

顔検出と教員の発話による

授業中の学習者の集中度評価システムの開発

左座祐之助^{*1}, 中村伊吹^{*1}, 尾崎剛^{*1}, 広瀬啓雄^{*1}

^{*1} 公立諏訪東京理科大学大学院

Development of an Evaluation System for Learners' Concentration Level in Class by Face Detection and Teacher's Speech

Yuunosuke Zouza^{*1}, Ibuki Nakamura^{*1}, Takeshi Ozaki^{*1}, Hiroo Hirose^{*1}

^{*1} Suwa University of Science

Learners' facial orientation was photographed, and the relationship between facial orientation and the teacher's speech content was visualized to assess the learners' level of concentration. The results of the experiment showed that there was a correlation between the teacher's speech content and the learner's face orientation, indicating that the teacher's active speech increased the learners' interest and they tended to turn their faces. In addition, the time for learners to turn their faces includes the time for working on assignments and checking slides, and this information may also be useful for class improvement.

キーワード: 集中度, 学習支援, 顔検出, 文字おこし

1. はじめに

授業中, 教員が学習者の注意を引き続けることは, 学習成果の向上や, モチベーションを維持するうえで必要不可欠である. しかし, 教員が常に注意をひくことは容易ではなく, また, 教室全体の集中度を測るのも困難である.

左座(2022)では, 学習者のモチベーションを把握及び維持・向上させ, 授業改善の支援方法を明らかにすることを目的とした研究が行われた. この研究では, 達成動機理論を用いて学習者のモチベーションの数値化し, ARCS 動機づけモデルを活用した授業改善支援システムの構築及び授業改善を行をした. それぞれ学習者にアンケートに回答してもらい, 結果を教員に提示した. 構築したシステムを教員が利用し, 授業改善をした結果, 学習者のモチベーションが向上したことが確認できた. しかし, 授業を重ねるごとにアンケート

回答者数が減っており, マンネリ化などが問題となり, また, アンケート手法は主観的な結果になってしまうという問題が挙げられた.

また, 田中(2010)では, プロジェクタを使用した講義において, 表示切り替え時における学生の動きや顔の向きを検出し, 相関を調べることで, 学生の授業に対する集中度を客観的に評価するシステムを開発した. また, この手法を板書で行った場合の研究がされた. 結果として, ある程度以上の大きさで写っており, 受講態度がはっきりしている学生に対しては, 目視による判定に対して 70%程度の精度で判定できることが確認された. しかし, プロジェクタを使用した表示切り替えに対して, 板書では検出精度が低い結果となった.

そこで, 本研究ではARCS動機づけモデルの「注意」の側面に焦点を当て, 授業中の学習者の顔を撮影し,

顔の向きから集中度を評価する。また、教員の発話内容を文字おこしし、発話した文字数と学習者の顔の向きとの関係を可視化することを目的とする。

2. システム概要

2.1 RetinaFace

RetinaFace は、顔検出や顔の特徴点検出のための深層学習ベースのアルゴリズムである。顔の位置を検出するだけでなく、顔の主要な特徴点（目、鼻、口）を検出することも可能である。また、小さな顔やスケールの異なる顔、マスク着用している場合でも検出できる。

2.2 顔の角度

RetinaFace で検出された顔の特徴点の座標をもとに顔の向きを計算する。顔の向きによって、目、鼻、口の位置が変わる。それを利用し、特徴点間の距離を求め、そこから角度を求めるようにした。図 1 は、右目・左目・鼻の特徴点から角度を求める例である。同様に、右目・鼻・右口、左目・鼻・左口、鼻・右口・左口の 3 か所の角度も求める。また、本研究では正面を向いている学習者を検出するため、顔の角度に閾値を設定し、顔の向きを判別することにした。

