

多人数ビデオ会議における 話者交替のための視線提示手法

飯塚 陸斗^{1,a)} 川口 一画^{2,b)}

概要：多人数ビデオ会議において、円滑な話者交替を行うことは対面会議と比較して困難である。その原因として、多人数ビデオ会議では話者交替に大きく影響を及ぼす非言語情報である視線配布がされていないことが考えられる。多人数会話において、自分の担う役割である参与役割を理解することによって、円滑な話者交替が実現される。また、参与役割の理解は視線配布によって促される。本手法では、多人数ビデオ会議において視線配布を再現することにより、参与役割の理解を支援し、話者交替を促進するシステムを提案する。本手法を実現するために、我々は矢印を用いた視線配布表現、およびビデオウィンドウのサイズ変更を用いた視線配布表現を行うビデオ会議システムを実装した。また、本手法が話者交替時の課題に与える影響の調査を目的とした実験の設計を述べる。

1. はじめに

近年、Zoom や Microsoft Teams といったビデオ会議システムの需要が高まっている。しかし、3人以上の多人数によって行われるビデオ会議においては、話者交替が対面会議と比べて難しく、発話衝突や沈黙が多く発生することが課題として挙げられる [1]。ここで発話衝突とは、現在の話し手が発話を終了した時、2人以上の聞き手が同時に話しはじめることにより発話が衝突することである。また、現在の話し手が発話を継続している状態において、他の聞き手が発話を開始し、発話が衝突することである。それに対して、沈黙とは、現在の話し手の発話が終了した時、他の聞き手が次に誰が発話するべきかを判断できず、誰も発話を行わない状態である。

対面会議においては、ビデオ会議と比べて発話衝突や沈黙が起こりづらく円滑な話者交替が可能である。これは会議の参加者が互いの参与役割を把握しているためである。ここで参与役割とは、現在話している話し手、次に発話権を得るべき受け手、会話に参加しているが受け手とはならない傍参与者等、会話における各参加者の役割を示したものである [2]。そしてこのような参与役割を把握する上では非言語的情報が重要であり、中でも視線配布（各参加者が誰を見ているかの情報）は特に重要な役割を果たすことが知られている [3]。具体的には、話し手は発話の終



図 1 矢印により視線配布を表現したシステム。視線配布を行なっている参加者から行われている参加者に対して矢印が表示されている。話し手による視線配布が明確に認識できる。

了前に視線配布を行うことにより、聞き手に発話権譲渡する意志を伝達する。また、聞き手は話し手に視線配布を行うことにより、会話への興味や関心を伝達している。

しかし、多人数ビデオ会議において、視線配布は共有されない。そのため、互いの参与役割を把握することができず、話者交替時の発話衝突や沈黙の発生につながると考えられる。

そこで本研究では、多人数ビデオ会議において、視線検出プログラムを用いて各参加者が誰を見ているかを検出し、それを可視化することによって視線配布を再現するシステムを提案する (図 1)。これにより、多人数ビデオ会議における話者交替時の発話衝突、および沈黙を減少させることを目的とする。

¹ 筑波大学情報メディア創成学類

² 筑波大学システム情報系

^{a)} iitsuka@iplab.cs.tsukuba.ac.jp

^{b)} kawaguchi@cs.tsukuba.ac.jp

2. 関連研究

本章では、まず対面状況における話者交替時の視線配布の役割について述べる。次にビデオ会議における話者交替を支援するために非言語情報を利用した研究について述べる。

2.1 対面状況における話者交替時の視線配布

Goffman [2] は会話の参加者が会話に参加する手続きを述べた。会話においては、現在話している話し手、次に発話権を得べき受け手、会話に参加しているが受け手とはならない傍参加者等、各参加者が参与の度合いに応じた役割（参与役割）をもつ。そして話者交替が起きる際には話し手や受け手などの参与役割が動的に変化する。そして会話の参加者は互いの視線に基づいて自分の参与役割を理解していることが知られている [4, 4-7]。

Kendon [5] は話し手が発話の切れ目において受け手を注視し、受け手が注視を受け入れ話し手に注視を返すことによって話者交替が成立することを示した。これに関連して坊農ら [6] は、話し手の発話終了前の視線配布は発話権譲渡の意思を示し、聞き手が自らの参与役割を意識することを促すことを示した。一方、聞き手の視線配布には、発話に対する興味、関心、および発話欲求が表出されており、話し手による次話者の選択に大きく影響しているという [4, 7]。

