



# テレプレゼンスチャットアプリにおける指差しの情報価値と 応用についての考察

A Study of the Information Value of Pointing and its Applied Value in Telepresence Chat Application

上岡玲子<sup>1)</sup>, 赤塚大典<sup>1)</sup>

Ryoko UEOKA and Daisuke AKATSUKA

1) (株)zeroinon (〒174-0045 東京都板橋区西台 1-51-5, postmaster@zeroinon.jp)

**概要:** 誰もが使えるデバイスで気軽に空間共有しながらビデオチャットができる CoMADO の開発を行った。これは共有画面で指差しをしながら会話できることを特徴とするが、指差しが同時に人の関心・興味を反映した画像の記録手段にもなる。こうして収集された記録情報の価値について、インタラクション・体験要約を中心に事例を元に分析を行い考察する。

**キーワード:** テレプレゼンス, 指差し, 主観と客観

## 1. はじめに

コロナ禍によるリモートワークの浸透で我々の生活様式は変化し、特に Zoom に代表されるリモート会議システムが普及している。筆者らはコロナ禍にリモート会議システムでは実装されていない、ライブ空間映像に指差しアイコンを重畳させオンラインで指示語を使いコミュニケーションできる仕組みを備えた可用性のあるテレプレゼンスチャットアプリ「ここそこ」を開発、それをリモートフィールド調査に特化したプロダクトとして発展させ、2022 年 11 月に CoMADO をリリースし運用している [1]。

本論ではアプリの特徴の一つである「指差し」に注目した。指差しの行為は一般に興味や関心を他者に伝える共同注意や指差し点検の行為など、主観行為を客体化する身体行為と考えられる。筆者らは CoMADO によりデジタル化された指差しはデジタル情報として行為を客体化していると考えた。そこでデジタル指差しの情報価値やその価値応用の可能性について、3 種の事例をもとに考察する。

## 2. テレプレゼンスチャットアプリ CoMADO

CoMADO はオンラインフィールド調査ツールの Web サービスで、WebRTC をベースにしたシステムである。複数参加者の内 1 名が「プレゼンター」となりスマートフォン背面カメラを使い映像を共有し、全ての参加者が指示語の代わりにアバター付きの指差しを映像上に表示することで、リモート参加者同士での空間情報の共有とコミュニケーション体験の向上を実現している。

図 1 は CoMADO のセッション中の PC 画面を示す。プレゼンターが共有するライブ動画を表示する画面にタップ/クリックすると指差しアイコンが表示され、同時にスクリーンショットが保存される。その他、メモ作成やハートのアニメーションの LIKE アクションをフラグにスクリーンショット

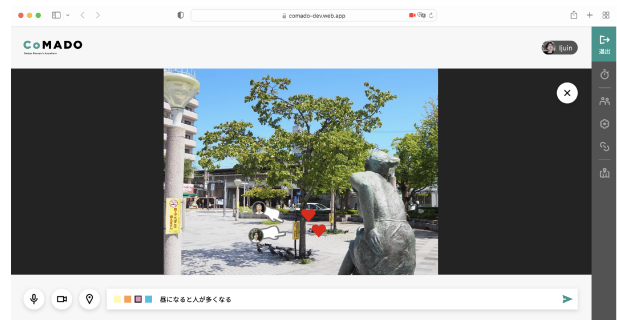


図 1: CoMADO 実行画面

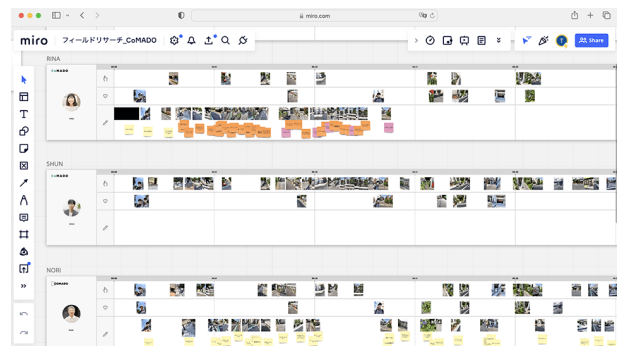


図 2: ログデータを配置したオンラインホワイトボード

トが保存、セッションログとして記録され、参加者ごとの時系列記録を配置したオンラインボードが図 2 のように自動生成される。画像、テキストの他、動画録画機能も備わる。

## 3. 事例概要

表 1 に示す通り、CoMADO 利用の 3 種の事例から本論ではデジタル指差しの価値について分析する。特に従来のフィジカルな指差しを人の関心・興味を示す主観的な定性情報とした時、デジタル指差しにより客観的定量情報とし

表 1: 3 事例概要表

番号	事例	利用人数	実施時間
1	リモート散歩 1	7 人	30 分
2	リモート散歩 2	3 人	25 分
3	リモート散歩 3	3 人	30 分
4	活動記録 1	1 人	150 分
5	活動記録 2	1 人	160 分
6	フィールド調査	1 人	30 分

