



会話ノードマッピングと視線検出を利用した会話支援技術の有効性評価

Evaluating the Effectiveness of a Dialogue Support System with Conversation Node Mapping and Gaze Detection

山下純平¹⁾, 齊藤寛人¹⁾, 高下修聡¹⁾, 櫻田国治¹⁾, 門内靖明¹⁾, 稲見昌彦¹⁾

Jumpei YAMASHITA, Hiroto SAITO, Shuto TAKASHITA, Kuniharu SAKURADA, Yasuaki MONNAI, Masahiko INAMI

1) 東京大学 (〒113-8656 東京都文京区本郷 7-3-1, {j.yamashita | hiroto.saito | shuto.takashita | kuniharu.sakurada | monnai | drinami}@star.rcast.u-tokyo.ac.jp)

概要: 本研究ではハイコンテキスト文化における、会議のオンライン化による非言語情報の欠落とそれに伴う情報伝達効率の低下を課題として設定した。これに対し、既存の会話ノードマッピング手法に、マップ上に対する視線検出を利用した話者へのフィードバックを統合する手法を提案した。実験によって、話し手が聞き手の注目などの意図を推定可能であり、安心感を得られることが明らかになったが、同時にその推定した意図が本当に聞き手のそれと合致しているかは未解明である。今後、聞き手の注目の理由を推定・分類するシステムを開発し、信頼性の高いシステムを目指す。

キーワード: 非言語コミュニケーション; 会話構造ノードマッピング; アイトラッキング

1. はじめに

コミュニケーションは人間が集団で協力し社会活動をしていく上で不可欠なものである。この二者の関係性を説明するものとして共有メンタルモデル(Shared Mental Model:SMM)[1]という概念が挙げられる。これは、目標、タスク、役割、プロセス、環境といった業務の主要な要素に関して、チームメンバー間で共有されている理解のことである。一方で、自分の持っている知識を相手も知っているはずだと感じる「知識の呪い」という認知バイアスなどから生まれる齟齬や誤解によって SMM の構築が困難になる[2][3]。こうした齟齬や誤解の解消のためには、コミュニケーションの失敗を相手の視線や表情、態度といった非言語情報やメタ認知の明示的な確認などから理解することが重要である[4]。

しかし、コロナ禍に伴い、多くの学校や企業においてリモートワークやオンライン授業・会議環境が普及し、人々は情報メディアを通じてコミュニケーションを取ることが多くなった。こうした非対面形式のコミュニケーションにおいては、ジェスチャーやアイコンタクトといった SMM の構築に重要である非言語情報が欠落する。こうした非言語情報の欠落に対して言語情報でコミュニケーションすることでチームとしてのパフォーマンスを維持することが示唆されている(Virtuality Paradox[5])。しかし、この手法は「空気を読む」「目は口ほどに物を言う」といった

ハイコンテキスト文化の強い日本などでは相性が悪く、依然として非効率性を生みやすくなっている [6]。

こうした非言語情報をはじめとした情報の欠落による SMM の構築困難性という課題に対し、少数派を考慮した AI による要約やファシリテーション等の技術が開発されている[7]。その一つに LLM を用いた話題毎の論理構造を可視化するノードマッピングという方法がある[8][9]。このような自動ノードマッピングに注目し、その有効性を調査した研究では認知負荷の低減[8]や会話における対立の解消[9]といった効果が報告されている。一方で、会話中に併用するこうしたツールからの情報取得は、相手の表情等の非言語情報の取得とトレードオフの関係にあり、SMM の構築を支援するために十分であるとはいえない。

本研究では、議論の論理構造と非言語情報を統合した新たな会議支援のフレームワークの提案し、その有効性を評価することを目的とする。具体的には、ノードマップで表現できる情報と非言語情報をユーザが同時に取得することを可能とするため、ユーザの感情や考え、あるいはその表れである非言語情報をマップ上に重畳することが有効であるという仮説を検証する。そして、齟齬が起きやすい場面として異なる専門分野を持つ話し手と聞き手のプレゼンテーション形式のコミュニケーションという場面を設定し、非言語情報の一つを計測するアイトラッキングを利用した実験を実施する。

