

オンラインミーティングを対象とした会議の質評価システムの設計と構築

曾根田悠介[†] 荒川 豊^{†‡} 安本 慶一[†][†] 奈良先端科学技術大学院大学 [‡] JST さきがけ

1. はじめに

現在日本では、働き方改革と叫ばれるように勤務時間の減少が求められるほか、育児や介護の両立のために在宅からのリモートワークの推奨が積極的に行われている。しかし、会議室等で机を囲い直接顔を合わせてのミーティングが主流であり、オンラインでのミーティングは以前として一般的ではない。対面でのミーティングと比べてオンラインでのミーティングであることにメリットを付加させることにより、オンラインミーティングの普及に大きく貢献する事が可能と考える。

また、仕事の質の向上を求める上で、多くの時間を占めるミーティングの質の向上は必要不可欠である。ミーティングの評価に関するガイドラインも存在するが、定性的な指標がほとんどであり定量的にミーティングの評価を試みている研究は乏しいというのが現状である。会議に関するサービスや研究は、ミーティングのログを記録し、後から参照できるといったものが多くを占めるが、ミーティングの質の評価までは行われていない。

本研究では、オンラインミーティングを対象とし、ミーティングの質の定量的な評価を行うためのシステムの設計と構築を行った結果について報告する。オンラインミーティングに限定することでユーザ毎の音声情報と映像情報を容易に取得できる。本研究の狙いは、これらの情報を解析することによって、どのようなデータがミーティングの定量的な評価に繋がる指標になるかを明らかにすることである。

本稿の構成は以下の通りである。第2章では、関連研究について述べる。第3章では、ミーティングから抽出する指標の検討について述べる。第4章では、構築するシステムについて述べる。最後に、第5章で本稿のまとめと今後の課題を述べる。

2. 関連研究

Samiha^[1]らの研究によると、2回ミーティングを行いそのインターバルに1回目のミーティングから得られた情報を提示することによって、2回目のミーティングでは参加者全員が発言に参加できるような行動変容が促されたと報告されている。「全会話時間中の支配率」、「笑顔の回数」、「会話の割り込みされた回数、割り込みを行なった回数」、「笑顔や驚きといった表情」、「発言の内容がポジティブであるかネガティブであるか」、「発言者の移り変わり」の6つを音声情報と映像情報から取得し、ミーティング参加者に提示している。しかし、この提示方法ではリアルタイムでのミーティングの質の向上には繋がらず、同じ人物同士のミーティングが前提となる。また、提示された6つの情報のうちどの情報が影響しているのかといった評価なされていない。解析される顔の表情も「集中」「笑顔」「驚き」といった感情をWEBカメラから取得しているが、ミーティング中に発生する感情は多様であり、より細かい表情を取得する必要がある。

ある。

本研究で提案するシステムでは、仕草といった身体の動きを含めて解析し、ミーティングの理解度、集中度を求めることによって、ミーティングの質の評価を行うためのシステムの設計と提案を行う。

3. ミーティングから抽出する指標の検討

ミーティングを定量的に評価するために検討した指標案を表1に示す。いずれも、ミーティングの音声情報と映像情報から抽出するが、音声情報に関しては、プライバシー問題およびコンテンツ依存性を抑えるため、発言内容は扱わず音量や声のトーンといった外形情報のみ取り扱う。

表1 ミーティングから抽出する指標の検討結果

特徴名	詳細について
会話支配率	音声情報からミーティングに参加している者が、全体の時間に対して発言した時間の割合を会話支配率と定義する。どの人物がミーティングにおいてどのような影響を及ぼすのか推定を行う。
発言者の移り変わり	音声情報を用いてミーティング参加者がどのような順番で発言を行なったかを解析する。ミーティング参加者の関係性や、ミーティングにおいての役割が明らかになると期待される。
詳細な表情	Samiha ^[1] らの研究では、Affdex ^[2] を用いて表情を「集中」「笑顔」「驚き」「怒り」に分割し評価を行なっている。当研究では、ミーティングにおける表情変化はさらに多様なものと考え、理解度や集中度を顔の動きから推定する。
仕草	Peter ^[3] の研究によると、傾きといった身体の仕草はその人物の心理状況を表現したり、人の印象や心情の表現に大きく影響し、非言語コミュニケーションとして重要であるという結果が報告されている。3.1節では、さらに仕草について詳細に記述する。

3.1 仕草についての詳細

表1の中で、特に仕草には様々な種類が含まれ、仕草を行う者の心情が無意識に表れている。また、仕草を受け取る者にもそれぞれ異なる印象を及ぼす。ここでは、Peter^[3]の論文を参照し、どのような仕草がミーティング参加者の心情に影響しているかを調査し、認識を試みる仕草について記述する。

3.1.1 話題に対して興味を持っていることを表現する仕草

身体を前に傾ける、発言中に足を後ろに引く、傾くという仕草はその人物が話題に対して興味を持っている事を表す。特に傾きは、話者にとって話を聞いているという印象を与え、発言しやすさを促すという効果がある。

