

1 対多遠隔コミュニケーションにおける聴衆反応の集合的提示

前田 清洲* 暦本 純一*†

概要. オンラインでのコミュニケーションの機会は増えているが、特にオンライン講義やウェビナーといった1対多の状況では、対面コミュニケーションに比べて話者が聴衆の反応を一度に確認することができず、話しづらさを感じてしまう。既存の研究では、聴衆がアイコンボタンを押すことでリアクションを送る方法や、視線を中心とした非言語情報の計測と提示が行われてきたが、前者は聴衆が明示的に操作しなければならない、後者は聴衆側に専用の機材を設置する必要がある、導入が困難だった。本研究では、webブラウザで利用できる視線追跡技術を使用することで、特別な装置を必要とせず聴衆の視線を取得し、多人数の聴衆の視線位置をヒートマップとして集合的に話者に提示する手法を提案する。この手法により、話者が画面共有している資料のうち聴衆がどこを見ているかなどを直感的に把握できるようになる。本手法を用いたオンラインでの実験を実施して、話者の話しやすさが改善されるかどうかを検証した。

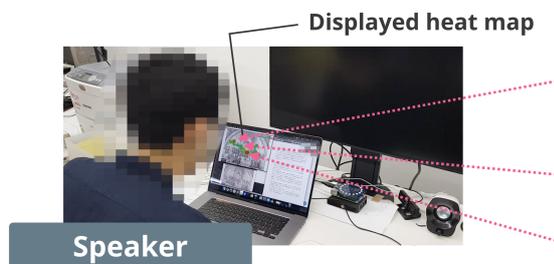


図 1. ヒートマップを見ながら話す人

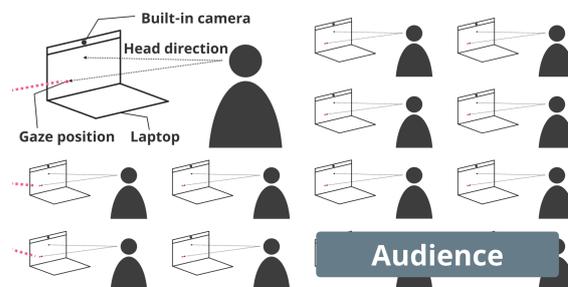


図 2. 視線をトラッキングされながら話を聞く聴衆

1 はじめに

近年我々はオンライン会議ツールなどを利用して遠隔でのコミュニケーションをする機会が増えている。オンラインコミュニケーションは物理的制約を受けずどこからでもアクセスできる利点もある一方で、対面でのコミュニケーションと比べ視線やジェスチャーといった非言語情報が欠落することで相手の反応を読みづらいといったデメリットも存在する[1]。特にオンライン授業やウェビナーといったリアルタイムでの1対多のコミュニケーションでは聞き手はビデオをオフにすることがほとんどで、話者は視線や表情、ジェスチャーから読み取れる聴衆の反応を知ることができない。こうした状況では、話者はオフラインと比べて話しづらさを感じるようになる。

遠隔コミュニケーションツールによってはテキストや絵文字を用いて話者に対してリアクションを送ることができるが、こうしたリアクションは聞き手の明示的なアクションが必要であるため聞き手にとっ

て面倒となる。一方で視線やジェスチャーのような非明示的な非言語情報は聞き手に負担を強いることなく取得することが可能である。

遠隔講義システムにおいて視線を利用することによって集中力といった生徒の状態を把握できることが知られている[26]。ただしこうした研究で生徒の視線情報は主に講義視聴完了後の分析か、集中力が途切れたことを生徒に提示するために使用される。生徒の反応を知るためにリアルタイムで教師や話者へフィードバックされることはなかった。遠隔コミュニケーションにおけるリアルタイムな視線情報の共有によるコミュニケーション支援の試みがなされてきたが、多くの場合1対1の状況に限られる[3]か、ノートPCに加えてアイトラッカーが必要であるため聴衆の数が数人に限定される[25]。大人数の聴衆が同時に視聴するようなオンライン授業やウェビナーで使用することは難しい。

そこで本研究では図1,2のように、ノートPCに内蔵されるカメラのみを使用するwebベースの視線トラッキングライブラリを用いることで大人数の聴衆の視線を話者側にリアルタイムでヒートマップとして集合的に表示する手法を提案する。このシス

Copyright is held by the author(s).

