

グラフ構造に基づく 多重参与の発現を促すオンライン対話環境のデザイン

市川 雅也^{1,a)} 竹内 勇剛^{1,b)}

概要：一般に対面状況における教室やオフィスでは、単一の空間に複数の対話場が同時に存在し、その空間にいる人は複数の対話に多重的に参加していることがある。このような多重参与は、空間あたりの発話量を増加させるとともに、対話間での情報の流通を通じて、他者の視点の獲得や創造的な発想の創出を促進すると考えられる。ところが、多くのオンライン対話環境では、原則として単一チャンネル内で複数の対話が同時に進行することや、参加者が複数の対話に同時に参加することは許容されておらず、こうした機会が喪失している。その一方で、対面状況とは異なり、オンライン対話では対話へのアクセスや情報の流通が参加者や対話場の位置・距離といった地理的要因によって制約されないという特徴をもつ。本研究は、オンライン対話環境の利点を踏まえた複数の対話への多重参与を可能とするオンライン対話システムの開発を目的とする。具体的には、Goffman (1981) によって提案された参与構造をグラフ構造で表現し、これに基づく身体性を前提としないオンライン対話環境を設計した。本稿では開発システムの設計・実装の概要を示すとともに、予備的な利用を通して観察された事象について報告する。

1. はじめに

一般に、対面状況の教室やオフィスでは、単一の空間内に複数の対話の場（対話場）が同時に存在している（図 1(a)）。このような環境では、人はしばしば複数の対話場に同時に参与（多重参与）することがある。その結果、授業や会議の進行中であっても、近傍の者同士が内々に対話したり、それらを傍聴・傍観したりすることが可能である。これにより、空間内の参加者は他者の優れた発想や異なる視点に触れる機会を得ることができ、授業や会議全体の創造性および理解性の向上につながると考えられる。

ところが、従来の多くのオンライン対話環境では、単一のチャンネル内で複数の対話を同時に成立させることが難しい。そのため、ある時点における話者を 1 人に限定するか、対話ごとにセッションを分割する必要がある（図 1(b)）。しかしこの場合、聞き手もまた複数のセッション同時参加することが一般には想定されておらず、結果として利用者は一つの対話に排他的に関与せざるを得ない。この構造的制約により、多様な意見に同時に触れたり、それらの相互作用から新たな発想や理解が生じたりする機会は生まれにくくなっている。

対面状況の対話において、人はその身体配置 [1], 姿勢 [2],

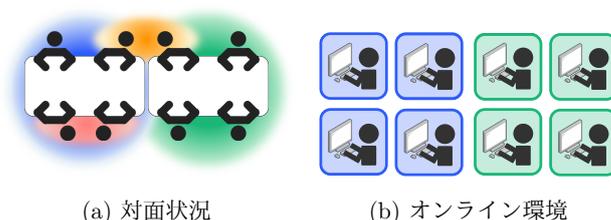


図 1: 対面／オンラインの対話場の構造の違い
青, 緑: 準備されていた対話場. 赤, 黄: 参加者間で偶発した対話場.

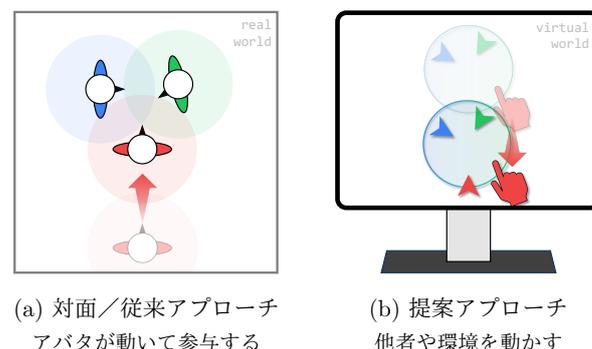


図 2: 対話環境の可動性と参与のアプローチ

物理的な距離 [3], 視線 [4], [5] など様々な身体動作を器用に調節することで対話を維持していることが報告されており、こうした非言語情報に依拠する情報がオンライン対話で伝達困難であることが指摘されている (e.g., アイコンタクト [6], [7], 空間的な位置 [8], [9]). こうした背景から、