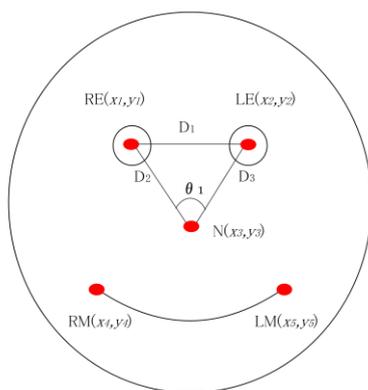


図 1 顔の向き概要

2.3 文字おこし

文字おこしは、Python の音声認識ライブラリである SpeechRecognition を使用し、Google の音声認識 API を利用した。本研究では、授業中に教員が話している内容を録音しておく。その後、SpeechRecognition ライブラリを使用し、1 分間隔で音声を取り出し、Google の音声認識 API を利用し、音声をテキストに変換し

csv に保存した。

3. 実験

3.1 実験方法

3.1.1 分析対象科目

本実験を行うにあたり、公立諏訪東京理科大学工学部情報応用工学科で開講している「データベース論及び演習」を分析対象とした。

「データベース論及び演習」は、3 年次選択科目であり、100 分×2 コマで 1 回分の授業となっている。この授業では、IoT などから収集されるビッグデータを、データベースに格納し、AI や機械学習および統計的手法で分析するために、必要なデータを検索する。さらに、分析可能な状態の前処理の基礎を学習する。また、SQL, PHP, Python を用いて授業を行っている。

3.1.2 実施方法

対象科目の授業中に、教室全体が映るように、カメラをモニターと同じ高さに配置した（図 2）。カメラには広角で教室全体が撮影できる Gopro10 を利用した（図 3）。また、撮影中に教員の音声のみを録音できるように、教員にトランスミッター（図 4 左）をつけてもらい、Gopro10 にチューナー（図 4 右）を接続した。



図 2 教室撮影風景



図 3 Gopro10



図 4 トランスミッターとチューナー

3.2 実験結果

3.2.1 分析間隔 1 分間ごと

「データベース論及び演習」の第 2 回で、1 分間隔で顔の向きは考慮せずに検出を行った。結果は図 5 のようである。横軸は時間（分）を表しており、縦軸は文字数と顔の数を表している。また、棒グラフが 1 分間ごとの教員の発話した文字数であり、赤色の折れ線グラフが 1 分ごとに検出した顔の数、緑色の折れ線グラフが移動平均である。図 5 より、顔を検出しただけでは、顔の数に変化が見られなかった。1 分間ごと検出しているため、検出のタイミングで下を向いており、検出されていない可能性も考えられる。