* 東京大学

† ソニーコンピュータサイエンス研究所京都研究室

テムでは視線トラッキングのための追加の機材を必要としないため、低コストで導入が容易であり大規模なオンラインでの使用に適している。そのため多数の聴衆の視線を同時に取得し話者に提示することができる。また今回は視線位置に加えて聴衆の頭の動きも提示した。これにより話者は聞き手のうなずきといった反応も知ることができる。実際の使用場面として話者がディスプレイ上で何らかの資料を共有し講義や発表、説明を行う場面を想定した。そうした場面で聴衆が共有した資料のどこを覗いているかが分かることによって聴衆の反応や興味を持っている部分を理解でき、話者の話しやすさが改善されるかを実験により検証した。

2 関連研究

本項では、視線を共有することによる遠隔コミュニケーションの促進、MOOC(massive open online course)のようなオンライン講義システムにおける視線の活用について関連研究を紹介する。Johansenの分類[10]に基づけば、前者は「同じ時間・違う場所」、後者は「違う時間・違う場所」におけるコミュニケーション形態を指す。

2.1 遠隔コミュニケーションにおける視線共有

対面でのコミュニケーションではお互いに相手が見ている場所を知ることができる[12]が遠隔での会話では難しい。ここではリモートコミュニケーションにおいて1対1でのどちらか一方、もしくは両方の視線共有と1対多、もしくは三人以上の遠隔コミュニケーションでの多から1への視線共有に関する研究について紹介する。

お互いの視線を共有することは1対1の遠隔コミュニケーションで有効であることが報告されており、これまではリモートペアプログラミング[3]、特定の物体を共同で探す作業[2, 8]などで視線の共有が行われている。

1対多のオンラインコミュニケーションでプログラミング課題を解いている生徒の視線を教師側に表示することによって教師が生徒の理解度を把握するのに役立つことを明らかにした研究[25]や、講義と演習中に教師・生徒の視線をお互いに共有した研究[22]が存在する。また、GAZE Groupware System[23]は、複数人のオンラインコミュニケーションにおいて視線は誰が誰に話しかけているかを把握することのできるシステムである。

しかし、既存研究では視線トラッキングのためにノートPC以外に機材を必要としており聴衆の人数を増やすことは困難である。実際にAngeloら[25]やSpakovら[22]は聞き手の人数が4人であり人数が増えた際の技術的な難しさに言及している。また人数が増えることで視線表示は話者にとって気の散るものになり得る。

本研究ではwebベースの視線トラッキング手法を用いることで人数の限界を超えて大人数の聴衆の視線を話者に表示するシステムを提案する。

2.2 オンライン講義における非言語情報の取得

近年、MOOCをはじめとするオンライン講義システムが充実している一方で、集中力の低下などの理由からごく少数の学生しかオンライン講義を最後まで視聴しないといった学習効率の悪さも課題となっている[11]。

そこで視線と認知状態の関係性[17]に基づき、講義ビデオを視聴中の生徒の集中力の低下を視線を用いて検知する研究が多くなされてきた[9, 24]。

またMOOCにおける生徒の視線は生徒が授業についてこれているかを判断する手がかりにもなり得ることが分かっている[21]。

以上のような既存研究ではリアルタイム性とプライバシーの侵害に課題があることが指摘されている[19]。リアルタイム性とプライバシーの保護を志向していくつかの研究ではwebブラウザ上で完結する視線トラッキングライブラリを利用してMOOC視聴者の集中力の低下を検知している[26]。

