

視線アドレスの可視化と音響定位によるオンライン会議支援の研究

日比野 徹心^{1,a)} 竹内 勇剛^{2,b)}

概要: オンライン会議では、対面会話で重要な役割を果たす視線に基づく次話者予測が成立しにくく、その結果として沈黙や発話の衝突が生じやすいという課題がある。本研究では、視線情報をそのまま可視化するのではなく、参加者の視線行動をもとに会話の注目関係を推定し、それを空間的な表現として共有することで、次話者予測を支援するオンライン会議システムを提案する。4人1組でワードウルフ課題を用いた実験を行い、視線に基づく表現の有無および音響定位の有無を組み合わせた4条件で評価を行った。分析の結果、次話者の選択は、発話直前の視線状態そのものよりも、視線行動をもとに会話の焦点として空間的に示されていた参加者と一致する傾向が見られた。このことから、視線に基づいて共有された注目構造が、次話者予測の手がかりとして用いられていた可能性がある。一方、音響定位は次話者を直接決定する要因ではなかったが、視線に基づく注目構造の理解を補助し、次話者予測を安定化させる役割を果たしていた。以上より、オンライン会議における次話者予測においては、視線方向を個別に提示するよりも、視線行動から抽象化された注目関係を共有することが、話者交替の円滑化を支援する上で有効であることが示唆された。

1. 背景と研究目的

新型コロナウイルス感染症の拡大を契機として、オンライン会議システムは教育・研究・ビジネスなど多様な領域に急速に普及した。現在では流行のピークを過ぎたものの、地理的制約を越えて人々を結ぶ手段として、オンライン会議は重要な役割を担い続けている。一方で、対面会議と比較すると、オンライン会議では話者交替が円滑に行われにくく、沈黙や発話の衝突が生じやすいことが指摘されている [1]。

この要因の一つとして、対面環境では自然に共有される視線や身体の向きといった非言語的の手がかりが、オンライン環境では十分に伝達されない点が挙げられる。Sacksら (1978) の話者交替理論は、会話における次話者選択が円滑に行われることで会話の進行が保たれることを示しており、その過程において非言語情報が重要な役割を果たすとされている [2], [3]。特に視線は、注意の方向付けや参与関係の形成を通じて、話者交替の判断に強く関与することが知られている [4]。Kendon (1967) は視線の機能を認知・フィードバック・調整・表現の4つに整理し、会話における

多面的な役割を理論化している [5]。また、榎本ら (2009) は、多人数対面会話において、話し手から視線を向けられている聞き手が次話者になりやすいことを実証的に示している [6]。

しかし、ビデオ会議では、カメラ位置とディスプレイの不一致により、対面会話において自然に伝達される視線方向が正確に共有されにくいという制約がある。その結果、話し手が誰に向けて発話しているのか、また聞き手がどの程度注目しているのかが曖昧になり、話者交替が滞りやすくなる。このような問題に対し、参加者間の視線配布を可視化することで参与構造の理解を支援する試みが行われてきた。例えば、飯塚ら (2022) は、視線方向を矢印で提示する手法や、注視対象の表示を強調する UI を提案し、それらが話者交替に影響を与えることを報告している [7]。また、複数スクリーンを用いた視線伝達手法により、発話の重複や誤解が減少することも示されている [8]。

さらに、音声の空間的の手がかりである音響定位も、話者同定や注意の切り替えを支える重要な要素である。対面環境では、誰の声かどの方向から聞こえるかという情報が自然に共有されるが、多くのオンライン会議では音声は左右対称に提示され、方向情報が乏しい。このような状況では、話者同定に余分な認知的負荷が生じうる。Nowakら (2023) は、ビデオ会議に空間音響を導入することで、ター