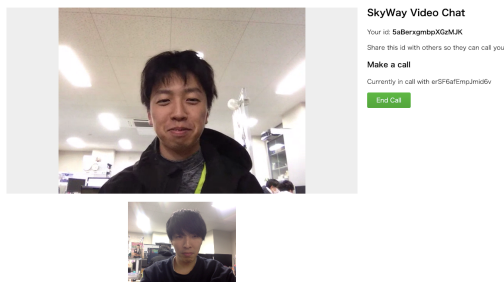


図1 WebRTC を用いて構築したオンラインミーティング

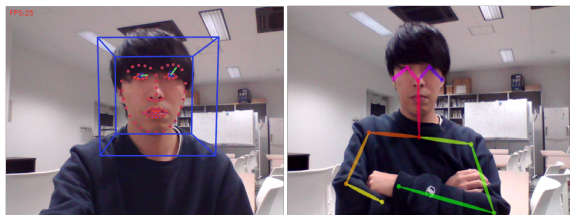


図2 OpenFace と OpenPose を用いて解析した図

3.1.2 話題に対して興味を持っていないことを表現する仕草

身体を後ろに傾ける、片手で頭を支える、頭を落とす、頭の位置を常に動かす、腕の位置を常に動かすという仕草はその人物が話題に対して興味を持っていないことを表す。

4. 構築するシステムについて

Google 社が主に開発している WebRTC^[4] というブラウザ間で Peer to Peer 通信を可能にするフレームワークを使用し、オンラインミーティングの基盤になる通信システムを構築する。また、WebRTC を使用することによって、PC に内蔵されているカメラとマイクの情報を取得することが可能である。「会話支配率」と「発言者の移り変わり」に関しては、各自の PC 上で取得した音声情報を解析することによって求める。実際に構築した WEB ミーティングの画像は図 1 に示す。WebRTC を用いることによって、本実験に適した WEB ミーティングのシステムを容易に構築する事が可能になる。

本研究では、映像情報から得られる特徴を OpenFace^[5] と OpenPose^[6] というオープンソースのソフトウェアを用いて抽出を行う。OpenFace と OpenPose を用いて解析した時、図 2 のようになる。一般的に、顔や身体の詳細な動きを解析するには、前田ら^[7]の研究のように Kinect のような深度センサを用いる事が多いが、深度センサが組み込まれている PC は普及していない。したがって、本研究では、PC に内蔵されている単眼カメラで解析が行える OpenFace と OpenPose を用いる。

これらの二つのソフトウェアは、ブラウザ上では実行することが不可能なため、PC のインカメラで取得した映像情報を解析するために、OpenFace と OpenPose を実行するサーバを用意する。以上の提案したシステムに関する構成図を図 3 に示す。

4.1 OpenFace

OpenFace は、顔画像を入力とし顔の特徴点のポジションや視線の向きを解析するオープンソースのソフトウェアである。顔の特徴点は 67 点、目の特徴点は 55 点の特徴点を持ち、目、口、鼻の輪郭を認識することが可能となる。

Samiha らが用いた Affectex では喜びや怒り、驚きといった大きな表情とフレームごとの解析のため静止画で求められる情報のみ取得できる。しかし傾きといった、時系列に影響する情報に関して取得する事は不可能である。OpenFace を用いることで、傾きといった時系列に依存する情報とさらに詳細な顔の動

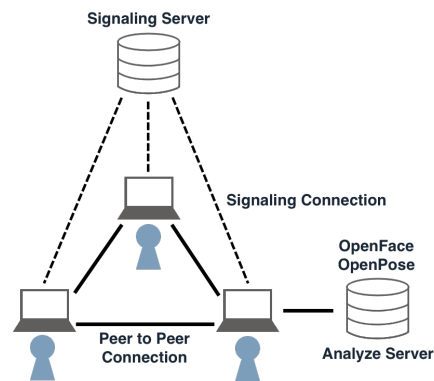


図3 システム構成図

きの取得を試みる。

4.2 OpenPose

OpenPose は、腕や脚といった関節の位置を解析するオープンソースのソフトウェアである。出力される特徴点は 15 点で座標の情報が求められる。出力された情報から身体の姿勢や仕草を認識することが可能になる。

オンラインミーティングでは上半身のみ撮影が可能のため、主に頭の傾き、腕の位置、身体の傾きを取得する。

5. おわりに

本研究では、オンラインミーティングを対象とし、音声と映像情報から参加者の集中力や理解度を推定できる要素の定義とその解析システムの提案と構築を行った。

推定を行う仕草を本研究では定義したが、実際に収集された映像情報からさらにどのような仕草を推定するかについて考察する。

ミーティングの質の評価を行うことを目標とし、さらにどのような情報を提供すれば参加者の支援を行えるか検討していく。

謝辞

本研究は、JST さきがけ、および、Society 5.0 実現化研究拠点支援事業、の支援のもと実施されている。

参考文献

- [1] Samiha Samrose, Ru Zhao, Jeffrey White, Vivian Li, Luis Nova, Yichen Lu, Mohammad Rafayet Ali, and Mohammed (Ehsan) Hoque. Coco: Collaboration coach for understanding team dynamics during video conferencing. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 1, No. 4, 2017.
- [2] Affectiva. affectex. <https://affectiva.jp/>.
- [3] Peter E Bull. *Posture and Gesture*. Pergamon Press, 1987.
- [4] Google LLC. <https://webrtc.org/>.
- [5] Openface. <https://github.com/TadasBaltrusaitis/OpenFace>.
- [6] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.
- [7] Naoki Maeda, Yuko Hirabe, Yutaka Arakawa, and Keiichi Yasumoto. Cosms : Unconscious stress monitoring system for office worker(demo). *The ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 329–332, 2016.