

コミュニケーションにおける理解や共感の促進にむけた取り組み

Efforts to Enhance Comprehension and Empathy in Communication

齊 藤 功 樹, 榎 本 真, 銭 尾 春 仁

要 約 コミュニケーションにおいて理解や共感は重要な役割を果たしており、それらを促進することで信頼関係が醸成され、生産的で効率的な関係を構築できる。本研究ではコミュニケーションにおける理解や共感の促進に向けて、実際のコミュニケーションを模した実験を行い、対話内容の可視化と相手の理解度や共感度などの状態推定に取り組んだ。その結果、対話内容の可視化においては絵や図などで構造化するグラフィックレコーディングの一部を自動化して表示することにより、聞き手の理解を促進する可能性が示唆された。相手の状態推定においては、聞き手と話し手の生体情報の同期度を用いることで聞き手の理解度や共感度を推定できることが明らかになった。

Abstract In communication, people can build trust which leads to productive and efficient relationships when they comprehend or empathize with others. This study explores methods to enhance comprehension or empathy in communication by visualizing the content of conversations and estimating the other person's internal states through experiment that simulates actual communication. The results showed that automated visualization of conversation contributed to improving comprehension. In estimating listener's internal states, it became clear that the level of comprehension or empathy could be grasped by biometric synchronization between speaker and listener.

1. は じ め に

人はコミュニケーションなしには生きられず、他者と関わり合いを持ちながら、社会生活を営んでいる。対人関係を円滑にするコミュニケーションは重要であり、well-being（身体的・精神的・社会的に良好な状態）の向上に繋がる^[1]。

価値観や考えが多様化する現代において、多様な背景を持つ人たちとのコミュニケーションの機会が増しており、そのような人たちと事業を進めていくためには信頼関係が重要である。社会の潤滑油として働く信頼が欠如すると、自己の利益を最優先しようとする機会主義的行動の可能性が高まると報告されている^[2]。信頼関係が構築されている場合は、情報交換が活発になり、累積的かつ相乗的な相互作用が生じ、生産的な関係を構築できるとも報告されている^[3]。このようなことから特にビジネスの場において信頼関係は非常に重要な役割を果たすものの、利害関係の問題や背景の違いからその構築は容易ではない。

信頼関係の構築において、理解や共感といった人の内的状態が重要な役割を果たすことが知られている^{[4]-[6]}。自身が相手に対して理解や共感を示すことで相手への好意が増し^[5]、好意が媒介して相手との信頼関係の構築に寄与することが報告されている^[6]。また、ビジネスの場では、国際化が進み、より多様な国籍、文化、タイムゾーンの人たちとの協働が求められ、共感がよりよい労働環境作りにおいて重要な要因であると報告されている^[7]。

しかし、コミュニケーションにおいて相手の発言に対して理解/共感しづらい、相手が理解/共感してくれているのか判別できないなどの課題が存在するため、それらを促進する支援が求められる。そこで、本研究では、言語と非言語の二つの観点で、理解/共感を促進する取り組みを実施した。

言語の観点では、相手との対話内容を図や絵などの視覚表現（以降、グラフィック）を用いて自動的に可視化することで、理解/共感の促進を試みる。話し手が説明する内容をグラフィックで記述し、構造化する手法であるグラフィックレコーディングを用いることで、受け手が共感を覚えやすいことが報告されている^{[8], [9]}。しかし、その効果はグラフィックを描画する人（以降、レコーダー）の能力に依存するため、多くの人が納得し、共感するようなグラフィックを生成することは容易ではない。そこで、グラフィックレコーディングの工程の一部の自動化を試みた。

非言語の観点では、コミュニケーションにおいて興味関心を抱いた聞き手は話し手に対して脳波、心拍、体動及び瞬目（瞬き）などの生体情報が同期することが報告されていることから^{[10]-[14]}、生体情報の同期を用いることで理解/共感の推定を試みる。さらに、相手から理解/共感されたと感じることによって、相手への好意が増したり、自分と相手が重なっているように感じやすくなったりすることが報告されている^[15]。そのため、相手が理解/共感してくれているもののそれが表情などからはわかりにくい状況において、相手の理解度や共感度を推定して提示することで、相手から理解/共感されたという知覚を促進することができ、信頼関係構築に繋がれると考えた。

本稿では、実際のコミュニケーションを想定した説明動画を用意し、動画視聴時の参加者の生体情報や主観的質問票（以降、質問票）の結果を基に、自動化したグラフィックレコーディングの効果を確認し、生体情報の同期度と理解/共感度の関連性を調査した取り組みについて述べる。まず2章で調査方法を説明し、3章で調査結果を示して、4章で考察を述べる。

なお、本稿は人工知能学会全国大会（第39回）^{[16], [17]}にて発表、Proceedings of The 47th Annual Meeting of the Cognitive Science Society (CogSci2025)^[18]にて発表の論文を加筆・修正したものである。

2. 調査方法

本章では、調査のために行った実験の内容と、話し手が説明する内容を可視化するグラフィック描画プログラム、体動や瞬目の同期を検出する仕組みについて説明する。

2.1 実験

実験の参加者（以降、参加者）は三つの説明動画を視聴し、その後に視聴した動画を基にアイデア発散ゲームを行い、質問票に回答した。アイデア発散ゲームでは、拡散的思考タスクである Alternative Uses Task^[19]を用いて、説明動画内で紹介した研究の活用アイデアを、できるだけ多く回答してもらった。グラフィックレコーディングはアイデア発散に効果があることが報告されているため^[9]、それらの効果を確かめる目的でアイデア発散パートを実施した。動画視聴時の参加者の上半身部分を録画し、生体情報として用いた。なお、動画の中で説明する人物（以降、発話者）は一人である。

