

図 1: Work Engagement, Burnout, Workaholic の関係性

ある [11]. Oerlemans が構築した仕事に関連した主観的幸福感の二次元的な見方 [12] を簡略化した形式で Work Engagement, Burnout, Workaholic の関係性を図 1 に示す。

2.2 参加者へのフィードバック

遠隔会議における参加者へのフィードバックをする手法として, Samrose らは参加者ごとにパーソナライズされたダッシュボードである MeetingCoarch の構築を提案している [2]. COVID-19 以降, 遠隔会議システムは以前よりもはるかに多くの企業にとって不可欠となった一方で, 参加者の表情や反応を相互に伝達する際に問題が残る. MeetingCoach は, 参加者にコンテキストおよび行動の会議情報を要約して提供する. また, 話者交替, 感情などの信号を識別する. MeetingCoarch の設計は, 実際の遠隔会議の参加者によって評価された. 彼らは, 4 週間にわたって 8 チームとの電話会議を記録するためのツールを構築した. ダッシュボードを作成することにより, 参加者の参加意欲の向上に役立つことを示した. 包括的な課題としては, (1) データとダッシュボードにアクセス可能な者を設定することや, (2) 機密性の高い会議の内容を外部分析から除外する手法が挙げられる. また設計の課題としては, (1) 実装の観点から, 提案は必然的にコンテキストに依存するため, すべての会議に提案を適用することは困難である, (2) 例えば, 話者交替のパターンは, 計画に関する会議, プレゼンテーションなど, 会議の種類によって大きく異なる, (3) 会議のコンテキストを理解するために, 単に会議のオーディオビデオ信号をキャプチャーするだけでなく, 会議の種類とチーム文化に関する詳細情報を必要とするということが挙げられる.

2.3 表情分析手法

表情を分析する手法として, Baltrušaitis らは OpenFace を提案している [13]. OpenFace は, 顔のランドマー

ク検出, 頭部のポーズ推定, 顔の Action Unit の認識, 視線の推定が可能なオープンソースあツールである. OpenFace はリアルタイム性能を備えており, 特別なハードウェアを必要とせず, 単純なウェブカメラから実行することが可能である. 顔のランドマークの検出と追跡には Conditional Local Neural Fields (CLNF)[14] を使用している. 彼らが提案した CLNF モデルでは, 顔にある 68 個のランドマークをすべてまとめて検出する. また, 顔の Action Unit の認識は, [15, 16] のフレームワークに基づいている. 実験では 100 ピクセル以上の顔画像に対して適切な結果が得られることが示されている. また, 感情の表し方は文化や環境により異なり, 普遍的なものではない [17, 18].

2.4 音声からの話者交代推定手法

音声処理アルゴリズムに関して, Tashev は音声と雑音の異なる確率密度関数に基づいたオフライン音声活動検出器 (Voice Activity Detector: VAD) を提案した [19]. 提案された VAD アルゴリズムは, 周波数領域で動作し, 各オーディオフレームの各周波数に対する音声信号の存在確率と, 各フレームの音声存在確率を推定し, さらにフレームごとに二値判定を行う. これによりグループディスカッション参加者の話者交代を検出することが期待される.

2.5 心拍からの感情推定手法

Harper らは, 単峰性心拍時系列から感情の価値を分類する End-to-end の深層学習モデルを提案している [20]. また, これらの価値予測に対する不確実性をモデル化するためのベイジアンフレームワークを提案している. これらの結果は, 非侵襲的なデータ収集と予測の確実性が重視されるヘルスケアなどの実世界の領域における感情コンピューティングの応用の基礎となると述べている.

3. 対象とするグループディスカッションデータ

本章では, 本研究で対象とするグループディスカッションデータについて説明する. 本研究では, Fitbit Ionic^{*}, ラップトップに搭載されたウェブカメラ, マイク, 遠隔会議サービス Zoom[†], グループディスカッション参加者による他者の発言に対するアノテーションから収集されたグループディスカッションデータを活用して, 以下に説明する Work Attitude を認識, 分析することを目的としている.

