

映像対話型行動支援におけるインタラクションの一貫性に関する考察

小泉 敬寛[†] 小幡佳奈子^{††} 渡辺 靖彦^{†††} 近藤 一晃^{††} 中村 裕一^{††}

[†] 京都大学 工学研究科, 〒 606-8501 京都市左京区吉田本町

^{††} 京都大学 学術情報メディアセンター, 〒 606-8501 京都市左京区吉田本町

^{†††} 龍谷大学 理工学部, 〒 520-2194 大津市瀬田大江町横谷 1 番 5

E-mail: †{koizumi,obata,kondo,yuichi}@media.kyoto-u.ac.jp, ††watanabe@rins.ryukoku.ac.jp

あらまし 身体にカメラを装着した作業者と遠隔地にいる支援者との間で映像を使って対話しながら作業を進める形態を「映像対話型行動支援」と呼ぶ。本研究では、このような場におけるコミュニケーションの円滑さを評価するために、発話やコミュニケーション行動の一貫性を指標化する方法を検討した。その方法として、映像対話に頻出するパターンを抽出し、そこに含まれている基本的な呼応パターンを抽出し、その時間的性質とコミュニケーションの一貫性との関係について調査した。その結果、良好な相関があることがわかった。

キーワード 映像対話型行動支援, インタラクション分析, マルチモーダルコミュニケーション, コミュニケーションの一貫性

Discussion on Interaction Coherency in Working Support through FPV Communication

Takahiro KOIZUMI[†], Kanako OBATA^{††}, Yasuhiko WATANABE^{†††}, Kazuaki KONDO^{††}, and Yuichi NAKAMURA^{††}

[†] Dept. of Engineering, Kyoto University, Japan

^{††} ACCMS, Kyoto University, Japan

^{†††} Faculty of Science and Technology, Ryukoku University, Japan

E-mail: †{koizumi,obata,kondo,yuichi}@media.kyoto-u.ac.jp, ††watanabe@rins.ryukoku.ac.jp

Abstract “Working Support through FPVC (First Person Vision Communication)” is a working support style in which a person wearing a camera is working under the guidance of an experienced person who is monitoring the FPV at a distant place. In this research, we investigated the method of evaluating the coherence of interactions. For this purpose, we first extracted frequent communication patterns in FPV communications, chose response patterns as sub-patterns included in them, and examined the relationship between temporal characteristics of those patterns and the smoothness of the communication. Our experiment shows the indices based on those response patterns have tight relationships to the smoothness of the communication.

Key words First Person Vision Communication, Interaction analysis, Multimodal communication, Coherence of communication

1. はじめに

作業者と支援者が映像を用いて対話しながら作業を進める形態を我々は「映像対話型行動支援」と呼ぶ。ここでは、図1に示したように、作業者がカメラやその他のセンサを装着し、無線を介して映像等を遠隔地にいる支援者に送る。支援者は映像を見ながら作業者と対話し、指示や質問などを行う。

このような映像対話型行動支援は様々な用途、例えば、狭い場所や危険な場所で、作業者が外にいる支援者と情報を共有し

ながら作業を行う場合や、直接現地に呼ぶことが難しい専門家に遠隔地から指示を出してもらい場合等に活用できる。今後、様々なデバイスの高機能化や小型化が一層進み、様々な形態が増えることが予想される。

このような背景から、我々は、映像対話型行動支援で起こる作業者と支援者の間のコミュニケーションと作業の円滑さや失敗しやすさの関係を調査してきた [1]。その方法として、作業者と支援者のインタラクションを時系列のパターンとして記述し、データマイニングの一手法 (PrefixSpan) を適用することに

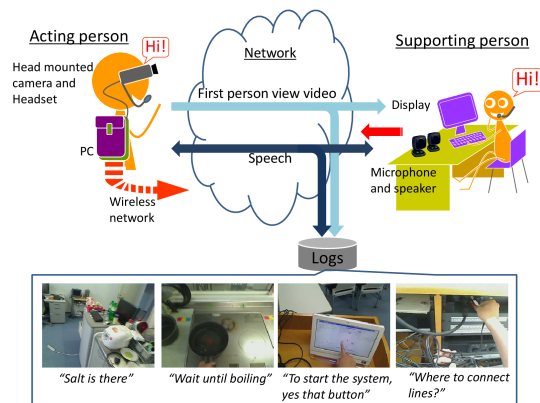


図 1 映像対話型行動支援の概要

よって、コミュニケーションの基本的なパターンの抽出を行った。また、作業員・支援者ペアの振る舞いの特徴がこのようなパターンの頻度や時間的構成に現れることなどを示した。

これらの結果を踏まえ、本研究ではさらに、個々の状況における作業員と支援者のコミュニケーションの円滑さや良さを表す指標について検討した。具体的には、作業員と支援者のインタラクションが自然に噛み合っている状態を一貫性が高い状態であると考え、表層的な振る舞いの基本的なパターンやそこから逸脱によってその程度を推定することを試みた。

以下本稿では、2. で映像対話型行動支援におけるインタラクションの表層的な一貫性について議論し、次に 3. で、一貫性を推定する指標について議論する。さらに、4. で指標の計算方法を、5. で実際のデータに適用した実験結果について述べる。

2. 映像対話における一貫性

2.1 研究の背景と目的

作業が円滑に進むためには、作業員と支援者の間で同一の対象に注意が向けられ、それぞれの意図が正確に伝達されることが望ましい。このようなコミュニケーションが円滑に行われているかどうかを評価するための指標を得ることが本研究の目的である。これが実現すれば、作業が円滑に行われるようなコミュニケーション方法の訓練をしたり、失敗や事故の原因を探ったり、予防する助けとすることが可能となる。さらに、それらが自動化・実時間化されれば、作業の場で、状況の良さを作業員や支援者に知らせたり、問題点を注意したりすることも可能になる。

このような観点から、上述したように、我々は作業員・支援者ペアのコミュニケーションの特徴をとらえるための指標の設定を行ってきた。それによりこれまでわかってきたことは、インタラクションの統計的性質に作業員・支援者ペアの振る舞いの特徴や失敗の起こしやすさなどがよく現れることである。

