

# 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

飯塚 陸斗<sup>1,a)</sup> 川口 一画<sup>2,b)</sup>

**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。これは多人数ビデオ会議において、参与役割（誰が次の話者であるか等）を把握することが困難であることに起因すると考えられる。そこで本研究では、多人数ビデオ会議において参与役割の把握を支援し、話者交替を促進するシステムを提案する。我々はシステムの実装にあたり、参与役割の推定および提示方法を検討した。まず、対面会議では参与役割の把握に視線配布が用いられるという知見に基づき、各参加者の視線情報を取得・提示するシステムを実装した。提示手法として、各参加者の視線情報より参与役割を推定しウィンドウサイズを変化させる手法、および視線配布を直接矢印としてビデオウィンドウに重畳表示する手法を実装した。比較実験の結果、参与役割に応じてウィンドウサイズを変化させる手法において、参与役割の把握および話者交替が促進されることが示唆された。一方で、視線は不随意的な運動を伴うため、ウィンドウサイズやベクトル方向が頻繁に変化し会議への集中度が低下すること、参与役割の調整のための能動的な制御が困難であること、さらにアイトラッキングの精度不足等の課題が示唆された。そこで我々は、視線と同様に対面会議において参与役割の把握に用いられる身体配置（各参加者の位置および方向）を再現するシステムを提案する。

## 1. はじめに

近年、Zoom [1] や Microsoft Teams [2] 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、多人数ビデオ会議において、対面会議と比較して話者交替が難しく、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 [3,4]。ここで発話衝突とは、会議の参加者が同時に話し始めてしまうことである。これに対して沈黙とは、次に誰が発話するべきか判断できず、誰も発話を行わない状態である。

対面会議において、発話衝突や沈黙が起きることは少なく、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 [5]。参加者が参与役割を把握する際、非言語情報が用いられる。非言語情報のうち、特に視線配布（各参加者が誰を見ているかの情報）は話者交替に重要な役割を果たすことが知ら



図 1 提案する視線配布表現を伴う多人数ビデオ会議システム。矢印による視線配布表現（左）およびビデオウィンドウの大きさによる視線配布表現（右）を示す。

れている。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、参加者は互いの参与役割を把握することが難しく、話者交替時の発話衝突や沈黙が多く発生すると考えられる。この課題に対し、3次元CG空間を用いて視線配布を共有することにより、会話を促進する研究は多く行われている [6,7]。

本研究において、2次元的にビデオウィンドウが配置される多人数ビデオ会議において参与役割の把握を促し、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。そのために各参加者の視線情報を取得・提示するシステムを提案した [8]。提示手法として、視線情報を直接矢印としてビデオウィンドウに重畳表示する手法、および各参加者の視線情報により参与役割を推定しウィンドウサイズを変化させる手法を実装した（図1）。

<sup>1</sup> 筑波大学 情報理工学位プログラム

<sup>2</sup> 筑波大学 システム情報系

<sup>a)</sup> iitsuka@iplab.cs.tsukuba.ac.jp

<sup>b)</sup> kawaguchi@cs.tsukuba.ac.jp

実装したシステムを用いて、比較実験を行った結果、参与役割に応じてウィンドウサイズを変化させる手法において、参与役割の把握が促進され、話者交替が促進することが示唆された。一方で参与役割の提示方法、および推定方法に対する課題が示唆された。そこで本研究において、システムの改善手法として、視線配布と同様に参与役割の把握に用いる各参加者の位置および方向といった身体配置を再現するシステムを提案する。

## 2. 関連研究

本研究は、多人数ビデオ会議において話者交替を促進することを目的とし、そのために視線配布を利用する。本章では、多人数会議に関する社会学的知見、およびビデオ会議において非言語情報を利用した関連研究について説明した後、本研究の方針について述べる。

### 2.1 多人数会議に関する社会学的知見

非言語情報はコミュニケーションの中で重要な役割を果たしている [9]。ここで非言語情報とは、コミュニケーションを行う際の言語以外の情報のことを指し、主に対人距離、体の動き、表情、視線配布、接触、準言語、嗅覚作用、および人工物などが挙げられる [10]。ここで本研究では視線配布に着目する。これは、視線配布が話者交替において重要な役割を持つため、そして通常のビデオ会議ではその共有が行われていないためである。視線配布が共有されないビデオ会議において、話者交替に関して視線はほとんど機能していない。よって視線配布を共有することによってビデオ会議において他の参加者の状態を把握するための判断材料が増え、話者交替を促進できるのではないかと考えられる。