2.1.1 参加者

参加者は、社内での募集を通じて自発的に参加した BIPROGY 株式会社の従業員 22 名（男：8，女：14），20 歳代：4，30 歳代：8，40 歳代：7，50 歳代：3 であり，参加者の視力は正常（矯正含む）であった．実験はヘルシンキ宣言の原則^{*1}に基づいて行われ，実験計画は BIPROGY 株式会社の倫理審査の承認を得た．参加者は実験に先立ち，インフォームド Consent に署名した．

2.1.2 実験に用いる動画と実験の流れ

参加者が視聴する動画は，研究テーマについて発話者が説明する動画である．三つの研究テーマごとに説明原稿を用意して，それぞれの研究テーマに対して三つの形式の動画を作成した．一つ目の動画形式は，発話者がその原稿を基に口頭で約 3 分間説明する姿を上半身のみ撮影した動画（図 1（A））である．二つ目は，この口頭説明の動画を元に，同一のレコーダーによる描画過程を合成した動画（図 1（B）），三つ目は，動画原稿を元にしてグラフィック描画プログラム（2.2 節で詳説）の描画過程を合成した動画（図 1（C））である．

なお，三つの研究テーマは全て同一の発話者が説明する．また，参加者を三つのグループに分けることにより，視聴する動画の研究テーマと動画形式が重複しないように配慮した．



図1 動画形式のイメージ，(A) 描画なし（口頭説明），(B) 描画あり（レコーダー），(C) 描画あり（プログラム）

参加者は実験時にウェブ会議に参加して，実験方法についての説明を受けた後，PC で実験プログラムを起動して実験を開始した．安静時計測で PC の画面を眺め（図 2 ①），アイデア発散の例題を実施（図 2 ②）した後，内容と形式の異なる三つの動画について，動画視聴（図 2 ③）とアイデア発散ゲーム（図 2 ④）と質問票回答（図 2 ⑤）に取り組み，安静時計測で画面を眺めて（図 2 ⑥）実験を終了した．アイデア発散の例題は，アイデア発散ゲームに慣れるために一度だけ実施した．

安静時計測時と動画視聴時に PC 付属のカメラで参加者の上半身部分を録画した．ビデオ視聴後，参加者はそれぞれ，主観的質問票に，話し手の印象や研究テーマに対する理解，興味，共感，支援について，5 件法（5 段階評価）で回答した．

実験結果の分析には，対話内容の可視化では，アイデア発散パートの結果と質問票回答パートで取得した回答を，相手の状態推定では，動画視聴パートで描画なし（口頭説明）の動画を視聴した際に参加者の上半身部分を録画した動画と質問票回答パートで取得した回答を使用した．

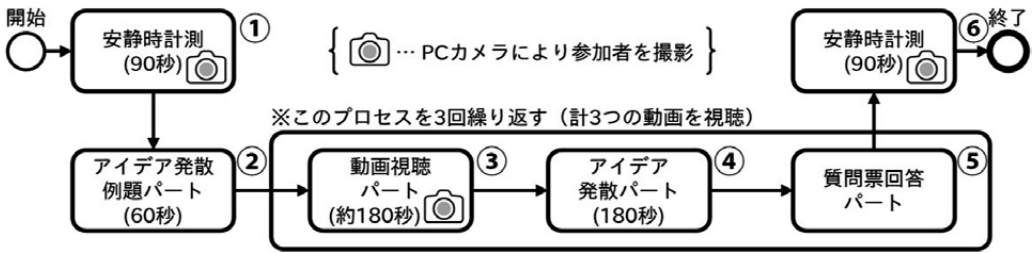


図2 実験の流れ

2.2 対話内容の可視化

2.1.2 項で述べた三つ目の動画形式 (図1 (C)) を作成するため、画面上で選択した文のテキストから名詞句と名詞を抽出し、名詞句を文字で、名詞をグラフィックで描画する処理を自動化したグラフィック描画プログラムを作成した。このプログラムの入力となるテキストは、発話者が説明する内容が音声認識によりテキスト化され、人またはプログラムによって誤認識が修正され、文の区切りが適切に行われた状態を想定している。

2.2.1 グラフィック描画プログラム

グラフィック描画プログラムの画面構成は、左上部に原稿の編集と描画に関する操作領域、その下にグラフィック描画の対象となる文を表示し選択するための原稿領域を配置し、右上部にタイトル領域、その下にグラフィックの描画領域を配置した。グラフィック描画領域は2列とし、1列に8行、各行は、先頭に名詞句表示領域を一つ、その後にグラフィック表示領域を四つ設けた (図3)。

