LLMを用いた複数人議論における視線推定

脇田 健照[1]　嶋田 和孝[2]

概要：遠隔環境の複数人議論では，視線の錯誤により話者が聞き手の関心を把握しにくい．この課題に対処するには聞き手の視線推定が必要となるが，遠隔環境ではカメラや画面の位置が参加者ごとに異なるため，顔映像に基づく従来手法は補正処理を要し，高コストである．そこで本研究では，議論中の聞き手の視線は，話者による発話によって変化するという仮定の下，映像を用いずに議論中の発話から聞き手の視線（話者を見ている・自分自身を見ている・他の聞き手を見ている・誰も見ていない）を推定する．具体的には，視線情報付き対面議論データセットを分析用 LLM に入力し，発話に伴う視線変化の傾向を自然言語で抽出する．その傾向と遠隔議論データセットの発話文を推定用 LLM に与えることで，発話を用いた視線推定を実現する．

キーワード：議論支援，視線推定，大規模言語モデル（LLM）

Gaze Estimation in Multi-person Discussions using LLM

Abstract: In remote multi-party discussions, gaze misalignment makes it hard for speakers to understand listeners' attention. While listener gaze estimation can mitigate this,conventional face-video-based methods are costly in remote settings because participant-specific camera and screen layouts require calibration and correction. Under the assumption that listeners' gaze changes with the speaker's utterances, we estimate gaze states from speech alone (looking at the speaker, self, another listener, or no one). We first use an analysis LLM on a gaze-annotated face-to-face discussion dataset to extract natural-language tendencies of gaze shifts triggered by utterances, then feed these tendencies and utterances from a remote discussion dataset to an inference LLM to achieve speech-based gaze estimation.

Keywords: Debate Support, Gaze Estimation, Large Language Model (LLM)

1. はじめに

議論中における視線の役割は多岐にわたる．Argyle ら [1] は，アイコンタクトには情報探索機能や関係性の確立等の役割があることを示した．また Kendon [2] は，視線機能を「知覚情報の取得」「感情の表現」「会話の発言権の調整」という 3 つの機能に分類した．このように，議論中の人間の視線には非言語情報としてさまざまな意味が含まれている．特に発話中の話者は，聞き手からの視線を把握することで，自身の発話内容に対する聞き手の関心や理解度を推測している．

[1] 九州工業大学大学院 情報工学府
Department of Creative Informatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820–8502, Japan
[2] 九州工業大学 大学院情報工学研究院 知能情報工学研究系
Department of Artificial Intelligence, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820–8502, Japan

しかし近年は，コロナ禍や働き方改革の影響で討論や議論は対面のみならず遠隔でも行われるようになった．遠隔環境における議論では視線の錯誤が発生しやすく，話者が聞き手の視線を把握することが困難となる．例えば図 1 のように，遠隔議論において話者の B を聞き手の C が見ているとする．しかし B にとって，C が自分を見ていると判断することは困難となる．これは発話の交代を困難にし，発話の衝突や参加者全員の沈黙などを引き起こす．このような問題を解決するため，遠隔議論中の聞き手の視線推定を行い，話者の補助を行うシステムが求められている．

従来の視線推定は，主にリアルタイムで行われている議論の参加者の顔映像を処理する手法が提案されている [3]．しかし，遠隔議論の場合は体の位置，カメラの位置，遠隔議論に使用されている画面の位置など，様々な要因（パラメータ）が存在し，正しい視線を検出するためには，これ

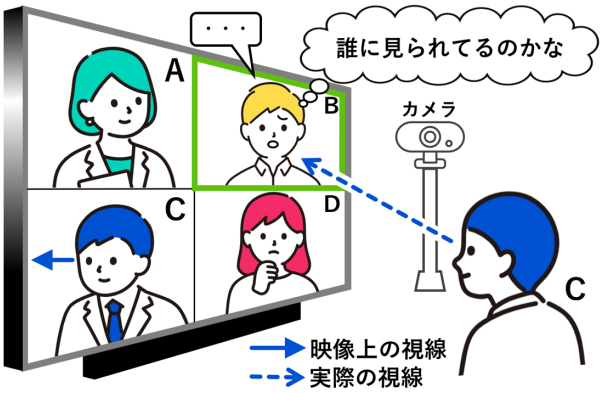

図 1: 映像上の視線と実際の視線の差異

らパラメータのキャリブレーションが必要となり，非常に高コストである．例えば図 1 において，この状況から C が視線はそのままで画面が右に平行移動したとする．この場合，画像上の視線方向は変化していないにも関わらず，C は映像上の A を見ていることになり，本来 B を見ていたとする推定結果と異なってしまう．このように，遠隔議論における映像を用いた視線推定は困難な場合が多い．