Goffman [5] 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、会話には参与せずに話のみを聞くのみである傍観者および盗み聞き者等、各参加者が参与の度合いに応じた役割（参与役割）を担う。さらに話者交替が起き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配布や身体配置に基づいて自分の参与役割を理解している [11–14]。Kendon [11] 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（相互注視）によって話者交替が成立することを示した。また、話し手の発話終了前の視線配布は発話権譲渡の意思が表出し、聞き手が自らの参与役割への意識を促すことを示した。このことから会話の参与役割は話し手によって主体的に選択されることがわかる。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 [13, 14]。

### 2.2 話者交替支援のためにビデオ会議において非言語情報を利用した研究

ビデオ会議において非言語情報を用いて話者交替支援を行った研究は多く存在する。玉木ら [15] 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。Vertegaal ら [6, 16] は、アイコンタクトを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。また、Vertegaal ら [17] は 3 人のグループに視線情報を共有しながら協調作業を行わせた。その結果、協調作業においても視線を共有することは話者交替を促進することを明らかにした。Okada ら [18] は 3 人でのビデオ会議において等身大の相手画像と視線を一致させ、湾曲スクリーンに投影することにより、3 人で机を囲んでいるようにビデオ会議を行った。

### 2.3 本研究の方針

対面で行う多人数会議において、参加者は視線配布から参与役割を把握することによって円滑な話者交替を行うことができる。しかし、ビデオ会議において視線配布は共有されないため、話者交替時に自分の参与役割が把握しにくく、発話衝突や沈黙が発生すると考えられる。これまでの研究において、ビデオ会議において、視線配布を共有することによって会話を促進する研究が多くされていた。しかし、それらの研究は実空間、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いた。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を提案する。また、視線配布を提示することによって多人数ビデオ会議においても参与役割の把握を促し、話者交替を促進する手法を提案する。

## 3. 提案システム

本研究では、多人数ビデオ会議システムにおいて視線配布を提示することによって参与役割の把握を促し、話者交替を促進する手法を提案する。本章では、視線配布表現、およびシステム構成について述べる。

### 3.1 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であり、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。

#### 3.1.1 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 2）。こ



図 2 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る。話し手による視線配布が明確に認識できる。話者交替前における想定される話し手のビデオ画面 (左), および傍参加者のビデオ画面 (右) を示す。



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。話者交替前における想定される傍参加者のビデオ画面 (左), および受け手のビデオ画面 (右) を示す。

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、ビデオウィンドウに矢印が重畳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。

### 3.1.2 ビデオウィンドウの大きさによる視線配布表現

視線配布が多く集ま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)。例えば他の参加者が自分に視線配布を向けている場合は、自分のビデオウィンドウが拡大され、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。ビデオウィンドウの大きさは違いが明確に知覚できるように、視線配布が 1 人集まる毎にそのビデオウィンドウの面積を 2.25 倍の大きさに拡大した。誰が見ているかという情報は共有されず、誰が視線配布を多く集め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。つまり、視線配布がビデオウィンドウの大きさとして間接的に共有されている。またこのビデオウィンドウの大きさが参与役割を表している。

## 3.2 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 [19] 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 [20] 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か点を見ているかという視線情報を検出し、多人数ビデオ会議時にその点がど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した。検出された視線配布を SkyWay を用いて各参加者に送信する。受信した視線配布をそれぞれの視線配布表現に従って表現する。

## 4. 実験

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、および提案する 2 種類の表現のうち、より参与役割の把握を促進する手法を明らかにすることを目的として実験を行った。本章では、実験設計、仮説と評価、結果、考察、および議論と今後の課題を述べる。

### 4.1 実験設計

本実験では、3 人の実験参加者がビデオ会議システムを用いて会議を行う実験設計とした。会議について話者交替が頻繁に行われる創造会議を実験タスクとして設定した。実験参加者は 9 名 (男性 8 名, 女性 1 名, 平均 21.7 歳,  $SD = 0.67$ ) であり、参加者内配置を用いた。参加者はそれぞれ 3 名 1 組の 3 グループに分けられた。グループ内における参加者はお互いに親しい知り合いであった。参加者は日常的に多人数ビデオ会議を行っていた。

