライン会議の円滑な進行を支援することを目的として，オンライン会議の内容理解度の可視化を目指す．オンライン会議参加者の会議内容の理解度推定を行う．まずそのために，顔画像の特徴選択について実験的に検討した．本稿では，頭部運動情報が理解度推定において有効な手がかりとなるか検討を広げた．

2. 関連研究

Miao ら[4]は，講義視聴中の学生の顔表情に基づいて，無関係な聴覚刺激に対する反応時間を機械学習により予測する手法を提案した．反応時間を注意状態の客観的指標として用いることで，エンゲージメントの定量的な推定を実現している．同研究では，顔表情の動きを示す Action Unit（以下，AU）[5]および頭部姿勢を OpenFace[6]により抽出し，LightGBM モデルを用いて反応時間を回帰的に予測した．また，SHAP を用いた特徴量重要度の分析により，特定の AU が反応時間の予測に大きく寄与していることが示された．さらに，顔表情と反応時間の関係性には個人差による有意な違いが見られたことから，汎用モデルによる予測が困難であり，個人差を考慮したモデル設計の必要性が指摘されている．

著者らは[7]，オンライン会議における理解度推定のために，個人差を考慮した有効な特徴量の検討した．同研究では，OpenFace を用いて抽出した AU を特徴量とし，LightGBM モデルにより，すべての AU を用いた場合と，被験者ごとの特徴量重要度上位 5 つに頻出する AU を抽出した場合とで，精度の比較を行った．表 1 に使用した AU の対応表を示す．その結果を表 2 から表 4 に示す．理解時に表出される表情には個人差が存在することが明らかとなったため，k-means 法によるクラスタリングを実施した．表 5 よりクラスタ内で特徴量重要度の高い AU の方向性の一致率が高い場合には，予測精度が向上することが確認された． AU だけでは十分な精度が得られなかったことから，他の情報も利用して精度向上を図る．今回，参加者の頭部運動に着目した．

本研究では，オンライン会議における理解度推定において頭部の回転角（pitch，yaw，roll）に着目して，その影響について分析を行った．図 1 に頭部の回転角の図を示す．

表 1 AU の対応表

AU	説明
AU01	眉毛内側を上げる
AU02	眉毛外側を上げる
AU04	眉毛を下げる
AU05	上の瞼を上げる
AU06	顔の頬を下げる
AU07	瞼をひきしめる
AU09	鼻をしかめる
AU10	上の唇を上げる
AU12	口元を引っ張る
AU14	えくぼを作る
AU15	口元を下げる
AU17	あごを上げる
AU20	唇を伸ばす
AU23	唇にしわをつくる
AU25	唇を離す
AU26	あごを下げる
AU45	瞬く

表 2 Group1 の精度評価

選択法	Accuracy	Precision	Recall	F1
全 AU	51.3%	51.1%	61.4%	55.8%
上位 OR	50.5%	50.5%	51.6%	51.0%

表 3 Group2 の精度評価

選択法	Accuracy	Precision	Recall	F1
全 AU	55.1%	53.7%	72.8%	61.8%
上位 OR	52.2%	51.1%	96.7%	66.9%

表 4 Group3 の精度評価

選択法	Accuracy	Precision	Recall	F1
全 AU	51.1%	50.6%	96.2%	66.3%
上位 OR	50.1%	50.1%	95.1%	65.6%

表 5 グループごとの AU の向きの一致率

Group	AU07	AU17	AU25	AU26
Group1	57.1%	85.7%	71.4%	71.4%
Group2	83.3%	100%	83.3%	83.3%
Group3	75.0%	62.5%	87.5%	87.5%