ここで，議論中の聞き手の視線は，話者による発話で変化しうると考えられる．例えば，図 2 のように，話者が重要な発言をした場合には，聞き手はその話者を見ると想定され，逆に，分かりにくい発言の場合には，聞き手の視線は話者から逸れると考えられる．また，ある発話が直前の話者ではなく特定の参加者への返答である場合，聞き手の視線はその参加者に向けられる可能性がある．このような発話と視線の関係性に着目し，本研究では発話情報のみを用いた視線推定を試みる．

この推定を実現するには，まず発話と視線の関係性に関する傾向を獲得し，それを視線推定モデルに組み込む必要がある．しかし，目的とする遠隔議論データセットから人手でこの傾向を獲得するには多大なコストを要する．また，視線推定モデルとして従来の機械学習モデルを用いる場合，獲得した傾向をモデルに組み込むには，複雑な特徴量化が求められるという問題がある．そこで本研究では，視線情報が付与された既存の議論データセットを LLM に入力し，発話と視線の関係性に関する傾向を自動的に獲得する．さらに，本手法では LLM を推定モデルとして採用する．これは，獲得した傾向を LLM に自然言語の指示（プロンプト）として直接与えることで，視線推定を行うことが可能であるためである．この方法により，複雑な特徴量設計や補正規則を必要とせず，発話の内容および議論の流れを統合的に考慮した視線推定が実現できると考えられる．

2. 提案手法

本研究では，発話に応じて視線ラベルがどのように変化するかという対応関係に見られる傾向を「視線傾向」と定

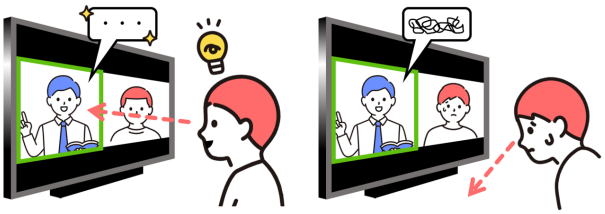

図 2: 発話の内容と聞き手の視線の変化

義し，この視線傾向を大規模言語モデル（LLM）によって自動獲得することを目指す．近年の LLM は，入力データに存在する規則性や判断基準を抽出・整理し，説明が可能であることが示されている [4,5]．そこで本研究では，視線情報付きの議論データから「発話と視線の関係性」を LLM に分析させることで，視線傾向を自然言語として獲得する．

さらに本研究では，獲得した視線傾向を用いた視線推定器としても LLM を採用する．従来の機械学習モデルを推定器として用いる場合，獲得した視線傾向をモデルに組み込むために，視線傾向を数値的な特徴へ落とし込むことや，確率補正を実装することが必要となり，結果として複雑な特徴量化や追加のモデル設計が求められる．一方，LLM を推定器として用いれば，獲得した視線傾向を自然言語の指示としてプロンプトに直接組み込み，その指示に従って分類を行わせることができると考えた．このような自然言語による指示で分類器を形成する枠組みは，プロンプトベースの学習として示されている [6]．また，大規模事前学習に基づく文脈内学習により，少数例や自然言語の指示のみで様々なタスクに適応できることが報告されている [6,7]．以上より，視線傾向の獲得と利用を，特徴量設計や確率補正を人間が介さず一貫して実現できる手法として，LLM を用いたの視線推定を提案する．

提案手法の全体像を図 3 に示す．まず，LLM に文脈を考慮した分析を行わせるため，LLM に「視線情報付き議論データセット」の対話データ全体（話者，発話内容，各発話の視線ラベル）を一度に入力する．これにより，発話と視線の関係性を分析させ，視線推定に用いる傾向を獲得する．次に，この抽出された傾向を補助的な知見としてシステムプロンプトに組み込み，視線推定モデルである LLM に与える．最後に，推定対象である「遠隔議論データセット」の対話データ（話者，発話内容）を一度に LLM へ入力し，各発話における聞き手の視線ラベルを推定させる．

3. 遠隔議論データセット

本研究では．波多野ら [8] によって作成された Kyutech Remote コーパスを遠隔議論データセットとして用いる．本コーパスは，複数人対面議論コーパスである Kyutech コーパス [9] を基に，遠隔環境での議論を対象として構築されたものである．