実験条件として下記の 3 つの条件を用いた。

- 1). **視線配布無し条件**: 既存のビデオ会議を想定した条件である。視線配布の共有を行わないビデオ会議システムを用いて会議を行う。
- 2). **矢印条件**: 矢印による視線配布表現を伴うビデオ会議システムを用いて会議を行う条件である。誰から誰へ視線配布が行なわれているかが矢印によって直接的に表現される。
- 3). **ウィンドウ条件**: ビデオウィンドウの大きさによる視線配布表現を伴うビデオ会議システムを用いて会議を行う条件である。視線配布が多く行われる参加者のビデオウィンドウが大きく表示される。

3 つの実験条件において、21 分間会議を行い、7 分間毎に条件を変更した。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」とした。各参加者はそれぞれ与えられた部屋に分かれて会議を行った。各条件が終了した後に、参加者に直前に行っていた会議についての印象評価をアンケートとして回答させた。アンケートの項目を表 1 に示す。

また全ての条件の終了後、システム、および実験に関する最終アンケートを回答させた。分析のために、会議にお

表 1 各条件間に参加者に回答させたアンケートを示す。

番号	項目
A-1	他の参加者が自分の話を聞いているように感じた
A-2	自分がちゃんと聞いていることが相手に伝わっているように感じた
A-3	自分が話し終える時、次に誰が話し始めそうかわかった
A-4	他の人が話し終える時、次に誰が話し始めそうかわかった
A-5	他の人が話し終える時、自分が話したい時に話すことができた
A-6	話す人が変わる時、話し始めが他の人とぶつかった
A-7	話す人が変わる時、話し始めが他の人とぶつかったことがストレスに感じた
A-8	話す人が変わる時、誰も話し始めず沈黙が発生した
A-9	話す人が変わる時、誰も話し始めず沈黙が発生したことがストレスに感じた
A-10	スムーズに議論が進んだ
A-11	議論に集中できた
A-12	誰が話しているかが明確にわかった

ける音声、およびビデオ画面を記録した。実験後に音声およびビデオ画面の記録を見返し、評価項目のカウントを行った（評価項目については次節参照）。

#### 4.2 仮説と評価

本実験では次の仮説の検証を行った。

- **仮説 1**：矢印条件、およびウィンドウ条件は視線配布無し条件と比較して参与役割を把握しやすくなる
- **仮説 2**：矢印条件、およびウィンドウ条件は視線配布無し条件と比較して話者交替が促進され
- **仮説 3**：矢印条件、およびウィンドウ条件は視線配布無し条件と比較して会議が促進される

これらの仮説を検証するために会話分析を行った。仮説 2 を検証するため、発話衝突確率、沈黙の発生確率、1 回あたりの沈黙時間の増減、および話者交替が不成立となった確率を評価する。仮説 3 を検証するため、参加者の発話数を評価する。また、アンケートによる会議の主観評価を行った。A-1 から A-5 を仮説 1、A-6 から A-9 を仮説 2、そして A-10 および A-11 を仮説 3 の評価として用いる。

#### 4.3 結果

本節では会話分析、およびアンケートによる印象評価について述べる。

##### 4.3.1 会話分析

まず各条件における発話衝突確率の平均を図 4 に示す。グラフのエラーバーは標準誤差を示す。発話衝突確率は会議間における発話衝突の起きた回数を会議間全ての話者交替の回数（発話衝突・沈黙等により成立しなかった場合を含む）で割って求めた。発話衝突確率はウィンドウ条件において一番低く、視線配布無し条件および矢印条件はあまり変わらなかった。一元配置分散分析を行ったところ、群間