¹ 静岡大学創造科学技術大学院

^{a)} ichikawa.masanari.18@shizuoka.ac.jp

^{b)} takeuchi@inf.shizuoka.ac.jp

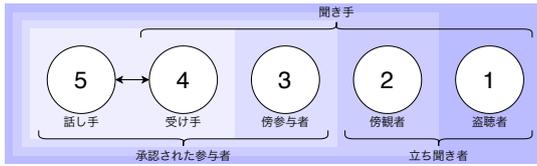


図 3: 参与構造

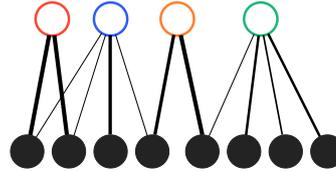


図 4: 参与構造グラフ

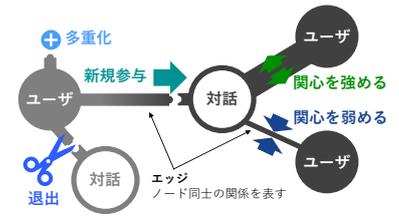


図 5: 参与構造グラフの操作

プロジェクションマッピング [10], AR やアバターロボット [11] などを用いて対面状況の身体性を忠実に模倣するアプローチが盛んに行われており, VRChat [12] や ovice [13] などすでに社会で広く利用されているものもある. このようにオンラインでの音声対話インタラクションやその環境については, 対面状況との比較において議論されることが多かった一方で, 標準で身体を有さない認知環境や通信ネットワークの離散性, 視聴覚情報の伝達制御の柔軟性などオンライン対話特有の環境特性に着目した対話環境のデザインについて十分な議論や検証が行われてこなかった.

本研究ではオンライン対話環境における多重参与を可能とするために, その離散性に着目し, 図 2(a) に例示するような対面状況的なアプローチではなく, 図 2(b) に示すような物理的な身体を前提としないアプローチによるオンライン対話環境の開発を目的とする. 具体的には人の参与活動の基盤として広く理解されている参与構造 (図 3) をグラフで記述し, これを元にしたオンライン対話環境を独自に開発した. 本研究の成果は, オンライン対話を単に遠隔地からの対話参加手段として捉えるのではなく, 身体的・空間的制約から解放することによって, 従来の音声対話の限界を拡張しようとする点に意義がある.

2. 参与構造とグラフ

2.1 参与構造

参与構造は Erving Goffman (1981) が指摘した対話の場における聞き手の役割の階層性に関する概念である (図 3) [14]. Goffman の参与構造によれば聞き手の役割は話し手になる権利がある「承認された参与者」と, そうでない「立ち聞き者」に大別される. 承認された参与者とは一般的な意味での会話の参加者を指し, 中でも発話をダイレクトに差し向けられている者を「受け手」, そうでない者を「傍参与者」と呼ぶ. 立ち聞き者の内訳は「傍観者」と「盗聴者」であり, その名の通り対話場の外縁から遠巻きに関与する者である. このように参与構造は同じ聞き手の中にも役割の階層性があることを示している.

2.2 参与構造のグラフによる表現

本研究ではオンライン対話環境において参与構造をデザインするにあたり, これまでにグラフによる参与構造の非言語的な記述を試みてきた [15], [16]. グラフとは, ノード

とエッジからなる離散的な要素間の関係を表す数学的な構造である. グラフを用いることで, 参与者と対話場という離散的な要素同士の関係, すなわち参与の程度を表現することが可能となる.