2. 研究手法

2.1 システム概要

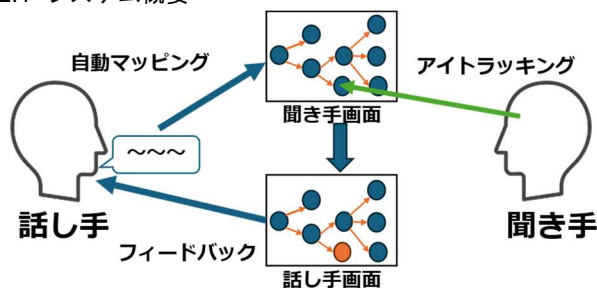


図 1. システム概要図

本システムは、LLM を利用したノードマッピングソフトウェアと Tobii Pro Nano を利用したアイトラッキングによって構成される(図 1)。

ノードマッピングソフトウェアは OpenAI の Whisper を使用して文字起こしを実施する。文章中には誤変換や聞き取りミスが多いため o4-mini モデルを使用して文章の整形を実施し、同時にノードマップとして JSON 形式に変換する。JSON 形式のデータをもとに、Matplotlib を使用してノードマップとして表示する。

実験では話し手と聞き手で別の画面を提示する。まず、聞き手用インターフェース画面は二画面で構成される。ひとつはノードマップウィンドウであり、前述したシステムにより生成されたノードマップを表示する(図 2 左)。もう一つはノードマップをクリックした際にその話題の詳細を表示するウィンドウである(図 2 右)。そして、話し手用インターフェース画面は一つのウィンドウで表示される(図 3)。こちらでは聞き手と同じノードマップに重畳して、現在の注目点と直近 10 秒間の聞き手の注目点を注目時間/計測時間に基づいて色の濃さで表示する。ここで、注目点の判定は、左画面上でそのノードを見ている場合、あるいは

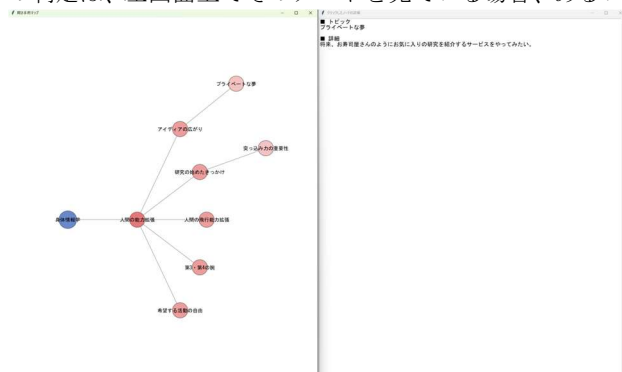


図 2. 聞き手用インターフェース画面

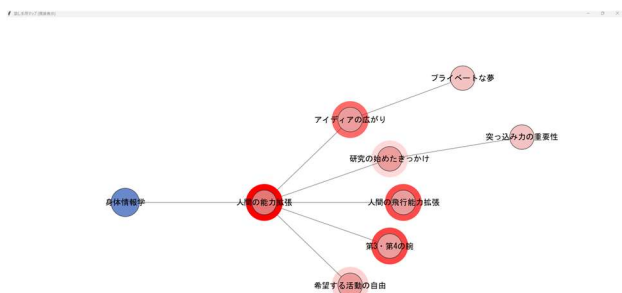


図 3. 話し手用インターフェース画面

は右画面上でそのノードの詳細を見ている場合とした。

2.2 実験環境

実験においては、オンライン会議を想定し、会話環境として Discord を使用する。話し手はオンラインで参加し、画面共有機能を通して話し手用インターフェースを提示される。聞き手は著者と同じ環境から参加し、聞き手用インターフェースを提示される。同時に、聞き手に対しては Tobii Pro Nano を利用したアイトラッキングを実施する。