¹ 静岡大学大学院総合科学技術研究科

² 静岡大学創造科学技術大学院

a) hibino.tesshin.21@shizuoka.ac.jp

b) takeuchi@inf.shizuoka.ac.jp

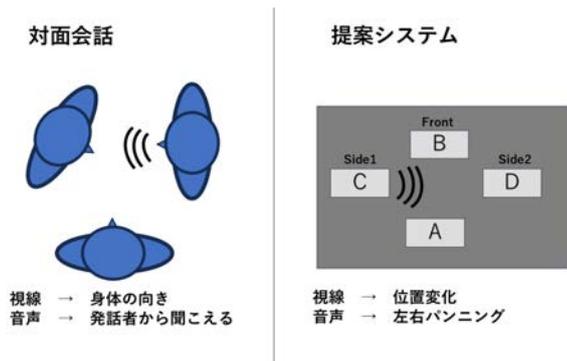


図 1 対面会話において視線・身体配置・音声方向が統合的に共有されている状況(左)と、提案システムにおいて参加者配置の変化および左右音響定位によってそれらを代替的に再構成した状況(右)

ンテイキングや社会的存在感の知覚に影響が生じることを示しており、左右の音響定位が話者交替の支援に寄与する可能性を示唆している [9].

一方で、対面会話において共有されている空間的手がかりは、単一の要因によって次話者を一意に決定するものではなく、視線や身体配置を通じて形成される会話全体の構造理解が重要であることが指摘されている [10], [11]. オンライン会議支援においても、空間的手がかりを提示すること自体の効果だけでなく、それらが会話の焦点や参与関係の理解をどの程度促進するののかについては、十分に検討されていない。

そこで本研究では、視線情報や発話の向きに基づいて参加者の表示位置が変化し、左右の音響定位を伴って音声提示されるオンライン会議システムを用い、位置変化と音響定位の組み合わせが話者交替や発話タイミングに与える影響を定量的に分析する。特に、これらの空間的表現が、対面会話における視線アドレスや音響定位の代替として機能しうるかを明らかにすることを目的とする。

2. 提案システム

2.1 システムの概要

本研究では、オンライン会議における話者交替を支援するため、参加者の表示位置が動的に変化し、かつ左右の音響定位を伴って音声提示されるオンライン会議システムを提案する。本システムは、対面会話において自然に共有されている視線方向や身体配置、および音声の方向といった空間的手がかりを、オンライン環境において代替的に再構成することを目的としている。提案システムでは、会話中の注目関係や発話の向きに応じて、画面上に表示された参加者の配置が動的に左右と前方に変化する。さらに、各参加者の音声は、その表示位置に対応した左右の音響定位を付与して再生される。これにより、参加者は視覚的・聴覚的な両面から話者を把握でき、次話者の判断や話者交替が行いやすくなると考えられる。

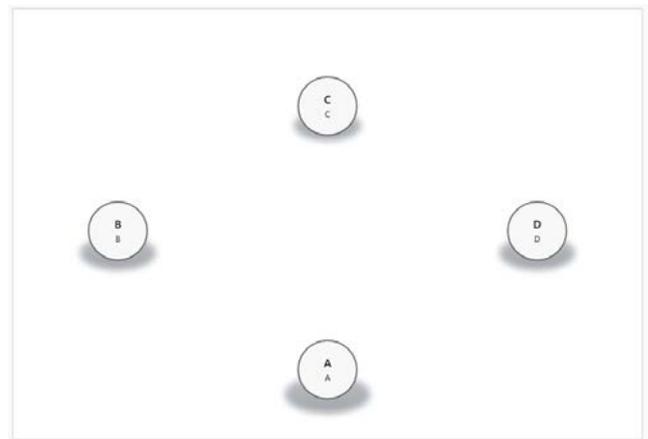


図 2 デモ画像. 初期状態.