本研究では、それらに加え、個々の場面におけるコミュニケーションの円滑さに関する指標を得ることを検討した。その際の前提として、深い意味解析を行わず、インタラクションの表層的な役割や時間的性質を用いることとする。

そのために、本研究ではコミュニケーションの状態を表す「一貫性 (coherence)」に着目した。すなわち、作業員と支援者の

注意が同一の対象に向けられ、その対象に関する言動が順にタイミング良く行われていることを指標化することを試みる。コミュニケーションの性質としては、「結束性 (cohesion)」や「凝集性 (cohesiveness)」も良く用いられる。一貫性と結束性とは似た概念であるが、結束性は修辭的な文脈構造を含んで議論されることが多いため、ここでは一貫性と呼ぶことにした。また、凝集性は要素間の関連性だけでなく、数や量が多いことも要件とするため、別途議論することにする。

このような解析を映像対話に対して行うためには、マルチモーダルな特徴を対象とする必要がある。それには Norris による “modal density” の考え方が参考となる [3] [4]。Norris は、“high level action” と呼ぶ意味的に高次な行為 (目的) に対して、複数のモダリティが協調して働くこと (modal complexity) に注目し、その相互作用の密度を modal density として表現している。modal density の高い状況では相手が直接リアクションを起こしやすいこと、その結果良好なインタラクションが行われやすいことなどが示されている。ただし、modal density に対する定量的な指標や自動処理などは議論されていない。

本研究では、このようなモダリティ間の対応関係を含んだインタラクションの一貫性を定量的にとらえる指標やそれを得るための自動処理について検討する。

2.2 典型的な例と一貫性

映像対話型行動支援におけるコミュニケーションの性質を考えるために、典型的な状況を考えてみよう。

表 1 の (a) と (b) は円滑な意志の疎通が行われている例である。(a) では、両者の注意が一つの対象に対して向いている、つまり、注意 (焦点) を共有している状態である。良好に作業が進行している場合にはこの状態が多く見られる。発話だけでなく、見回す行動などもコミュニケーションとして重要な役割を果たしている。(b) では、関連性のない 2 つの焦点に対して注意が向けられているが、作業員・支援者間で注意は共有されている。これらに対し、(c), (d) では注意が共有できていない例となっている。(c) では、作業が始まっているにも関わらず、B の応答が省略されているために、A が不安になっていることが推測される。(d) では、(S4-2) で起こった突発的な問題を B が A に伝えられなかったため、最後まで A には焦点が特定されていない。

以下では、これらの典型例を基に、コミュニケーションの円滑さと一貫性について検討する。

2.3 コミュニケーションの円滑さと一貫性

作業員と支援者が活発に発話したり、お互いを意識した行動をしていることが意志の疎通のための重要な要素となるが、言動の量が多いだけでは十分な条件とは言えない。前節の例 (c), (d) のように、注意の対象が複数あり、情報や意図が共有されていない場合もある。

そのため、本研究では、(I) 個々の言動がお互いに呼応している (明示的・非明示的に関わらず、相手の要求に込えている) ことを円滑なコミュニケーションの要件として考える。また、それには、(II) 作業員と支援者の注意 (焦点) が同じ対象に向けられていることが前提となる。

表 1 典型的な映像対話の例

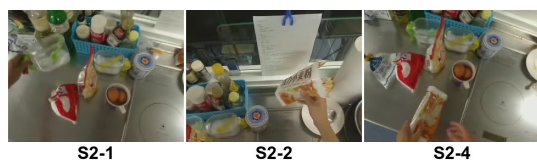
(a) 焦点一つ, または, 複数が密に関係

- (S1-1) A: 「卵を 1.5 個用意して下さい」
 (S1-2) B: 「はい」 見回しながら「卵 1 個はありますが, もう 1 個もしくは 0.5 個分, はどこですか?」
 (S1-3) A: 「その瓶に 0.5 個分が入っています」
 (S1-4) B: (見つける) 「はい」 作業を始める



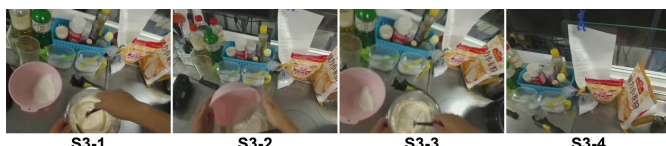
(b) 複数焦点. お互いに関連なし

- (S2-1) A: 「薄力粉を 60g 量って下さい」
 (S2-2) B: 振り向く 「この強力粉も量ります?」
 (S2-3) A: 「えっと, それは型に振る用なので量らなくて構いませんので, 薄力粉だけ計量して下さい」
 (S2-4) B: 「分かりました, では量ります」 作業を始める



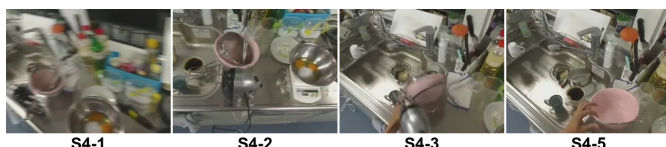
(c) 複数焦点. 応答・確認不足

- (S3-1) A: 「粉を入れて切るように混ぜて下さい」
 (S3-2) B: 「オープンの余熱終わったかな」 作業を始める
 (S3-3) A: 「…終わったんじゃないかな. とこで, 切る, じゃなくて練るように混ぜちゃってますよ」
 (S3-4) B: (見まわす) 「うーん, ゴムベラをグルグル回し過ぎたかな」



(d) 焦点が共有できていない

- (S4-1) A: 「卵を軽く溶きほぐしてから湯煎にかけて下さい」
 (S4-2) B: 「ハンドミキサーを使えば良いかな…」 (ハンドミキサーのケーブルを積んであるボウルや泡立て器に引っ掛け, それらを流しに落とす)
 (S4-3) B: 「あーっ!」
 (S4-4) A: 「何?」
 (S4-5) B: 「大惨事が!」



例えば, 前節の例 (a) では (S1-1) の「〇〇して下さい」という依頼に対して, (S1-2) の「はい」という応答とともに, 見回す行為が呼応している. (b) では, (S2-1) に対して (S2-4) が, (S2-2) に対して (S2-3) の言動が呼応して現れていることからわかる. これらの呼応が上述したように, 注意や焦点を共有していることを表している. また, 2.1 で述べた modal density

のように, 複数のモダリティが一つの対象に向けられていることも重要な現象である.

