

# 映像対話型行動支援における頻出パターンに基づいた コミュニケーションの分析

小泉 敬寛<sup>1,a)</sup> 小幡 佳奈子<sup>1</sup> 渡辺 靖彦<sup>2</sup> 近藤 一晃<sup>1</sup> 中村 裕一<sup>1</sup>

受付日 2014年6月30日, 採録日 2014年11月10日

**概要:** 本論文では, 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「映像対話型行動支援」とその記録データである「映像対話型行動記録」の解析について述べる. 映像対話型行動支援では, 作業の対象や状況についての様々な説明や指示, 報告などが少数のモダリティを介して伝達される. 本研究の目的は, このような状況を記録し, コミュニケーションのパターンを調査することによって, 作業者, 支援者の振舞いの特徴や, 失敗につながる傾向を明らかにすることである. 本研究では, 実際に調理作業に対する映像対話型行動支援を記録し, その性質を調べることによって, 提案手法の有効性を確かめた.

**キーワード:** ライフログ, 映像対話型行動支援, 個人視点映像, データマイニング, 頻出パターン

## Communication Analysis for Working Support through First Person Vision Communication Based on Frequent Patterns

TAKAHIRO KOIZUMI<sup>1,a)</sup> KANAKO OBATA<sup>1</sup> YASUHIKO WATANABE<sup>2</sup>  
KAZUAKI KONDO<sup>1</sup> YUICHI NAKAMURA<sup>1</sup>

Received: June 30, 2014, Accepted: November 10, 2014

**Abstract:** This paper addresses the analysis of “Working Support through FPV (First Person Vision) Communication”, which is a working style in which a person with a head-mounted camera is working under the guidance of an experienced person monitoring the FPV at a distant place. Communications in this working style contain various types of significant interactions concerning intentions, instructions, explanations as well as usual information on working. The purpose of this research is to analyze the working situations and to examine the characteristics of working and supporting persons, and to clarify the relationships to the failure or near failure situations. In this research, we applied the analysis to the records on cooking with FPV communications, and verified the potential of our methods.

**Keywords:** lifelog, working support through FPV, first person vision, data mining, frequent patterns

### 1. はじめに

映像対話型行動支援では, 図 1 に示したように, カメラ, マイク, その他のセンサを作業者が体に装着し, 無線を介して支援者に映像とセンサデータを送る. 支援者は送られ

てくる映像や計測データを見ながら作業者と対話し, 指示などを行う. 映像対話型行動記録はこのような映像対話を記録したものである. 映像対話型行動支援は様々な用途, たとえば, 狭い場所や危険な場所で, 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う場合などに活用できる. 実際に, 遠隔医療指導や救急医療補助などへの応用 [1] や, ヘルメットに付けたカメラの映像を携帯電話の回線を用いて送るシステム [2] などが実用化されている. 今後, 種々のデバイスの高機能化や小型化がいつそ

<sup>1</sup> 京都大学学術情報メディアセンター  
ACCMS, Kyoto University, Kyoto 606–8501, Japan

<sup>2</sup> 龍谷大学理工学部  
Faculty of Science and Technology, Ryukoku University,  
Otsu, Shiga 520–2194, Japan

a) koizumi@media.kyoto-u.ac.jp

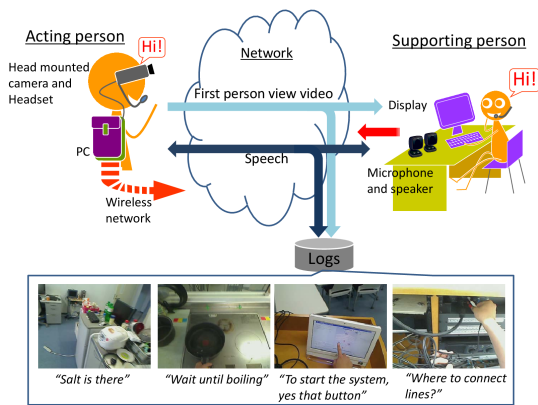


図 1 映像対話型行動記録の概要

Fig. 1 Overview of the working support through FPV communication and that logs.

う進み、種々の行動支援の形態が増えることが予想される。

このような背景から、映像対話型の行動支援で起こる現象を調査することは興味深く、重要な課題となる。たとえば、作業が円滑に進んでいる状況や失敗が起こった状況を比較することによって、失敗や事故が起こりやすい状態を明らかにしたり、その防止方法を検討したりすることが期待される。また、作業者と支援者間のコミュニケーションの良さを評価したり、将来的には、それをリアルタイムにフィードバックしたりすることにより、作業をより円滑に進める補助を自動的に行うことも考えられる。

このような問題に対し、本研究では、作業者と支援者の振舞いのパターンと、意思疎通の円滑さ、また、それらと作業の円滑さや失敗との関係を解析する手法を検討し、実際にデータの解析を行った。