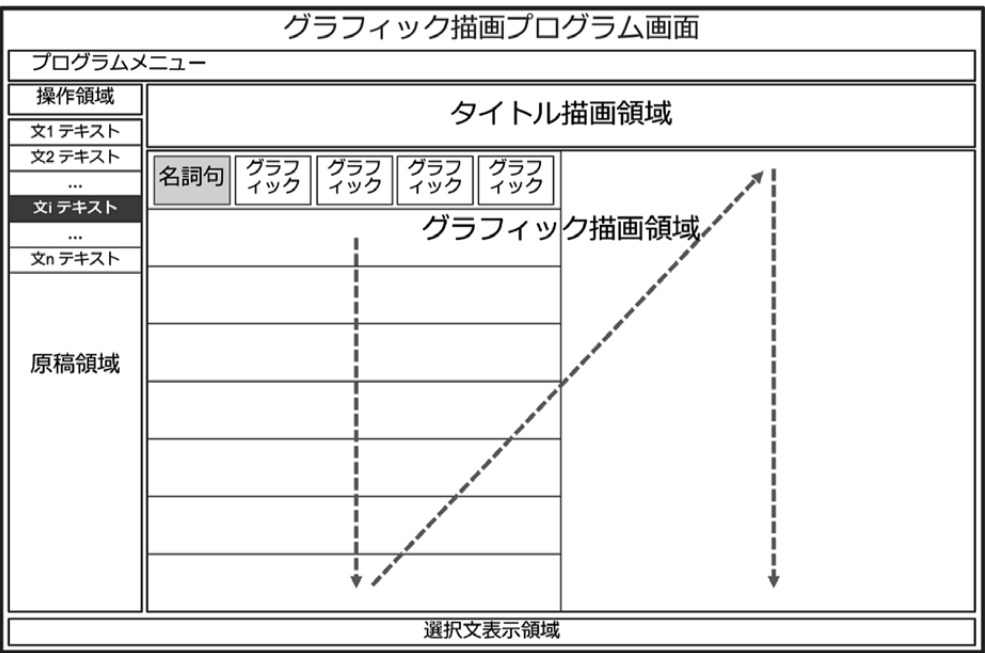


図3 グラフィック描画プログラムの画面構成イメージ

グラフィック描画プログラムは、原稿のテキストファイルを読み込んだ後、原稿領域でリスト表示された文を選択すると最下部の選択文表示領域に文を表示する。操作領域にあるボタンでグラフィック描画を指示する操作を行うことにより、名詞句のテキスト描画と名詞に紐づいたグラフィックが存在すればその描画を行う。文中の名詞に紐づくグラフィックが四つを超えた場合は次の行のグラフィック表示領域に描画する。新たなグラフィック描画指示が行われた場合は次の行に移動して処理を行う。1列目の最終行を超える場合は2列目の1行目に移動して描画を行い、16行を超える描画は行わないこととした。グラフィックのイメージは、名詞のテキストとそれを説明する絵で構成し、プレゼンテーションソフトウェアで作図し画像ファイルを作成した。

2.2.2 実験の準備

実験に用いる動画を用意するため、グラフィック描画プログラムに研究テーマを説明する原稿を読み込ませて、「描画なし（口頭説明）」形式の動画を視聴しながら、発話者が話し終えた文に対してグラフィック描画の指示を行い、グラフィック表示されていく画面をキャプチャして動画として記録した。この記録した動画のタイトル描画領域とグラフィック描画領域の範囲を「描画なし（口頭説明）」形式の動画（2.1.2項の図1（A））と合成し「描画あり（プログラム）」形式の動画を作成した。

2.3 聞き手の状態推定（体動の同期）

説明動画および参加者の録画からそれぞれ頭部姿勢を検出し、体動の同期を算出する手法について説明する。

2.3.1 動画処理

説明動画および参加者の録画から頭部姿勢を推定して取得した。頭部姿勢はオイラー角（ヨー、ロール、ピッチ）（図4）で表される。推定には Google-MediaPipe^[20]、OpenCV^[21]、および perspective n-points（カメラの位置姿勢の推定を行ってから実空間の物体の座標を推定すること）^[22]を用いた。

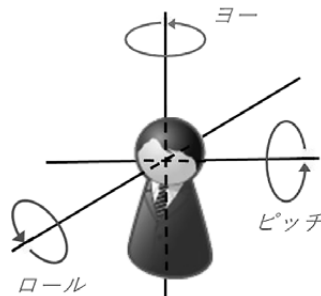


図4 頭部姿勢

2.3.2 時系列リサンプリング

説明動画のフレームレートは24Hzであり、説明動画のヨー、ピッチ、ロールもサンプリング

グ周波数が24Hzである時系列データとして表される。参加者は各自PCのカメラにて別々に撮影したため、動画のフレームレートは16-30Hzのようにばらつきがある。次項で述べる相互相関の計算のために、視聴動画についてはヨー、ピッチ、ロールを抽出後の時系列データに対して3次スプライン関数を用いてリサンプリングを行い、説明動画と同じ24Hzに周波数を合わせた。

2.3.3 同期の検出

同期の度合として既存研究^[23]と同様に相互相関を用いた。すなわち姿勢をオイラー角（ヨー、ピッチ、ロール）で表し、それぞれに対応する説明動画の変量を $x(t)$ 、参加者の変量を $y(t)$ とすると、相互相関は式(1)であらわされる。

$$R_{xy}(\tau) = \frac{1}{(N_t - 1)} \sum_{t=1}^{N_t - \tau} \frac{(x(t + \tau) - \bar{x})(y(t) - \bar{y})}{\sigma_x \sigma_y} \quad (1)$$

N_t は時系列データの長さ（おおよそ4,320（180秒×24Hz））を、 $\bar{x}$ 、 $\bar{y}$ はトレンドを、 σ_x と σ_y はそれぞれトレンド除去後の標準偏差を表す。データの平滑化はSavitzky-Golayフィルター^[24]を用いてwindow幅：241（10秒間）、次数：2として計算した。

2.4 相手の状態推定（瞬目の同期）

ビデオ視聴時の参加者の顔動画を基に瞬目を検出し、その後、話し手（ビデオの中の発話者）と参加者の瞬目の同期を計算した。本節では瞬目検出と同期算出手法について説明する。