ただしこれらの研究では録画された講義ビデオを閲覧する学生が対象であることがほとんどであるため、視聴者に対して集中力が低下したことなどをその場で知らせる[18, 20]ことはあっても話者へのフィードバックについては考慮されていない。リアルタイムで行われるオンライン講義やウェビナーでは話者が聞き手の非言語情報から聞き手の反応を知ることが極めて重要な要素となる。

3 システム設計

3.1 聴衆の視線・顔の向き の推定

大人数の聴衆の視線を話者に提示するためにウェブベースのオープンソース視線トラッキングライブラリであるWebGazer.js¹[16]を採用した。WebGazer.jsの使用例として、遠隔講義を視聴している生徒の集中力低下の検知[26]やWebGazer.jsを拡張したライブラリであるSearchGazer.jsを用いてweb検索中のユーザの視線を追跡した研究[15]、またタッチタイピストと非タッチタイピストの違いに焦点を当てて視線とタイピングの関係を調査した研究[14]などがある。Tobii²などのアイトラッカーと比べて精度は高くないもののノートパソコンの内蔵カメラで使用できる点、リアルタイム性に優れる点から大人数での講義に適していると考えられる。

リアルタイムで推定される聞き手の視線位置はその精度からちらつきの多いものとなる。この問題を解決するために、話者へと送信する視線位置は聞き

¹ <https://webgazer.cs.brown.edu>

² <https://www.tobii.com/ja/product-listing/>

手の数フレーム分の推定された視線位置の平均値とした。これにより話者は聞き手が注視している場所をより把握しやすくなる。また聞き手と話者とのディスプレイの大きさの違いを考慮して送信される視線位置はディスプレイの相対位置となる。

WebGazer.jsを使用する場合、視線位置とディスプレイ上のカーソル位置の関連性に基づいてwebブラウジング中にユーザが継続的に行うマウスと視線の動きからキャリブレーションを行い精度を高めることができる。しかし今回は聴衆が受動的に話を聞く場合を想定しているため継続的なキャリブレーションは行わない。代わりにユーザは最初にキャリブレーションを行う。図3のようにWebGazer.jsを組み込んだwebアプリケーションを立ち上げたのち画面上に配置された8つの円をマウスの位置と視線を合わせながら数回クリックすることでキャリブレーションを行う。

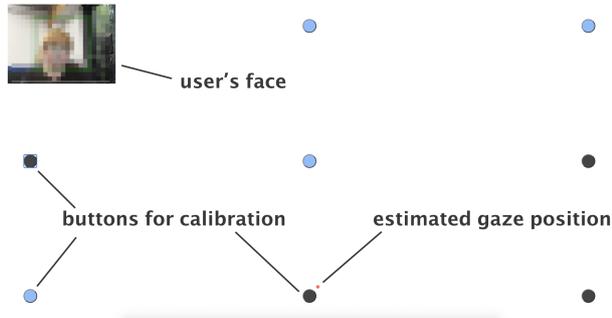


図 3. 視線トラッキングの様子

また、聞き手の顔の向きを推定するためにWebGazer.jsで取得される顔の特徴点の位置を利用した。図3の左上のように聴衆の顔にはリアルタイムで特徴点が重ね合わせて表示されている。この中から鼻の先端位置を元に顔の向きを推定した。

3.2 話者への視線・顔の向きの表示

相手の視線を表示する方法にはヒートマップ、カーソル、お互いに同じ場所を見ている時のみ表示されるマーカーなどがある[5]。相手の視線がリアルタイムで画面に表示される場合、話者の気を散らす可能性があることが指摘されている[4]。今回は聞き手の数が大人数であることを想定しているため、カーソルやマーカーで表示するだけでは話者の認知的負荷を上げ気の散るものとなる。本研究では多数の聴衆の見ている場所を図4のようにヒートマップとして話者に提示することにする。赤い部分はより多くの聴衆が同時に見ている部分であり、逆に緑や青、透明な部分は見ている聴衆があまりいない部分となる。ヒートマップ表示をする利点として話者が気を散らすことなく聴衆の反応を確認しやすいこと、全体的な傾向として見ている人が多い部分が直感的