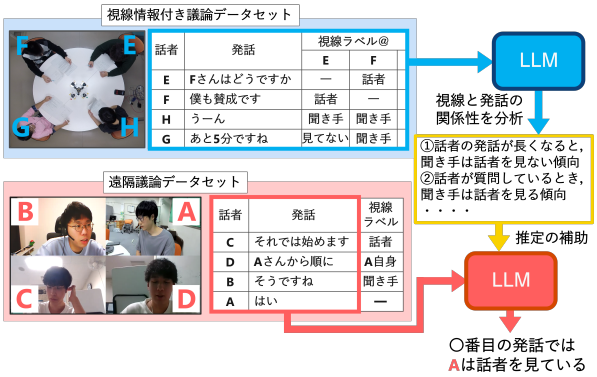

図 3: 本手法の概要図

図 4: 実際の遠隔議論における被験者 A の PC 画面

3.1 収録環境とタスク設定

議論は，Zoom[*1]を用いた遠隔形式で実施され，議論の様子は映像及び音声として収録されている．参加者は大学生 4 名（A，B，C，D）で構成され，そのうち A のみが指定の対話収録用 PC を使用し，視線情報のための被験者となっている．他の 3 名（B，C，D）は各自が用意した PC を使用して議論に参加している．被験者 A の PC 画面内には，被験者 A が対話収録用 PC の画面上のうちどこを見ているかという視線情報が，アイトラッカーによって画面内に白い円で可視化されている．なお，被験者 A を撮影するカメラは，被験者の正面ではなく，画面の右上方向に設置されている．そのため被験者 A の映像からでは，視線や眼球の運動を安定して捉えることが難しく，顔映像のみから被験者 A が画面上のどこを見ているかを高精度に推定することは困難であると考えられる．このような撮影条件は，遠隔議論環境において一般的に生じうる状況であり，映像に依存しない視線推定手法の必要性を示している．実際の遠隔議論における被験者 A の PC 画面を図 4 に示す．

議論は，4 名の参加者が架空の都市のショッピングモールの経営者であるという設定のもと，レストラン街に新たに出店させる店舗を 3 つの候補の中から 1 つ選択するというタスクで行われた．参加者には，出店候補の店舗情報や既存店舗の情報，周辺環境に関する資料が事前に配布され，一定時間の黙読後に自由討論形式で議論が行われている．本コーパスでは，出店候補のテナント店などの情報が異なるシナリオによって，議論が計 5 議論収録されている．

また，本コーパスでは，各議論ごとに参加者の組み合わせが異なっており，視線情報の取得対象となる被験者 A も全 5 議論においてすべて異なる人物である．この収録環境により Kyutech Remote コーパスは，特定の個人に依存しない形で，議論内のある 1 名の視線情報を含む複数の遠隔議論データとして構築されている．

3.2 書き起こしデータと視線アノテーション

本コーパスでは，収録された遠隔議論の映像および音声に基づき，各発話の書き起こしデータが作成されている．書き起こしデータは，話者 ID，発話開始時間，発話終了時間，および発話内容から構成されており，これらの情報を基に，発話単位での分析が可能となっている．

表 1: 書き起こしデータおよび視線アノテーションの一例

話者 ID	発話内容	A が見ている人物
C	3,40 歳とかなんですかね	A
A	お年寄り何食うんですかね	A
C	お年寄り食べる	D
B	お年寄りおしゃべりの方が	B
B	要してるイメージある	B
B	井戸端会議じゃないけどさ	X
D	ファミリープレート	A
C	作りますね	X

さらに，被験者 A が画面上のどの参加者を見ていたかというラベルが，アイトラッカーによって取得された視線情報を基に，人手でアノテーションされている．このラベルは発話を最小単位とし，各発話に対して 1 つ付与されている．また，ラベルの種類として，参加者 A，B，C，D のいずれか，または誰も見ていない状態を表す X の計 5 種類のラベルが定義されている．なお，アノテーションにおいて，ある発話区間内に被験者 A が一瞬でも特定の参加者の顔を見た場合，その発話において被験者 A は当該参加者を見ていたものとして扱われている．また，同一の発話区間内で複数の参加者を見た場合には，見た時間が最も長い参加者をその発話における注視対象としてラベルが付与されている．表 1 に，本研究で用いる書き起こしデータおよび視線アノテーションの一例を示す．

4. 視線推定の定義

Kyutech Remote コーパスには，各発話に対して被験者 A の視線ラベルが付与されているが，本研究では，被験者 A が聞き手である発話のみを推定対象とする．すなわち，被験者 A 自身が話者である発話については，視線推定の対象から除外する．これは，本研究の目的が，「話者が発話を

[*1] https://www.zoom.com/ja

表 2: 視線推定タスクにおけるデータの一例

話者 ID	発話内容	A が見ている人物	視線ラベル
C	3,40 歳とかなんですかね	A	A 自身
A	お年寄り何食うんですかね	A	ー
C	お年寄り食べる	D	他の聞き手
B	お年寄りおしゃべりの方が	B	話者
B	要してるイメージある	B	話者
B	井戸端会議じゃないけどさ	X	誰も見ていない
D	ファミリープレート	A	A 自身
C	作りますね	X	誰も見ていない

行っている際に，聞き手がどこに注意を向けているか」を推定することであり，話者自身の発話中の視線行動とは性質が異なるためである．また，聞き手の視線は，話者の発話内容や議論の構造とより密接に関係すると考えられることから，聞き手に限定した視線推定を行うことで，発話と視線の関係性をより明確に分析できると考えた．