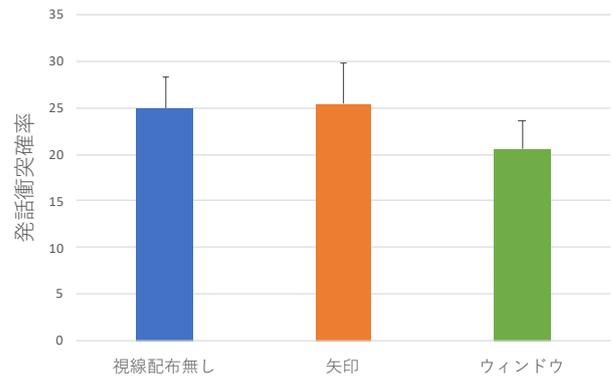


図 4 各条件における発話衝突確率。

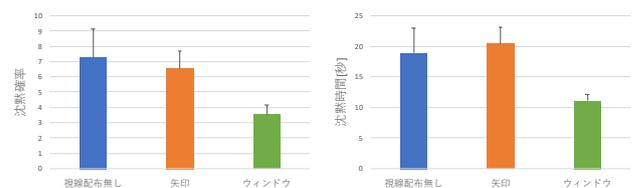


図 5 各条件における沈黙発生確率と発生時間。

自由度 2、群内自由度 6、 $p$  値 0.2453 で主効果はなかった。

次に各条件における沈黙発生確率、および発生した沈黙 1 回あたりの時間の平均を図 5 に示す。沈黙は話者交替時に不自然に間が空いたときに沈黙が発生したとカウントした。沈黙発生確率は会議間における沈黙の起きた回数を成立したかに関わらず会議間全ての話者交替の回数で割って求めた。沈黙発生確率の平均について、視線配布無し条件において一番高く、ウィンドウ条件において一番低かった。1 回あたりの沈黙時間の平均について、ウィンドウ条件において一番短かった。

沈黙発生確率について、一元配置分散分析を行ったところ、群間自由度 2、群内自由度 6、 $p$  値 0.0511 ( $<0.10$ ) であり、有意水準 10% で主効果があり、条件によって沈黙発生確率は異なる結果が得られた。Bonferroni-Holm の方法により有意水準を補正し、 $t$  検定を用いて多重比較したところ、各条件間で有意差は見られなかった。1 回あたりの沈黙時間について、一元配置分散分析を行ったところ、群間自由度 2、群内自由度 6、 $p$  値 0.6252 で主効果はなかった。

また、発話衝突または沈黙が発生し、話者交替が不成立となった確率の平均を図 6 に示す。話者交替が不成立となった確率の平均について、ウィンドウ条件において一番低く、視線配布無し条件および矢印条件はあまり変わらなかった。一元配置分散分析を行ったところ、群間自由度 2、群内自由度 6、 $p$  値 0.0870 ( $<0.10$ ) であり、有意水準 10% で主効果があり、条件によって話者交替が不成立となった確率は異なる結果が得られた。Bonferroni-Holm の方法により有意水準を補正し、 $t$  検定を用いて多重比較したところ、視線配布無し条件およびウィンドウ条件間において、有意水準 10% でウィンドウ条件が有意に低い結果が得ら