に分かるので視線トラッキングの精度が高くないという欠点を補えることが挙げられる。聴衆の顔の向きは聞き手の人数分のカーソルを用いて図5のように表現した。個々のカーソルがまとまって動くことによって話者はうなずきや否定のリアクションを確認することができる。例えば図5の左の場合は顔の向きがばらけているため聴衆の反応を読み取るのは難しい。一方で図5の右の場合は全体的に顔の向きが縦に振れているため多くの聞き手が頷いていることが分かる。顔の動きを確認することは1対1の遠隔コミュニケーションにおいて研究されていたが[7]、1対多の遠隔コミュニケーションにおいて聞き手の集会的な顔の向きの提示が話者にとって有益か検証する必要がある。

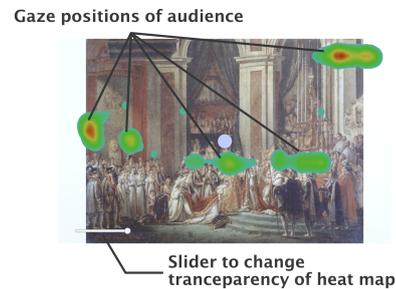


図 4. 話者側の視線ヒートマップ

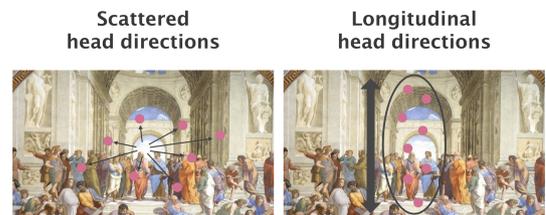


図 5. (左) ばらけている顔の向き (右) 縦方向に動く顔の向き

各ピクセルの色は以下のように決定される。ディスプレイの左上の角から数えて横方向に x 個分、縦方向に y 個分だけ進んだ位置にあるピクセル $P(x, y)$ について、まず n 人分の視線位置との距離を考える。ある視線位置 g_i とピクセル P の近さを表す指標を p_i とし、 p_i を

$$p_i = \frac{\max((r - d_i), 0)}{r} \quad (1)$$

のように定義する。ここで d_i は g_i と P の距離、 r は任意の視線位置が考慮するピクセルとの距離の最大値である。この p_i を用いてピクセル P が全視線から受ける影響 P_{xy} は

$$P_{xy} = \sum_{k=1}^n p_k = \sum_{k=1}^n \frac{\max((r - d_k), 0)}{r} \quad (2)$$

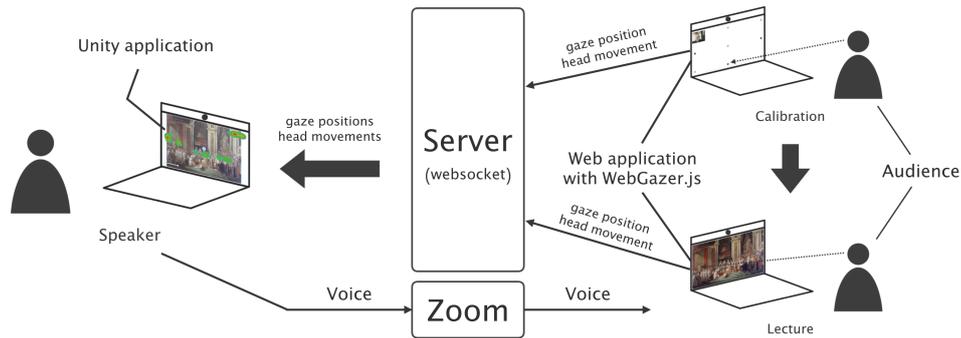


図 6. システム構成

と表すことができこの P_{xy} がある閾値 t より大きい場合、ピクセル P は赤で描画される。 t より小さくなるにつれてピクセルの色は青色に近くなる。

実際には新しい視線位置を受信した場合、 n 個の視線位置の中から最も古いものと入れ替え (2) の計算を行う。計算対象のピクセルは新しい視線位置と最も古い視線位置との距離が r 以下のものだけでよい。話者はディスプレイ左下にあるスライダーを用いてヒートマップの濃さを調整できる。