これらに対し, 前節の例 (c), (d) ではお互いに呼応していない言動が目立つ. (S3-1) の A の指示に対して (S3-2) で B が関連性のない話題を返したため, (S3-3) でもう一度 A が (S3-1) に関連する発話をしている. (d) では, (S4-1) の依頼に対して (S4-2) で実際に作業が始まっているが, (S4-3) の B の応答「あーっ!」が A の期待したものではなかったため, (S4-4) の質問となっている.

このような呼応を指標化することが本研究の目的となる. 深い意味解析は難しい問題であるため, まず本研究では 上記 (I) の一部をとらえる表層的な特徴について検討し, 指標化された一貫性が高い状態を円滑なコミュニケーションが行われている状態だと考える.

ただし, 一貫性が高いことが作業として必ずしも良い状態とは言えない場合もある. 臨機応変な行動をするためには, 対話の相手の言動だけでなく, 周囲に注意を払っておく必要があり, 状況に応じて相手の言動に割り込んだり, それまでの言動に呼応していない言動を起こす必要がある. これらについては今後の課題とする.

3. 映像対話における一貫性の指標

3.1 基本的な考え方

一貫性の検出には以下の方法をとることにした.

(1) インタラクションの記録データから得られた頻出パターンに含まれる呼応パターンを選ぶ.

(2) 選ばれた呼応パターンに対し, 映像対話の各時点でのインタラクションが適合する度合いを数値化し, その時間的変化や統計的な性質を求める.

頻出パターンを用いるのは以下のような考察による.

- 良好に作業が進んでいる状況を集めれば, 共通の対象に注意が向けられ, 意志の疎通が良好な状況が多く含まれる. つまり, 一貫性の高いパターンがその中で頻出している.

- 一般的な状況でも一貫性のないコミュニケーションが現れることもあるが, このようなパターンは一貫性の高いパターンに比べて頻度が低い.

また, これまでの調査により, 頻出パターンの各要素が共起する際の時間差に特徴的な偏りがあることがわかっている. 従って, 各状況で呼応パターンが成立するか (存在するか) どうかだけでなく, その時間的な性質を調べることにより, 一貫性のより詳細な指標を得られることを期待する.

映像対話型行動支援では, 視点の自由度が小さいこと, 嗅覚や力覚が使えないこと等, コミュニケーションのチャンネルが大きく制限されている. そのため, 多くの情報が明示的に言語や行動で説明される特徴がある. また, 作業者と支援者の間で伝わる情報をほぼ網羅的に記録できるという利点があるため, 以上の考え方で多くの重要な性質を捉えられることが期待できる.

3.2 頻出パターンの検出

頻出パターンの抽出については概要のみを説明する. 詳細は [1] を参照されたい.

表 2 画像からの特徴抽出

| 特徴 | 備考(検出条件) | 略号 |
|------|-------------------|-----|
| 視線停留 | カメラモーションが小さな状態が続く | V:h |
| 見直し | カメラモーションが見直す動きを示す | V:l |
| 移動 | カメラモーションの前進を示す | V:m |

表 3 発話から抽出する特徴

| 特徴 | 検査対象 | 条件 | 値 | 略号 |
|-------|------|----------|----------|------|
| 話者 | 話者 | 発話者の役割 | 作業者, 支援者 | W, S |
| 発話的役割 | 文 | 発話における役割 | 説明, 定型応答 | D, P |

表 4 頻出パターンの例

| パターン | 出現数 |
|-----------------|-----|
| (W:P)(S:D)(W:P) | 112 |
| (W:P)(S:D)(S:D) | 112 |
| (V:l)(W:D)(W:D) | 110 |
| (V:l)(W:D)(S:D) | 110 |
| (V:l)(W:D)(V:h) | 109 |
| (V:l)(V:h)(S:D) | 109 |
| (V:l)(V:h)(W:D) | 108 |
| (V:l)(S:D)(W:D) | 108 |
| (V:l)(S:D)(S:D) | 107 |
| (V:m)(S:D)(W:D) | 107 |
| (V:m)(S:D)(V:h) | 107 |

まず、発話や行動のモダリティからコミュニケーションの要素を抽出する。これらを「特徴」と呼ぶことにする。将来的な自動処理を想定し、画像処理、自然言語処理が可能な範囲内での特徴を設定した。画像特徴については表 2 に示したものをを用いた。視線停留、見直し・視線移動、移動は、カメラの動きから判断する。発話特徴には、表 3 に示す発話者と文タイプのみを用いる。

これらの特徴を発生時刻順に並べたトランザクションデータから頻出パターンを抽出する。抽出アルゴリズムには、時系列パターンのマイニングで良く用いられる PrefixSpan [2] を用いた。PrefixSpan では複数の系列データ入力とし、そこから抽出される数が閾値以上になる部分系列パターンを出力する。得られた頻出パターンの例を表 4 に示す。ここで用いられている記号は、”.” の前の”W” が作業者の発話、”S” が支援者の発話、”V” が作業者の行動(画像より得られる)を意味し、”.” の後の”P”, ”D” 等が表 2, 3 にあげた特徴を表す。例えば、”(V:l)(W:D)(S:D)” は、作業者が周囲を見回した後、支援者が何らかの説明をし、作業者がさらに何らかの説明をすることを表す。