て変換され得るのか、また、その情報の意味や応用について結果から考察する。

### 3.1 リモート散歩

屋外に 1 名が赴き、リモートで複数人が参加する散歩体験を実施した。散歩はプレゼンターが主に現地を紹介、リモート参加者とコミュニケーションをとりながらルートを散策した。

### 3.2 ワークショップ活動記録

筆頭著者を含む 3 名の講師で実施した計 2 回のワークショップの活動記録を実施した。ワークショップの様子を実時間で記録し、ワークショップ後に他講師メンバーとオンラインボードに配置された記録をもとに実施者視点でワークショップを振り返り、ワークショップでの学びや課題を議論し、ブログ記事にまとめた [2]。

### 3.3 フィールド調査

フィールド調査を模して東京大学本郷キャンパス内の屋外展示探索を実施した。

## 4. 結果

### 4.1 リモート散歩

3 回のリモート散歩のアクション数から指差しの客観化を評価する。CoMADO では指差しの他にメモや画面長押しで LIKE アクションなど 3 種類のアクションが可能だが、アクションの中では指差しが他のアクションと比較し圧倒的に多く、3 回の散歩の平均アクション数比率は 79:9:12(指差し:メモ:LIKE) で指差しのアクション数が多かった。これは他のアクションに比較し指差しが短時間で行え、習得しやすいアクションであるためと考えられる。図 3 は散歩 2 の 5 分間隔の 3 人の参加者毎のゆびさし、メモ、LIKE の累積数を示す。前半 15 分はコンスタントに指差しされるが、指差し数が減少する時間帯があることから、参加者の空間への関心の薄れが推測でき、指差しの簡単なアクションから全員の主観の集合知としてどこに興味を持っているのか可視化されていることがわかる。またこれは各個人の興味、関心の可視化としても有効である、

### 4.2 ワークショップ活動記録

ワークショップ活動記録では、指差しはすぐに記録したい時、メモは発表や参加者の発話などコメントを残したい時

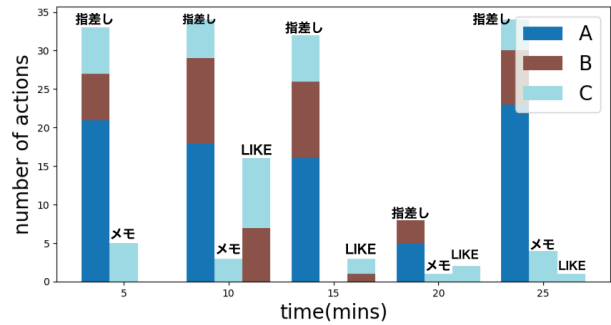


図 3: リモート散歩 2 の各アクションの参加者毎累積数

に使分け利用した。(平均アクション数比率は 67:33(指差し:メモ)) 記録者が興味をもったシーンの記録はワークショップのハイライトシーンと重なる点が多く、約 8 時間のワークショップの流れが図 2 にあるようなログデータを共有することですぐに把握でき、どこで何をしたかの検索性も高かったので記録を用いた振り返りが効率的に実施できた。

### 4.3 フィールド調査

調査中アクションから取得されたスクリーンショットを主観データ、動画記録から等間隔で静止画抽出した画像を客観データと定義し、事前学習済の画像エンコーダーと大規模言語モデル(以下 LLM) をブートストラップし画像からキャプションを自動生成する Blip2-t5 モデルを使用し画像からキャプションを自動生成した。次にキャプション群とメモ情報と LLM(ChatGPT-4) を GPT Index の Llamaindex[3] を用い接続し、主観・客観データそれぞれと全体から体験要約を行なった。LLM の文章表現力から要約はどれもそれらしくまとめられていたが、客観データのみで生成された体験要約は、等間隔でのシーンキャプションから生成された。そのため、体験者が何に関心を持ったのか明確にはわからない。一方、主観データからの体験要約は体験者のアクションに即し体験要約がされている印象であった。ただし、主観、客観データ両方を利用することでより全体の文章要約の完成度が高かったことから、LLM での体験要約については体験者のアクションに基づいたデータと周辺情報を合わせることで体験者のリアルな体験に即した要約がされた。

## 5. 考察

デジタル指差しは、自身や他者の興味・関心といった主観的な定性データを反映した定量データであり、人やコンピュータが識別できる主観の集合知として可視化、情報圧縮、検索に役立つ可能性が確認できた。画像内の指差しの座標や GPS 座標など CoMADO で記録されている他のデータを組み合わせることでデジタル指差しの主観情報としての価値向上の可能性について引き続き検討する。

### 参考文献

- [1] CoMADO サイト: <https://comado.co>
- [2] note ブログ: <https://note.com/zeroinon/n/nd165658d16cf>
- [3] LlamaIndex: [https://github.com/jerryliu/llama\\_index](https://github.com/jerryliu/llama_index)