ヒートマップの描画には Unity2018.4.22f1 を使用し、ウィンドウを透過させることによって話者が共有する資料と重ね合わせて表示させることができる。

3.3 全体的な構成

システムの全体的な構成は図 6 のようになる。聴衆は WebGazer.js を組み込んだ web アプリケーションを立ち上げてキャリブレーションを行った後、web アプリにあらかじめ埋め込まれた話者の資料を閲覧できる (図右部分)。web アプリケーションにより取得された聴衆の視線位置と顔の向きは数フレームおきにサーバを経由して話者へと送信される。話者側では Unity アプリケーションがデータを受信するたびにヒートマップを更新する (図左部分)。通信には websocket を使用し、プライバシーの観点からサーバへ送る情報は視線位置と顔の向きのみとした。聞き手側から何秒おきにデータを送信するかは事前に設定することができる。また、音声はオンライン通話ツールである zoom を用いて伝達される。

4 実験

本研究で提案するシステムを用いて聴衆の反応を見られるようになることで話者が話しやすくなるかどうかを検証するためにオンラインで実験を行った。被験者は話者役 2 人と聞き手役 6 人に分かれており、いずれの被験者も 21~24 歳の大学生である。話者はどちらも普段オンラインでのコミュニケーションの経験はあるが、オンラインで不特定多数の聴衆に話した経験はなかった。

4.1 実験の概要と結果

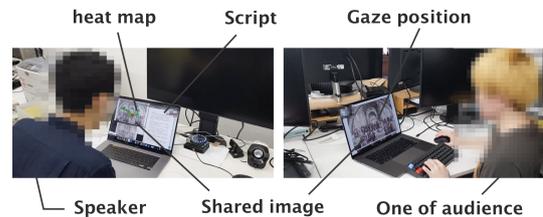


図 7. 実験の様子 (左) 話者 (右) 聴衆の一人

実験の概要は図 7 のようになる。話者はある画像に登場する人物について説明を 2 回行う。1 回目は聴衆からのフィードバックなし、2 回目は聴衆の視線と顔の向きがリアルタイムで画像に重ね合わせて提示される。この際聴衆は説明を聞きながら web アプリ上で画像を観ている。画像はパブリックドメインの絵画で話す内容は決まっているが、話者は重ね合わせて表示された反応から聞き手が画像のどこに興味を持っているかを類推して話す順番を自由に変えたり、聴衆の視線を誘導したりできる。実験は 2 人の話者にそれぞれ一回の説明あたり 5 分間の説明時間が与えられ、聞き手は合計 10 分間画像の説明を受けた。今回一人の聞き手からのデータ送信頻度は 0.3 秒間に一回とした。そのため話者側は一秒間に 20 個のデータを受信することになる。

実験終了後には話者に対して以下の 4 つの質問に回答してもらった。

- Q1: システムは聴衆の反応を知るのに役に立ったか
- Q2: 聴衆の視線・顔の向きの表示は邪魔にならなかったか?
- Q3: 視線と顔の向きが表示されない場合と比べて話しやすさを感じたか?