3.1 概要

本研究の目標は遠隔会議の参加者それぞれに対し, 他者の発言に対するアノテーション付きグループディスカ

^{*}<https://www.fitbit.com/>

[†]<https://zoom.us/>

ションデータから抽出したマルチモーダルデータを用いて、感情推定モデルを複数の機械学習アルゴリズムにより構築・評価可能な Work Attitude 推定手法について検討する。

3.2 対象とするグループディスカッションデータ

本研究では、日本大学酒井元気准教授から提供いただいたグループディスカッションデータセットを用いる。このデータセットには、(1) 遠隔会議サービスを用いた被験者3から4人が写ったグループディスカッションのギャラリー映像、(2) 発話者のみが写った映像、(3) 各被験者の音声、(4) 音声を用いて遠隔会議用議事録音サービス Talk to CSV[‡]により生成した議事録、(5) ウェアラブルデバイスから取得された各被験者の心拍データ、(6) 他者の発言に対するアノテーションが含まれる。

(1) ギャラリー映像

遠隔会議サービスの表示におけるギャラリーモードを用いて実現している。グループディスカッション実験に参加している被験者全員が表示される映像をMP4形式で保存している。2020年11月11日15時15分から15時30分までに行われた遠隔会議サービスを用いたグループディスカッションのギャラリー映像から抽出した1フレームの例を図2に示す。

(2) 発話者のみが写った映像

遠隔会議サービスの表示におけるスピーカーモードを用いて実現している。グループディスカッション実験に参加している被験者が発言した際に画面全体に表示される映像をMP4形式で保存している。

(3) 各被験者の音声

遠隔会議サービスの「参加者ごとに個別のオーディオファイルで録音」モードを用いて実現している。グループディスカッション実験に参加している被験者の音声をM4A形式で保存している。上記グループディスカッションのうち被験者Aの音声の例を図3に示す。

(4) 議事録

Talk to CSVにより録音した各被験者の音声をテキスト変換することで実現している。CSV形式で保存している。カラムインデックスは時間、発話者、発話内容となっており時系列で保存している。

(5) 心拍データ

各被験者にウェアラブルデバイスを装着してもらい、そこから心拍を取得することで実現している。Fitbit SDKにおけるWeb APIを用いて心拍デー



図 2: 映像データ

表 1: 感情アノテーション

時間	感情ラベル	評価対象の参加者
2020-11-11 15:15:11	3	D
2020-11-11 15:15:21	2	B
2020-11-11 15:15:53	5	D
2020-11-11 15:15:59	4	C
2020-11-11 15:16:08	5	D
2020-11-11 15:17:08	3	C
2020-11-11 15:17:26	2	B
2020-11-11 15:18:20	5	B
2020-11-11 15:18:32	4	C
2020-11-11 15:18:53	3	B

タを抽出し、JSON形式を変換してCSV形式で保存している。上記グループディスカッションのうち被験者Aの心拍波形の例を図4に示す。

(6) 他者の発言に対するアノテーション

遠隔会議サービスによるグループディスカッション収録後に、発話者のみが写った映像を用いて、各被験者がグループディスカッションに同席していた他の被験者の発言に対して、1点から5点で評価してもらう。悪ければ1点、良ければ5点の評価をつけることでCSV形式で実現している。上記グループディスカッションのうち被験者Aによる、他者の発言に対するアノテーションの例を表1に示す。

4. Work Attitude 自動評価システム

本章では、グループディスカッション参加者の Work Attitude 自動評価システムについて説明する。本システムでは、遠隔会議サービスやウェアラブルデバイス、グループディスカッション参加者が他者に対して行ったアノテーションから得られたデータから、自動的に参加者の Work Attitude の評価を行う。理由としては、カメラ