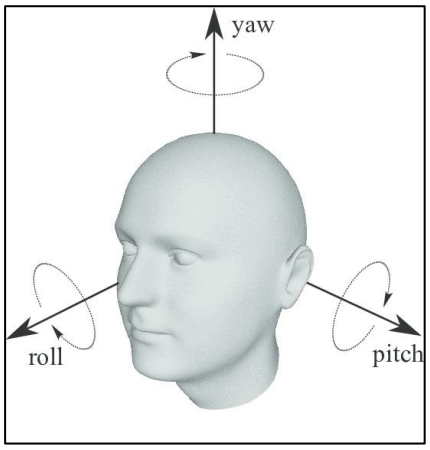

図 1　頭部の回転角[8]
pitch が x 軸周り，yaw が y 軸周り，roll が z 軸周り

3. 実験

3.1. 実験の概要

本研究では，オンライン会議中の顔画像を取得することを目的として，クラウドワークスを通じて募集した 21 名の被験者を対象にオンライン会議実験を実施した．

表情分析には，顔表情の特徴量である AU，三次元ベクトルの回転角を推定可能なアプリケーションである OpenFace を用い，各被験者の表情データを取得した．

先行研究では，1 フレーム単位で Action Unit の特徴量重要度を分析し，判別精度を評価した．これに対し本研究では，取得した顔映像データの頭部の回転角に着目した．被験者 Q が頭部運動の有意差が見られたため，被験者 Q について分析した．

3.2. データの取得方法

図 2 にデータ取得のフロー図を示す．実験では，参加者を 4 人 1 グループとし，1 グループあたり 17 分間のオンライン会議を実施した．図 3 にオンライン会議実験中の画像を示す．参加人数が規定に満たないグループについては，代役を配置して実験を行った．オンライン会議のテーマは「若者の選挙投票率を向上させるための案の検討」とし，参加者が「理解している表情」および「理解していない表情」を自然に表出できるよう設計した．

会議中の映像からは表情データを取得し，実験終了後には，参加者自身に自身の会議中の映像を見返してもらった．その際，アノテーションツールである ELAN[9] を用い，理解していると感じた時間帯には「理解している」，理解していないと感じた時間帯には「理解していない」とラベル付けをしてもらった．

取得した映像データに対しては，OpenFace を用いて 1 フレームごとに AU の推定値を算出した．最終的に，「理解している」および「理解していない」とアノテーションされた時間区間に対応する AU の推定値を，それぞれ抽出した．

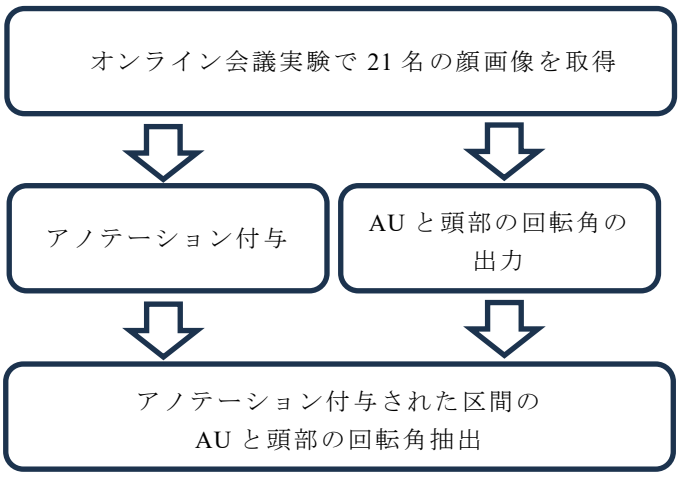

図 2 データ取得のフロー図

図 3 オンライン会議実験画像

3.3. 分析方法

オンライン会議中に収集された被験者の頭部の回転角に基づき，理解／不理解の推定を試みた．pitch,yaw,roll の 3 種類の特徴量を用いた．

各時刻における理解／不理解で付与した教師データを使用した．これらの特徴量とラベルに基づき，分類器として SVM と Random Forest を構築し，理解／不理解の 2 値分類を行った．分類精度は Precision, Recall, F1 スコアの指標に基づいて評価した．

なお，評価は 5-fold クロスバリデーションにより行い，学習と検証を分離した上での性能測定を実施した．可視化にあたっては，頭部の回転角それぞれの時間変化を折れ線グラフとして描画した．