- Q4: 視線と顔の向きが表示されることで話し方はどのように変化したか？

Q1, Q2, Q3 は最高 7 点としてリッカート尺度を用いた。Q4 は自由記述で回答してもらった。

結果は以下の表 1 のようになった。Q4 で話者は、「視線が表示されることで人物に注目しながら読むことを意識できた」や「自分の発言に聴衆が追いついているかどうか、大まかに分かった」と回答した。

表 1. システムに関する質問の回答結果 (7 段階リッカート尺度による)

	speaker1	speaker2
Q1	5	6
Q2	7	7
Q3	6	5

本研究が提案する大人数の聴衆の視線と顔の向きを重ね合わせて表示するシステムは話者にとって説明する際に邪魔にならず、聴衆の反応を知るのに役立った。話者は、画像右側の人物について説明した後左側の人物の説明に移った際にヒートマップも右から左へ移ったことによって聴衆の反応を知ることができたことも報告した。既存の研究 [22] と同様に話者の話し方の変化を示す客観的尺度は用意されなかったが、今回のこうしたヒートマップの移動は今後話者の話し方の変化を分析する上で重要な手がかりとなる可能性がある。

今回話者役の被験者はオンラインで不特定多数の人に話す経験のない人であった。今後の実験ではオンライン講義経験のある教師にシステムを使用してもらい話しやすさを評価する必要がある。実際、SyncClass[6] では、オフラインでの大人数の授業中に生徒が授業にどれだけついていけているかをモニターに可視化して教師にフィードバックするシステムの有用性は教師の指導経験によって異なることが明らかになっている。オンラインでも同様に話者に対する聴衆の非言語情報の可視化の有用性は講義経験によって異なると考えられる。

5 議論

5.1 精度とキャリブレーションによる課題

WebGazer.js は継続的なクリックによるキャリブレーションで精度を高める事が可能であるが、今回は事前のキャリブレーションしか行わなかった。そのため、聞き手の姿勢が変化することなどを考慮すると長時間の使用に向かない。キャリブレーション後にも共有された資料を聞き手がクリックするなどしてキャリブレーションを行えるようにすると長い講義でも使用でき実用性が増すと考えられる。

5.2 視線以外の非言語情報の取得

本研究では大人数の聴衆の非言語情報として視線と顔の向きだけに注目したが、今後の研究では表情などの視線以外の非言語情報から感情 [13] といったより複雑な情報も話者へ提示できると考えられる。1 対 1 の遠隔コミュニケーションにおいて Facetop[7] のようにワークスペースと相手の顔やジェスチャーを重ね合わせて表示するシステムが存在するが、1 対多の文脈でこれらの情報を集的に話者に提示する方法についてはまだ検証されていない。複雑な非言語情報を話者に提示することで話者はより話しやすくなるのか、それとも認知的負荷が上がるのか検証する必要がある。

5.3 聞き手に対する他の聴衆の非言語情報の提示

今回大人数の聴衆の視線と顔の向きを提示するのは話者のみであったが、聞き手に対してもリアルタイムに提示することで聞き手は他の聴衆が注目している部分分かるため、集中力が低下したり聞き手は授業についていけなくなったりすることが減る可能性がある。この他者の非言語情報の可視化が自分自身の非言語情報の可視化 [20] と比較してどの程度有効なのか調査する余地がある。

またリアルタイムでの使用にとどまらず、講義ビデオを視聴中の人に対して過去にそのビデオを観た人たちの視線位置の蓄積が提示されることによって講義中の注目しなければならないタイミングや場所を把握することができると考えられる。

6 むすび

本研究では、web ベースの視線トラッキングライブラリを使用してオンラインコミュニケーションにおいて大人数の聴衆の視線と顔の向きを話者へ集的に提示するシステムを提案し実験を行った。このシステムにより話者は気を散らされることなく聴衆の反応を知ることができるようになった。今後は視線と顔の向きに加えて表情といった複雑な非言語情報を集的に提示すると同時に聞き手自身に対する非言語情報のフィードバックについても検証することも目指す。

参考文献

- [1] K. Betts. Lost in translation: Importance of effective communication in online education. *Online Journal of Distance Learning Administration*, 12(2), 2009.
- [2] S. E. Brennan, X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3):1465–1477, 2008.