2.4.1 瞬目検出

瞬目は、ビデオから検出した3次元のランドマークを基に算出する個人最適化されたEye Aspect Ratio（EAR）を用いて検出する^[25]。3次元のランドマークはGoogle-MediaPipe^[20]を用いて検出する。目の周辺は16点のランドマークで構成されている（図5（A））。算出されたランドマークの各点を基に、EARをKraft et al.^[26]が考案した式(2)にて算出する。両目それぞれでEARを計算し、二つの平均値を用いる。目が閉じた状態では、EARはゼロに近づき、開いている状態では1に近づく。

$$EAR \stackrel{\text{def}}{=} \frac{\|P3 - P13\|_3 + \|P4 - P12\|_3 + \|P5 - P11\|_3}{3 \cdot \|P0 - P8\|_3} \quad (2)$$

目が閉じている時のEARを式(3)、目が開いている時のEARを式(4)で定義する。瞬目の有無を判別する閾値として、Personalized EARを式(5)で定義する。

$$EAR_{Closed} \stackrel{\text{def}}{=} \frac{\min(\|P3 - P13\|_3) + \min(\|P4 - P12\|_3) + \min(\|P5 - P11\|_3)}{3 \cdot \max(\|P0 - P8\|_3)} \quad (3)$$

$$EAR_{Open} \stackrel{\text{def}}{=} \frac{\text{mean}(\|P3 - P13\|_3) + \text{mean}(\|P4 - P12\|_3) + \text{mean}(\|P5 - P11\|_3)}{3 \cdot \min(\|P0 - P8\|_3)} \quad (4)$$

$$Personalized\ EAR_{Threshold} \stackrel{\text{def}}{=} (EAR_{Open} - EAR_{Closed}) / 2 \quad (5)$$

その後、ピークを計算して瞬目の直前のピークを瞬目開始点と定義する (図 5 (B)).

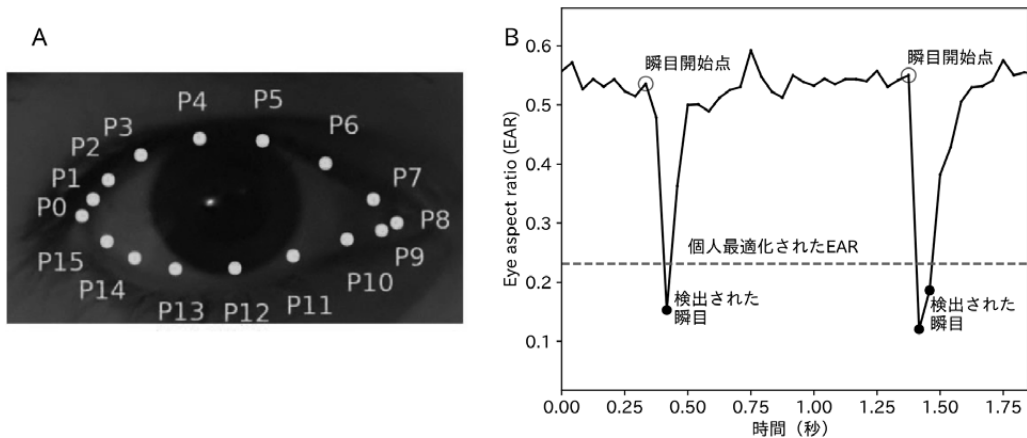


図5 (A) 目の周辺の16個のランドマーク位置, (B) EARを用いた瞬目検出の例

2.4.2 瞬目同期

瞬目同期は Nakano and Miyazaki^[12]によって開発された手法を用いて計算した。開発手法では、ペアを作りテストとリファレンスに分けられ、両者の瞬目開始点の差分 (リファレンスからテストを引いた差分) を計算する。計算した瞬目開始点の差分を基に ± 2.4 秒 (ビンの幅: 0.6秒) の範囲内でヒストグラムを作成することで、同期度合 (以降、同期度) を算出した。その際に、ヒストグラムの値をZスコア化した。Zスコア化では、瞬目と瞬目の間隔を保ったまま、テストの瞬目開始点をランダムにシャッフルした1,000個のサロゲートデータを用いて実施した。Zスコアが正であれば、発生確率がランダム確率よりも高いことを示す。

話し手 (ビデオ) と聞き手 (参加者) の間の瞬目同期の計算において、話し手をテストに、聞き手をリファレンスとした。聞き手の瞬目は、話し手の瞬目から数百ミリ秒遅れて同期することが報告されているため^{[12], [27]}, 0.6秒 (0.3-0.9秒) におけるZスコアを聞き手の瞬目同期とした。

3. 調査結果

質問票調査の結果からグラフィックレコーディングとグラフィック描画プログラムの効果について確認した。頭部振動や瞬目についてはそれぞれの同期度と共感、支援、理解の項目について関連があることが分かった。

3.1 対話内容の可視化

実験への参加者は22名であったが、計3回のアイデア発散ゲーム全てで無回答の参加者3名のデータを除外し、19名のデータを分析対象とした。参加者の年代と性別の内訳は、20代女性4名、30代男性1名、30代女性5名、40代男性4名、40代女性2名、50代男性2名、50代女性1名である。

3.1.1 質問票調査の結果（理解度、興味度、共感度）

質問票調査の理解度、興味度、共感度の主観評価回答について平均値を計算し、「描画あり（プログラム）」および「描画あり（レコーダー）」と「描画なし（口頭説明）」との間でt検定を実施した。対話内容の可視化においては支援度の分析は行わなかった。

その結果、理解度、興味度、共感度で「描画あり（レコーダー）」と「描画なし（口頭説明）」との間で有意差が確認できた（理解度： $t_{17}=3.77$, $p<0.01$, 興味度： $t_{17}=3.62$, $p<0.01$, 共感度： $t_{17}=2.28$, $p<0.05$ ）。また、理解度において「描画あり（プログラム）」と「描画なし（口頭説明）」について片側t検定を実施した結果、 $t_{17}=1.59$, $p=0.065$ という結果が得られた。理解度、興味度、共感度の箱ひげ図を図6に示す。