[‡]<https://skybeje.net/talktocsv/chatowner/>

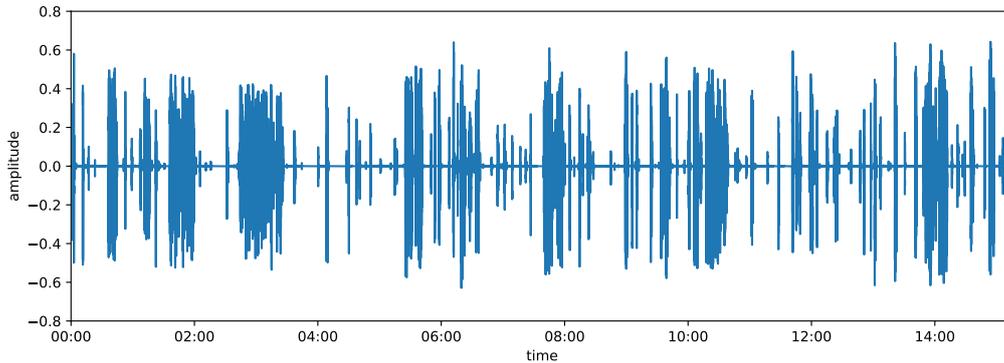


図 3: 音声データ

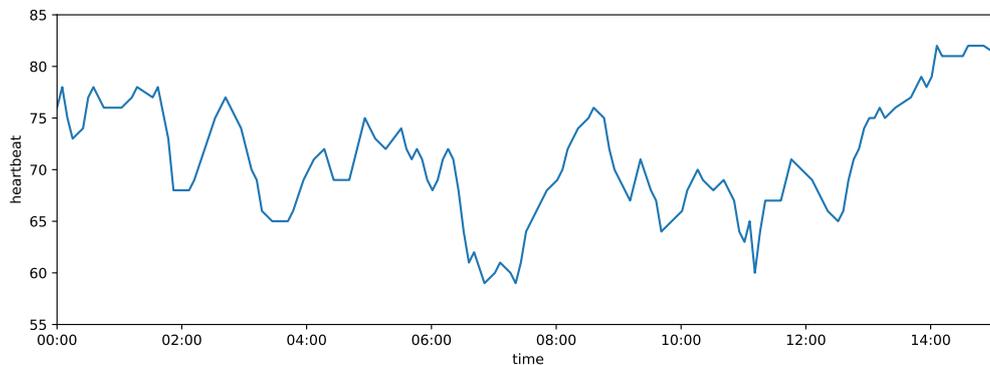


図 4: 心拍データ

などを用いる手法ではカメラの設置や画角の調整といった手間が生じてしまうからである。しかし、本支援システムでは、カメラを準備する手間が省けるため手軽に棒体操を行うことが可能である。我々は本研究の最終目標として、遠隔会議サービスやウェアラブルデバイス、グループディスカッション参加者が他者に対して行ったアノテーションを用いて、自動的に参加者の Work Attitude を評価するシステムの実現を目指している。

本システムは、3つのフェーズで構成されている。(1) 遠隔会議サービスとウェアラブルデバイスを用いてグループディスカッションを実行・記録するフェーズ、(2) グループディスカッション参加者に他者の発言に対するアノテーションを記録された音声・映像をもとに行なってもらふフェーズ、(3) 音声・映像・心拍・テキスト解析をすることでマルチモーダルにグループディスカッション参加者の Work Attitude を評価するフェーズである。

5. 実験

本章では、2020年11月11日15時15分から15時30分までに行われた遠隔会議サービスを用いたグループディスカッションの際に取得された、(1) 音声データ

のテキスト抽出、(2) 映像データの表情検出を行った結果について述べる。

5.1 音声データのテキスト抽出

Talk to CSV を用いて、個別に保存されていた被験者4人の音声データをテキスト化し、発言した時間に合わせて統合を行った。被験者Aの音声データを図3に示す。図より、発話のタイミングが確認することができる。また、この音声データをフーリエ変換するなどすることにより、声質などを取得できるようになると考える。また、4人の被験者がグループディスカッションをした内容を実際に統合して生成したテキストデータを表2に示す。テキストデータを確認することで、与えられたテーマに沿った発言をしているかどうかを判断することや、ポジティブ・ネガティブ分析が可能となり、感情アノテーションと統合して発言の適切性を確認することが可能となる。

5.2 映像データの表情検出

OpenFace の映像内に複数人が映っている場合に表情認識を行う FaceLandmarkVidMulti を用いて、複数人の顔のランドマーク検出を行った。4人の被験者の映像

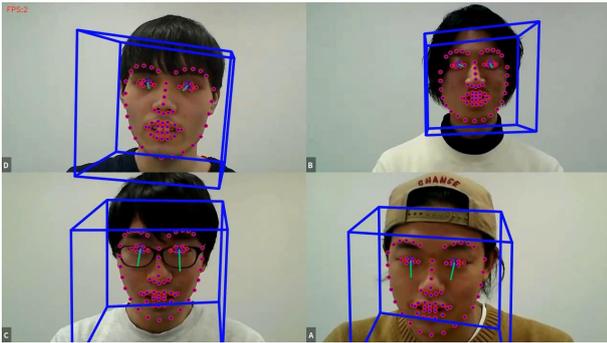


図 5: OpenFace で映像データ解析した結果

の 1 フレームを図 2 に示す。また、4 人の被験者の映像を実際に OpenFace を用いて顔のランドマーク検出を行って生成した映像を図 5 に示す。ランドマーク検出をすることにより、顔の傾き、目線、傾きがわかるため、それらの情報を用いることで感情アノテーションとの比較をすることが可能となる。

6. まとめ

本稿では、遠隔会議サービスを用いたグループディスカッションにおける work Attitude 評価システムの実現に向けた初期段階として、OpenFace による映像解析を行った。また、テキスト解析のために音声からの議事録データ作成を行った。今後は、機械学習を用いて、音声・映像・心拍・テキストデータをマルチモーダルに用いて遠隔会議サービスを用いたグループディスカッションにおける work Attitude 評価システムの実現を目指す。