3.2 節で述べたように，本コーパスでは各発話に対して，被験者 A が見ている人物を表す A，B，C，D，X の 5 種類からなるラベルが付与されている．本研究では，これらのラベルをそのまま用いて「被験者 A が誰を見ているか」を人物単位で推定するのではなく，以下の 4 種類の視線状態に分類して扱う．

- 話者：現在の発話を行っている参加者を見ている状態
- A 自身：被験者 A 自身を見ている状態
- 他の聞き手：現在の話者および被験者 A 以外の参加者を見ている状態
- 誰も見ていない：いずれの参加者も見ていない状態

以上の 4 種類からなる状態を視線ラベルとする．そして，各発話に対して付与された視線ラベルを推定する 4 値分類タスクを，聞き手の視線推定タスクとして定義する．データセットの各発話に視線ラベルを付与した例を表 2 に示す．このように新たに視線ラベルを設けた理由は，本研究が発話の情報のみを用いて視線推定を行うためである．本データセットでは，被験者 A が聞き手である状況において，話者と A 以外に聞き手は 2 名存在することになる．この 2 名を区別するための手がかりは発話の情報からは得られないと判断したため，「他の聞き手」という 1 つのクラスにまとめて扱うこととした．

表 3 に，各議論における視線ラベルの発話数を示す．表より，視線ラベルの分布には偏りが見られ，特に「話者」と「誰も見ていない」の発話数が他のラベルに比べて顕著に多いことが分かる．一方で，「A 自身」および「他の聞き手」は相対的に少なく，議論によってはほとんど出現しない．

5. 視線情報付き議論データセット

本研究では，Kyutech Debate コーパス [10] に視線情報を付与したものを「視線情報付き議論データセット」[11] として用いる．また，このコーパスに付与された視線情報を，遠隔議論データセットの「視線ラベル」と区別するため「視線ラベル DB」と呼ぶ．Kyutech Debate コーパスは，議論 ID が付与された 5 つの命題（T0 から T4）について 4 名が対面環境で議論する様子を，映像データおよび発話書き起こしテキストとして収録したコーパスである．参加者は大学生・大学院生計 13 名で，この 13 名から 4 名のグループを 5 個作り，各グループに命題を与えて議論を行う．各議論では参加者 ID として A，B，C，D が割り振られる．議論は，命題に対して賛成派 2 名・反対派 2 名に分かれて 20 分間の討論を行い，続いて合意形成を 20 分間行う．1 命題につき討論と合意形成の 2 議論が存在するため，全体で 10 議論が収録されている．収録時，参加者は円卓に A，B，C，D の順で均等に着席し，卓上には資料と筆記具が配置されている．議論中の様子は上半身カメラ，頭上カメラ，360 度カメラなどで撮影されている．

表 3: 各議論における視線ラベルの発話数

議論	話者が A	話者	A 自身	他の聞き手	誰も見ていない	合計
1	105	100	10	29	133	377
2	62	149	18	29	207	465
3	161	57	0	32	179	429
4	147	117	17	15	179	475
5	164	397	2	84	193	840
合計	639	820	47	189	891	2586

Kyutech Debate コーパス の全発話は，0.2 秒以上の無声区間を転記単位の区切りとして人手で書き起こされており，各発話には話者 ID，発話開始時刻，発話終了時刻，発話内容が記録される．このように書き起こされた Kyutech Debate コーパス の総発話数は 7449 発話であり，視線ラベル DB はこの 7449 発話それぞれに対して付与されている．

5.1 視線情報（視線ラベル DB）の付与方法

視線ラベル DB の付与では，発話を最小単位として，話者側の視線を表す「話者ラベル」と，3 名の聞き手それぞれの視線状態を表す「聞き手ラベル」を付与する．ここで「視認」とは，発話中に一度でも対象人物の顔を見た場合と定義される．

話者ラベルは，各発話に対して 1 つ付与され，話者が発話中に視認した人物を表す．ラベルは {A,B,C,D,0} であり，A から D は対応する参加者を視認していたこと，0 は誰も視認していなかったことを表す．同一発話内で「誰も視認せずに発話した時間」の方が長い場合でも，一度でも聞き手を視認した場合はその聞き手の ID を話者ラベルとし，同一発話内で複数の聞き手を視認した場合は視認時間がより長い聞き手の ID を話者ラベルとする．

表 4: 視線情報付き議論データセットの例

話者	発話	話者ラベル	聞き手ラベル A	B	C	D
B	あれはあまり知るところじゃないんだけどえっと	0	3	話者	3	3
B	どうしよう	0	3	話者	4	3
B	むずかしいな	0	3	話者	4	3
A	とりあえずタブレット導入した際の	D	話者	1	4	1
A	メリットについては	0	話者	3	3	3
A	なんかまぁ	C	話者	1	3	1
A	教育上良いっていうのは	C	話者	1	1	1
C	うーん	A	1	3	話者	2
A	みんな	C	話者	1	1	2
D	はい	B	2	2	2	話者