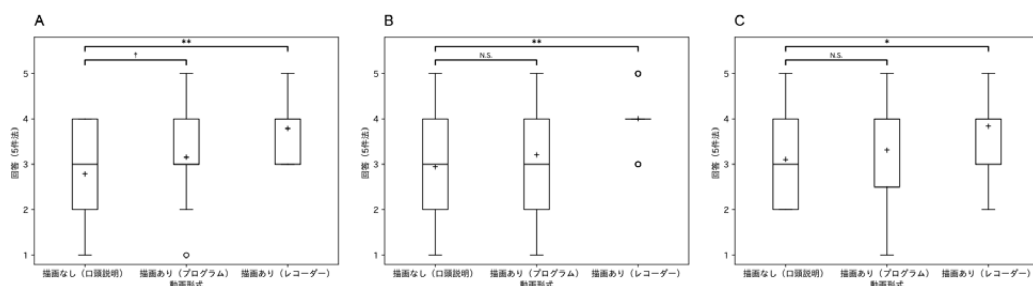


図6 箱ひげ図, (A) 理解度, (B) 興味度, (C) 共感度 (†: $p<.1$, *: $p<.05$, **: $p<.01$)

3.2 相手の状態推定（体動）

検出した相互相関について全参加者の平均的な相互相関を算出した結果と同期の遅れについて説明する。次に個別の同期度を定義して、この同期度と質問票調査との関連について記述する。

3.2.1 同期の検出

説明動画と参加者の動画について-1～1秒の間のヨー、ロール、ピッチの相互相関の全参加者の平均を図7(A)に示す。遅延時間が正の場合に参加者が発話者に対して遅れ、負の場合に先行していることを意味している。既存研究によれば同期は-1～1秒で発生しているので、この時間で相互相関により同期を評価した。相互相関が正のピークにおいては、参加者の同期時間の遅れはヨー：0.54秒となり、既存研究^[23](0.45秒)とほぼ一致した。またピッチ：-0.21秒、ロール：-0.85秒であり、参加者が発話者に対して先行している。一方向のコミュニケーションにおいては、受け手の体動が送り手に対して遅れて同期することが報告されている^[23]ため、本実験においても参加者は発話者に対して遅れて同期すると想定される。合わせて、ピッチとロールにおいてはヨーと比較してピークが顕著にみられないため、同期していないと考えられる。ロールにおいては相互相関が負のピークがみられるものの、同様にピークが顕著にはみられないため同期していないと考えられる。そのため、以後はヨーの同期について考察する。

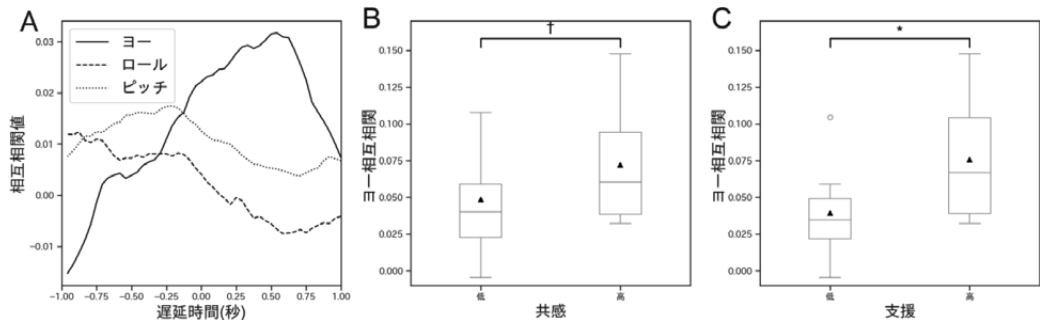


図7 頭部振動の相互相関。(A) 頭部姿勢の全参加者の相互相関の平均、(B) ヨー相互相関と共感の関係、(C) ヨー相互相関と支援の関係 (†: $p < .1$, *: $p < .05$)

3.2.2 同期と共感

参加者の同期度を以下の方法で定義した。ヨーの相互相関のピーク値をとる遅延時間について平均 m と標準偏差 σ を算出し ($m - 2\sigma$, $m + 2\sigma$) を相互相関のうち同期が起こっている遅延時間の区間とする。この区間について参加者別に相互相関の最大値を算出し、それを同期度とした。最大値はピークと一致する 경우가多いが、区間の端の値である場合もある。

参加者を質問票調査の理解度、興味度、共感度、支援度の主観評価回答に応じて、高と低の2グループに分け、同期度を比較した。2以下と回答した参加者を低に、3以上と回答した参加者を高とした。その結果、共感度においては2以下(低; 9名)と3以上(高; 13名)の群に分けられ、片側t検定を行った結果、有意差はなかったものの有意傾向が得られた ($t_{20} = 1.52$, $p = 0.071$) (図7 (B))。支援においては2以下(低; 8名)と3以上の群(高; 14名)に分けて片側t検定を行った結果、2群の間に有意な差が得られた ($t_{20} = 2.36$, $p = 0.014$) (図7 (C))。共感や支援が高い場合に、ヨーが同期する傾向にあることがわかった。一方、理解度と興味度については有意な差は見られなかった(理解度: $t_{20} = 0.445$, $p = 0.33$; 共感度: $t_{20} = 1.12$, $p = 0.14$)。

3.3 相手の状態推定(瞬目同期)

話し手と聞き手の瞬目同期の平均値を図8 (A) に示す。瞬目開始点の時間差分が正の場合には、聞き手が遅れており、負の場合には先行している。0.6秒(0.3-0.9秒)に顕著なピークが発生しており、聞き手は話し手から0.3-0.9秒遅れて同期していた。この結果は、既存研究^[12]と同様である。