- [3] S. D'Angelo and A. Begel. Improving communication between pair programmers using shared gaze awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6245–6290, 2017.
- [4] S. D'Angelo and D. Gergle. Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2492–2496, 2016.
- [5] S. D'Angelo and D. Gergle. An eye for design: gaze visualizations for remote collaborative work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.
- [6] K. Fujii, P. Marian, D. Clark, Y. Okamoto, and J. Rekimoto. Sync Class: Visualization system for in-class student synchronization. In *Proceedings of the 9th Augmented Human International Conference*, pp. 1–8, 2018.
- [7] K. Gyllstrom and D. Stotts. Facetop: Integrated semi-transparent video for enhanced natural pointing in shared screen collaboration. *May*, 15:1–10, 2005.
- [8] K. Higuch, R. Yonetani, and Y. Sato. Can eye help you? Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5180–5190, 2016.
- [9] S. Hutt, J. Hardey, R. Bixler, A. Stewart, E. Risko, and S. K. D'Mello. Gaze-Based Detection of Mind Wandering during Lecture Viewing. *International Educational Data Mining Society*, 2017.
- [10] R. Johansen. *Groupware: Computer support for business teams*. The Free Press, 1988.
- [11] K. Jordan. Initial trends in enrolment and completion of massive open online courses. *International Review of Research in Open and Distributed Learning*, 15(1):133–160, 2014.
- [12] A. F. Monk and C. Gale. A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3):257–278, 2002.
- [13] M. Mukhopadhyay, S. Pal, A. Nayyar, P. K. D. Pramanik, N. Dasgupta, and P. Choudhury. Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System. In *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*, pp. 107–115, 2020.
- [14] A. Papoutsaki, A. Gokaslan, J. Tompkin, Y. He, and J. Huang. The eye of the typer: a benchmark and analysis of gaze behavior during typing. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 2018.
- [15] A. Papoutsaki, J. Laskey, and J. Huang. Searchgazer: Webcam eye tracking for remote studies of web search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pp. 17–26, 2017.
- [16] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence-IJCAI 2016*, 2016.
- [17] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.
- [18] T. Robal, Y. Zhao, C. Lofi, and C. Hauff. IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking. In *Proceedings of the 29th on Hypertext and Social Media*, pp. 106–114, 2018.
- [19] T. Robal, Y. Zhao, C. Lofi, and C. Hauff. Webcam-based attention tracking in online learning: A feasibility study. In *23rd International Conference on Intelligent User Interfaces*, pp. 189–197, 2018.
- [20] K. Sharma, H. S. Alavi, P. Jermann, and P. Dillenbourg. A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pp. 417–421, 2016.
- [21] K. Sharma, P. Jermann, and P. Dillenbourg. "With-me-ness": A gaze-measure for students' attention in MOOCs. In *Proceedings of international conference of the learning sciences 2014*, No. CONF, pp. 1017–1022. ISLS, 2014.
- [22] O. Špakov, D. Niehorster, H. Istance, K.-J. Rähä, and H. Siirtola. Two-Way Gaze Sharing in Remote Teaching. In *IFIP Conference on Human-Computer Interaction*, pp. 242–251. Springer, 2019.
- [23] R. Vertegaal. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 294–301, 1999.
- [24] F. Yang, Z. Jiang, C. Wang, Y. Dai, Z. Jia, and K. Hirota. Student Eye Gaze Tracking During MOOC Teaching. In *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, pp. 875–880. IEEE, 2018.
- [25] N. Yao, J. Brewer, S. D'Angelo, M. Horn, and D. Gergle. Visualizing gaze information from multiple students to support remote instruction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2018.
- [26] Y. Zhao, C. Lofi, and C. Hauff. Scalable mind-wandering detection for MOOCs: a webcam-based approach. In *European Conference on Technology Enhanced Learning*, pp. 330–344. Springer, 2017.