表 5 抽出された頻出パターンとそこに含まれる呼応パターン

| 呼応パターン | 発生数 | 頻出パターン | $T_{(i,j)}$ |
|-------------|-----|-----------------|-------------|
| (S:D)→(W:D) | 137 | (V:l)(S:D)(W:D) | 0.79 |
| (W:D)→(S:D) | 137 | (V:l)(W:D)(S:D) | 0.80 |
| (S:D)→(W:P) | 136 | (W:P)(S:D)(W:P) | 0.82 |
| (W:P)→(S:D) | 141 | (W:P)(S:D)(W:P) | 0.79 |
| (V:l)→(S:D) | 140 | (V:l)(W:D)(S:D) | 0.80 |
| (V:h)→(S:D) | 138 | (V:l)(V:h)(S:D) | 0.79 |
| (S:D)→(V:h) | 135 | (V:m)(S:D)(V:h) | 0.79 |

3.3 呼応パターンの設定

頻出パターンに含まれる部分パターンのうち、特に重要なものを呼応パターンとする。現在は 2 特徴のペアのみを考える。

選ぶ条件としては、以下のものをを用いた。

- 3 つ以上の頻出パターン(上位パターン)に含まれること
- 上位パターンの中で省略される割合が低いこと

前者はコミュニケーションとして意味のない組をとらないために必要となる。2 要素しか含まれない頻出パターンには、個々の特徴の頻度が高いために、呼応していないにも関わらず頻度が高くなるものがあるからである。

後者も前者と同様の考えに基づいており、3 要素あることが重要な上位パターンのうち、省略されることの少ない 2 要素を重要な呼応関係としている。その判定は以下のようにする。

表 4 に示された 3 要素の頻出パターンの出現数を $N(F_i^3)$ 、そこに含まれる 2 要素(特徴のペア)の出現数を $N(F_{i,j}^2)$ とし、その比率 $r_{(i,j)} = N(F_i^3)/N(F_{i,j}^2)$ を求める。この $r_{(i,j)}$ が閾値以上のものを採用する。

閾値を 0.79 にした場合に選ばれる呼応パターンとその上位の頻出パターンとの関係を表 5 に示す。以下の実験ではこれらを用いた。

次に、呼応パターンの時間的性質を考える。意味的に対応関係にある特徴同士であれば、起点となる側の特徴に相手が反応するための適度な間がある。この間が極端に短かければ、その応答は起点となった特徴の内容を良く把握していない可能性や、呼応していない(他の動機に基づく言動である)可能性が高い。また、長い間が空いている場合にも、呼応していない可能性が高い。呼応パターンに含まれる特徴間の生起時刻のずれの分布を調べた。その方法を、特徴カテゴリ F_a と F_b からなる呼応パターンについて調べる場合を例として説明する。まず、全データから特徴カテゴリ F_a に属する特徴 (f_a^i) のリスト $[f_a^0, f_a^1, \dots, f_a^n]$ を抽出する。それぞれに対して、その生起後最も近い時刻に現れる F_b に属する f_b^i を求める。その 2 つの特徴間の生起時刻の差 (ΔT_i とする) の頻度分布を求める。

得られた分布は図 2 のようになる。ほとんどの応答は 1~3 秒以内に現れることが多く、5 秒以内に行なわれている。ただし、S:D→W:D や S:D→V:h のように支援者からの発話に対しては明確なピークがない分布になる。これは、支援者からは作業の指示のように長めの発話があるためであると考えられる。

以下の実験では、これらの頻度分布を簡単な確率分布として近似し、それを用いる。具体的には、 F_a と F_b に属する特徴間の生起時刻差が Δt である確率分布を $P(F_b|F_a; \Delta t)$ としたとき、 $P(F_b|F_a; \Delta t)$ を上記の頻度分布から求める平均と分散に従う正規分布で近似する。