本研究で提案する参与構造グラフでは, 参加者を表す「参与者ノード」(塗りつぶした丸), 対話の場を表す「対話ノード」(枠線の丸), それらの関係を表す「参与エッジ」の 3 つの記号を定義する. すると図 4 に示すように, 多重参与を含めた図 1(a) の様相を, 重み (太さ) の異なる参与エッジによって記述できるようになる. さらに図 5 に示すようにエッジを繋いだり切ったり, その重みを増やしたり減らしたりすることで, グラフの操作によって参与者の各対話に対する状態を調整することができる. また多重参与状態についても同様に表現できる.

次章では, 以上で述べた参与構造グラフに基づき, オンライン対話環境において多重参与を実現するためのシステム設計について述べる.

3. グラフ構造に基づく対話環境の設計

3.1 システム概要と主要な構成要素

提案システムは, Windows のアプリケーションであり, 図 2(b) に示したように, 画面 (キャンバス) の中央に固定された「アバタ」に対して, 対話場オブジェクトである「ブース」をカーソル操作で近づけることで参与の程度を調整できるオンライン対話環境である (図 6).

アバタがブース内に位置づけられると, ネットワークユーザのブースに「ゴースト」が出現し, 対話場の参与者であることが表現される. アバタがブースの中心に近づけられると参与の程度を示す「エンゲージメント」が高まり,