2.2 参加者配置による空間的アドレッシング表現

対面会話では、話し手の視線や身体の向きによって、発話の宛先や注目対象が暗黙的に示される。本システムでは、この視線によるアドレッシング機能を、参加者の表示位置の変化として表現する。

画面上には4名の参加者を表すアイコン(A, B, C, D)が配置されており、発話者が見ながら話していると推定された参加者は、画面前方(front)に移動して表示される。一方、発話者自身は後方(back)に配置され、その他の参加者は左右(side1 / side2)に配置される。

この配置により、現在の発話が誰に向けられているのか、また誰と誰の間で会話が進行しているのかが、視覚的に把握しやすくなる。自分が会話に直接関与していない場合には、左右に配置された参加者同士が会話を行っている状態として表現され、参加者は自身の立場を直感的に理解できる。本システムの配置更新の流れを図3に示す。

このような空間配置によって、明示的な指名を行わなくとも会話の焦点や参与関係が共有され、対面会話において視線配布や身体配置が果たしている役割を、オンライン環境で代替的に再現することを目指す。

2.3 視線情報に基づく注目対象の推定

本システムでは、Webカメラ映像を用いた視線推定ライブラリ WebGazer.js を用いて、参加者の視線情報を取得する。取得された視線座標は、画面上に表示された参加者アイコンとの距離関係に基づいて処理され、視線が一定時間以上向けられた参加者を注目対象として推定する。

視線は瞬間的にも変動しやすく、短時間の視線一致は必ずしも意図的な注目や発話のアドレスを反映しないことが、対面会話研究および視線インタフェース研究において指摘されている。対面会話において次話者選択に寄与する視線は、一時的な視線移動ではなく、一定時間配布された持続的な視線であることが示されている。また、HCI分野における視線入力研究では、瞬間的な視線移動による誤検

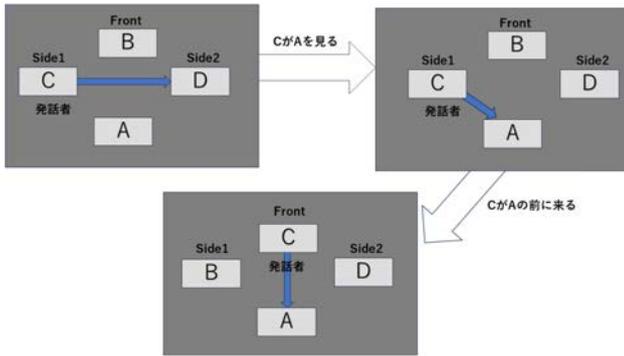


図 3 視線推定から、注視対象の確定、および参加者配置の更新・描画までの流れを示す。

出を抑制するため、500ms～2秒程度の dwell time を設ける手法が一般的に用いられている [12] [13]。

具体的には、視線位置が特定の参加者アイコン近傍に一定時間（2秒）以上安定して留まった場合に、その参加者が注目対象として確定される。この確定処理にはヒステリシスを設け、短時間の視線の揺れによる誤検出を防いでいる。視線推定の精度を向上させるため、9点キャリブレーションを行う。

2.4 左右の音響定位による話者同定支援

本システムでは、参加者配置による視覚的提示に加えて、音声に左右の音響定位を付与することで、話者同定を支援する。各参加者の音声は、画面上での表示位置に対応して左右のパンニングが行われ、左側に配置された参加者の音声は左方向から、右側の参加者の音声は右方向から聞こえるよう制御される。

この音響定位は Web Audio API を用いて実装されており、参加者の表示位置が変化すると、それに伴って音声の定位も更新される。これにより、視線を画面に向けていない状況でも、音声の方向から話者を把握することが可能となり、多人数会話における話者交替の判断を支援する効果が期待される。

2.5 システム構成

提案システムは Web システムを用いて実装されている。音声通信および参加者管理には、WebRTC を基盤とした SkyWay SDK を用い、低遅延な音声通信を実現している。音声のみを対象とした SFU 型の構成を採用し、各参加者の音声ストリームを個別に制御可能としている。参加者の配置および描画は HTML5 Canvas 上で行われ、位置の更新はフレームごとに補間処理を行うことで滑らかな移動を実現している。

3. 実験概要

3.1 実験目的

本実験の目的は、提案システムにおける参加者配置によ

表 1 実験条件

条件	配置	音響	条件の概要
cond1	あり	あり	注目関係に応じて表示位置が変化し、左右音響定位が付与される条件
cond2	あり	なし	表示位置は動的に変化するが、音声は左右対称に提示される条件
cond3	なし	あり	表示位置は固定され、左右音響定位のみが付与される条件
cond4	なし	なし	表示位置・音響定位ともに固定される条件