本研究では、多人数ビデオ会議においても視線配布を共有することにより、各参加者の参与役割の把握を促進し、円滑な話者交替が可能になると考えた。

2.2 ビデオ通話における非言語情報の利用

ビデオ会議において非言語情報を利用し会話を支援した研究は多い。

玉木ら [8] は、Web 会議において、人が発話前に行う発話予備動作を検出し、次に発話する可能性が最も高い参加者を強調表示することにより、円滑な話者交替を支援する手法を提案した。ここで発話予備動作とは、発話前に体を前に乗り出したり、大きく頷くなど発話欲求を表出する動作のことである [9]。玉木らの提案システムにおいては、発話予備動作として手を上げる、手を顔の周りに近づける、および頷く動作が用いられており、視線方向の検出・提示は行われていない。

多人数ビデオ会議において視線配布を伝達した研究として以下の研究が挙げられる。長谷川ら [10] は、発話意図表出に関連する無意識的身振りを伝達するテレプレゼンスロボットを提案した。この研究で用いられた無意識的身振りには頭部の回転も含まれ、テレプレゼンスロボットの顔部分が回転することにより、遠隔の参加者がどこを見ているかが伝達される。また Vertegaal ら [11] は、アイコンタ



図 2 予備実験の様子。3 人の実験参加者にビデオ会議システムを用いて創造会議を行ってもらい、各参加者の視線配布を調査した。

クトに参加者に伝える多人数ビデオ会議システムを提案した。このシステムではディスプレイ内に表示された 3 次元 CG 空間内に各参加者のビデオウィンドウが立体的に配置され、ビデオウィンドウが回転することにより視線配布を提示する。

これらの研究では視線配布が伝達されているが、実空間もしくは 3 次元 CG 空間内における 3 次元的な空間配置を利用しており、システム構成が複雑なものであった。そこで本研究では、Zoom のように各参加者のビデオウィンドウが 2 次元的に配置される多人数ビデオ会議システムにおいて、視線配布を伝達する手法を提案する。

3. 予備実験

3.1 概要

各参加者のビデオウィンドウが 2 次元的に配置される多人数ビデオ会議システムにおいて、視線配布を伝達する上での懸念点として、このようなシステムを用いる場合に対面会議と同様の視線配布が行われているかは明らかではなかった。そこで本研究では、多人数ビデオ会議における各参加者の視線配布を調査するための予備実験を行った。

予備実験では、3 人の実験参加者にビデオ会議システムを用いて創造会議を行ってもらい、話者交替時の各参加者の視線配布を調査した。ここで創造会議とは、決められた題材の元アイデアを創造する会議であり、参加者が活発に発話する必要があるため、話者交替が起りやすい [12]。使用するビデオ会議システムは WebRTC のプラットフォームである SkyWay を用いて Web アプリケーションとして実装した。システムを用いた実験の様子を図 2 に示す。参加者は互いに親しい知り合いである 3 名の 1 グループで行った。会議の議題は「スマートフォンの新機能を考える」とし、7 分間会議を行った。会議中は、音声、各参加者の視線方向、およびビデオ画面を記録し、話者交替時に参加者が誰のビデオウィンドウを見ているかを調査した。なお視線方向の検出には WebGazer.js [13] を用いた。

3.2 結果と分析

7分間行われた会議で25回の話者交替が起こった。話者交替前において受け手と傍参加者はそれぞれ80%、76%の確率で話し手を見ていた。しかし話者交替前に話し手が受け手を見ていた確率は40%であった。参加者ごとに見ると、参加者のうち1人は自分が話し手の際、受け手を見ている確率がその他の場所を見ている確率に比べて一番高かった、しかし他の2人はそれぞれ話し手である自分のビデオウィンドウを見ている確率、およびディスプレイから目を離している確率が一番高かった。また、話者交替後において、これまでの話し手が次の話し手を見ている確率は88%であった。一方で、話者交替後に次の話し手がこれまでの話し手を見ている確率、および自分自身を見ている確率は共に40%であった。また、話者交替後に傍参加者がこれまでの話し手を見ている確率、および次の話し手を見ている確率はそれぞれ36%、48%であった。

以上の結果をまとめると、多人数ビデオ会議において

- (1) 話者交換前、受け手、および傍参加者は話し手を見ている（対面と同等）
- (2) 話者交替後、これまでの話し手は次の話し手を見ている（対面と同等）
- (3) 話者交替前、および話者交替後、話し手は受け手、または話し手（自分）を見ている（対面と異なる）
- (4) 話者交替後、傍参加者はこれまでの話し手、または次の話し手を見ている（対面と異なる）

