

# 講義動画視聴時における受講者の微弱な反応推定のための データセットについて

今西 咲季<sup>†</sup> 菊地 真人<sup>†</sup> 大園 忠親<sup>†</sup>

名古屋工業大学情報工学科<sup>†</sup>

## 1 はじめに

講義動画を視聴するオンデマンド型の講義において、受講者の反応を確認することは容易ではない。そのため、受講者の理解度や集中度に合わせた講義動画の作成が課題となる。そこで、本研究では、講義動画視聴時の表情から反応を推定することを目的としている。しかし、講義動画視聴時の表情は対話時と比べ、微弱な表情であり、反応を推定することは困難である。その原因の1つは、従来の表情認識で使用されているデータセットは感情を誇張した強い表情であることが挙げられる。一方、微弱な表情に関する既存のデータセットは、収集における前提条件が本研究と異なるため、本研究で利用可能であるかを調べる必要がある。本研究では、講義動画視聴時の表情の強度を評価し、現在公開されているデータセットの中で表情の強度において適切なデータセットがあるか検討した。

## 2 データセットについて

講義動画視聴時を想定したデータセットについて説明する。Pengら[1]は、リアルタイムなオンライン講義中に問題を解いている状況を想定したデータセットを構築した。受講者のパソコン画面に映される講義スライドを通して問題を出し、解いている間の表情が収録されている。被験者は10人(男性6人, 女性4人)で、年齢は全員20代前半である。撮影デバイスは各自のパソコンの内蔵カメラでフレームレートは30fpsである。問題の内容は、様々な難易度のなぞなどである。

次に、講義動画視聴時のデータセットと比較するLSEMSW[2], CAS(ME)<sup>3</sup>[3]について述べる。我々は、講義動画を視聴する学生の表情が微弱であると仮定している。これらのデータセットは、微弱な表情を識別するためのデータセットである。以降、これらのデー

タセットについて説明する。

LSEMSWは、映画やドラマから収集したデータセットである。従来の表情認識に使用されているデータセットと比べ、微弱な表情が収録されている。画像には、幸福、不安等の13種類の感情ラベルが付与されている。

CAS(ME)<sup>3</sup>は、感情を表出させる動画視聴に基づくデータセットである。動画視聴時を想定しており、本研究と条件が近いといえる。微表情と呼ばれる、隠された真の感情を表し、無意識に短時間で現れる微弱な表情を認識することを目的としている。そのため、無表情を保つように被験者に指示をしている。微表情は0.5秒以内の長さで発生する表情で、それ以外の表情は0.5秒から4秒以内に発生する表情と定義されている。CAS(ME)<sup>3</sup>はこの定義に従い、微表情に感情ラベルを付与したもの(以下, micro-expression)と微表情以外の表情に感情ラベルを付与したもの(以下, macro-expression)の2種類のデータセットである。画像には、幸福、嫌悪等の7種類の感情ラベルが付与されている。

## 3 表情強度の測定方法

表情の強度の測定方法について説明する。測定する目的は、識別対象である講義動画視聴時の表情とLSEMSW, micro-expression, macro-expressionの表情の強度の対応関係を理解するためである。まず、動画形式の講義動画視聴時のデータセットから手動で表情表出画像を抽出した。その抽出過程で表情の強度の値の範囲が広いことがわかった。そこで、他のデータセットとの比較を簡単にするため、5段階の表情の強度で画像にラベルを付与し、データセットを細分化した。ラベル名は、強度が低い方から順に微弱, 弱, 中, 中強, 強とした。分類結果を表1に示す。特徴量として、顔面の筋肉に対応づけて動作を記述するAction Unit(以下, AU)を使用し、そのAUの強度を測定することによって表情の強度を計算した。動画からの顔画像取得やAUの強度の測定にはOpenFace[4]を用いた。

On a student weak reaction dataset on watching lectures

<sup>†</sup>Saki IMANISHI, <sup>†</sup>Masato KIKUCHI and <sup>†</sup>Tadachika OZONO  
<sup>†</sup>Dept. of Computer Science, Nagoya Institute of Technology.