る空間的アドレッシング表現および左右の音響定位が、オンライン会議における話者交替にどのような影響を与えるかを明らかにすることである。特に、対面会議において視線や音声の方向といった空間的手がかりが会話の円滑さに寄与している点に着目し、それらを代替的に再現した本システムが、話者移行の頻度や構造に影響を与えるかを検証する。

3.2 実験環境

実験は参加者が自身のノートパソコンを用いてオンラインで実施した。ノートパソコンのハードウェア・ソフトウェア要件は事前に確認し、最低限の動作要件を満たすものとした。画面サイズの違いによって表示倍率が異ならないよう、CSS でスケールを調整した。一回あたりの実験への参加者の人数は 4 人とした。

3.3 実験条件

本実験では、参加者配置による表現の有無および左右の音響定位の有無を要因とし、これらを組み合わせた 4 条件を設定した。表 1 に、本実験で設定した 4 条件を示す。

これにより、参加者配置および音響定位それぞれの効果、ならびにそれらの組み合わせ効果を比較可能とした。

3.4 実験課題

実験タスクにはワードウルフを用いた。ワードウルフは、参加者のうち 1 名のみが他の参加者と異なる単語を与えられ、議論を通じて少数派（ウルフ）を特定する会話課題である。各参加者は自身が多数派か少数派かを知らされず、発話内容や他者の反応を手がかりに推理を行う。

この課題では、単純な情報交換ではなく、発話の内容や向け先、反応の仕方が相互に解釈されながら会話が進行するため、固定的な進行役を持たず、話者交替や会話の焦点が自発的に変化しやすい。その結果、「誰が誰に向けて発話しているか」といった参与構造が、短時間のうちに動的に切り替わる。

以上の特徴から、ワードウルフ課題は、参加者配置による空間的アドレッシング表現や左右の音響定位が、会話の進行や次話者選択に与える影響を観察するのに適した課題

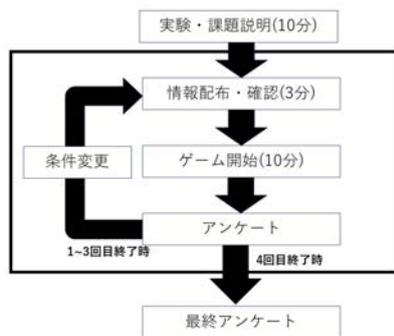


図 4 実験全体の手順。

であると考えられる。

3.5 実験手順

実験手順の概要を図 4 に示す。各参加者グループは、1 回の実験セッション内ですべての 4 条件を実施した。条件の実施順序は、順序効果や学習効果を抑制するため、参加者ごとにランダム化した。まず参加者に目的と手順の説明を行い、課題内容を理解させた上で、各条件ごとの会議を行った。各条件で 10 分ずつ会話を行い、合計 40 分の議論を実施した。

3.6 観測事項

実験中は、会話のタイムラインに基づき、各発話に対応する話者情報、次話者、発話間に生じた沈黙時間、ならびに発話終了時点での参加者配置 (front / side1 / side2) と視線状態をログとして記録した。

記録されたログを用いて、以下の指標を算出した。

話者移行回数

沈黙時間

次話者の選択傾向

参加者配置 (front / side1 / side2) の変化回数

これらの指標を条件間で比較することで、参加者配置および音響定位が、話者交替の頻度および次話者選択の構造に与える影響を分析した。

すべての会議の音声と映像は記録され、後の分析に用いられた。

3.7 仮説

本研究では、オンライン会議における次話者の選定が、視線情報の可視化、音響定位によって誘導されるかどうかを検討する。具体的には、発話者の視線が向いているように見える人物に対して次話者が移行する確率が高まるかどうかについて検討する。

