

Zoom でのプレゼンにおける表情認識を用いた聴衆の反応可視化システム

大谷 紀子 研究室

1872006 飯島 伊武己

1. 背景と目的

現在, 新型コロナウイルスの蔓延により Zoom でのプレゼンテーションが数多く行われている. 対面でのプレゼンテーションであれば, 聴衆の反応を直接見ることができるため, 全体を通してどのくらい興味を持たれていたかというフィードバックを直感的に受け取ることができる. しかし, Zoom を用いたプレゼンテーションでは, 聴講者が数百人を超える場合でも, 数名の反応しか確認することができないことから, フィードバックが受け取りにくいという問題点が存在する. また, 対面での聴衆を見渡しながらのプレゼンテーションとは違い, 聴衆に対して語り掛けている印象を与えるにはカメラを見ながら話す必要がある. 結果として, 画面上に表示されている聴衆の情報が受け取りにくくなるという問題点が存在する.

本研究では, Zoom でのプレゼンテーションの質を向上させることを目的とし, 聴衆の反応を確認しながら発表を振り返ることを支援するためのフィードバックシステムを構築する.

2. システム概要

上間ら[1], 安彦ら[2]により, 作業負荷が高い状態や, 集中が必要とされる条件下において, 瞬きの回数が減少することが確認されている. 本システムでは, 聴講者の姿を Web カメラで捉えて 30 秒当たりの瞬き回数を計測し, 聴講中の集中度として提示する. 併せて, よく表情として現れ, かつ発表の改善に繋がる無感情, 喜び, 悲しみ, 怒り, 驚きの 5 つの感情について提示する. 顔の検出, 分析には faceapi.js を用いる.

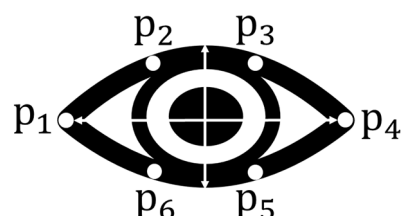


図 1: 瞬きの検出に利用する目のランドマーク

瞬きの計測に関しては, 検出された顔の中に 68 個のランドマークを取ることで検出器を用いる. 68 個のランドマークのうち, 図 1 に示す 6 点のランドマーク $p_1 \sim p_6$ を利用し, 目のアスペクト比である EAR を式(1)によって算出する.

$$EAR = \frac{P_2 - P_6 + P_3 - P_5}{2P_1 - P_4} \quad (1)$$

Tereza Soukupova ら[3]より, 式(1)によって導出される目のアスペクト比は, 目を開いている間はほぼ一定であるが, 目を閉じているときにはゼロに近づくことが確認されている. 本システムでは, 両目それぞれで導出された EAR の平均を算出し, 平均値が閾値以下になったとき, 瞬きをしていると判定する.

表情に関しては, 表情認識モデルを利用して分析した各感情の想起度合いを表す感情値が 0~1 の値で出力される. 表情認識モデルは, 公開されているデータセットに含まれるさまざまな画像や, Web からスクレイピングされた画像を用いた学習により獲得されたものである.

プレゼンテーション終了後には, 瞬き回数と 5 つの感情値を基にした分析結果を, 各データが最高値に達したときの時間帯, 時系列グラフの 2 つで提示する. 最高値に達したときの時間帯を提示

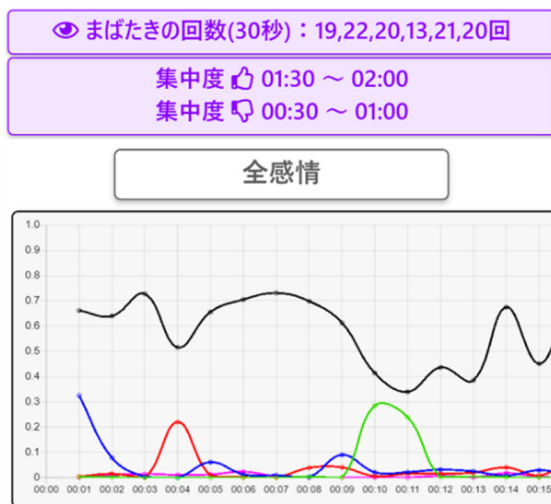


図 2 : 結果表示画面

することで、グラフすべてに目を通すことなく大まかなフィードバックを受け取ることが可能となる。結果表示画面の一部を図 2 に示す。

3. 評価実験

20 代の学生 5 名を対象として、評価実験を行った。被験者それぞれに今現在自身が行っている研究の内容について 3 分程度のプレゼンテーションをさせ、発表者以外の被験者全員の集中度と表情の変化を本システムで提示した。全員のプレゼンテーション終了後、被験者自身が行ったプレゼンテーションの分析結果を見せ、アンケートを実施した。アンケートでは、インタフェースは使いやすかったか、瞬きと表情の分析結果は適切であったか、どのデータが次のプレゼンテーションの改善に有用であると感じたか、などについて調査した。アンケートの結果、本システムを使用することによって、次回のプレゼンテーションの改善に繋げることが期待できるという肯定的な評価を被験者 5 人全員から受け取ることができた。特に、集中度のグラフが一番改善に役立つさうという意見が過半数を超える 3 人から挙げられた。

4. 考察

集中度のグラフは、カメラで一人一人の瞬きを計測することができる Zoom での発表ならではの分析結果であるので、集中度のグラフが改善に役立つさうという意見が多くなったと推察できる。

したがって、集中度に関しては Zoom でのプレゼンテーションにおける聴衆の反応が対面に比べてわかりずらいということを補う以上の結果が得られたといえる。また、被験者全員が有用なフィードバックを受け取ることができたと回答したため、本研究の目的であるプレゼンテーションの質向上のための支援ができたといえる。一方で、感情データの精度や個人差については否定的な意見も挙げられた。

現状の問題点としては、WEB カメラを用いた瞬き検出の精度に偏りがあることが挙げられる。原因は、人によって目の形や大きさや位置が違うため、ランドマークに誤差が生じることである。解決策としては、目のアスペクト比を用いた判定において、目を閉じていると判定される閾値をユーザの顔に合わせてパーソナライズするように改良することが挙げられる。また、被験者からの意見として「発表中は聴衆の顔が見たい」という意見も挙げられた。現状では分析にカメラが専有され、聴衆の顔を発表者に表示することができない。したがって、聴衆の顔が発表中に確認でき、かつバックグラウンドで分析するシステムに改良すると、リアルタイムな反応も考慮できるようになり、さらなるプレゼンテーションの改善が期待できる。

参考文献

- [1] 上間裕二, 小川剛史, 川島隆太, 高橋信, “メガネ型デバイス JINS MEME を用いたワークロード推定の基礎的検討,” ヒューマンインタフェースシンポジウム 2015, pp. 825-828, 2015
- [2] 安彦智史, 兜森仁志, 佐久田博司, 長谷川大, “web カメラを用いた瞬き検出による集中度評価,” 情報処理学会 第 77 回全国大会講演論文集, Vol. 2015, No. 1, pp. 931-32, 2015
- [3] Tereza Soukupova, Jan Cech, “Real-Time Eye Blink Detection using Facial Landmarks,” Proceedings of 21st Computer Vision Winter Workshop, 2016