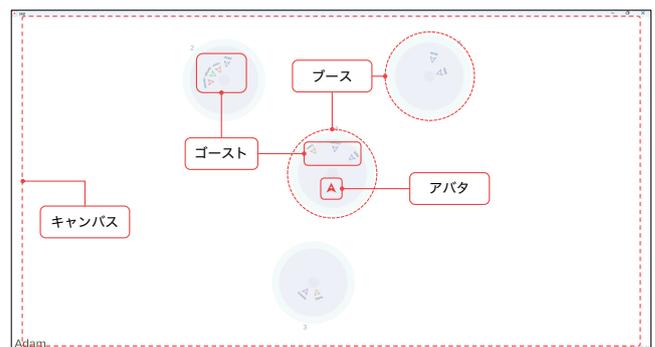


図 6: 提案システムの画面

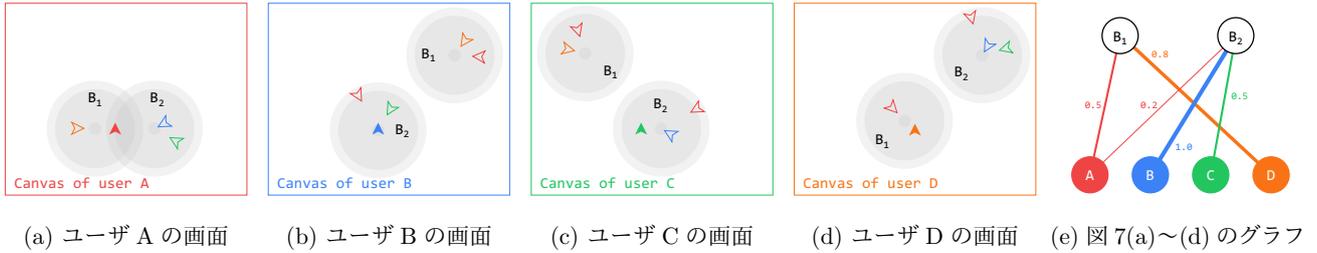


図 7: 提案システムにおける参与状況に対するユーザ別の視点

図 7(a)~(d) の矢尻形のオブジェクトのうち、塗りつぶされたものはアバタ、枠線のものゴースト、2つの円 B_1 , B_2 はブースを表す。

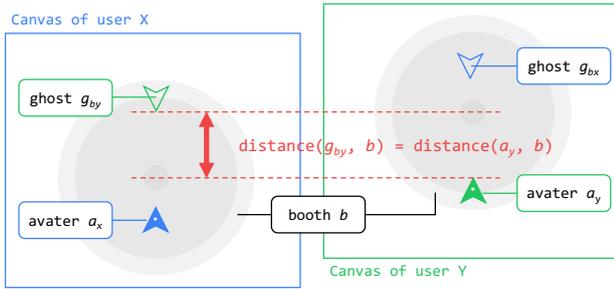


図 8: ブース半径に対するゴーストの位置はそのユーザのキャンバスにおけるブースとアバタの距離



図 9: メニュー

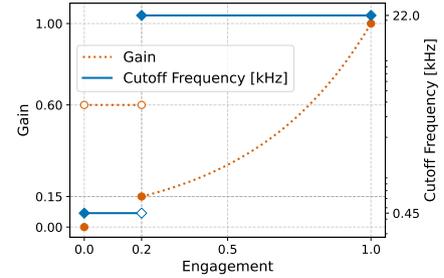


図 10: ゲイン倍率 (左軸) とローパスフィルタ (右軸) の決定関数

それに伴った音響フィードバックがなされるようになる。またブース中心とアバタとの距離は、ネットワークユーザのゴーストの位置と同期している (図 8)。

2.2 節で述べた参与構造グラフとの対応関係は次のとおりである。ローカルユーザ (操作者) の参加者ノードはアバタに、ネットワークユーザの参加者ノードはゴーストに、対話場ノードはブースにそれぞれ対応する。また、参与エッジの重みの変化はエンゲージメントとして表現される。これにより提案システムは、図 7 のような多重参与状態を含む参与構造グラフで記述可能な参与状態を一貫して表現することが可能である。

3.2 基本 4 機能と操作方法

(1) 参与する

キャンバス中央に固定されたアバタに、ドラッグで掴んだブースを重ねると対話場に参与したとみなされる。参与するとローカルユーザはブース内の音声対話を聞くことができ、アバタがブース中心に近いほど音が大きくなる。図 8 のように、アバタとブース中心の距離は、ネットワークユーザのブースの半径に対するゴーストの位置と同期している。これにより各ユーザが対話場に対してどの程度の強さで参与しているのかという非言語情報を可視化しており、状況把握や次話者選択などに役立てられる。

図 7(a) のように、複数のブースに対して多重参与する場合、図 7(b)~図 7(d) に図示されているように、他のユーザからは分身しているように見える。ユーザ間でブースの位置は同期しておらず、各ユーザの各ブースに対するエン

ゲージメントのみが同期している。これにより地理的に制約されることなく各自があらゆるブースに多重参与が可能となる。

(2) マイクをつける・発話する

参与中のブースを右クリックすると、マイクが有効/ミュートが切り替わり、そのブースに対する音声送信が可能になる。複数のブースに対して同時にマイクを有効にすることはできない。マイクが有効になっている参加者のゴーストにはバッチが付き、ただの参加者とは一線を画す話者になりうる者として視覚的に区別される。発話による音が検知されている間はローカルユーザのアバタとこれに対応するゴーストが明滅して強調される。

(3) 新しいブースを作成する・ブースを削除する

キャンバス上で右クリックすると図 9 のコンテキストメニューが出現し、任意のメンバーを選択して新しいブースを作成することができる。作成者自身のキャンバスでは新規ブースはアバタに重なる位置に出現する。招待されたユーザのキャンバスでは特定の位置でブースが生成されたのち、ブースが被招待者のアバタに重なる位置まで勝手に自走し、ドラッグ操作を介さずに参与状態となる。

図 9 下部のオプションから参加者以外には見えない「シークレットブース」としてブースを作成するか否かを選択できる。またシークレットブースかに関わらず、ブースは参加者が 0 人になると、自動的に削除される。

(4) 対話の場に誰かを招待する

各ブースには番号のラベルがついており、このラベル上で右クリックすると、招待用のコンテキストメニューが出

現し、既存のブースに他者を招き入れることができる。なお任意の他者を招待可能である一方で、既存の参加者を追い出す機能は搭載されていない。

3.3 媒介変数：エンゲージメント

ブースに対する参加の程度を表す量であるエンゲージメント E は、およそローカルユーザのアバタとブースとの距離に基づいて算出される。図 8 に見られるように、ブースは三重円の構造をしており、厳密にはその最も内側の円 (100%リング) とアバタとの距離 d に基づいて、式 (1) のように定義される。ただし 100%リングの半径を r_1 、ブース (の最も外側の円 = 0%リング) の半径を r_0 とする。

$$E = \begin{cases} 0 & \text{if } r_0 \leq d \\ 1 - \frac{d - r_1}{r_0 - r_1} & \text{if } r_1 < d < r_0 \\ 1 & \text{if } d \leq r_1 \end{cases} \quad (1)$$