そこで、次に 10 秒間隔で検出を行った場合と、顔の向きを考慮した検出を行った。

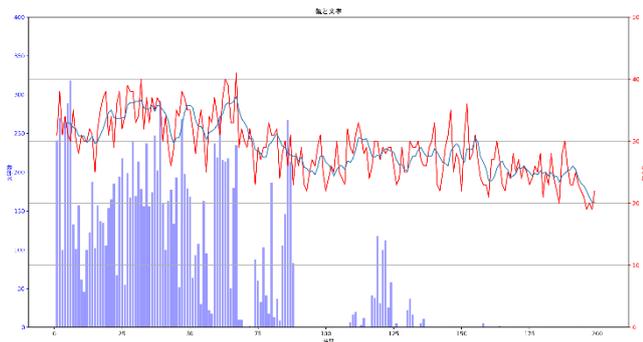


図 5 1 分間ごとの顔検出推移

3.2.2 分析間隔 10 秒ごと

10 秒間隔で検出を行った結果は図 6 のようになっている。縦軸、横軸は図 5 と同様であり、赤色の折れ線グラフが 10 秒ごとに検出した顔の数である。

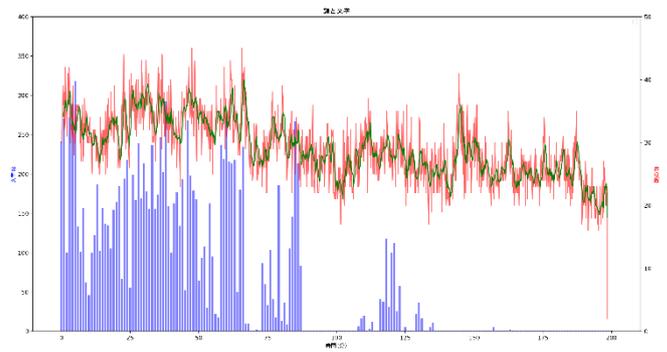


図 6 10 秒ごとの顔検出推移

次に 2.2 の顔の向きを考慮した結果は図 7 のようになっている。教員の発話数が多いと正面を向いている学生が多くなり、発話数が少ないと検出した顔数も減少した。第 3 回での結果は図 8 のようになっており、第 2 回と同様、発話数によって、顔の数に変化していることが分かる。

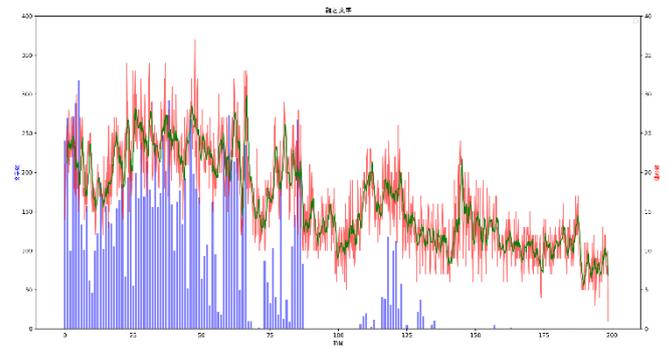


図 7 10 秒ごとに顔の向きを考慮した検出推移
(第 2 回)



図 8 10 秒ごとに顔の向きを考慮した検出推移
(第 3 回)

4. 考察

4.1 顔と文字数の相関

検出した顔の数と教員の発話した文字数の相関を調べた。なお、顔の数は顔の向きを考慮した結果を用いることとし、教員の発話した文字数とデータの数を同

じにするため、10秒ごとのデータを6回分の移動平均を算出した。第2回の授業での結果は図9のようになっている。相関係数は0.752と正の強い相関があった。また、第3回でも0.481と正の相関があり（図10）、教員が積極的に発話することにより、学習者の関心が高まり、顔を向ける傾向がある可能性が考えられる。しかし、第2回に比べて第3回は相関係数が低下している。この原因として、第3回では教員の発話数が多く、関心がそれてしまった可能性が考えられる。授業に抑揚を持たせるなどを行うことで、関心を維持できると考える。また、第3回目以降で相関係数が低下していれば、授業のマンネリ化なども考えられるため、引き続き調査していこうと考える。