聞き手ラベルは，各発話における 3 名の聞き手それぞれに付与され，聞き手が発話中にどこを視認しているかを表す．ラベルは以下の 4 種類である．

- ラベル 1：話者を見ている
- ラベル 2：話者と本人を除く 2 名の聞き手のいずれかを見ている
- ラベル 3：誰も見ていない
- ラベル 4：メモ等の作業を行っている

後の実験では，この聞き手ラベルを中心に発話と視線の対応関係（視線傾向）を LLM に分析させる．表 4 に，アノテーション済みデータセットの例を示す．各行が 1 発話に対応し，「話者」および「発話」に加えて，話者ラベルと，参加者 A から D の各列に聞き手ラベルが格納される．なお，話者本人に対応する列には，聞き手ラベルは付与されないため「話者」と記載されている．

5.2 アノテーションの規模と信頼性

視線ラベル DB のアノテーションは，Kyutech Debate コーパス に収録された 10 議論・7449 発話に対して実施されている．また，アノテーションの客観性を確認するために，T2 の一部 1 分間（36 発話）を対象として複数人で試験的にアノテーションを行い，Fleiss'κ が話者ラベルは 0.616，聞き手ラベルは 0.588 であったことが報告されている．本研究では，以上の手順で構築された視線ラベル DB 付き Kyutech Debate コーパス を，視線傾向を獲得するための入力データとして用いる．

6. 実験設定

6.1 LLM による視線傾向の獲得

本節では，5 節で説明した視線ラベル DB 付き対面議論データセットから，発話内容および談話状況と聞き手の視線行動の対応関係（視線傾向）を LLM に獲得させる方法を述べる．

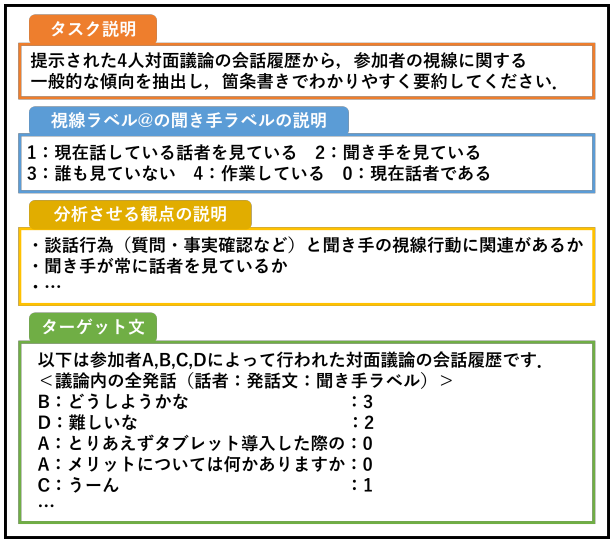

図 5: 視線傾向獲得のためのプロンプト

6.1.1 視線傾向獲得のためのプロンプト設計

図 5 に，視線傾向獲得のためのプロンプトを示す．まず，タスクの説明として視線の傾向を出力させるように指示する．そして，視線ラベル DB の聞き手タグの意味を定義する．さらに，分析してほしい観点を列挙し，観測の方向性を与える．このように観測の方向性を与えなかった場合，同じデータに対しても異なる傾向が出力された．そのため本手法では，分析観点の例を与えることで，視線傾向の獲得を安定化させる．具体的に与えた観点としては，

(1) 聞き手が話者注視を保つか否か
(2) 質問・相槌・意見表明など発話タイプと視線の関連
(3) 直前話者への注視の有無
(4) 意見の対立や衝突が生じた場面での視線傾向
(5) その他発見した視線傾向

最後に議論の発話を話者，発話テキスト，参加者ごとの聞き手ラベルの形式で提示し，LLM に分析させる．またこの際，LLM の文脈理解能力を活かすため，発話を個別に独立して入力するのではなく，1 つの議論を構成する発話（各発話の話者，発話テキスト，参加者ごとの聞き手ラベル）を，議論単位で LLM に入力した．これにより，議論全体の流れを踏まえたうえで視線傾向を分析できるようにした．

6.1.2 その他実験設定

本手法では，OpenAI 社の GPT-4o[*2] を LLM として採用した．また，分析タスクとしての安定性を保ちつつ，データから多様な規則性を引き出すため，生成の temperature は 0.5 と設定した．そして，Kyutech Debate コーパス の 1 議論ごとに LLM に入力を行い，合計 10 議論から視線傾向を獲得した．