4. 映像対話に対する一貫性の計算

呼応パターンの特徴間の生起時刻のずれから、ある時点でのインタラクションの一貫性を評価するための指標を設計する。

その基本的な考え方は以下の通りである。

- それぞれの呼応パターンの 1 つ目の特徴の生起毎に、それ以後最も近い時間に生起した 2 つ目の特徴を求め、その時間

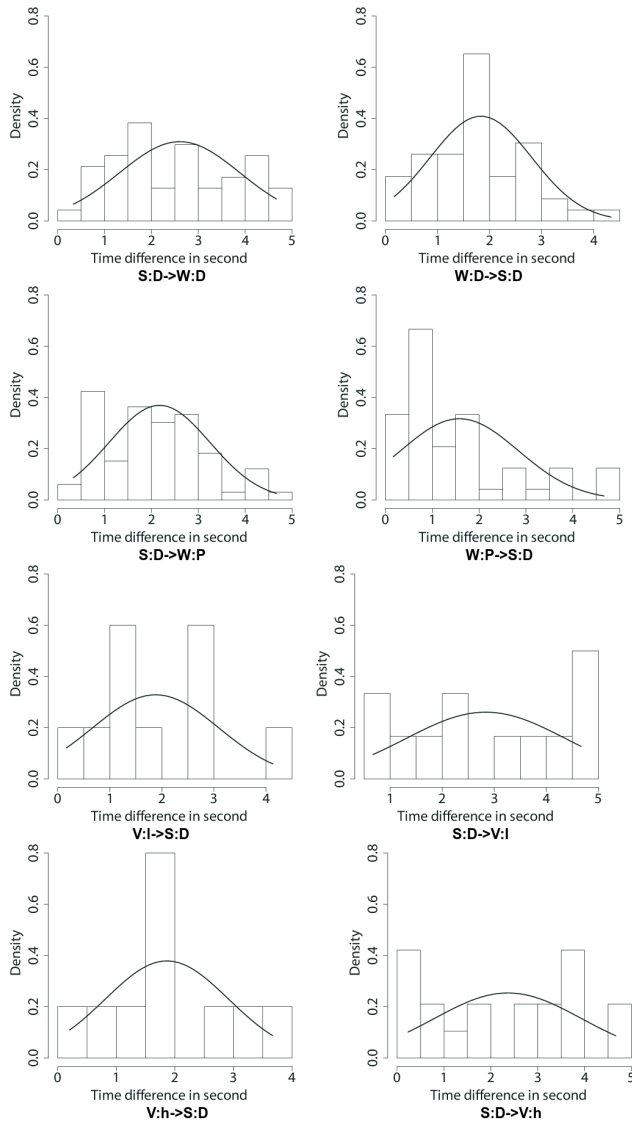


図 2 各呼応パターンの生起時刻ずれの分布と正規分布による近似カーブ

差によってスコア付けする。

- 各時刻に対して複数の呼応パターンが重複している場合には、各呼応パターン毎に求めたスコアの和、最大値などを用いて、各時刻における一貫性の指標とする。

前者のスコア付けは上で述べた確率分布を基にして計算する。前節で述べたように、 $P(F_b|F_a; \Delta t)$ の値が大きい部分で起こっている呼応は良い呼応関係であると考え、良いスコアを与える。そのため、呼応パターンの時間差が Δt の場合の正スコア (以下 s_p とする) を、 $P(F_b|F_a; \Delta t)$ とする。逆に、 $P(F_b|F_a; \Delta t)$ が小さかったり、呼応する特徴がない (一定時間内に現れない) 場合には、悪さを考える必要がある。そこで、呼応関係の悪さを表す負スコア (以下 s_n とする) を $P(F_b|F_a; \Delta t) - \max_{\Delta t} P(F_b|F_a; \Delta t)$ とする。正スコアが呼応関係の良さを表すスコアとなり、負スコアは呼応関係が最良の状態から外れている度合いを表すスコアとなる。

複数の呼応パターンが重複している場合には、個々の呼応パターンから得られるスコアを総合する必要がある。その際の観

表 6 実験システムの設定

| | |
|----------|--|
| 作業用システム | USB カメラ (ヘアバンドで固定), ヘッドセット (マイク付きヘッドフォン), ノート PC |
| 支援者用システム | ノート PC (内蔵マイク) |
| 映像伝送方式 | vp8 コーデック (約 15fps) |

点として、以下の 3 つを用いた。

(1) コミュニケーションの受け手が送り手に反応しているかどうかは、最も良く呼応しているパターンによってわかる。それを評価するには、時刻 t における各呼応パターン i による正スコアを $s_p(t, i)$ とする場合の $S_m(t) = \max_i s_p(t, i)$ が良い指標となる。

(2) 複数の呼応パターンが重複して現れることは modal density 等の観点から良い状態であると考え。それを評価するには、 $S_s(t) = \sum_i s_p(t, i)$ が良い指標となる。

(3) コミュニケーションの受け手が反応していないことは、最も良く呼応しているパターンの悪さによってわかる。それを評価するには、 $S_n(t) = \max_i s_n(t, i)$ が良い指標となる。

以上の 3 つそれぞれが異なる観点からの指標であるため、本研究では一つに絞り込むことをせず、実際のインタラクションの例に対する上記それぞれの指標を求め、その適切さについて検証する。

5. 実験例

データ収集

システムは表 6 の構成とした。作業者は USB カメラ (ヘアバンドを用いて額に装着) とヘッドセットマイクを装着し、それを QVGA の品質で支援者に伝送する。支援者側にはカメラを設置せず、音声のみが作業者に送られる。

作業には複数の調理タスクを選んだ。作業者は筆者らの研究室に設置されたシステムキッチンで調理を行い、支援者は別の部屋で映像を見ながらアドバイスを与える。作業にかかる時間は調理により異なるが、およそ 20 分から 30 分程度になる。

作業者と支援者の設定

表 7 被験者の組み合わせ

| ペア | 作業者 | 支援者 |
|----|------------|------------|
| a | 学生 (初心者) | 学生 (初心者) |
| b | スタッフ (熟練者) | スタッフ (初心者) |
| c | 学生 (初心者) | スタッフ (熟練者) |

作業者と支援者の立場やタスクへの熟練度の違いから現れるコミュニケーション傾向の違いを調べるために、表 7 のように異なる立場の被験者の組み合わせで実験を行なった。

相手への話しやすさや知識の差などの様々な要因でコミュニケーションに差異が現れると考えられる。例えば、ペア a は同学年の学生同士で普段から良く話す間柄であり、調理のスキルも同等であった。そのため、過不足なく良好なコミュニケーションをとりながら慎重に調理を進めていた。ペア b はスタッフ同士であり、作業者のスキルが高かったことから、支援者の指示を無視するような場面がいくつか見られた。ペア c は実際

に指導を受ける立場の学生が作業者となっている。しかし、学生側からコミュニケーションを取ることが少なかったため、それぞれが異なる対象に注目しながら作業を行なっているような状態になった。

今回は被験者をそろえることが難しく網羅的な組み合わせを試すことはできなかったが、これらのペア毎に提案した指標でどのような性質の違いとして現れるかの考察を行なった。

5.1 サンプルの抽出

収録されたデータから、コミュニケーションが円滑に行われている部分、また、種々の観点からコミュニケーションに問題がある部分を抜き出し、各状況における指標を確認した。

表8と表9は作業者と支援者の間で円滑なコミュニケーションが行われている例である。それに対し、表10は作業者が支援者への状況伝達をほとんど意識せずに作業を行なった例である。表11では、複数の注目対象があるためにインタラクションの一貫性がなくなっている例である。表12では作業者の発話に支援者が応答しているが作業者はそれを無視しているような形になっている。その他、紙面の都合上省くが、収録したデータには、一貫性の低いシーンが他にも多くある。これらについて網羅的に調べていくことは今後の課題である。

表8 一貫性の高いインタラクションの例

| id | 時刻(秒) | 特徴 | 内容 |
|-------|-------|-----|-----------------------|
| A-a1 | 162.0 | S:D | 「黄身と白身をむらなく良く混ぜる」 |
| A-a2 | 163.3 | V:m | 移動 |
| A-a3 | 164.7 | W:P | 「はい」 |
| A-a4 | 164.7 | V:h | 停留 |
| A-a5 | 166.3 | S:D | 「で、だし汁はそのボウルに入っているやつ」 |
| A-a6 | 167.6 | V:m | 移動 |
| A-a7 | 169.0 | S:D | 「ちょっともう見えんけど、それ、それ」 |
| A-a8 | 169.0 | V:h | 停留 |
| A-a9 | 170.7 | S:D | 「もうちょっと頭さげて、あ、それ、だし汁」 |
| A-a10 | 172.3 | W:P | 「はい」 |