4. 実験結果

理解／不理解で顕著な差が見られた被験者 Q に着目して，結果の詳細を示す．被験者 Q の pitch，yaw，roll の波形のグラフと Random Forest での精度評価の結果を以下に示す．青線が不理解とアノテーションが付けられた時間区間，赤線が理解とアノテーションが付けられた時間区間である．空白の時間区間は OpenFace が

認識できなかった時間区間である．

4.1. 被験者 Q の pitch, yaw, roll の波形のグラフ

被験者 Q の pitch，yaw，roll の時間変化を可視化した．その結果を図 4 から図 6 に示す．おおまかな傾向としては，pitch,yaw,roll に共通して理解時の変動が大きく不理解時には変動が小さいことが見て取れる．

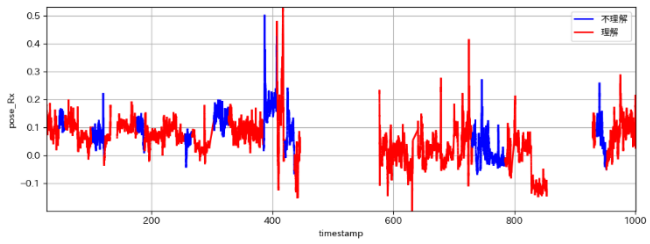

図 4　pitch の時間変化

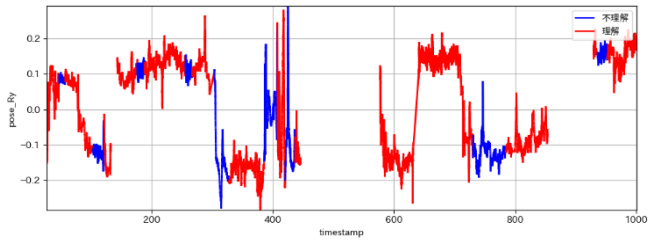

図 5　yaw の時間変化

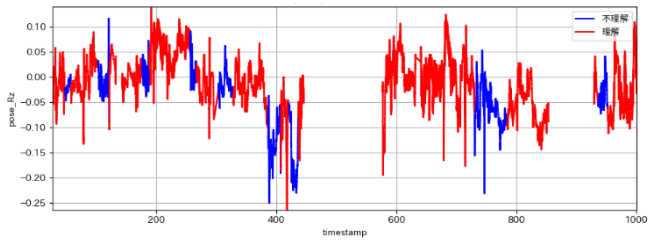

図 6　roll の時間変化

4.2. 精度評価結果

SVM と Random Forest を用いて被験者 Q の理解／不理解の 2 クラス判別器を構成し，精度評価を行った．その結果を表 6 に示す．どちらも理解の F1 は 90%以上の精度であった．しかし，SVM の不理解の F1 の結果が 57%に対し，Random Forest は 80%の精度であった．

表 6 SVM と Random Forest の精度評価結果

モデル	理解/不理解	Precision	Recall	F1
SVM	理解	84%	96%	90%
	不理解	80%	45%	57%
Random Forest	理解	92%	95%	94%
	不理解	85%	75%	80%

5. 考察

被験者 Q の pitch,yaw,roll が理解時の方が不理解時より変動が大きく，理解時に頷くような動作などが確認された．これは理解時の非言語的サインとして自然に頷く傾向があると示唆される．対して不理解時には，振れ幅が小さく頷きとみなされるような動作は見られなかった．不理解時には首を傾げるなどの動作が発生することを予想していたが，こうした傾向はみられなかった．この差異の原因として，対話中のメッセージとしては，理解したことを相手に伝えるのは良いが，不理解を敢えて伝えるのは相対的には望ましくはないとする儀礼上の配慮が働いていたと解釈することは可能であると考える．