式 (1) より、エンゲージメントは参加するブースの 0%リングの内側にアバタが位置した瞬間に正の値をとり、ブースの中心に近づくにつれて線形に増加する。そして 100%リング以内に位置した瞬間に最大値 1 をとる。またアバタがブースの外に位置するとき、最小値 0 をとる。

3.4 音響処理

提案システムではゲイン倍率、ローパスフィルタ、パンをエンゲージメントによって制御している。

(1) ゲイン倍率 (音量)

ゲイン倍率はパソコンで出力される音量の係数であり 0 以上 1 以下の値を取る。ゲイン倍率はエンゲージメントが増加するにしたがって指数関数的に増加する (図 10)。ただしローパスフィルタがかかる $0 < E < 0.2$ の範囲においては音量が小さくなると音の聞きとりが困難になるため定数を指定している。ゲイン倍率 G は次の式 (2) によって決定する。

$$G(E) = \begin{cases} 0.6 & \text{if } 0.0 < E < 0.2 \\ 10^{\frac{16.48E - 16.48}{20}} & \text{if } 0.2 \leq E \leq 1.0 \end{cases} \quad (2)$$

式 (2) における定数 20 は、デシベル表現と線形ゲインとの変換に用いられる係数であり、音圧や振幅量を扱う音響信号処理において一般的に用いられるものである。また、定数 16.48 は、本システムにおいて定義した最小ゲイン倍率 $G_{\min} = 0.15$ をデシベルフルスケールに変換した値の絶対値であり、 $20 \log_{10}(0.15) \approx -16.48$ に対応する。なお計算負荷を低減するために、実際にはゲイン倍率は長さ 101 の配列に格納されており、エンゲージメントが 1/100 変化するたびに、対応するゲイン倍率を返す設計になっている。

(2) ローパスフィルタ (高周波成分のカット)

$0 < E < 0.2$ のとき、すなわち、アバタを 0%リングとブース中間の円 (20%リング) の間に位置付けたとき、450Hz 以上の高周波成分がカットされる (図 10)。これにより、ブースの外縁部で聞こえる音は、壁を挟んで聞こえる音のような籠もった音になる。この処理は複数の対話に多重参加したときに、より高いエンゲージメントで参加している対話音声より低いエンゲージメントの方が邪魔しないようにするための設計である。

(3) パン (音像定位)

本システムでは、音像定位に Head-Related Transfer Function (HRTF) を用いる。 $0.2 \leq E \leq 1$ のとき、参加者が知覚する音像は、各ゴーストの空間的位置に基づいて定位する。一方で、 $0 < E < 0.2$ の場合、音像はブースの中心に定位する。エンゲージメントが高い状態では、誰が発話しているかを明確に知覚させることが重要であるが、低い状態では、対話が存在しているという事実のみを知覚させることに焦点を当てている。

3.5 開発環境・ミドルウェア・プラグイン

本研究で使用した開発環境を表 1 に示す。

表 1: 開発環境

種別	名称	バージョン
ゲームエンジン	Unity	6000.0.40f1
ネットワーク SDK	Photon PUN 2	2.51
音声通信 SDK	Photon Voice 2	2.59
空間音声プラグイン	Stem Audio Spatializer	-

4. 試験利用

4.1 調査目的

本調査は今後実施される提案システムを用いた実験に先駆けて行った。本調査は、図 2(b) ならびに 3 章に示した参加の程度が調整可能なオンライン対話環境かつ身体がない状況におけるユーザの振る舞いの観察を目的に実施した。