表9 一貫性の高いインタラクションの例2

| id | 時刻(秒) | 特徴 | 内容 |
|------|-------|-----|--------------------|
| B-c1 | 243.1 | W:D | 「で？」 |
| B-c2 | 244.0 | S:D | 「で、余熱」 |
| B-c3 | 245.4 | W:D | 「いやまだ粉ふってない」 |
| B-c4 | 246.0 | V:h | 停留 |
| B-c5 | 246.5 | W:D | 「これまだ、まだふってないですよ？」 |
| B-c6 | 247.7 | S:D | 「ああ、どっちでも良いです」 |

5.2 各サンプルに対する指標

円滑にコミュニケーションが行われている状況(表8)に対する指標は図3となる。ここでは、支援者の指示に対して作業者が適切に応答しているため、 S_s , S_m の指標は高い値となっている。その中でも(A-a1)の指示に対しては(A-a3)で発話による応答に加え(A-a4)で視線を停留させて現場の状況を映像でも伝えられている。そのため、この時の S_s が特に高い値になって

表10 作業者が支援者と状況を共有しないまま作業を行なっている例

| id | 時刻(秒) | 特徴 | 内容 |
|-------|-------|-----|------------|
| C-b1 | 128.0 | W:D | 「うーん」 |
| C-b2 | 129.0 | V:l | 見まわし |
| C-b3 | 130.3 | W:D | 「おさら」 |
| C-b4 | 130.3 | V:m | 移動(食器棚へ) |
| C-b5 | 133.3 | V:l | 見まわし |
| C-b6 | 134.7 | V:m | 移動(シンクへ戻る) |
| C-b7 | 136.7 | V:h | 停留 |
| C-b8 | 137.8 | V:l | 見まわし |
| C-b9 | 138.7 | V:m | 移動(食器棚へ) |
| C-b10 | 141.0 | S:D | 「ん、おさら？」 |
| C-b11 | 141.0 | V:h | 停留 |
| C-b12 | 145.3 | S:D | 「おちるおちる」 |

表11 注目対象が重なってしまっている例:(c-1 → c-2 → c-4) と (c-3 → c-5)

| id | 時刻(秒) | 特徴 | 内容 |
|------|-------|-----|-------------------|
| D-c1 | 143.3 | S:D | 「粉が全部綺麗に一面に付くように」 |
| D-c2 | 147.5 | W:D | 「粉?」 |
| D-c3 | 148.3 | W:D | 「あーここに粉ふるんですか」 |
| D-c4 | 149.1 | S:D | 「強力粉」 |
| D-c5 | 149.7 | S:P | 「はい」 |
| D-c6 | 150.0 | V:l | 見まわし |
| D-c7 | 151.0 | W:D | 「あ、ほんとだ」 |
| D-c8 | 151.1 | V:h | 停留(レシピ確認) |

表12 作業者が支援者の指示を無視して作業を行なっている例

| id | 時刻(秒) | 特徴 | 内容 |
|------|-------|-----|----------------------|
| E-c1 | 56.2 | W:D | 「ミキサー」 |
| E-c2 | 56.8 | S:D | 「あれ？」 |
| E-c3 | 57.5 | W:D | 「ミキサー」 |
| E-c4 | 58.0 | S:D | 「その…上に置いてありませんでしたっけ」 |
| E-c5 | 60.0 | W:D | 「ミキサー」 |
| E-c6 | 60.7 | S:D | 「棚の所にはありませんか？」 |
| E-c7 | 63.0 | W:D | 「え、ミキサー」 |
| E-c8 | 63.7 | S:D | 「オープン棚の下？」 |
| E-c9 | 64.8 | S:D | 「オープンの下」 |

いる。同様に、表9に対する指標は図4のようになる。(B-c4)のように、対象を映像で相手に伝える行為が S_s の値の高さとして現れていることがわかる。

表10に対する指標は図5のようになった。(C-b3)で作業者は皿を取りに行くことを発話しているが、支援者はそれに反応できていない。その後(C-b10)に対して作業者も返事をしていない。そのため、 S_n の値が低くなっている。特に作業者の発話である(C-b3)に対する応答がないことが S_n のスコア低下に現れている。(C-b7)と(C-b8)の視線移動に反応して(C-b10)があることで S_n は一旦改善しているが、その後作業者の応答がないことで、再び S_n は悪化している。

表11に対しては図6となった。両者の発話について、それぞれでは良好な応答が現れている。しかし、(D-c2)から(D-c3)で作業者が質問を重ねたために(D-c2)に対する答えである