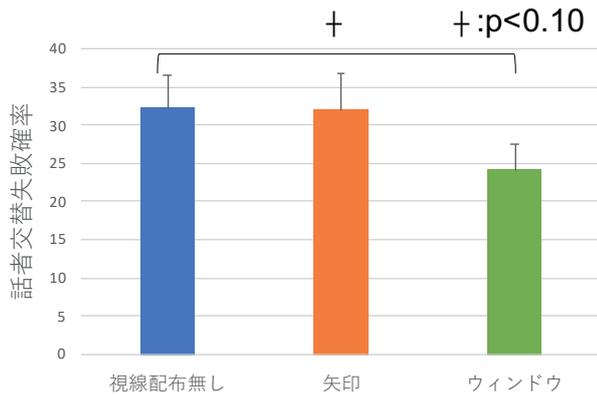


図 6 話者交替が不成立となった確率。視線配布無し条件とウィンドウ条件間に有意傾向があった。

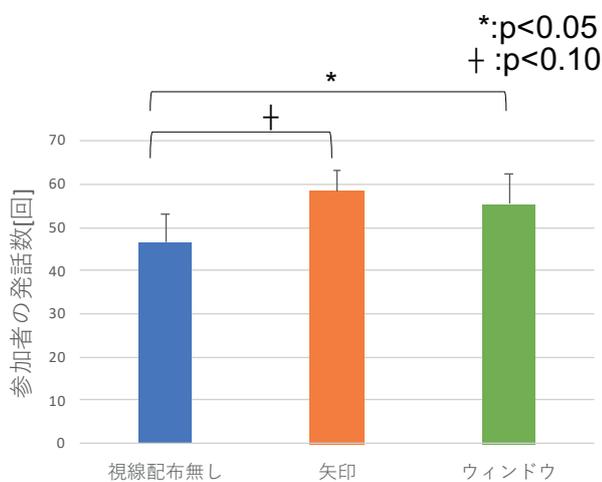


図 7 参加者の発話数の平均について示す。視線配布無し条件とウィンドウ条件間で有意差があり、視線配布無し条件とウィンドウ条件間で有意傾向があった。

れた。

次に、参加者の発話数について各条件の平均を図 7 に示す。矢印条件、およびウィンドウ条件の両方で視線配布無し条件と比較して発話数が多い。一元配置分散分析を行ったところ、郡間自由度 2、郡内自由度 24、 $p$  値 0.0296 ( $<0.050$ ) であり、有意水準 5% で主効果があり、条件によって発話数は異なる結果が得られた。Bonferroni-Holm の方法により有意水準を補正し、 $t$  検定を用いて多重比較したところ、視線配布無し条件およびウィンドウ条件間において、有意水準 5% でウィンドウ条件が有意に発話数が多いという結果が得られた。また、視線配布無し条件と矢印条件間において、有意水準 10% で矢印条件が有意に発話数が多いという結果が得られた。

#### 4.3.2 アンケートによる印象評価

各条件におけるアンケートについて、その結果を図 8、および図 9 に示す。

各アンケート項目について 3 条件間で Friedman 検定を

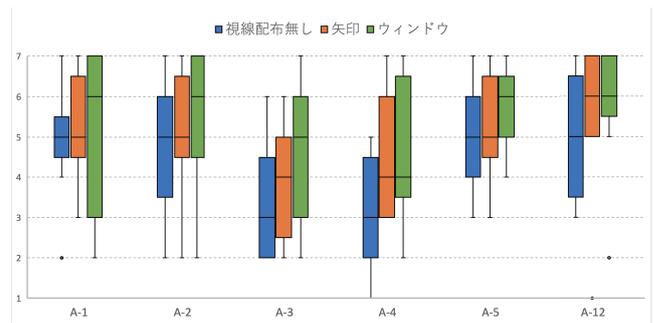


図 8 各条件後に行ったアンケートの結果を示す。これらは参与役割の把握を促せたかどうかの主観評価を行うための項目であり、A1 から A5 まで、および A12 の結果を示す。

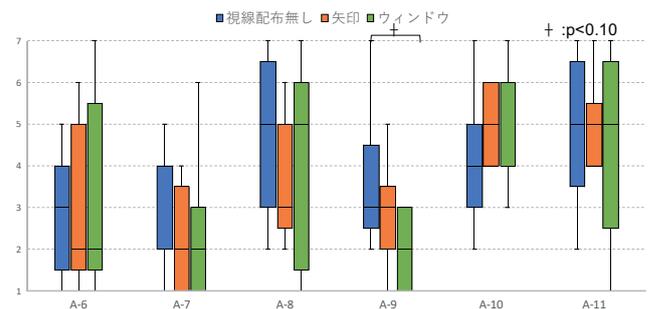


図 9 各条件後に行ったアンケートの結果を示す。これらは話者交替が促進されたか、および議論が促進されたかどうかの主観評価を行うための項目であり、A6 から A11 までの結果を示す。

行った。ここでは主効果のあった 3 つの項目について述べる。まず「(A-4) 自分が話終える時、次に誰が話し始めそうかわかった」について、検定の結果、自由度 2、カイ二乗値 5.2414、 $p$  値 0.07275 ( $<0.10$ ) であり、有意水準 10% で主効果があった。そこで Bonferroni の方法によって有意水準を補正し、Wilcoxon の符号順位検定を用いて多重比較を行ったところ、条件間に有意差は見られなかった。

次に「(A-9) 話す人が変わる時、誰も話し始めず沈黙が発生したことがストレスに感じた」について、検定の結果、自由度 2、カイ二乗値 9.1724、 $p$  値 0.01019 ( $<0.050$ ) であり、有意水準 5% で主効果があった。そこで Bonferroni の方法によって有意水準を補正し、Wilcoxon の符号順位検定を用いて多重比較を行ったところ、視線配布無し条件およびウィンドウ条件間に有意水準 10% の有意傾向が見られた。

最後に「(A-12) 誰が話しているかが明確にわかった」について、検定の結果、自由度 2、カイ二乗値 6.9286、 $p$  値 0.0313 ( $<0.050$ ) であり、有意水準 5% で主効果があった。そこで Bonferroni の方法によって有意水準を補正し、Wilcoxon の符号順位検定を用いて多重比較を行ったところ、条件間に有意差は見られなかった。