話し手と聞き手は異なる専門分野を持ち、話し手の内容は自身の研究内容等、専門的な内容を話す。

2.3 実験手続き

最初に Tobii Pro Nano のキャリブレーションを実施した。

話し手と聞き手は Discord を通じて会話し、条件に応じて前述したそれぞれのインターフェース画面を提示された。加えて、話し手には聞き手の表情も表示した。

次に、話し手が自身の専門に関わる内容についてスピーチ形式で話した。この時、以下の 3 条件について、先の 2 条件をランダムな順番で実験した後、3 つ目の条件で実験を実施した。

1. 顔映像に加え聞き手用画面のみを使用する場合
2. 顔映像に加え両者にそれぞれの画面を提示する場合
3. 顔映像のみを提示する場合

各条件で会話終了後、評価プロセスに移行した。

2.4 評価

会話終了後、「両者にツールの利用しやすさ」、「話し手の話の理解を補助するためのツールの利用方法」、「聞き手の考えていた内容の推論の可否と方法」、「ツールの改善提案」について質問する半構造インタビューを実施した。インタビューは話し手→聞き手の順番で実施された。また、議論の内容がどの程度伝わっていたかを明らかにするため、参加者間でプレゼンテーションの内容に対する質問や確認などのディスカッションを実施した。

3. 結果

3.1 分析方法

本研究では、ノードマップを用いた対話における参加者の主観的経験を解明するため、質的調査法の一つであるテーマ分析 (Thematic Analysis) を採用した。分析対象は、実験参加者 (N=2 組) に実験直後に実施した半構造化インタビューの逐語録である。

分析は、Braun & Clarke が提唱した手順[10]に準拠し、以下のステップで進めた。

1. データへの精通: まず、全ての逐語録データを繰り返し熟読し、参加者がノードマップの利用についてどのように語っているか、全体的な傾向を把握した。
2. コーディング: 次に、テキストデータから、ノードマップの利用経験に関連する発言を意味のある単位として抽出し、その内容を示すラベル (コード) を付与した。
3. テーマの形成: 最後に、生成されたコード群を分類・整理し、データ全体を貫く共通のパターンや意

味のまとまりを「テーマ」として抽出した。

3.2 分析結果

インタビューと対話内容のテーマ分析の結果、参加者のノードマップ利用経験は、主に3つのテーマに集約されることが明らかになった。それは、(1) 聞き手の注目点という非言語フィードバックの可解釈性、(2) 対話構造の俯瞰と今の話題の立ち位置の把握、および、(3) 会話とノードマップの確認というマルチタスクによる作業負荷の増大である。

3.2.1 聞き手の注目点という非言語情報の解釈可能性

聞き手の現在の視点や10秒間の注目点ヒートマップという機能は本研究のシステム独自のものである。話し手の多くがこれを用いて興味・疑問の所在を類推していたと述べた。また、話し手の一部は聞き手の傾聴、理解、混乱といった状態の類推も実施し、安心感を得られていたと述べた。

この機能の利用について、2人の話し手はそれぞれ以下のような経験を報告した。

「(聞き手1)さんの意図は伝わったんじゃないかなと思います。(中略)どのように感じているのかをリアルタイムに確かめながら話すのはすごく話しやすかった」

「ある程度わかりました、この話題を追っているんだなっていうのを。(中略)今どこを見てらっしゃるか、こっちからこっちまで移った理由とまではわからないんですけど、ここに注目してらっしゃるんだなというところはわかりました。」

また、これに関連して、カメラで撮影された顔の情報に対して、話し手はノードマップ利用の有無に応じた違いを感じられなかったと報告している。

3.2.2 対話構造の俯瞰と今の話題の立ち位置の把握

上記のような本研究独自の機能に加え、先行研究において報告されたノードマップの基本的な利点も本システムにおいて確認された。多くの聞き手が、ノードマップを用いることでプレゼンテーションの全体構造における現在の話題の関連性を把握しやすくなったと報告した。