参加者を質問票調査の理解度、興味度、共感度、支援度の主観評価回答に応じて、高と低の2グループに分け、瞬目同期を比較した。2以下と回答した参加者を低に、3以上と回答した参加者を高とした。その結果、理解度が高いグループは、低いグループと比較して有意に瞬目同期が高かった(図8 (B), $t_{20} = 2.01$, $p < 0.05$)。一方、興味度、共感度及び支援度においては、瞬目同期に有意な差は得られなかった(興味度: $t_{20} = 0.15$, $p = 0.44$; 共感度: $t_{20} = 0.28$, $p = 0.39$; 支援度: $t_{20} = 0.37$, $p = 0.36$)。したがって、理解度が高い人は、低い人と比べて瞬目が同期する傾向にあることがわかった。

さらに、理解度が低い参加者は1.8秒(1.5-2.1秒)にピークが発生しており、理解度が高い参加者より遅れて同期している可能性がある。

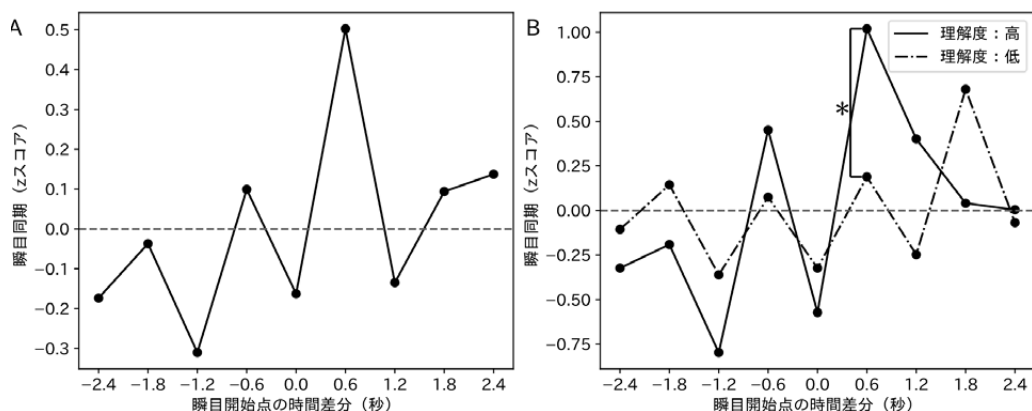


図8 聞き手と話し手の瞬目同期, (A) 全参加者での瞬目同期の平均値,
(B) 理解度でグループ分けした瞬目同期の平均値 (*: $p < .05$)

4. 考察

3章の調査結果を基に, 対話内容の可視化, 相手の状態推移についての考察を述べる.

4.1 対話内容の可視化

質問票調査の回答の分析結果から, レコーダーによりリアルタイムに描画されたグラフィックレコードは, 一方向のコミュニケーションにおける理解度, 興味度, 共感度を向上させる効果があることが確認できた. また, グラフィック描画プログラムによるリアルタイムでの描画は理解度の向上に寄与する可能性が示唆されたが, プログラムによる描画はレコーダーによる描画ほどの効果は確認できなかった. これは, グラフィック描画プログラムがグラフィックレコーディングの工程の一部分を自動化したものであるため, 多くの機能が不足していることが原因であると考えている. グラフィックレコーディングの自動化に向けて, Listen-Think-Draw のそれぞれの工程について考察する.

4.1.1 グラフィックレコーディングの自動化についての考察

Listen の工程では, 音声認識によりトランスクリプトを作成するが, 誤認識, 専門用語の認識の難しさといった課題があり, 音声認識システムとしてこれらへの対応が求められる. また, 口語では省略や代名詞の多用, 文の区切りの不明確さ等があるため, 照応解析等により文章を補完したり文の区切りを行ったりして次の Think の工程へ進める.

Think の工程では, 重要なキーワードやフレーズの選定, 重複情報の排除や要約, 知識にない情報の収集と理解, 物事の関係性の整理等を行うことで, 重要な情報が重複なく抽出され, 物事の関係性が整理される. この工程での処理の品質向上が, プログラムが描くグラフィックレコードの品質向上に繋がるため, 知識情報の整備と知識情報処理機能の検討が不可欠である.

Draw の工程では, キーワードやフレーズからのグラフィックへの翻訳, 類似イメージの想起, グラフィックの配置調整等を行う機能が欠かせない. グラフィック描画プログラムでは, 名詞に対応するグラフィックを事前に用意し, 文章中の名詞に対応する画像があった場合に描画する方法を採った. 他にも, 重要なキーワードやフレーズを人による指示やプログラムによる重要度判定で選定してから, 知識ベースや概念辞書を用いて関連する情報を収集して描画に

利用するような方法や、生成 AI に描画内容を指示するプロンプトを与えてグラフィックを動的に生成する方法等が考えられる。

4.1.2 今後に向けて

視覚文法^[28]の考え方をいれれば、グラフィックの精緻な形式化ができるようになり、より理解しやすい表現を提供できるだろう。今後は、効果的なグラフィックについて検討していく。

今回の実験では、発話者が説明する情報量とグラフィック描画プログラムの描画領域を限定したが、より情報量が多い対話の場面を想定すると、さらに大きな描画領域が求められる。そのため、表示インターフェースとして、仮想的なホワイトボードやウォールのような表示機能を検討しなければならない。また、話題の遷移を検出して、話題のまとまりの範囲や話題の流れを表現する方法も検討すべきである。そして、自動グラフィックレコーディング機能を多くの対話の場に提供するために、コミュニケーションツールが持っているような情報管理基盤等の整備も検討していく。