[*2] https://platform.openai.com/docs/models/gpt-4o

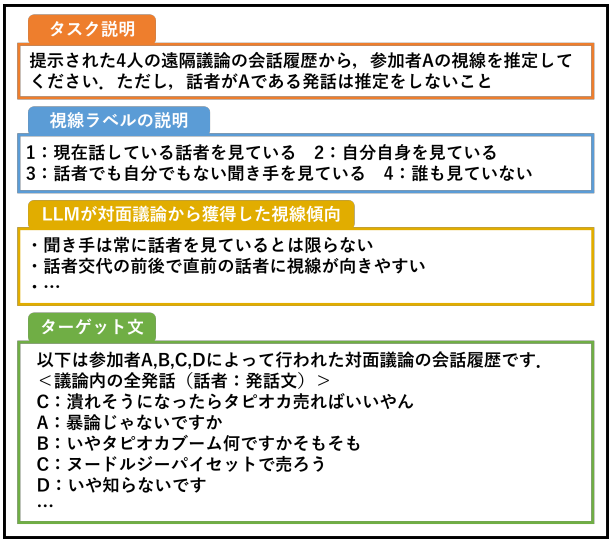

図 6: 視線推定のためのプロンプト

6.2 LLM による視線推定モデル

本節では，6.1 節で獲得した視線傾向を自然言語の指示として組み込み，LLM を用いて遠隔議論における参加者 A の視線を推定するモデルについて述べる．

6.2.1 視線推定のためのプロンプト設計

図 6 に，LLM を用いた視線推定のためのプロンプトを示す．まず，タスクの説明および視線ラベルの定義を与え，参加者 A の視線を 4 値に分類することを指示する．ここで，本手法の中心は，前節で対面議論から獲得した視線傾向を，推定時の判断材料として提示する点にある．そこでプロンプト内に判断材料として視線傾向を箇条書きで埋め込み，推定時に参考とする情報として与える．ここで提示する視線傾向は，6.1 節の結果を文章化したものを使用する．最後に，議論の発話を話者と発話テキストの形式で提示し，LLM に参加者 A の視線ラベルを推定させる．なお，遠隔議論では発話数が多い傾向にあり，単一プロンプトで全発話を入力するとトークン上限を超えるケースが発生し得る．そこで本研究では，トークン数に基づくチャンク分割を行い，複数回の LLM 推定結果を統合する．

6.2.2 その他実験設定

本手法では，OpenAI 社の GPT-4o を LLM として採用した．また，視線推定は分類タスクであり，再現性を重視するため，temperature を 0.0 とし，推定結果の揺れを抑える．なお，LLM が推定した視線ラベルが 1 から 4 の範囲外である場合や，出力が存在しない場合は，エラーとしてログを保存し，その発話の推定を失敗扱いとする．

7. 実験結果と考察

7.1 LLM が獲得した視線傾向の分析

本節では，6.1 節で述べた方法により LLM が獲得した視線傾向について述べる．また，Kyutech Debate コーパスの視線ラベル DB を付与することで構築された視線情報付きデータセットの作成時に，脇田ら [11] に考察された内容と照らし合わせながら，LLM の出力内容を分析する．なお，この節で述べるラベルというのは，5.1 節で説明した視線ラベル DB の聞き手ラベルを指す．

(1) 聞き手は常に話者を見ているとは限らない．
LLM は，聞き手が話者（ラベル 1）だけでなく，他の聞き手（ラベル 2）や誰も見ない（ラベル 3），作業（ラベル 4）に分散することを指摘した．この点は，対面議論データセットにおいて「聞き手全員が話者を見ている発話の割合は少ない」という観察とも整合する．したがって，視線推定では「聞き手は基本的に話者を見る」という単純な仮定だけでは不十分であり，複数の視線状態が自然に起こる前提で扱う必要がある．

(2) 発話タイプによって，聞き手が話者を見る度合いが変わる可能性がある．
LLM は，質問や短い相槌の場面では話者を見る（ラベル 1）が増え，長い意見表明や事実確認では資料・思考（ラベル 3）や作業（ラベル 4）が増える，という視線傾向示した．データセット構築時の考察でも，名前を呼ぶ発話や注意を引く発話では聞き手が話者を見る場面が見られた一方で，淡々とした説明では視線が外れる可能性が述べられており，LLM の出力はこれらの観点を捉えたと解釈できる．

(3) 話者交代の前後で，直前の話者に視線が向きやすい．
LLM は，話者が交代する局面で「直前に話していた人を見る」傾向を挙げた．聞き手（ラベル 2）は，話者以外を見るときに話者交代を促す意思表示が含まれるかを分析する目的で定義されている．この定義を踏まえると，ラベル 2 の一部は，「次の話者や周囲の反応を確認する視線」として解釈でき，LLM がその点をまとめた可能性がある．

(4) 意見が対立する場面では，話者以外にも視線が向きやすい．
LLM は，議論が対立すると，話者だけでなく他の参加者へ視線が移る，つまり聞き手（ラベル 2）を見る機会が増えるという視線傾向を示した．データセット構築時の考察でも，発話内容だけでは決まらない要因として，意見の立場が視線に影響する可能性が述べられている．この点から，LLM は発話の流れの中で「対立や確認」が起きそうな箇所を手がかりに，視線の分散を仮説としてまとめた可能性が考えられる．