#### 4.4 考察

本節では実験の結果を踏まえて、各仮説に対する考察を

行う。

#### 4.4.1 仮説 1：矢印条件、およびウィンドウ条件は視線配布無し条件と比較して参与役割を把握しやすくなる

図 8 の結果より、全ての項目において矢印条件および視線配布無し条件間、およびウィンドウ条件および視線配布無し条件間に有意差は見られなかった。よって仮説 1 は支持されなかった。しかし、中央値を比較すると、まずウィンドウ条件および視線配布無し条件間において、全ての項目においてウィンドウ条件の評価が高かった。また矢印条件および視線配布無し条件間において、「(A-1) 他の参加者が自分の話を聞いているように感じた」、「(A-2) 自分がちゃんと聞いていることが相手に伝わっているように感じた」、および「(A-5) 他の人が話し終える時、自分が話したい時に話すことができた」の項目における中央値は等しく、他の項目においては矢印条件の方が高かった。本実験においては、参加者が 9 人と少なかったため、これらの項目について有意差が見られなかった可能性がある。今後、参加者数を増やした再実験を行うことにより、前述の各条件間の差について有意差が見られる可能性がある。

#### 4.4.2 仮説 2：矢印条件、およびウィンドウ条件は視線配布無し条件と比較して話者交替が促進される

仮説 2「矢印条件、およびウィンドウ条件は視線配布無し条件と比較して話者交替が促進される」について、図 9 の結果より、主観評価において 1 つの項目を除いて矢印条件および視線配布無し条件間、およびウィンドウ条件および視線配布無し条件間に有意差は見られなかった。また、会話分析において、発話衝突確率 (図 4)、沈黙発生確率および発生時間 (図 5) において、矢印条件および視線配布無し条件間、およびウィンドウ条件および視線配布無し条件間に有意差は見られなかった。しかし、図 6 の結果より、話者交替の不成立となった確率についてウィンドウ条件および視線配布無し条件間において有意差が見られた。よってウィンドウ条件において話者交替は促進され、仮説 2 は部分的に支持されたと考えられる。

#### 4.4.3 仮説 3：矢印条件、およびウィンドウ条件は視線配布無し条件と比較して議論が促進されている

仮説 3「矢印条件、およびウィンドウ条件は視線配布無し条件と比較して議論が促進される」について、まず図 9 の結果より、主観評価において矢印条件および視線配布無し条件間、およびウィンドウ条件および視線配布無し条件間に有意差は見られなかった。しかし図 7 の結果より、参加者の発話数について矢印条件および視線配布無し条件間、およびウィンドウ条件および視線配布無し条件間に有意差が見られた。よってウィンドウ条件および視線配布無し条件間において、会話は促進されたと考えられる。よって仮説 3 は部分的に支持されたと考えられる。

## 4.5 議論と今後の課題

本節では実験により得た考察をもとに議論、および今後の課題を述べる。

### 4.5.1 提案手法について

矢印条件において、話者交替は促進されなかった。また、参与役割の把握は促進されなかったと考えられる。一方でウィンドウ条件において、参与役割の把握を促進した可能性があった。また話者交替の不成立が減少し、発話数が増加した。よって多人数ビデオ会議において、視線配布の共有は話者交替を促進し、会議を促進すると考えられる。ここで実験参加者数が少なく明確な差は見られなかったが、矢印条件に対してウィンドウ条件において、より話者交替の促進に効果があることが示唆された。以上のことから、システムにより参与役割を提示することで話者交替および参与役割の把握の促進に効果があると考えられる。一方で実験終了後に実施した自由記述のアンケートにおいて「ビデオウィンドウが高頻度で動くため、目が疲れた」という意見が見られた。これは参与役割をウィンドウサイズのみで提示したことから、参与役割の切り替わりに応じて頻繁にサイズが変更されたためであると考えられる。よって今後はサイズの変更だけでない参与役割を提示するシステムが必要であると考えられる。

### 4.5.2 システム設計について

本実験において、「議論に集中できた」という項目の主観評価において、条件の中でばらつきが大きかった。また、実験終了後に実施した自由記述のアンケートにおいて「矢印が出たり消えたりするのが邪魔であった」および「ビデオウィンドウが瞬間的に変わるのがストレスに感じた」等の意見が見られた。この原因について、実験参加者によってアイトラッキングを用いた視線検出において、キャリブレーションの精度にばらつきがあり、参加者によって視線情報の誤検出が細かく発生したことに起因すると考えられる。また会議参加者は他の参加者の様子を確かめるため、無意識的にビデオウィンドウに一瞬視線を向けることがある。その際に、ウィンドウサイズや矢印が頻繁に変化するため、それが会議への集中を妨害する可能性がある。またそれにより、視線配布を向けるべき相手に向ける参与役割の能動的な制御も困難であると考えられる。以上のことから、本システムにおいて視線情報を用いることによって多くの課題が発生した。よって今後はより精度の高く会議への集中を妨害しない、参与役割の把握を促進するシステムが必要であると考えられる。