4.2 相手の状態推定

体動と瞬目による、相手の状態推定についての考察を述べる。

4.2.1 体動による相手の状態推定

体動の実験結果から、ヨー方向では、参加者の頭部振動が発話者に対して遅れて同期した。一方、ロール、ピッチ方向の頭部振動の相互相関のピークはヨーよりもかなり小さいことから、同期していないと考えられる。既存研究^[23]では垂直方向の体動の同期は見られないが、前後左右の体動は同期している。既存研究では体動の同期は立位で測定され、本実験では座位にて行われている。このため横方向への姿勢を変化させるロールは、ピッチやヨーよりも動きの程度が大きく、また座位は姿勢を元に戻す動きは少ない。それらはトレンドとして除去されているが、全体的な特徴として同期は観測しづらいと考えられる。また平均を取ると、動きの大きい参加者の影響が大きいため、動きの大きい参加者がうまく同期していない場合は、全体の同期の傾向がうまく捉えられないことも考えられる。

4.2.2 瞬目による相手の状態推定

瞬目の実験結果から、理解度の高い聞き手は話し手の瞬目に対して無意識に遅れて同期することが分かった。既存研究では、興味度に応じて瞬目が同期することが報告されていたが^[12]、本実験では興味度においては瞬目同期に違いは見られなかったため、瞬目は状況に依存して異なる認知状態を反映している可能性がある。既存研究では、日常的な製品に対する説明を聞く状況であったのに対して、本実験は研究に対する説明を聞くため理解が求められる状況であった。そのため、理解が必要な状況においては、興味度ではなく理解度に応じて瞬目が同期する可能性があることが分かった。さらに、理解度が低い場合に、高い場合と比較して瞬目同期がより遅れており、理解度に応じて瞬目同期の遅れが変化する可能性が示唆された。瞬目同期は理解に要するタイムラグによって起こる可能性が報告されており^[12]、理解度が低い聞き手は、より理解に時間がかかっていると考えられ、瞬目同期の遅れが理解度の指標として活用できる可能性があることが分かった。

4. 2. 3 今後に向けて

体動の同期においては支援や共感との相関が、瞬目の同期においては理解との相関がみられ、生体情報の同期を用いることで理解度、共感度及び支援度といった人の内的状態の推定ができることが分かった。しかし、体動と瞬目の同期共に、動画全体での同期度を評価しており、実際のコミュニケーションにリアルタイムでの活用は難しい。今後はより短い時間での同期度の計測や、リアルタイムでの同期度の算出手法の開発に取り組みたい。

5. お わ り に

本研究では、コミュニケーションの円滑化に向けて、対話内容の可視化と相手の状態推定を試みた。対話内容についてはグラフィックレコーディングの一部を自動化することで可視化を行い、理解度の向上に寄与できる可能性が示唆された。相手の状態推定においては、生体情報の同期度を用いることで、理解度や共感度が推定できることが分かった。

本研究の成果を用いることで、コミュニケーションでの理解/共感を促進できる可能性が示唆されたものの、実際の状況で使用する上では課題が残っている。対話内容の可視化においては事前に用意した原稿を基に描画を行っており、グラフィックレコーディングの自動化に向けては音声認識技術の導入が必要である。また、考察で述べたような検討すべき課題がある。相手の状態推定においては、動画全体での同期度を計算しており、リアルタイムでの推定は困難である。今後は、実際のコミュニケーションでの適用を見据えて、リアルタイムでの対話内容の可視化や状態推定を目指す。

* 1 世界医師会で採択されたヒトを対象とする医学研究の倫理的原則。

- 参考文献** [1] 小川一美, “対人コミュニケーションに関する実験的研究の動向と課題,” 教育心理学年報, vol. 50, pp. 187–198, 2011, doi: 10.5926/arepj.50.187.
- [2] F. Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, 1995.
- [3] 佐々木宏, 鈴木秀一, and S. T. DAVIS, “ビジネス・エコシステムと信頼,” 日本情報経営学会誌, vol. 39, no. 2, pp. 75–84, 2019, doi: 10.20627/jsim.39.2_75.
- [4] F. Derksen, J. Bensing, and A. Lagro-Janssen, “Effectiveness of empathy in general practice: a systematic review,” *Br. J. Gen. Pract.*, vol. 63, no. 606, p. e76 LP-e84, Jan. 2013, doi: 10.3399/bjgp13X660814.
- [5] N. J. Goldstein, I. S. Vezich, and J. R. Shapiro, “Perceived perspective taking: When others walk in our shoes,” 2014, *American Psychological Association*, Goldstein, Noah J.: 110 Westwood Plaza, Suite A-412, Los Angeles, CA, US, 90095, noah.goldstein@anderson.ucla.edu. doi: 10.1037/a0036395.
- [6] C. Y. Nicholson, L. D. Compeau, and R. Sethi, “The role of interpersonal liking in building trust in long-term channel relationships,” *J. Acad. Mark. Sci.*, vol. 29, no. 1, pp. 3–15, 2001, doi: 10.1177/0092070301291001.
- [7] W. Rahman, “Empathy and trust: Into a better workplace environment,” *J. Bus. Econ.*, vol. 7, no. 12, pp. 2025–2034, 2016.
- [8] 安武伸朗, “デザイン思考におけるレコーディングの効果についての考察,” 常葉大学造形学部紀要, no. 13, pp. 19–23, 2015.
- [9] 石井陽子, 中谷桃子, 渡辺昌洋, “共創的な対話におけるグラフィックレコーディングの効果,” ヒューマンインタフェース学会論文誌, vol. 24, no. 1, pp. 37–52, 2022, doi: 10.11184/his.24.1_37.