本研究では，1 名の被験者を対象に，頭部回転運動の動的変化が理解状態の推定に寄与し得るかを検討した．その中で，SVM およびランダムフォレストの 2 つの分類手法を用いて，理解と不理解の分類を行った．その結果，SVM においては理解状態に対する F1 スコアは高かった一方，不理解状態では Recall が低く，F1 スコアも 57%にとどまった．一方で，ランダムフォレストは両クラスに対して理解時が 94%，不理解時が 80%と高い F1 スコアを示し，Precision と Recall のバランスが比較的良好であった．これは，被験者の頭部運動に一定の理解不理解の傾向が含まれており，分類器がそれを捉えている可能性を示唆している．

6. おわりに

本研究では，オンライン会議の参加者の理解度を推定することを目的として，オンライン会議実験を実施し，理解度推定をするために頭部運動の有効性を検討した．特に，被験者 Q のデータに対して詳細な検討を行った結果，理解時には pitch, yaw, roll といった頭部回転角の変動が大きくなる傾向が見られた．これは，理解時に非言語的なサインとして頷く行動が自然に生じている可能性を示唆している．

さらに，SVM および Random Forest による分類器を構築し，理解／不理解の二値分類を行ったところ，Random Forest では両クラスに対して高い F1 スコアが得られ，頭部運動が理解度推定において有効な手がかりとなる可能性が示された．

ただし，本結果はあくまで 1 名の被験者における動作パターンをもとにしたものであり，統計的な一般性を持つ結論を導くには不十分である．したがって，ここで得られた F1 スコアは，分類性能を評価するための確定的な指標というよりも，頭部運動が非言語的な理解状態の手がかりとして機能し得る可能性の一端を示すものとして位置づけられるべきである．

今後は，複数の被験者に対して同様の分析を行い，個人差を考慮したモデルの一般化可能性を検討する必

要がある．また，頭部運動情報に加えて表情変化や時系列情報などさらに音声特徴などのマルチモーダルな特徴も含めた分析を進めたい．

7. 謝辞

本研究の一部はJSPS科研費JP21K12075の助成を受けたものです．

文　献

[1] 総務省. (2023). 令和5年版 情報通信白書. 第2部 第4章「ICT市場の動向」,『ICT機器・端末関連の動向』節, pp.180‐185. (参照 2025-7-14)
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/nd241100.html

[2] 国土交通省. (2024). 令和6年度 テレワーク人口実態調査. テレワーク実施状況・導入時期・今後の意向に関する統計. (参照 2025-7-14)
https://www.mlit.go.jp/toshi/kankyo/telework_index.html

[3] 宮内佑実, 遠藤正之, “オンライン会議とオフライン会議の意思疎通の比較,” 経営情報学会　全国研究発表大会要旨集，2020年全国研究発表大会，pp.144-147，Nov.2020.

[4] R. Miao, H. Kato, Y. Hatori, Y. Sato, and S. Shioiri, “Analysis of facial expressions to estimate the level of engagement in online lectures,” IEEE Access, vol. 11, pp. 88801–88813, 2023.

[5] P. Ekman and W.V. Friesen,“The facial action coding system: A technique for the measurement of facial movement,” Consulting Psychologists Press,1978.

[6] T. Baltrušaitis, A. Zadeh, Y.C. Lim, and L.-P. Morency, “Openface 2.0: Facial behavior analysis toolkit,” Proc. 2018 13th IEEE Int. Conf. Automatic Face Gesture Recognition (FG 2018), pp.59–66, 2018.

[7] 佐藤矯汰郎, 齊藤寛己, 大和淳司, ” オンライン会議での顔画像によるユーザの理解度推定,” 電子情報通信学会　クラウドネットワークスロボット研究会，vol.123，pp.82-85，Nov.2023.

[8] A. Fernández Villán, R. Usamentiaga, J. Carús Candás, and R. Casado, “Driver distraction using visual-based sensors and algorithms,” Sensors, vol. 16, p. 1805, Oct. 2016.

[9] H. Sloetjes and P. Wittenburg, ” Annotation by category - ELAN and ISO DCR, ” Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), May.2008.