## 5. 改善指針

前章において、ウィンドウサイズのみで参与役割を提示した際に高頻度でウィンドウサイズが切り替わる課題が示された。また、視線情報を用いることにより、集中が低下する、能動的な制御が困難である、そして視線検出の精度

が不足している等の課題が示された。そこで本章では、上記の課題を解決するため、参与役割の提示方法および推定方法を改善するための提案を行う。まず、参与役割の提示方法の改善について述べ、そして推定方法の改善について述べる。

### 5.1 参与役割の提示方法の改善

提案手法において、ウィンドウサイズのみで参与役割を提示した。その結果、高頻度でウィンドウサイズが切り替わり、目が疲労する課題が生じた。この課題の改善のため、ウィンドウサイズに加え、2次元的な配置も組み合わせる表現する手法を提案する。2次元的な配置の表現として対面で用いられる身体配置を再現する。対面会議において、各参加者の位置および方向といった身体配置は視線配布と同様に参与役割の把握に用いられる。本システムにおいて、身体配置を再現するために、F 陣形を利用する。F 陣形とは会話構造を理解するための分析単位であり、2人以上の複数人が集まって会話する際の空間が維持される現象を説明している [21,22]。F 陣形において、空間を複数人の会議参加者の下半身方向によって円陣の中央にできる O 型の空間 (O 空間)、O 空間を構成するために参加者が身体を配置することによって作られる O 空間の外円的な空間 (P 空間)、そして P 空間の外側の空間 (R 空間) の 3 つに分ける。会話の途中から参与する人物はまず、R 空間に留まり、会話に参与するタイミングを見計らう。その後、P 空間に身体を運び入れることで会話の参与者であることを他の会話の参与者に表明する。ここで話者交替は P 空間で行われる。つまり、P 空間には話し手、受け手および傍参加者が存在し、O 空間に傍観者・盗み聞き者が存在する。会議参加者はこのようにして身体配置から参与役割を把握している。本システムにおいて、F 陣形を再現するため、まずビデオウィンドウを円形に配置する。また、参与役割に応じて中心からの距離を変化させることにより、P 空間および R 空間の移動を再現する。さらに、P 空間において、ビデオウィンドウサイズを変更することによって、参与役割である話し手、受け手、および傍参加者等の把握を促進する。

### 5.2 参与役割の推定方法の改善

提案手法において、視線配布のみに応じて参与役割を推定し、表示した。しかし、視線情報を用いることにより、能動的な制御が困難であることや、視線検出の精度が不足している等の課題が発生した。そこで本システムにおいて、非言語情報のうち、発話欲求の表出する動作 (発話予備動作) を用いて参与役割を変化させる。対面会議において、参与役割は、傍観者・盗み聞き者から傍参加者、傍参加者から受け手、受け手から話し手へ、またはそれらの逆向きに状態が変化する。よって本システムにおいて、参与役割

は対面と同様に段階的に変化させる。まず、傍観者・盗み聞き者から傍参加者への変化は発話予備動作を用いる。オンライン会議における発話予備動作として、口の開き、相槌、頭部運動、手の動き等が挙げられる [23-25]。本研究ではこれら 4 つの発話予備動作に、直前の発話の有無を加えた 5 つの要素に重み付けを行い、利用することを考えている。そして、それぞれの要素が行われた回数に応じたスコア (発話欲求度) を算出し、ある閾値を超えた場合に傍参加者として変化させる。発話欲求度は時間経過で減少する。傍参加者から受け手について、話し手の頭部角度を用いた選択を行う。参与役割において、受け手は話し手の選択によって決定する。また、複数人会話において頭部運動は受け手との関連が高いことが報告されている [23]。よって傍参加者のうち、話し手の頭部角度が向けられている参加者を受け手として変化させる。最後に受け手から話し手について、一定時間以上の発話が行われた場合に話し手として変化させる。

## 6. 終わりに

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。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。視線配布の共有においては、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。

実装したビデオ会議システムを用いた実験の結果、ビデオウィンドウの大きさによる視線配布表現においては話者交替を促し、会議を促す可能性があった。一方で視線情報をを用いたことにより、会議への集中や能動的な制御が困難、および視線検出の精度が不足といった課題が生じた。