- [10] R. Hari and M. V. Kujala, "Brain Basis of Human Social Interaction: From Concepts to Brain Imaging," *Physiol. Rev.*, vol. 89, no. 2, pp. 453–479, Apr. 2009, doi: 10.1152/physrev.00041.2007.
- [11] D. N. Saito *et al.*, "Stay Tuned: Inter-Individual Neural Synchronization During Mutual Gaze and Joint Attention," *Front. Integr. Neurosci.*, vol. 4, 2010, doi: 10.3389/fnint.2010.00127.
- [12] T. Nakano and Y. Miyazaki, "Blink synchronization is an indicator of interest while viewing videos," *Int. J. Psychophysiol.*, vol. 135, pp. 1–11, 2019, doi: <https://doi.org/10.1016/j.ijpsycho.2018.10.012>
- [13] T. Yokozuka, E. Ono, Y. Inoue, K.-I. Ogawa, and Y. Miyake, "The Relationship Between Head Motion Synchronization and Empathy in Unidirectional Face-to-Face Communication," *Front. Psychol.*, vol. 9, 2018, doi: 10.3389/fpsyg.2018.01622.
- [14] P. Pérez *et al.*, "Conscious processing of narrative stimuli synchronizes heart rate between individuals," *Cell Rep.*, vol. 36, no. 11, p. 109692, 2021, doi: <https://doi.org/10.1016/j.celrep.2021.109692>
- [15] 鈴木雄大, "他者からの視点取得の表明に関する心理学的検討," 2022.
- [16] M. Enomoto, K. Saito, H. Zenio, T. Hoshino, and E. Takahashi, "The relationship between empathy and head motion synchronization of presenter and viewer while viewing explanatory video," *Proc. Annu. Conf. JSAI*, vol. JSAI2025, 2025.
- [17] H. Zenio, K. Saito, M. Enomoto, T. Hoshino, and E. Takahashi, "Consideration towards automatic graphic recording," *Proc. Annu. Conf. JSAI*, vol. JSAI2025, 2025.
- [18] K. Saito, M. Enomoto, H. Zenio, T. Hoshino, and E. Takahashi, "Relationship between comprehension and blink synchronization," in *The 47th Annual Meeting of the Cognitive Science Society (CogSci2025)*, 2025.
- [19] J. P. Guilford, *The nature of human intelligence*. New York, USA: McGraw-Hill, 1967.
- [20] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," *CoRR*, vol. abs/1906.0, 2019, [Online]. Available: <http://arxiv.org/abs/1906.08172>
- [21] G. Bradski, *The OpenCV Library*, vol. 25, 2000.
- [22] E. Candeloro, "Real Time Driver State Detection," 2024, [Online]. Available: <https://github.com/e-candeloro/Driver-State-Detection>
- [23] S. Okazaki *et al.*, "Unintentional Interpersonal Synchronization Represented as a Reciprocal Visuo-Postural Feedback System: A Multivariate Autoregressive Modeling Approach," *PLoS One*, vol. 10, no. 9, p. e0137126, Sep. 2015, [Online]. Available: <https://doi.org/10.1371/journal.pone.0137126>
- [24] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964, doi: 10.1021/ac60214a047.
- [25] K. Saito, "Eye blink detection using personalized eye aspect ratio based on 3D landmark," *Proc. Annu. Conf. JSAI*, vol. JSAI2024, pp. 4I3GS702–4I3GS702, 2024, doi: 10.11517/pjsai.JSAI2024.0_4I3GS702.
- [26] D. Kraft, F. Hartmann, and G. Bieber, "Camera-Based Blink Detection Using 3D-Landmarks," in *Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence*, in iWOAR '22. New York, NY, USA: Association for Computing Machinery, 2023, doi: 10.1145/3558884.3558890.
- [27] T. Nakano and S. Kitazawa, "Eyeblink entrainment at breakpoints of speech," *Exp. Brain Res.*, vol. 205, no. 4, pp. 577–581, 2010, doi: 10.1007/s00221-010-2387-z.
- [28] 清水淳子, "グラフィックレコーディング習得のために必要な視覚文法," 日本デザイン学会研究発表大会概要集, vol. 64, p. 100, 2017, doi: 10.11247/jssd.64.0_100.

※ 上記参考文献に示した URL のリンク先は、2025 年 7 月 14 日時点での存在を確認。

執筆者紹介 斉 藤 功 樹 (Koki Saito)

2009 年日本ユニシス(株)入社。金融機関向けのバックシステムの開発・保守を担当。2013 年に総合技術研究所に異動。大規模データ処理技術、衛星画像のデータ処理・データ分析、視線情報を用いた文章読解に関する研究に従事し、現在は生体情報を用いたコミュニケーション支援研究に従事。博士(知識科学)。

**榎 本 真 (Makoto Enomoto)**

1994 年日本ユニシス(株)入社。客先システム開発、金融機関向けパッケージ開発等に従事した後、データマイニング製品の開発・適用とそれを利用したデータ分析等に取り組む。2020 年より総合技術研究所にて AI 技術を利用した社会課題の解決や生体情報を用いたコミュニケーション支援の研究に従事。

**銭 尾 春 仁 (Haruhito Zenio)**

1989 年日本ユニシス(株)入社。技術主管部門で OS リリース、特殊機器接続等の業務に従事した後、ヘルプデスクシステム開発、米国製 CRM パッケージの国内リリース、情報システム部門で自社コールセンターシステム構築業務に携わる。その後、金融部門でミドルウェア製品商品主管、技術コンサルティング業務、商品企画部門でビッグデータ、AI 技術関連の企画業務に携わり、2022 年より総合技術研究所にて、天空光源シミュレーション、自然言語処理の研究に従事。