聞き手の一人は以下のように話した。

「(ノードやその詳細の)内容自体を理解しようというよりは、今どういう話をして、それがどういう位置づけなのかっていうのを、多分ノードマップだとわかりやすいと思うので、(中略)そういった形で利用しました。」

これを利用した改善の提案として、事前に発話内容の原稿や概要を入力しノードマップを作成しておいたり、会話の冒頭で全体のあらすじを述べてから本題に入ったりするなどの手法が参加者から提案された。

3.2.3 会話とノードマップの確認というマルチタスクによる作業負荷の増大

一方で、このような利点と同時に、その利用に伴う課題も先行研究と同様に確認された。話し手、聞き手の両方の参加者が、プレゼンテーションタスクとマップの操作を同時に行うことによる認知的な負荷の高まりを報告した。

ある話し手は「原稿を見ること」「マップの表示を確認すること」「相手のカメラ映像を確認すること」を同時に行う状況について、努力を要したことが報告した。

「聞き手の方の視線を見るのと、ノードマップを見るのと、自分で話すのと、なかなかきれいに連動できなかった」

また、聞き手の二人についても、会話を聞きながらノードマップの詳細など文章を追うことは困難だと報告した。

「追っていくのが結構割と作業負荷になったので、それを頭の中のマップと統合して、というのを考えると結構難しかったので、余計混乱しそうだった」

「詳細は読もうとしても頭に入ってこないというか(中略)話している内容と同時に理解することが難しい、そもそも読んでもちょっとわからない、内容自体が抽象的で(中略)理解は難しかった」

4. 考察

以下では、(1)話し手に対するフィードバック機能、(2)ノードマップの設計の2点について、それぞれ考察する。

4.1 話し手に対するフィードバック機能の有効性評価

本研究では、非対面環境における知的共同作業において、聞き手の視線情報を可視化するシステムを提案し、その有効性を実験的に評価した。実験の結果、聞き手の視線位置をノードマップ上で話し手に提示することが、話し手の発話情報に対する聞き手の興味・傾聴といった非言語フィードバックとして作用した。また、一部の話し手にとっては主観的な安心感を向上させるというフィードバックが得られた。この結果は、非対面コミュニケーションにおいて物理的な身体情報が欠落する状況下でも、相手の注意の所在という限定的な情報を共有するだけで、話し手の心理的安定に寄与しうることを示唆する。これは、対話におけるラポール形成[11]やグラウンディング理論[4]における、注意の共有が果たす役割を支持する一つの証左と位置づけられる。

4.2 ノードマップの有効性評価と設計方針

インタビューの解析より、ノードマップが聞き手に対して「現在の話題が全体構造のどこに位置づけられるか」を即座に示す効果が確認された。これは、先行研究[8][9]の報告と一致する。一方で、詳細を表示する機能については活用が難しいという結果になった。これは参加者が述べたように、現在音声情報として入力される情報と、システム上に表示される詳細の文字情報を並列処理することが困難なためであると考えられる。

以上の評価より、会話の支援においてはノードマップを用いて詳細を表示する機能ではなく、会話の構造を直感的にとらえるという機能に重点を置くべきである。これを踏まえ、ノードマップを用いた会話支援システムでは幾何学的構造の固定化やノードに表示される情報の簡略化などが有効に働くことが示唆された。加えて、事前の話し手の原稿や概要の入力や、会話の冒頭にあらすじを述べてから本題に入るなどの方法により、会話の序盤から全体の構造

を提示する方法が有効であることが示唆された。

4.3 本研究の限界と展望

本研究の限界として、二点挙げられる。

第一に、提案システムは聞き手の視線の位置情報を提示するが、その視線の意味内容、すなわち注目している理由（例：トピックへの興味、内容の理解、あるいは単なる理解困難など）を伝達するようには設計されていない。視線の時空間的なパターン等から何らかの意味内容が話し手に伝わった可能性は否定できないものの、本実験の範囲では、その効果の有無や内容を明らかにすることはできなかった。したがって、視線が持つ意味の解釈は、話し手の主観に委ねられるという課題が残る。本実験では、提示された視線情報に対する話し手の解釈と、聞き手の実際の意図との一致度を測定していないため、システムを介した相互理解の質については評価不能である。