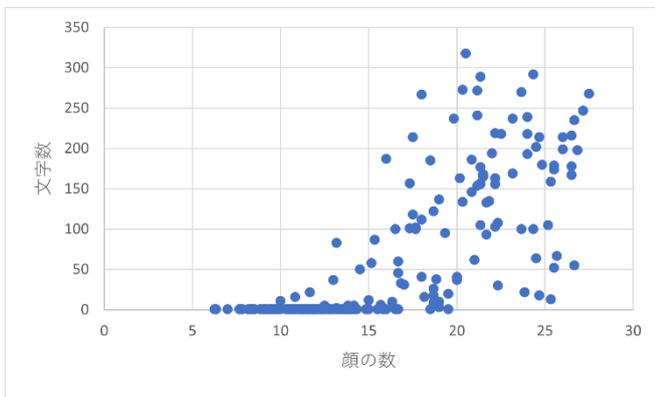


図9 発話文字数と顔の数の相関（第2回）

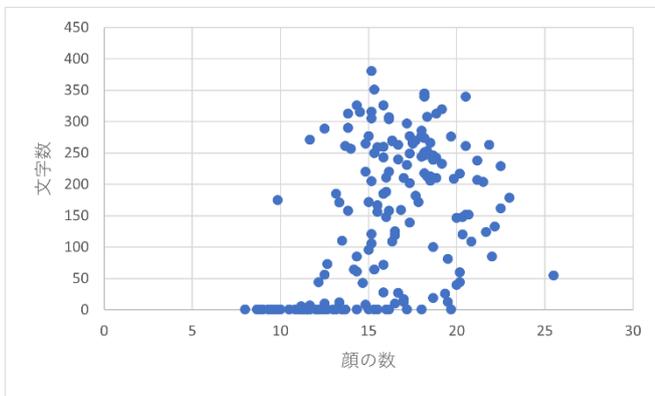


図10 発話文字数と顔の数の相関（第3回）

4.2 話してないのに顔の数が多いとき

図7、図8において、教員が発話していないが、発話時と同数顔を検出している。これは課題に取り組んでいる時間に、わからない箇所を正面のモニターでスライドや学習した内容を確認しているからだと考えられる。学習者の顔だけでなく、スライドも撮影することで、学習者がどこで苦戦しているかなども把握することができ、授業改善に役立つと考えられる。

4.3 学習者の顔検出率

第2回、第3回での学習者の出席者数は表1のようになっている。出席者数に対して顔の検出率は、第2回で39%、第3回で35%が最大であった。PCを見ていたり、別の方向を見ている学習者もいるが、前の人に顔が隠れていることや教室の奥に座っている学習者の検出ができていないことが検出率の低さの1つの要因だと考える。カメラの位置を変更することや、カメラを複数設置することで検出率が上がるのではないかと考える。また、本研究では約120名を対象にしたため広い教室での調査となったが、一回り小さい教室で対象が少なければ、より検出精度が上がると思う。

表1 授業の出席者数

授業回数	出席者数（人）
第2回	115
第3回	115

5. まとめ

本研究では、ARCS 動機づけモデルの「注意」の側面に焦点を当て、授業中の学習者の顔を撮影し、顔の向きと教員の発話数の関係を可視化することを目的とした。

授業中に教室全体が映るように、カメラをモニターと同じ高さに配置し、撮影中に教員の音声のみを録音できるようにマイクを設置した。

顔の検出には RetinaFace という、顔検出や顔の特徴点検出のための深層学習ベースのアルゴリズムを使用し、得られた特徴点から顔の向き（角度）を計算し、正面を向いているかどうか判別するようにした。

文字おこしは、Google の音声認識 API を利用し、音声をテキストに変換し csv に保存した。

10秒間隔で顔の向きを計算し、文字おこしは1分間隔で行うようにした。その結果、発話数によって、顔の数が増えていることがわかった。また、検出した顔の数と発話した文字数には正の相関があり、教員が積極的に発話することにより、学習者の関心が高まり、顔を向ける傾向があることが確認できた。

今後の課題として、本研究では教員の発話した文字数と顔の向きから集中度を測定したが、発話した内容

を分析し、どのような発言に学習者が興味を示すのかを調査していきたいと考えている。また、学習者が座っている位置によって顔の向きも変わってくるため、座席によって閾値を変化させていこうと考える。

謝辞

本研究は JSPS 科研費 22K02893, 23K02715 の助成を受けたものです。

参 考 文 献

- (1) RetinaFace | InsightFace: an open source 2D&3D deep face analysis library,
<https://insightface.ai/retinaface> (2023 年 6 月 1 日確認)
- (2) GitHub-UberI/speech_recognition: Speech recognition module for Python, supporting several engines and APIs, online and offline,
https://github.com/UberI/speech_recognition (2023 年 6 月 1 日確認)
- (3) 左座祐之助, 広瀬啓雄, 尾崎剛: “学習者と教員の授業に関するモチベーションを可視化するシステム構築及びその効果”, 第 84 回全国大会講演論文集 2022(1), pp.135-136 (2022)
- (4) 尾形直哉, CHAKRABORTY Goutam, 馬淵浩司, 松原雅文: “角度の変化に対応した顔認識手法”, 情報科学技術フォーラム, FIT 2007 号, 一般講演論文集第 3 分冊, pp.61-62 (2007)
- (5) 田中敏光: “学生の動きと教示情報との相関から講義中の学生の集中度を評価するシステムの開発”, 科研費 20500814, 2010 年度研究成果報告書 (2010)