4.2 調査環境

調査は 3 名 1 組を対象に行った。参加者はブリーフィングを受けたあと、異なる部屋に案内され、それぞれの部屋から図 11 に示す調査環境にアクセスした。調査環境にはあらかじめ 16 のブースが設置されていた。

4.3 課題：リスニング課題

参加者の 3 名 1 組は、制限時間 10 分間で 1~15 のブースで再生される計 15 問の 4 択リスニング問題に、3 名の合計点がより高くなるように取り組むことを指示された。参加者にはあらかじめ問題文と選択肢が書かれた用紙を配布

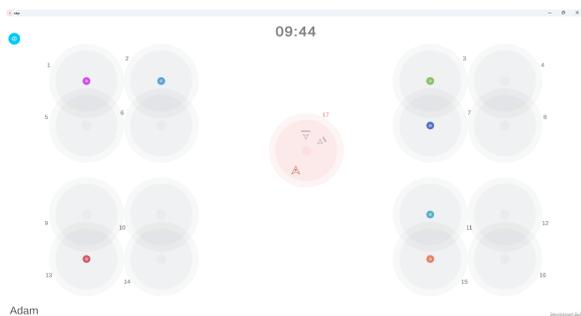


図 11: 調査環境

16 のリスニング音声放送ブースは参加者数に関わらず消滅しない

しており、各問題で何が問われるかをあらかじめ把握した上で課題に取り組んだ。また参加者には課題中にオンライン対話環境上で回答の情報共有を含め自由に会話して良いこと、制限時間 10 分に対して課題音声の総再生時間が 26 分 30 秒程度であることがあらかじめ伝えられていた。リスニング音声には日本語能力検定 N1 級の聴解問題を利用した。参加者はブース 16 を用いて操作練習を行ってから調査に臨んだ。

4.4 観察項目

本課題は、学会の平行セッションや就職活動における合同説明会のように、複数のセッションが同時進行する状況において、意図した情報を効率的に収集できるかという点に着目したものである。その戦略として、多重参加が実際にどのように用いられるかを主な観察項目とした。具体的には、以下の 2 点に注目した。

第一に、課題音声ブースに対して多重参加が実際に生じるかどうかである。本課題では、音声コンテンツは参加者の参与開始とは無関係に常時進行しているため、多くの場合、参加者は音声の途中から聴取を開始することになる。そのため、音声内容の把握に一定の見通しが得られた段階で、現在のブースへの参与を維持したまま、次のブースへ新たに参与するという行動が生じる可能性がある。

第二に、リスニング課題の聴解と、その内容に関する情報共有が並行して行われるかという点である。図 2(a) に示すような身体的・地理的制約を受ける対面環境では、参加者が分散して聴解に取り組む場合、情報共有のための対話を同時に維持することは困難であり、いずれかの時点で合流する必要がある。一方、身体的・地理的制約を前提としないオンライン対話環境では、リスニング用ブースと情報共有用の対話場に常時多重参加することが可能であり、必要な情報が得られた時点で即座に共有できると考えられる。

4.5 結果

今回の試験利用の結果を図 12 に示す。4.1 節で言及した実験用の各種機構は、いずれも問題なく動作した。

観察項目 1 について、図 12(a)~(c) は、制限時間 10 分間における各参加者の各ブースに対するエンゲージメントの遷移を示している。各図中の折線 $B_1 \sim B_{15}$ はリスニング問題放送ブースに対するエンゲージメントを表し、赤紫で示された折線 B_{21} は参加者間の対話に用いられたブースに対するエンゲージメントを表している。参加者 1 および参加者 3 では、 B_{21} とは別に複数のリスニング問題放送ブースに対するエンゲージメントが同時に高まる時間帯が確認され、多重参加が生じていたことが分かる。

観察項目 2 について、図 12(a)~(c) より、参加者 1 および参加者 3 は制限時間中を通して対話用ブースに参与し続けていたことが確認できる。一方で、図 12(d) に示す赤帯はマイクが ON であった時間帯、白点は実際に発話が生じていた時間帯を示しているが、参加者らは実験時間の大部分をミュート状態で過ごしており、3 者間の情報共有は主として制限時間終盤に集中的に行われていた。