(5) 視線には個人差がある．
LLM は，特定の参加者が他の聞き手を見る割合が高い，あるいは資料を見る割合が高い，といった個人差を指摘した．データセット構築時の考察でも，視線行動には個人の習慣や立場が影響する可能性が強く述べられており，LLM の出力はこれらの観点を捉えたと解釈できる．

(6) 視線の「持続時間」は，発話が長いと視線が固定されにくい．
LLM は，発話が長いと視線が固定されにくいという説明を出力した．しかし，本データセットのラベルは発話区間を要約したものであり，「視認」は発話中に一度でも顔を見たかどうかで定義される．そのため，視線が発話中にどれくらいの時間向けられていたか，あるいはどれくらい頻繁に動いたかは，ラベルだけからは直接は分からない．したがって，この出力は，ラベルに含まれない情報を LLM が一般的な対話知識から補って説明した可能性がある．

以上の結果から，LLM には，ラベルの定義（話者を見る，他者を見る，見ない，作業）と発話内容を対応づけ，参加者ごとの偏りや，話者交代の前後といった文脈を手がかりに，視線傾向を表現できると考えられる．特に，「聞き手全員が話者を見ている発話は少ない」というデータセット側の観察や，「発話以外の要因（立場や個人的な習慣）も視線に影響しうる」という構築時の考察と整合する形で出力されていた．これは，LLM が単に発話の表層だけを見るのではなく，複数発話にまたがる流れや参加者間の関係をまとめて捉えられる可能性を示している．

一方で，LLM が出力した「持続時間」のように，入力ラベルだけでは確かめにくい性質も含まれていた．これは，LLM が一般的な対話知識をもとに補って説明している可能性があり，必ずしもデータだけから導かれたとは限らない．

以上より本研究では，LLM が抽出した視線傾向をそのまま事実とみなすのではなく，視線推定に有用な仮説として扱う．そして，視線推定のプロンプトに組み込んだ際の精度を評価することで，LLM が視線傾向を正しく判断できているかを検証する．

7.2 視線推定結果

本節では，7.1 節で述べた視線傾向を，6.2.1 節で述べたプロンプトに組み込み，遠隔議論データセットの視線ラベルを推定した結果について述べる．なお，ベースラインとして，GPT のプロンプトに視線傾向を与えずに推定した場合の結果（GPT 単体）を比較対象として示す．

表 5 に各手法の全体結果（Macro 平均）を示す．表内の太字はそれぞれの評価指標のうち最高指標のものを示す．F1-score に注目すると，GPT 単体は 0.176 と低く，視線傾向を与えない状態では視線カテゴリを適切に使い分けられていないことが分かる．一方，GPT+視線傾向では Macro F1-score が 0.250 まで上昇した．この結果は，対面議論から抽出した視線傾向を判断材料として与えることで，LLM が特定ラベルへ偏って出力する状況が緩和され，少数ラベルを含めたラベル間のバランスが改善したことを示唆する．実際，GPT 単体では一部ラベル，特に「誰も見ていない」がほとんど出力されないという極端な偏りが見られたのに対し，視線傾向を与えることでラベル選択が分散し，Macro F1-score が改善したと解釈できる．

表 5: 視線推定結果（全体, macro average）

手法	Precision	Recall	F1-score	Accuracy
GPT 単体	0.1330	**0.2580**	0.1760	0.3350
GPT + 視線傾向	**0.2870**	**0.2580**	**0.2500**	0.3630

7.3 ラベル別精度比較

表 6 に，ラベル別の結果を示す．表内の太字はそれぞれのラベルにおける評価指標のうち最高指標のものを示す．以降では，ラベルごとに誤りの傾向を分析する．

7.3.1 「話者を見ている」

GPT 単体は Recall が非常に高い一方で Precision が低く，話者注視を過剰に選択していることが分かる．この傾向は，視線傾向を与えない場合に，LLM が「聞き手は話者を見る」という単純な仮定に寄りやすいことを示している．GPT+視線傾向では Precision が上昇し，話者への過剰な偏りが緩和された．一方で Recall は低下しており「話者」と判断すべき発話の一部が他ラベルへ移ったと考えられる．

7.3.2 「自分自身を見ている」

両手法とも「自分自身を見ている」を一度も正しく予測できていない．このラベルはデータ数が少なく，加えて，発話テキストから「自分自身を見た」ことを直接示す手がかりが乏しいため，テキストのみの推定では特に難しいと考えられる．また，本研究の視線傾向は対面議論から抽出しているため，「自分自身を見る」という遠隔特有の視線状態を十分に説明できていない点も影響している可能性が大いにある．よってこのラベルを扱うには，遠隔議論に特化した視線傾向を新たに獲得するほかないと考えられる．