対面会話では、視線や音声の方向といった空間的手がかりが統合的に用いられることで、話者交替が円滑に行われている。本研究で用いる参加者配置および音響定位は、こ

れらの手がかりをオンライン会議環境において代替的に再現するものであり、その結果、話者同定や発話タイミング判断を支援する役割を果たすと予想される。

また、見ながら話している相手が前方 (front) に配置される表現や、自分が会話に直接参加していない場合に左右の参加者同士が会話している状態が示される表現により、front に配置された参加者が次話者として選択される確率が高くなると予想される。これは、対面会話において視線や身体の向きが次話者選択を方向付けることに対応する効果であり、参加者配置が話者交替の構造そのものに影響を与える可能性を示すものである。

4. 実験結果

本実験は、8 グループ (参加者 32 名) を対象に実施し、各グループは 4 条件すべてを 1 回ずつ実施した。全条件合計で 7,144 件の発話直前イベントが観測され、内訳は cond1: 1,820 件, cond2: 1,760 件, cond3: 1,768 件, cond4: 1,796 件であった。本研究で用いた発話直前イベントは、発話開始時点ごとに定義された分析単位であり、話者移行に加えて同一話者による連続発話も含んでいる。

4.1 条件別の行動指標の比較

本研究における「視線共有」とは、視線方向そのものの提示ではなく、視線行動から推定された注目関係の共有を指す。表 2 は、各条件における配置操作 (front)、視線共有、音響の有無と、主要な行動指標の確率を示したものである。本研究では、発話直前のイベント単位データを用い、発話者注視を「発話者に向けられた視線割合が 20%以上」で定義した。

発話者注視の発生確率は、配置および視線共有が行われる cond1 および cond2 において、それぞれ 50.83% および 52.16% と、配置を伴わない cond3 (43.73%) および cond4 (44.74%) よりも高かった。この結果は、参加者配置および視線共有が、会話中の注意配分を高める方向に作用していることを示している。

次に、配置操作が行われる cond1 および cond2 において、front 配置時と非 front 配置時の次話者率を比較した (図 2)。front 配置時の次話者率は、cond1 で 27.01%、cond2 で 26.15% であり、非 front 配置時 (cond1: 21.30%、cond2: 21.87%) と比較して、それぞれ +5.71%、+4.28% の増加が見られた。

一方、配置操作を伴わない cond3 および cond4 では、front に基づく次話者率の比較は定義できず、発話者注視の発生確率も低かったことから、視線行動が次話者選択に与える影響は限定的であった。

以上より、参加者配置および視線共有は、次話者を直接的に決定する強い要因ではないものの、会話中の注意配分を高めることで、front に配置された参加者が次話者となる

表 2 条件別の行動指標

条件	$P(\text{gaze} \geq 0.2)$	$P(\text{next} F)$	$P(\text{next} \bar{F})$	Δ
cond1	50.83%	27.01%	21.30%	+5.71%
cond2	52.16%	26.15%	21.87%	+4.28%
cond3	43.73%	-	-	-
cond4	44.74%	-	-	-

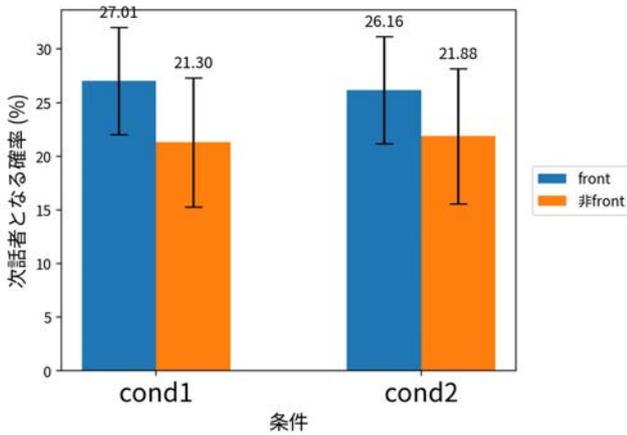


図 5 cond1 および cond2 における front 配置時と非 front 配置時の次話者率の比較. エラーバーは 95%信頼区間である. front に配置された参加者は、いずれの条件においても非 front の場合と比較して、次話者となる確率が高い。