なお、今回の参加者 3 名のグループは、制限時間 10 分間で 15 問中 13 問に着手し、正解数はいずれの参加者も 10/15 であった。誤答した問題は 3 名で共通していた。

5. 議論

本試験利用では、提案システムが想定どおり動作し、多重参加を含む参与状態が実際に発現することが確認された。特に観察項目 1 に関して、参加者 1 および 3 において、複数のリスニング問題放送ブースに対するエンゲージメントが同時に高まる場面が見られたことは、単一の対話場に排他的に関与する従来のオンライン対話環境とは異なる参与のあり方が実際に可能であることを示唆している。これは、参与構造をグラフとして記述し、その重みを操作可能にするという本研究の設計方針が、ユーザの行動レベルにおいても一定程度妥当であったことを支持する結果であると考えられる。

一方で観察項目 2 に関しては、リスニングの聴解と情報共有が常時並行して行われるという理想的な状態は必ずしも実現しなかった。参加者は制限時間の大部分をミュート状態で過ごし、情報共有は終盤に集中的に行われていた。この結果は、提案システムが多重参加によって「聴解しながら対話できる」構造的可能性を備えている一方で、参加者が必ずしもその可能性を自発的・戦略的に活用するとは限らないことを示している。

本調査の結果を解釈するにあたり、いくつかの制約と今後の課題が挙げられる。第一に、本調査は 3 名 1 組・1 セッションという単一事例に基づくものであり、一般化可能な知見を得るにはサンプル数を増やした検証が不可欠である。今後は複数グループ・複数条件による実験を行い、多重参加の発現頻度やその戦略的利用の違いを定量的に検討する必要がある。

第二に、本研究では図 2(b) に示した身体性を前提とし

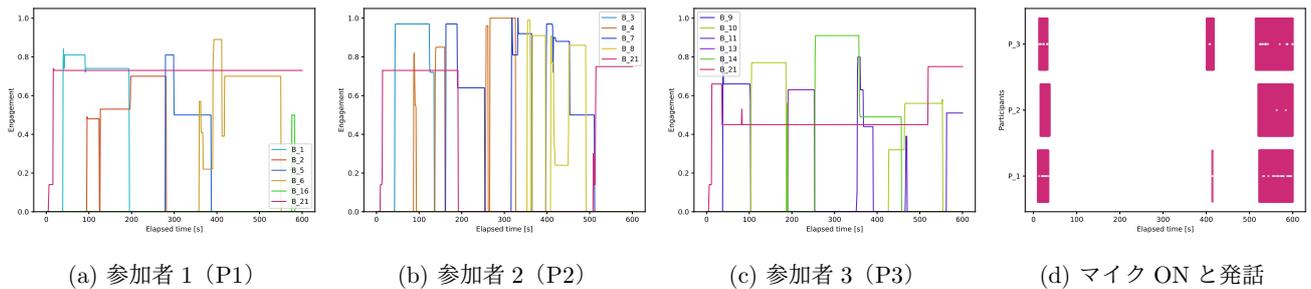


図 12: 試験利用の結果 ((a)~(c) は各参加者の時系列エンゲージメント推移, (d) はマイク ON と発話の時間帯)

ないアプローチのみを対象としており, 図 2(a) に示すような, アバタの移動や距離に基づく対面状況的アプローチとの直接比較は行っていない. 今後は, 同一課題を用いた比較実験を通して, 「身体を模倣する設計」と「身体を前提としない設計」との違いが, 多重参与の生じやすさや情報共有のタイミングにどのような影響を与えるのかを検討する必要がある.