対面と異なる結果となった理由として、(3)については、多人数ビデオ会議では対面とは違って自分の姿を確認しながら話すことが可能であることに起因すると考えられる。

(4)については、多人数ビデオ会議では、話者交替時に話し手による視線配布の共有が行われていないことに起因すると考えられる。

4. 提案システム

前章の予備実験の結果より、話者交替前後の話し手の視線配布、および話者交替後の傍参加者の視線配布について、対面会議とは異なる視線配布が見られた。よって提案システムの実装にあたっては、これらの視線配布が対面会議と同等になるよう誘導した上で、視線配布の可視化を行う必要がある。

4.1 対面同等の視線配布の誘起

まず、話者交替前、および話者交替後の話し手の視線配布について、受け手を見ている場合の他に、自分のビデオウィンドウを見ている場合があった。これは対面会議において、自分のことを見ながら話すのは不可能であるのに対し、多人数ビデオ会議では自分を確認しながら話すことが可能であるためであると考えられる。このことから、提案する多人数ビデオ会議システムにおいては、自分のビデオ



図3 ビデオウィンドウの大きさにより視線配布を表現したシステム。多く視線配布が行われている参加者のビデオウィンドウは拡大され、視線配布が行われていない参加者のビデオウィンドウは縮小される。

ウィンドウではなく他の参加者に対して視線配布を行うよう誘導する必要がある。よって自分のビデオウィンドウを画面の下部に小さく配置し、他の参加者のビデオウィンドウを画面中心付近に大きく配置した。これにより、対面と同様話者交替時に話し手から受け手への視線配布が行われるようになると考えた。

また、話者交替後の傍参加者の視線配布について、話者交替前の話し手を見ている場合と話者交替後の話し手を見ている場合があった。これらは、話者交替時に話し手の視線が共有されておらず、話者交替を察知することが難しいからであると考えられる。このことから提案する多人数ビデオ会議システムでは、誰に視線配布を行っているかを2つの手法で表現する（詳細は次節）。

以上により、対面会議とは異なる視線配布を対面会議と同様の視線配布へと誘導できると考えられる。

また、視線配布を可視化することにより、各参加者が明確に自分の参与役割を理解することが可能になると考えられる。これにより、発話衝突、および沈黙を減少させ円滑な話者交替を実現することが可能であると考えられる。

4.2 視線配布表現

本研究では、視線配布を可視化する手法として2種類の表現を提案する。1つ目は矢印による視線配布表現、2つ目はビデオウィンドウの大きさを変更することによる視線配布表現である。

4.2.1 矢印による視線配布表現

矢印によって各参加者の視線配布を表現する（図1）。これは視線配布を行っている参加者から視線配布が行われている参加者に対して、ビデオ画面上で矢印が表示される手法である。誰が誰を見ているかという情報が直接的に共有されるため、特に話し手による視線配布が明確に認識できる。

4.2.2 ビデオウィンドウの大きさによる視線配布表現

多く視線配布が行われている参加者のビデオウィンドウを拡大して表示し、視線配布が行われていない参加者のビ

デオウィンドウを縮小して表示する。(図3)例えば自分に多く視線配布が行われている場合は、自分のビデオウィンドウが拡大、他の参加者のビデオウィンドウが縮小されて表示される。誰が見ているかという情報は共有されず、誰が多く見られているかという情報が強調して共有される。この表現は、視線配布が多く集まっている参加者が一番大きく表示され、聞き手にとって視線配布すべき相手が明確である。

4.3 システム構成

本研究では視線配布表現を伴う多人数ビデオ会議システムをWebアプリケーションとして実装した。

ビデオ通話の実装について、WebRTCのプラットフォームであるSkyWayを用いた。また、検出した視線方向の送受信についても、SkyWayを用いた。視線方向の検出にはJavaScriptのライブラリであるWebGazer.js [13]を用いた。WebGazer.jsによりノートパソコンに備わっているカメラからディスプレイ上のどこの点を見ているかという視線方向を検出し、多人数ビデオ会議時にその点がどこの範囲内にあるかを判定し誰を見ているかという視線配布を検出する。検出された視線配布情報をSkyWayを用いて各参加者に送信する。受信した視線配布情報をそれぞれの視線配布表現に従って表現する。

5. 今後の予定

前章の提案システムにより、実際に話者交替が促進されるかを明らかにすることを目的に実験を行う予定である。3人の実験参加者にビデオ会議システムを用いて、3つの条件で創造会議を行う。