表 3 次話者選択に関するロジスティック回帰分析結果

変数	OR	95% CI	p
front	1.12	[0.90, 1.39]	.33
発話者注視	0.99	[0.80, 1.22]	.93

確率を数%程度上昇させる方向に作用することが示された。

4.2 ロジスティック回帰分析結果

次話者選択を目的変数とし、front 配置および発話者注視の有無を説明変数とするロジスティック回帰分析を行った。分析は、発話直前のイベント単位データを用いて実施した。

表 3 に示すように、front に配置された参加者は、非 front の場合と比較して次話者となるオッズが高い傾向を示したが、統計的に有意な効果は確認されなかった (OR = 1.12, 95% CI [0.90, 1.39], $p = .33$)。また、発話者注視についても、次話者選択に対する有意な主効果は認められなかった (OR = 0.99, 95% CI [0.80, 1.22], $p = .93$)。

4.3 沈黙時間の比較

また、沈黙時間については、配置および視線共有がある条件ほど沈黙時間がわずかに短縮される傾向が示された。条件ごとの平均値および中央値を表 4 に示す。沈黙時間について線形混合効果モデルを用いて分析した結果、参加者配置の主効果 ($p = .422$)、音響定位の主効果 ($p = .737$)、およびそれらの交互作用 ($p = .846$) はいずれも統計的に

表 4 条件別の次話者移行前沈黙時間

条件	平均沈黙時間 (ms)	中央値 (ms)	イベント数
cond1	504	496	486
cond2	510	516	540
cond3	556	560	443
cond4	589	588	508

表 5 沈黙時間に対する線形混合効果モデルの結果

効果	推定値 (ms)	標準誤差	z 値	p 値
切片	589.0	69.56	8.47	< .001
参加者配置	-79.0	98.37	-0.80	.422
音響定位	-33.0	98.38	-0.34	.737
配置 × 音響	27.0	139.13	0.19	.846

有意ではなかった。

5. 考察

5.1 視線アドレッシングが次話者選択に与える影響

本研究では、front に配置された参加者が次話者となる確率は、非 front 配置時と比較して高い傾向を示したものの、その増加量は限定的であり、ロジスティック回帰分析においても統計的に有意な主効果は確認されなかった。この結果は、視線に基づく空間的アドレッシングが、次話者を直接的に決定する強い指示ではなく、次話者選択に関する弱い手がかりとして機能していたことを示唆している。

対面会話研究においても、視線は次話者を一意に決定する要因ではなく、参与構造や注意配分を調整する役割を担うことが指摘されている。本研究の結果は、こうした知見と整合的であり、オンライン環境においても、視線を空間的に再構成することで、次話者選択に影響を与えうることが示している。

5.2 発話タイミング (沈黙時間) への影響

沈黙時間については、条件間で短縮傾向が観測されたものの、線形混合効果モデルによる分析では統計的に有意差は確認されなかった。一方で、参加者配置や視線強調表現が付与された条件では、平均沈黙時間が短くなる傾向が見られた。

会話における沈黙は、誰が次に発話すべきかという不確実性によって生じることが多い。front 配置によって会話の焦点や参与関係が空間的に明示されることで、参加者が発話開始のタイミングを判断しやすくなった可能性がある。このことから、本研究の手法は、次話者を制御するのではなく、会話進行の予測可能性を高めることで、発話タイミングを支援していたと解釈できる。

5.3 音響情報が果たす補助的役割

本研究の結果から、音響定位は次話者選択を直接的に左右する要因ではないものの、視覚的に提示された空間的アドレッシング構造を聴覚的に補完し、会話理解を安定化さ

せる役割を果たしていたと考えられる。音響定位が付与された条件では、front 配置や発話者注視の効果が比較的安定して観測された。

オンライン会議では、複数の参加者の音声が同一方向から知覚されやすく、話者同定に負荷が生じることがある。左右の音響定位は、発話者の位置情報を聴覚的に補完することで、視覚的な配置表現との対応づけを容易にし、会話の焦点把握を支援していた可能性がある。一方で、音響情報単独による大きな効果は確認されず、視覚情報と組み合わせられることで初めて有効に機能していたと解釈できる。