第三に, 本調査で用いた課題は, 情報がリスニング問題放送ブースから対話用ブースへと持ち込まれる状況を主に想定しており, 情報が対話場から外部へ流出する状況や, 複数の対話場間で双方向に情報が行き交う状況を十分に検討できていない. また, 人同士の対話場が複数同時に成立する状況も扱っていない. 今後は, 一人対話場が 2 つ以上存在する条件や, 情報の流入・流出が複雑に絡み合う課題設定を通して, 多重参与が対話間の情報循環に与える影響をより詳細に検討することが求められる.

加えて, 提案システムは, 3.2 節 (3) で示したシークレットブースへの多重参与や, (4) で示した招待によって被招待者が即座に参与状態になる機構など, 物理的な身体を前提としないオンライン環境特有の相互作用を実現している. これらは, 意図的に対話を「漏れ聞こえ」させることによる視点の揺さぶりや, 特定の他者を招き入れることで過去の文脈や知識を即座に共有するなど, 対面状況では実現が難しい新たな参与の形を生み出す可能性をもつ.

以上より, 本研究で提案したオンライン対話環境は, 対面状況の単なる模倣にとどまらず, 身体的・空間的制約から解放された認知環境として, 多重参与を前提とした新たな音声対話および協働の形を設計・探究するための基盤となりうる.

謝辞 3.4 節における各種音響パラメータの設計について貴重な助言を賜ったヤマハ株式会社研究開発統括部に深く感謝の意を表す. また, 本研究は JST RISTEX の研究開発プロジェクト「ローカルエコーチェンバーをステアリングするトラスト調和メカニズムの認知的検討」(課題番号: JPMJRS23L3) の支援を受けて実施された.

参考文献

- [1] Kendon, A.: *Conducting interaction: Patterns of behavior in focused encounters*, Vol. 7, CUP Archive (1990).
- [2] Schegloff, E. A.: Body torque, *Social Research*, pp. 535–596 (1998).
- [3] Hall, E. T.: *The hidden dimension*, Vol. 609, Garden City, NY: Doubleday (1966).
- [4] Duncan, R. B.: Characteristics of organizational environments and perceived environmental uncertainty, *Administrative science quarterly*, pp. 313–327 (1972).
- [5] Goodwin, C.: Conversational Organization, *Interaction between speakers and hearers* (1981).
- [6] Sellen, A. J.: Speech patterns in video-mediated conversations, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 49–59 (1992).
- [7] Okada, K.-I., Maeda, F., Ichikawa, Y. and Matsushita, Y.: Multiparty videoconferencing at virtual social distance: MAJIC design, *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 385–393 (1994).
- [8] Isaacs, E. A. and Tang, J. C.: What video can and can't do for collaboration: a case study, *Proceedings of the first ACM International Conference on Multimedia*, pp. 199–206 (1993).
- [9] Hirata, K., Harada, Y., Takada, T., Aoyagi, S., Shirai, Y., Yamashita, N. and Yamato, J.: The t-room: Toward the future phone, *NTT Technical Review*, Vol. 4, No. 12, pp. 26–33 (2006).
- [10] Zhang, Y., Yang, J., Liu, Z., Wang, R., Chen, G., Tong, X. and Guo, B.: Virtualcube: An immersive 3d video communication system, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 5, pp. 2146–2156 (2022).
- [11] Victor, L., Porter, L. and Lawler, E.: Expectancy theories, *Organizational Behavior 1*, Routledge, pp. 94–113 (2015).
- [12] VRChat: VRChat Inc. <https://hello.vrchat.com/>.
- [13] oVice Inc.: ovice, <https://ovice.in/ja/>.
- [14] Goffman, E.: *Forms of talk*, University of Pennsylvania Press (1981).
- [15] Ichikawa, M. and Takeuchi, Y.: CallMap: A Multi-dialogue Participative Chatting Environment Based on Participation Structure, *International Conference on Human-Computer Interaction*, Springer, pp. 294–313 (2023).
- [16] 市川雅也, 竹内勇剛: グラフ理論に基づく対話構造に立脚した遠隔対話システムデザインの提案, 電子情報通信学会技術研究報告; 信学技報, Vol. 121, No. 438, pp. 61–66 (2022).