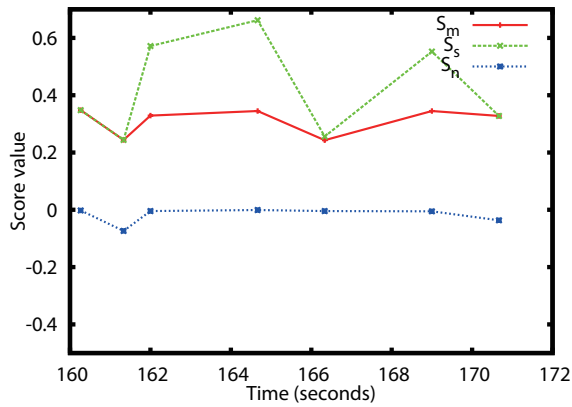


図3 表8の例に対する指標

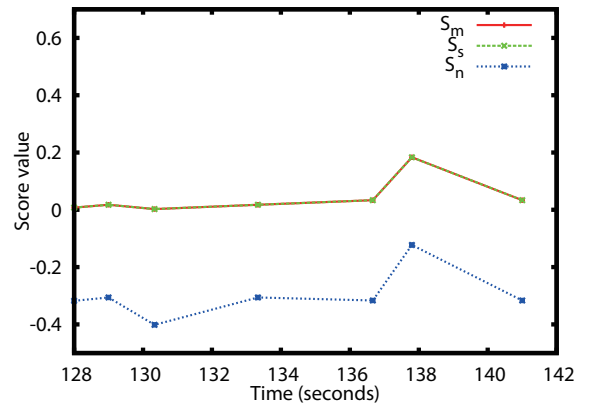


図5 表10の例に対する指標

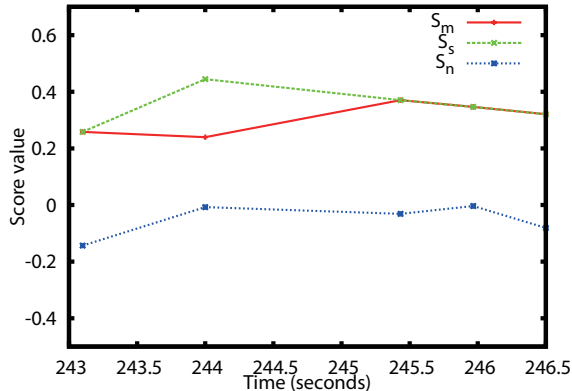


図4 表9の例に対する指標

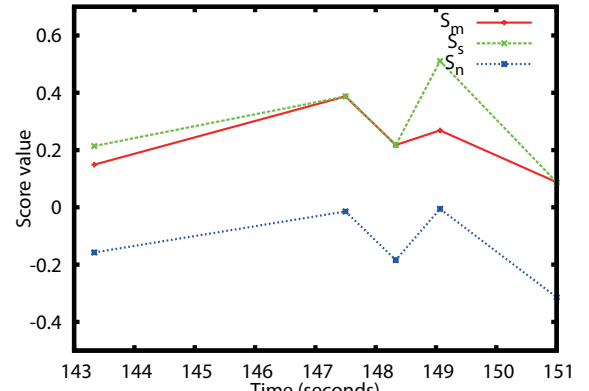


図6 表11の例に対する指標

(D-c4)が(D-c3)の直後に現れている。その時刻で S_n の値が一時的に低くなっていることがわかる。 S_m も相応に値が低下している^(注1)が、 S_s はその後良いコミュニケーションが取れていることで高い値となっており、一貫性の高いコミュニケーションが保たれていることがわかる。

表12では図7のようになった。こちらは作業者の発話(E-c1)~(E-c7)のそれぞれに支援者が応答している箇所 S_s 、 S_m 、 S_n が高くなっているが、支援者の応答に作業者が反応しないため、それぞれの応答間で特に S_n が低くあらわれており、ここでの一貫性が低くなっていることが伺える。

5.3 作業者・支援者ペアによる違い

作業者と支援者のペアによるコミュニケーション傾向の違いと出現する呼応パターンの数等を概観する。比較のために同一時間長となるようにしたが、それぞれのペアで調理内容は異なっている。

ペアaは双方で丁寧にコミュニケーションを取りながら作業を進めていた。一貫性の指標は図8のようになる。 S_s が高い値となる場面が多いことから、指示に従うだけでなく、映像もうまく使いながら一貫性の高いコミュニケーションをとれていることがわかる。表13にはここで出現した呼応パターンの数を示す。(S:D→W:P)と(W:P→S:D)の呼応パターンが良く現れており、支援者の発話に作業者が相槌をうちながら聞いていることが多いことが伺える。

ペアbでは全体的に一貫性の低いコミュニケーションが目

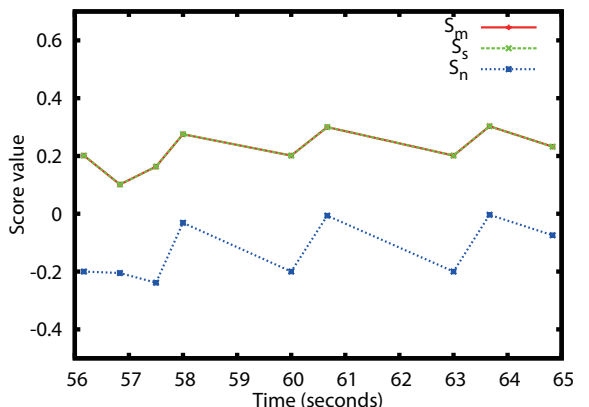


図7 表12の例に対する指標

立った。その指標は図9のようになる。多くの場合 S_s も S_m も低い値になっており、 S_n も低い値をとっている箇所が多い。そのため、特徴のペア自体がほとんど出ていないことがわかる。表13を見ると全体的に呼応パターンの出現数自体が少なく、他のペアと比べて発話回数が少ないことがわかる。

ペアcに対する指標は図10のようになる。その変動は大きい、 S_m が高い箇所が多く見られる。しかしペアaと比べると S_s の値が高くなる箇所が少ない。これは、作業者が自主的に作業を進める場面が多いため、映像に呼応する形で発話が見られる場面が少ないことが一因である。表13を見ると、動作を含む呼応パターンの出現数に比べて、発話による呼応パターンが多いことがわかる。しかし、(S:D)→(W:P)、(W:P)→(S:D)はそれに比べて出現数が少ない。これは、作業に関係のない雑談が多かったためである。

(注1)：今回の実験では(D-c5)の支援者の定型応答を含む呼応パターンを考慮していないことも影響しており、このように状況に応じた呼応パターンの一貫性への関係の調査は今後の課題である。