5.4 オンライン会議支援インタフェースとしての含意

以上の結果から、本研究で提案した参加者配置および視線強調表現は、会話の自律的な進行を妨害することなく、参加者の注意配分や役割理解を補助するオンライン会議支援インタフェースとして機能していたと考えられる。

重要なのは、本システムが次話者を強制的に指名するのではなく、弱い手がかりとして提示されている点である。そのため、次話者率の変化は限定的であった一方で、会話の自然さや参加者の裁量は維持されていた。この特性は、オンライン会議における円滑なターンテイキング支援を設計する上で、重要な示唆を与えるものである。

5.5 本研究の限界

本研究にはいくつかの限界がある。第一に、沈黙時間について条件間で短縮傾向は観測されたものの、線形混合効果モデルによる分析では統計的有意差は確認されなかった。したがって、本研究の結果は、沈黙時間短縮を定量的に実証するものではなく、傾向の提示にとどまる。第二に、本研究で用いた実験タスクはワードワルフに限定されており、課題特有の相互行為構造が結果に影響している可能性がある。他の協調的あるいは対立的タスクにおいても同様の傾向が見られるかについては、今後の検証が必要である。

謝辞 本研究を進めるにあたり、ヤマハ株式会社の関係者の皆様より、音響提示およびシステム設計に関して貴重な助言を賜りました。ここに記して感謝の意を表します。

参考文献

- [1] Vertegaal, R., Weevers, I., Sohn, C. and Cheung, C. : *Conveying Eye Contact in Group Video Conferencing Using Eye-Controlled Camera Directions*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2003), pp. 521–528 (2003).
- [2] Goffman, E. : *Forms of Talk*, University of Pennsylvania Press (1981).
- [3] Clark, H. H. : *Using Language*, Cambridge University Press (1996).
- [4] Argyle, M., Cook, M. and Cramer, D. : Gaze and Mutual Gaze, *British Journal of Psychiatry*, Vol. 165, No. 6,

- pp. 848–850 (1994).
- [5] Kendon, A. : Some functions of gaze-direction in social interaction, *Acta Psychologica*, Vol. 26, pp. 22–63 (1967).
- [6] 榎本美香 : 日本語における聞き手の話者移行適格場の認知メカニズム, *社会言語科学*, pp. 123–140 (2009).
- [7] 飯塚陸斗, 川口一画 : 多人数ビデオ会議における話者交替のための参与役割の提示手法, *情報処理学会研究報告*, 2022-HCI-199, No. 31 (2022).
- [8] 西村圭亮, 上野晃嗣, 坪井創吾, 下郡信宏 : テレビ会議において視線の伝達が話者交替に及ぼす影響の分析, *研究報告グループウェアとネットワークサービス (GN)*, 33(2009-GN-71), pp. 163–168 (2009).
- [9] Nowak, K., Tankelevitch, L., Tang, J. and Rintel, S. : *Hear We Are: Spatial Audio Benefits Perceptions of Turn-Taking and Social Presence in Video Meetings*, CHIWORK 2023: Annual Symposium on Human-Computer Interaction for Work, pp. 1–10, ACM (2023).
- [10] 小林哲則, 西田豊明, 阪口哲男 : 視線が次話者選択に与える影響 : 擬人化エージェントを用いた検討, *人工知能学会誌*, 23(4), pp. 545–551 (2008).
- [11] 坊農真弓 : 会話構造理解のための分析単位—F 陣形—, *人工知能学会誌*, 23(4), pp. 552–559 (2008).
- [12] Jacob, R. J. K. : The use of eye movements in human-computer interaction techniques, *Human-Computer Interaction*, Vol. 6, No. 2, pp. 69–102 (1991).
- [13] Ware, C. and Mikaelian, H. : An evaluation of an eye tracker as a device for computer input, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 183–188 (1987).