これに対し、我々は多人数ビデオ会議システムにおいて身体配置を共有することで参与役割の把握を促進する手法を提案した。

今後は、提案手法を用いて実験を行い、システムの評価を行う。

## 参考文献

- [1] Zoom Video Communications, I.: Zoom, <https://zoom.us>.
- [2] Microsoft: Microsoft Teams, <https://www.microsoft.com/ja-jp/microsoft-365/microsoft-teams/free>.
- [3] 鍛沢勇, 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).
- [4] 小磯花絵, 伝 康晴: 円滑な話者交替はいかにして成立するか, 認知科学, Vol. 7, No. 1, pp. 93-106 (オンライン), DOI: 10.11225/jcss.7.93 (2000).
- [5] 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).
- [6] Vertegaal, R.: The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, New York, NY, USA, Association for Computing Machinery, pp. 294–301 (online), DOI: 10.1145/302979.303065 (1999).
- [7] Mukawa, N., Oka, T., Arai, K. and Yuasa, M.: What is Connected by Mutual Gaze? User's Behavior in Video-Mediated Communication, *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, New York, NY, USA, Association for Computing Machinery, pp. 1677–1680 (online), DOI: 10.1145/1056808.1056995 (2005).
- [8] Itsuka, R., Kawaguchi, I., Shizuki, B. and Takahashi, S.: Multi-party Video Conferencing System with Gaze Cues Representation for Turn-Taking, *Collaboration Technologies and Social Computing* (Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A. and Ogata, H., eds.), Cham, Springer International Publishing, pp. 101–108 (2021).
- [9] Vargas, M. F.: *Louder than words : an introduction to nonverbal communication*, Iowa State University Press (1986).
- [10] 原岡一馬: 人間とコミュニケーション, ナカニシヤ出版 (1990).
- [11] 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).
- [12] 坊農真弓, 鈴木紀子, 片桐恭弘: 多人数会話における参与構造分析: インタラクシオン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).
- [13] 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, pp. 301–308 (online), DOI: 10.1145/365024.365119 (2001).
- [14] 佐藤良, 竹内勇剛: 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).
- [15] 玉木秀和, 東野豪, 小林 稔, 井原雅行, 岡田謙一: 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* (2012).
- [16] Vertegaal, R., Weevers, I., Sohn, C. and Cheung, C.: GAZE-2: Conveying Eye Contact in Group Video Conferencing 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, pp. 521–528 (online), DOI: 10.1145/642611.642702 (2003).
- [17] Vertegaal, R., Van der Veer, G. and Vons, H.: Effects of gaze on multiparty mediated communication, *Graphics interface*, pp. 95–102 (2000).
- [18] Okada, K.-I., Maeda, F., Ichikawaa, Y. and Matsushita, Y.: Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design, *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, New York, NY, USA, Association for Computing Machinery, pp. 385–393 (online), DOI: 10.1145/192844.193054 (1994).
- [19] Corporation, N. C.: SkyWay, <https://webtrc.ecl.ntt.com/>.
- [20] 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).
- [21] Kendon, A.: The role of visible behavior in the organization of social interaction, *Social communication and movement*, pp. 29–74 (1973).
- [22] 坊農真弓: 会話構造理解のための分析単位: F 陣形 (i連載チュートリアル*i*多人数インタラクシオンの分析手法 [第6回]), *人工知能*, Vol. 23, No. 4, pp. 545–551 (2008).
- [23] 山田楓也, 石田繁巳, 白石 陽: Web 会議におけるミクロな顔特徴に着目した発話予測手法の検討, *技術報告 19*, 公立はこだて未来大学大学院システム情報科学研究科, 公立はこだて未来大学システム情報科学部, 公立はこだて未来大学システム情報科学部 (2022).
- [24] 石井 亮, 大塚和弘, 熊野史朗, 大和淳司: 複数人対話における頭部運動に基づく次話者の予測, *情報処理学会論文誌*, Vol. 57, No. 4, pp. 1116–1127 (オンライン), 入手先 (<https://cir.nii.ac.jp/crid/1050564287860496256>) (2016).
- [25] 玉木秀和, 東野 豪, 小林 稔, 井原雅行: 発話がぶつからない Web 会議を実現するための発話欲求伝達手法, *情報処理学会論文誌*, Vol. 54, No. 1, pp. 275–283 (2013).