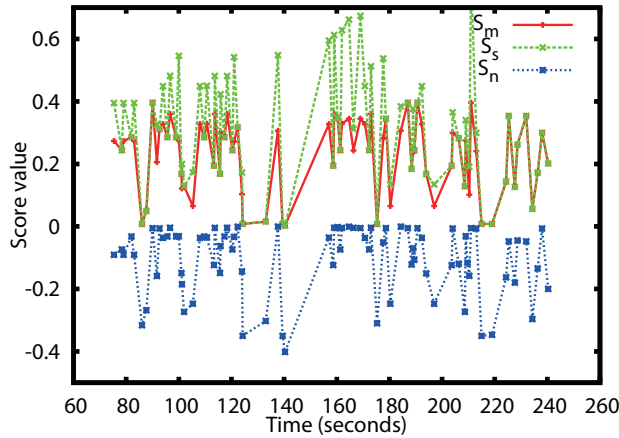


図8 ペア a の映像対話に対する指標

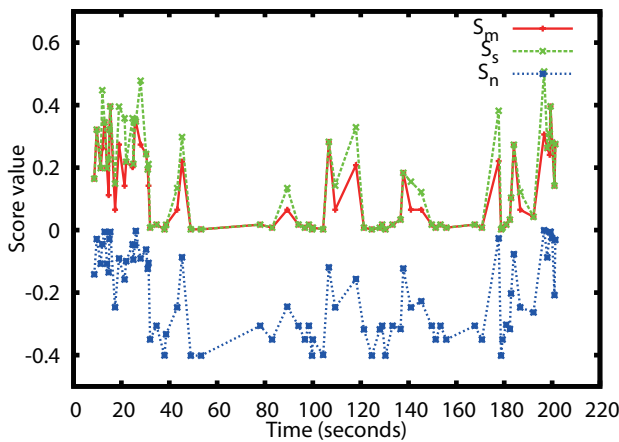


図9 ペア b の映像対話に対する指標

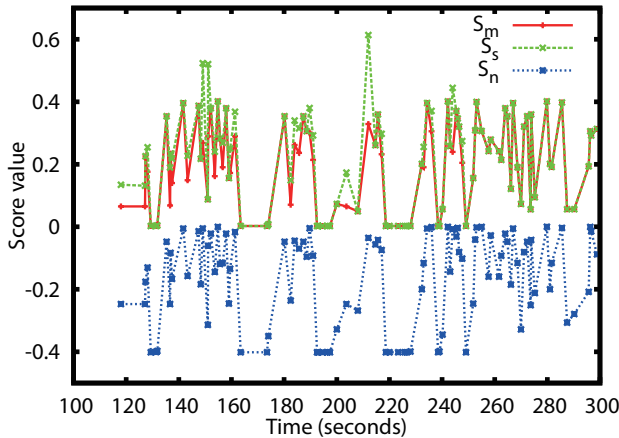


図10 ペア c の映像対話に対する指標

6. おわりに

本研究では、映像対話型行動記録から表層的に観測可能な特徴を用いて、コミュニケーションの円滑さを表す指標を提案し、実際のデータからインタラクションの性質との関係性を調査した。個々の特徴の呼応関係に基づく指標により、一貫性の高いインタラクションが行われていることや言動が呼応していないちぐはぐなインタラクションとなっていることを表すことが可能であることを確認した。

表 13 ペア毎の呼応パターンの出現数

| 呼応パターン | a | b | c |
|---------|----|----|----|
| S:D→W:D | 16 | 9 | 32 |
| W:D→S:D | 11 | 5 | 43 |
| S:D→W:P | 2 | 9 | 3 |
| W:P→S:D | 5 | 11 | 3 |
| V:l→S:D | 11 | 9 | 0 |
| V:h→S:D | 22 | 5 | 3 |
| S:D→V:h | 21 | 2 | 3 |

今後は様々な状況で実験を行ない、網羅的な調査を行う必要がある。また、今回の実験では2つの特徴の呼応パターンの、さらにその一部だけを用いて指標の検証を行なったが、他の呼応パターンとコミュニケーションの一貫性との関係についてもさらに検証をし、どのような呼応パターンがどのような一貫性の高いコミュニケーションに寄与しやすいのかを検証していく必要がある。

他に、本研究では良いコミュニケーションの条件に一貫性の高さを仮定していたが、実際は状況に合わせた臨機応変な対応が必要になる場合があり、一貫性の高さが必ずしもコミュニケーションの良さと直接結び付かない例もある。このような場面でどのようなコミュニケーションを取ることが良い支援に繋がるのかについても今後検討していく必要がある。

文 献

- [1] 小泉敬寛, 小幡佳奈子, 渡辺靖彦, 近藤一晃, 中村裕一, “映像対話型行動支援における作業者と支援者の態度の分析”, HCG シンポジウム 2013, pp. 425-432, 松山市総合コミュニティセンター, Dec., 2013.
- [2] J. Pei, J. Han, B. Mortazavi-asl, H. Pinto, Q. Chen, U. Dayal and M.-C. Hsu. PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. 17th International Conference on Data Engineering (ICDE '01), pp.215-224, 2001.
- [3] S. Norris: “Analyzing multimodal interaction”, Routledge, 2004
- [4] S. Norris: “Identity in (Inter)action”, De Gruyter Mouton, 2011
- [5] M. Billinghurst, S. Bee, J. Bowskill, H. Kato, “Asymmetries in Collaborative Wearable Interface”, The Third International Symposium on Wearable Computers, pp.133-140, 1999
- [6] R. Kraut, M. Miller, J. Siegel, “Collaboration in performance of physical tasks: Effects on outcomes and communication”, Proc of the ACM Conference on Computer Supported Cooperative Work (CSCW 1996), 1996
- [7] S. Fussell, R. Kraut, J. Siegel, “Coordination of communication: Effects of shared visual context on collaborative work”, Proc. of the ACM Conference on Computer Supported Cooperative Work (CSCW 2000), 2000
- [8] R. Kraut, D. Gergle, S. Fussell, “The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co- Presence”, Proc. of the ACM Conference on Computer Supported Cooperative Work (CSCW 2002), 2002
- [9] D. Gergle, R. Kraut, S. Fussell, “Action as language in a shared visual space”, Proc. of the ACM Conference on Computer Supported Cooperative Work (CSCW 2004), 2004
- [10] H. Clark, D. Wilkes-Gibbs, “Referring as a collaborative process”, Cognition, Vol.22, pp.1-39, 1986