7.3.3 「他の聞き手を見ている」

GPT 単体は「他の聞き手」の Recall が比較的高いが Precision は高くなく，他ラベルとの混同が多いと考えられる．GPT+視線傾向では Precision がさらに低下し，このラベルを「意見が対立する場面では，話者以外にも視線が向きやすい」といった視線傾向に引っ張られて選びやすくなった一方で，正確な使い分けができていない可能性がある．この結果からも，視線傾向はラベル選択を促進する効果がある一方で，遠隔議論に固有の状況に合わせた傾向でない場合，誤りも増やしうることが示唆される．

7.3.4 「誰も見ていない」

GPT 単体は「誰も見ていない」を全く出力できておらず，この点が F1-score 低下の最大要因である．これは，LLM にとって「誰も見ていない」という状態がテキストから説明しにくく，判断が難しいためであると考えられる．GPT+視線傾向では，「聞き手の視線は多様であり，資料・思考により話者を見ない状態も起こりうる」という視線傾向が判

表 6: 視線推定結果（ラベル別）

視線ラベル	手法	Precision	Recall	F1-score	正解データ数
話者	GPT 単体	0.3790	**0.7260**	**0.4980**	820
	GPT + 視線傾向	**0.4860**	0.4260	0.4540	
A 自身	GPT 単体	0.0000	0.0000	0.0000	47
	GPT + 視線傾向	0.0000	0.0000	0.0000	
他の聞き手	GPT 単体	**0.1530**	**0.3060**	**0.2040**	189
	GPT + 視線傾向	0.0690	0.2610	0.1100	
誰も見ていない	GPT 単体	0.0000	0.0000	0.0000	891
	GPT + 視線傾向	**0.5950**	**0.3460**	**0.4380**	

断材料として与えられるため，「誰も見ていない」を選択できるようになったと考えられる．

8. おわりに

本研究では，遠隔議論における聞き手の視線推定を対象とし，発話の情報から視線を推定する手法を提案した．具体的には，視線推定器の追加学習を行わず，プロンプトに与える判断材料によって推定を行う手法を提案した．まず，視線情報付きの既存対面議論データを入力として，発話内容・談話状況と聞き手の視線行動の対応関係（視線傾向）を獲得した．その後 LLM を推定器として用い，獲得した視線傾向をプロンプトに組み込むことで，視線推定を行った．

実験の結果，視線傾向を判断材料として与えることで，視線傾向を与えないモデルよりも精度が向上した．これは，視線傾向が推定時の根拠として機能し，偏りを緩和してラベル間のバランスを改善できたことを示している．ただし，遠隔固有の視線状態「A 自身を見る」の精度は依然として 0 であり，今後は視線傾向の改善や，遠隔議論特有の要因を考慮したプロンプト設計が必要である．

最後に，本研究で扱った視線行動は，発話の情報だけでなく，参加者の性格や思考の癖，能力，資料の参照頻度といった様々な個人差に大きく左右されるという前提を改めて強調する．この個人差が大きいままでは，特定の参加者や特定のグループに過学習した推定となりやすく，未知の参加者や異なる議論条件への適用が困難となる．したがって，実用的な視線推定を実現するには，個人差や議論形態を無視するのではなく，データセット作成の段階からこれらを明示的に扱うことが重要であると考える．具体的には，多様な参加者・議論条件を含む収録を進めてデータの偏りを低減するとともに，参加者ごとの特徴をメタデータとして記録することが考えられる．これにより，視線推定器が個人差や議論形態の影響を考慮できるようになり，より汎用的な視線推定が可能になると期待できる．

謝辞

本研究は科研費 23K11368 の一部です．

参考文献

[1] Michael Argyle and Janet Dean. Eye-contact, distance and affiliation. *Sociometry*, Vol. 28, No. 3, pp. 289–304, 1965.

[2] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, Vol. 26, pp. 22–63, 1967.

[3] Dan Witzner Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 3, pp. 478–500, 2010.

[4] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2022.

[5] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm, 2021.

[6] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing, 2021.

[7] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners, 2020.

[8] 波多野翔貴, 嶋田和孝. 複数人遠隔対話コーパスの構築と LLM を用いた取りまとめ役の特徴分析. 言語処理学会第 32 回年次大会 (NLP2026), pp. Q4–17, 2026.

[9] Takashi Yamamura, Kazutaka Shimada, and Shintaro Kawahara. The kyutech corpus and topic segmentation using a combined method. In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pp. 95–104, 2016.

[10] Tsukasa Shiota and Kazutaka Shimada. Annotation and multi-modal methods for quality assessment of multiparty discussion. In *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, pp. 175–182, 2022.

[11] Kensho Wakita and Kazutaka Shimada. An utterance is enough to the gaze? gaze detection from utterance information in multiparty discussion. In *2024 International Conference on Activity and Behavior Computing (ABC)*, pp. 1–8, 2024.