同時に、この解釈の不確実性にもかかわらず話し手の安心感が向上したという結果は、聞き手の真意とは無関係に、話し手が自身の認知的な枠組みの中でポジティブな解釈を行った可能性を示唆する。この点が、システムの有効性を評価する上での重要な留保事項となる。

第二に、本研究は2組のペアでのみ実施したため、結果の一般化可能性については注意が必要である。特に、詳細の表示機能を除去することは、内容の忘却や現在の話題への追従において課題が発生する可能性がある。

本研究で明らかになった限界点は、今後の研究の展望を示すものである。第一の課題の中心は、視線の意味内容をいかにして伝達するかという点にある。視線パターン分析や他の生体情報（例：瞬目、瞳孔径）を組み合わせ、聞き手の認知状態を推定するモデルの構築が考えられる。また、第二の限界について、詳細表示機能は今後の実験でも維持し、十分な組数で実施したのちも不要であると判断された場合視認性向上のため取り除く。

5. むすび

本研究では会話ノードマッピングと視線検出による話し手に対する聞き手の状態のフィードバック手法を提案した。実験によって、フィードバックにより話者が安心感や相手の意図類推が可能であることが示唆されたが、同時にその推定した意図が本当に聞き手のそれと合致しているのかは未解明である。

今後、被験者数を増やすだけでなく、聞き手の注視理由を特定し、それを効果的に話し手へ提示する手法を確立することを目指す。これによりシステムは単なるコミュニケーションの非対面化で失われる情報を補うツールとしてだけでなく、対面の非言語情報ですら困難な他者の認知プロセスの部分的な共有を可能にし、より高精度な共通理解の形成を支援する新たなコミュニケーション基盤となる

と期待される。

謝辞 本研究は、JST ムーンショット型研究開発事業（JP-MJMS2292）及び、公益財団法人セコム科学技術振興財団の助成を受けたものである。

参考文献

- [1] Maynard, M. & Gilson, Lucy. (2014). The Role of Shared Mental Model Development in Understanding Virtual Team Effectiveness. *Group & Organization Management*. 39. 3-32.
- [2] Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, 85(2), 273–283.
- [3] Camerer, Colin & Weber, Martin. (1989). The Curse of Knowledge in Economic Settings: An Experimental Analysis. *Journal of Political Economy*. 97. 1232-54.
- [4] Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association.
- [5] Kenda, Renata. (2018). Paradoxical Virtual Leadership: Reconsidering Virtuality Through a Paradox Lens. *Group & Organization Management*. 43. 752-786.
- [6] Lanigan, Richard. (1978). General and Theoretical: Beyond Culture. Edward T. Hall. *American Anthropologist*. 80.
- [7] Michael Henry Tessler et al., AI can help humans find common ground in democratic deliberation. *Science* 386, eadq2852 (2024).
- [8] Zhang, Rui & Zhang, Ziyao & Zhu, Fengliang & Zhou, Jiajie & Rao, Anyi. (2024). Mindalogue: LLM -- Powered Nonlinear Interaction for Effective Learning and Task Exploration.
- [9] Xinyue Chen, Nathan Yap, Xinyi Lu, Aylin Gunal, and Xu Wang. 2025. MeetMap: Real-Time Collaborative Dialogue Mapping with LLMs in Online Meetings. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW132 (May 2025), 35 pages.
- [10] Braun, Virginia, and Victoria Clarke. 2006. “Using Thematic Analysis in Psychology.” *Qualitative Research in Psychology* 3 (2): 77–101.
- [11] Tickle-Degnen, L., & Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1(4), 285–293.