表 1: 5段階の強度による分類結果

	微弱	弱	中	中強	強
画像枚数	13	15	36	66	40

また、測定した AU 値をそのまま使用すると、表情の癖で普段から口角が下がっている人は、口角の筋肉に対応づいた AU の強度が常に高くなる。表情表出による AU の変化を明らかにするために、被験者の無表情画像からの変化量を求めた。

#### 4 分析結果および考察

本研究では2つの目的で AU の値域を調べた。1つ目は、個人の特徴的な表情を考慮した AU 値の補正効果の調査である。2つ目は、講義動画視聴時の表情とデータセットの表情の強度の対応関係の調査である。

特に補正効果が高かった被験者の結果を、図1に示す。補正方法としては、15枚の無表情画像の AU 値の平均を差し引いた。図1は補正前と補正後の15枚の無表情画像に対する AU 値の分布を、AU 毎に示している<sup>1</sup>。補正前は、外れ値を除けば、AU04, AU12, AU14の3種類において、0.2を超えているが、補正後は、全ての値が0.2未満となった。補正前は3種類のAUによって、無表情であっても表情があると判定される可能性がある。補正後は、無表情をAU値がゼロ近傍である時として定義することが可能になる。

講義動画視聴時のデータセットである微弱、弱、中、中強、強と LSEMSW, micro-expression, macro-expression の表情の強度の分布を図2に示す。1枚の顔画像から17種類のAU値を検出したが、表情の強度の計算には、17種類のAUの平均値を使用した。微弱、弱、中、中強は micro-expression と macro-expression の分布内に含まれており、強は、macro-expression と LSEMSW の分布内に含まれている。よって、講義動画視聴時のデータセットの全ての表情の強度の分布は LSEMSW と macro-expression と micro-expression の分布内に含まれるため、表情の強度という観点においては、本研究で比較したデータセットは講義動画視聴時の表情認識に適切であるといえる。また、講義動画視聴時のデータセットの表情の強度は幅広いため、比較した3つのデータセットを再利用可能と考えられる。ただし、これらの3つのデータセットが、データの取得条件やラベルの種類などの表情の強度以外の観点で、講義動画視聴に適用可能か検討が必要である。

<sup>1</sup>ただし、AU04の補正前の平均値(1.8)が極端に大きいため、グラフから除外した。

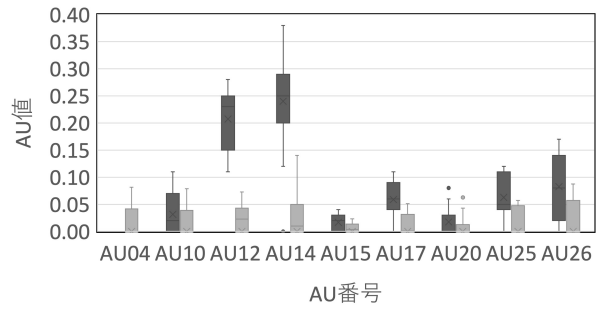


図 1: 無表情画像における AU 値

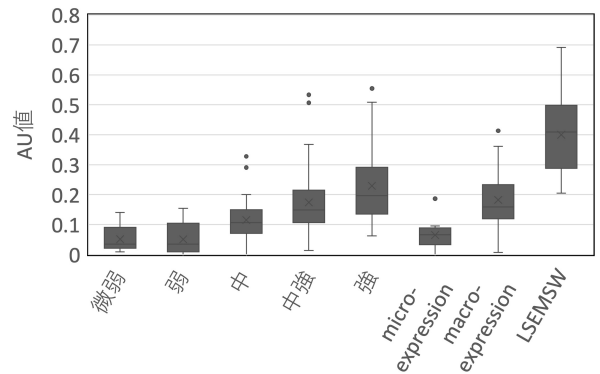


図 2: 表情の強度と AU 値の関係

#### 5 おわりに

本稿では、微弱な講義動画視聴時の表情を識別するにあたって、学習に使用するデータセットの表情の強度に着目し、本研究で収集した講義動画視聴時に関する表情データセットおよび既存データセットとの表情の強度を比較した。すなわち、本研究のデータセットと既存データセットでは収集における前提条件が異なるにも関わらず、微弱な表情の強度が同程度であることがわかった。よって、表情の強度という観点において、検討したデータセットは適切であるといえる。

謝辞 本研究の一部は JSPS 科研費 JP19K12266, JP22K18006 の助成を受けたものです。

#### 参考文献

- [1] Y. Peng, M. Kikuchi, T. Ozono, "Online Classroom Evaluation System Based on Multi-Reaction Estimation", WI-IAT '21, pp. 500-505, 2022.
- [2] G. Hu, et al, "Deep multi-task learning to recognise subtle facial expressions of mental states", ECCV, pp. 103-119, 2018.
- [3] Jingting Li, et al, "CAS(ME)<sup>3</sup>: A Third Generation Facial Spontaneous Micro-Expression Database with Depth Information and High Ecological Validity", IEEE TPAMI, 18p, 2022.
- [4] T. Baltrusaitis, et al, "Openface 2.0: Facial behavior analysis toolkit", FG 2018, pp. 59-66, 2018.