謝辞

本研究の一部は、Society 5.0 実現化研究拠点支援事業および科研費基盤研究 (B)(No.19H01719) の助成によって行った。

参考文献

- [1] Sandi Mann and Lynn Holdsworth. The psychological impact of teleworking stress, emotions and health. *New Technology Work and Employment*, Vol. 18, No. 3, pp. 196–211, 2003.
- [2] Samiha Samrose, Jina Suh, Sean Rintel, Daniel McDuf, Kael Rowan, Kevin Moynihan, Robert Sim, Javier Hernandez, and Mary Czerwinski. Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings. *Conference on Human Factors in Computing Systems*, Vol. 16, No. 252, pp. 1–13, 2021.
- [3] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. Meeting effectiveness and inclusiveness in remote collaboration. *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. 173, pp. 1–29, 2021.
- [4] Janet T. Spence and Ann S. Robbins. Workaholism: Definition, measurement, and preliminary results. *Journal of Personality Assessment*, Vol. 58, No. 1, pp. 160–178, 1992.
- [5] Herbert J. Freudenberger. Staff burn-out. *Journal of Social Issues*, Vol. 30, No. 1, pp. 159–165, 1974.
- [6] Christina Maslach and Susan E. Jackson. The measurement of experienced burnout. *Journal of Occupational Behaviour*, Vol. 2, pp. 99–113, 1981.
- [7] Christina Maslach, Wilmar B. Schaufeli, and Michael P. Leiter. Job burnout. *Annual Review of Psychology*, Vol. 52, pp. 397–422, 2001.
- [8] Hely Innanen, Asko Tolvanen, and Katariina Salmela-Aro. Burnout, work engagement and workaholism among highly educated employees: Profiles, antecedents and outcomes. *Burnout Research*, Vol. 1, No. 1, pp. 38–49, 2014.
- [9] Wayne E. Oates. On being a "workaholic". *Pastoral Psychology*, Vol. 19, pp. 16–20, 1968.
- [10] Lynley H. W. McMillan, Michael P. O'Driscoll, Nigel V. Marsh, and Elizabeth C. Brady. Understanding workaholism: Data synthesis, theoretical critique, and future design strategies. *International Journal of Stress Management*, Vol. 8, pp. 69–91, 2001.
- [11] Peter E. Mudrack and Thomas J. Naughton. The assessment of workaholism as behavioral tendencies: Scale development and preliminary empirical testing. *International Journal of Stress Management*, Vol. 8, pp. 93–111, 2001.
- [12] Wido Oerlemans. Subjective well-being in organizations. *The oxford handbook of positive organizational scholarship*, pp. 178–189, 2011.
- [13] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: An open source facial behavior analysis toolkit. *Winter Conference on Applications of Computer Vision*, 2016.

表 2: 議事録データ

時間	話者	内容
2020-11-11 15:18:30	C	お願いします
2020-11-11 15:18:33	D	明日鍋てる自分が書記やるんでチャットの方に入ってた相手はどんどん書き込んでいきます
2020-11-11 15:18:44	B	マイクロプラスチック女ためになっというテーマですよねちょっと
2020-11-11 15:18:46	D	レシピを英語で
2020-11-11 15:18:50	B	マイクロプラスチックって何でできるんですかね
2020-11-11 15:18:53	D	何からで金 5 mm かの
2020-11-11 15:18:55	B	芸人がわかんないなあ
2020-11-11 15:19:06	A	なんかそれこそ普通にペットボトルとかの飲み物がちっちゃくなってってそのまま海に流れるみたいな感じじゃないんですかね
2020-11-11 15:19:15	D	何かあれすかねその不法投棄されたものが溶けたりするんですかね受けて小さくなっ
2020-11-11 15:19:21	A	うん

- [14] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Constrained local neural fields for robust facial landmark detection in the wild. *International Conference on Computer Vision Workshops*, 2013.
- [15] Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson. Cross-dataset learning and person-specific normalisation for automatic action unit detection.
- [16] Jessica L. Tracy and David Matsumoto. The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *the National Academy of Sciences*, Vol. 105, No. 33, pp. 11655–11660, 2008.
- [17] José-Miguel Fernández-Dols and Maria-Angeles Ruiz-Belda. Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. *Cambridge University Press; Editions de la Maison des Sciences de l’Homme*, pp. 255–274, 1997.
- [18] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Alex M. Martinez, and Seth D. Pollak. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, pp. 1–68, 2019.
- [19] Ivan J. Tashev. Offline voice activity detector using speech supergaussianity. *Information Theory and Applications Workshop*, 2015.
- [20] Ross Harper and Joshua Southern. A bayesian deep learning framework for end-to-end prediction of emotion from heartbeat. *Transactions on Affective Computing*, p. 1, 2020.