5.1 実験概要

視線配布表現なし、矢印による視線配布表現、およびウィンドウサイズの大きさによる視線配布の3つの条件で比較実験を行う。男女18名の参加者を集め、3人1グループとして6回行う。参加者のグループメンバーは互いに親しい知り合いである。6つのグループは7分間の創造会議を各条件で1回ずつ行う。条件の順序はランダムとする。会議の議題は「グループメンバーで行く旅行の夜に行うレクリエーションを考える」とする。会議中の音声、ビデオウィンドウへの視線方向、およびビデオ画面を記録し、発話衝突の発生回数、沈黙の発生回数および発生時間、およびビデオウィンドウへの視線配布を調査する。また各条件の終了後、アンケートに回答してもらい、その条件に行われた会議についての印象評価を行う。

6. おわりに

本研究では、多人数ビデオ会議において発話衝突や沈黙を減少させ、円滑な話者交替を実現することを目的とした。

対面会議において、会議の参加者は互いの参与役割を把握することによって、円滑な話者交替が可能である。また、参与構造の把握には視線配布が使われている。そこで多人数ビデオ会議において、視線配布の共有を行い、各参加者へ参与役割の理解を促す手法を提案した。

多人数ビデオ会議において、対面会議と同様の視線配布が行われているか明らかでなかったため予備実験を行い、多人数ビデオ会議では一部で対面とは異なる視線配布が行われていることを示した。

本手法では視線配布が対面会議と同等になるよう誘導した上で、視線配布の可視化を行うビデオ会議システムを実装した。視線配布の可視化においては、矢印による視線配布表現、およびビデオウィンドウの大きさによる視線配布表現を実装した。

今後は、提案したシステムを用いてその有効性を確かめるために実験を行う予定である。

参考文献

- [1] 鍍沢勇, 滝川啓, 大久保栄, 渡辺義郎: 衛星通信を利用した画像会議におけるエコー及び伝搬遅延の影響, 電子情報通信学会論文誌 B, Vol. 64, No. 11, pp. 1281-1288 (1981).
- [2] Goffman, E.: Replies and Responses, *Language in Society*, Vol. 5, No. 3, pp. 257-313 (online), available from <http://www.jstor.org/stable/4166887> (1976).
- [3] Goodwin, C.: *Conversational Organization: Interaction Between Speakers and Hearers* (1981).
- [4] Vertegaal, R., Slagter, R., van der Veer, G. and Nijholt, A.: Eye Gaze Patterns in Conversations: There is More to Conversational Agents than Meets the Eyes, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, New York, NY, USA, Association for Computing Machinery, p. 301-308 (online), DOI: 10.1145/365024.365119 (2001).
- [5] Kendon, A.: Some functions of gaze-direction in social interaction, *Acta Psychologica*, Vol. 26, pp. 22 - 63 (online), DOI: [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4) (1967).
- [6] 真弓坊農, 紀子鈴木, 恭弘片桐: 多人数会話における参与構造分析: インタラクショナル行動から興味対象を抽出する, 認知科学, Vol. 11, No. 3, pp. 214-227 (オンライン), DOI: 10.11225/jcss.11.214 (2004).
- [7] 良 佐藤, 勇剛竹内: 多人数対話におけるロボットの視線行動に基づく発話権と対話場のデザイン, HAI シンポジウム 2013, pp. 219-228 (2013).
- [8] 秀和玉木, 豪 東野, 稔 小林, 雅行井原, 謙一岡田: 遠隔会議における発話衝突低減手法, 情報処理学会論文誌, Vol. 53, No. 7, pp. 1797-1806 (オンライン), 入手先 <https://ci.nii.ac.jp/naid/110009423549/> (2012).
- [9] Vargas, M. F.: *Louder than words : an introduction to nonverbal communication*, Iowa State University Press (1986).
- [10] 長谷川孔明, 中内 靖: テレプレゼンスロボットによる無意識的身ぶりの表出が発話交替に与える影響, 日本機械学会論文集, Vol. 80, No. 819, pp. DR0321-DR0321 (オンライン), DOI: 10.1299/transjsme.2014dr0321 (2014).
- [11] Vertegaal, R., Weevers, I., Sohn, C. and Cheung, C.: GAZE-2: Conveying Eye Contact in Group Video Con-

ferencing Using Eye-Controlled Camera Direction, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, New York, NY, USA, Association for Computing Machinery, p. 521–528 (online), DOI: 10.1145/642611.642702 (2003).

- [12] 智雄井上, 謙一岡田, 温 松下: テレビ会議における映像表現の利用とその影響, 情報処理学会論文誌, Vol. 40, No. 10, pp. 3752–3761 (オンライン), 入手先 (<https://ci.nii.ac.jp/naid/110002725123/>) (1999).
- [13] Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J. and Hays, J.: WebGazer: Scalable Webcam Eye Tracking Using User Interactions, *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, AAAI, pp. 3839–3845 (2016).