以下、本論文では、2章で背景や関連研究について述べ、3章では本研究での問題設定と目的について述べる。4章では映像対話型行動記録におけるデータの性質について述べる。5章で具体的な凝集性の検出方法について述べる。6章で実際に提案手法を用いて映像対話型行動記録から振舞いのパターンを検出する実験とその結果について述べた後に、7章でそれらを解析することで分かる実際のコミュニケーションとの関係について述べる。最後に8章で本論文のまとめを述べる。

## 2. 関連研究

ウェアラブルコンピュータによる映像通信とそれによる遠隔支援について、早くからその可能性が示唆されてきたが [3], [4], 作業支援の詳細な解析としては、Kraut らが先駆的な研究を行っている [5]。彼らは、自転車の修理を課題とした遠隔行動支援に対して、音声とマニュアルのみの場合とさらに映像を加えた場合の作業効率や作業者 (worker) と支援者 (expert) の振舞いについて比較した。映像対話が可能になることによって、両者の振舞いに大きな変化が現れること、特に支援者がプロアクティブに振る舞うこと

などが報告されている。Fussel らは、さらに、支援者が実空間で側についている場合を加え、映像通信がある場合と、音声だけの場合との比較を行った [6]。彼らは視覚情報の活用の仕方が3つのコミュニケーション形態で異なり、それによって発話が大きく変化することなどを報告している。

Kraut, Gergle らは、ジグソーパズルを題材としたオンラインの協力作業を設定し、作業者 (worker) と支援者 (helper) の振舞いを解析した [7], [8]。映像の遅延、映像共有の有無によるコミュニケーションの差が、現場参照の他に、理解の確認、作業後の確認などに顕著に現れることなどを統計的に示している。

このように、映像通信を用いた種々の作業支援の性質が調査されつつある。本研究の目的は、それをさらに進め、映像対話型行動支援におけるコミュニケーションの「良さ」を評価したり、失敗との関係を解析するための指標を作ることである。この点に関連して、Kraut, Gergle らの研究 [7], [8] では、高次の意味的な行動 (記号) を用いてコミュニケーションのパターンを解析している。それに対し、本研究では、自動的な解析を視野に入れ、直接観測可能な (低レベルの) 特徴を基にした定量的な解析に主眼を置いている点が特徴である。

## 3. 問題設定

本研究では、映像対話型行動支援におけるコミュニケーション記録に関し、以下を提案することを目的とする。

- (1) 作業者と支援者の振舞いの類型化と定量化の手法
- (2) (1) の手法が作業者と支援者の振舞いと作業の円滑さや失敗の関係を解析する指標として有用であること

(1) に関し、映像対話型行動支援の場では、作業や行動におけるノンバーバルな振舞いが重要な意味を持ち、それを解析の対象とする必要がある。(2) に関し、無知、誤判断、不注意、不遵守などが失敗の原因となることが知られているが [9], 作業者を支援する際のコミュニケーションが不十分な場合にこれらが顕在化することは明らかである。ただし、失敗や事故の生起数は少なく、また、意志の疎通が不十分な場合に必ず失敗するわけでもない。

このような考え方にに基づき、本研究では以下のような問題設定を行った。

- (a) マルチモーダルなコミュニケーションに対し、失敗を引き起こしていない状況を含めてパターンを解析し、頻出するパターンの検出とその定量的な解析を行う。
- (b) 意味的に深い解析を行うのではなく、データから自動的に抽出可能な表層的な特徴を用いる。多くのデータを解析することを容易にするとともに、将来的なりリアルタイム処理の可能性を探ることが目的である。

## 4. 行動とコミュニケーションの記述

### 4.1 モダリティの設定

映像対話型行動記録中の特徴とその相互関係を考えるために、まず、本研究で扱うモダリティとそれが伝える典型的な情報を以下のように整理する。

発話：(可視物の) 名前, 現場参照, 外観, 状況, 依頼・応答, 質問・説明, 報告, その他

見る行為：視線停留, 見回し, 見るための体の移動, その他  
 行動(作業・移動)：指示行動, 作業動作(手の動きなど), 移動, その他

「発話」以外の2つは主に画像から得られる情報であるが、それらが重要な意味を持つため、あえて別モダリティとしている。

### 4.2 典型的なシーンとインタラクションのパターン

映像対話型行動支援の典型的な場面を図2に示す。これらの例について、発話や行動の相互関係、および、情報がどのように伝わっていたか考えてみよう。

図2の(a)は円滑に作業が進んでいる理想的な例で、作業者と支援者の間で映像と発話を使って適切に情報交換が進められた結果である。たとえば、(a-3)では、支援者の発話における「その赤いボタン」が、現場指示(参照)、色情報、名前の情報を与えている。(a-4)で作業者はスイッチ周辺を見ることにより、スイッチの存在(位置)、外観、色に関する情報を得ている\*1。これは、映像を介して支援者にも伝わる。また、指さし(指示行動)が、発話の「これ」に対応し、(a-2)の「その赤いボタン」に呼応している。このように、複数のモダリティにおける情報が適切な順序かつ適切なタイミングで与えられていることにより、情報が正確に伝わっていることが分かる。

図2の(b), (c)は作業に問題が起こった例である。(b)では作業者が周辺の器具を落としてしまったが、その一因は作業者と支援者間のインタラクションにある。つまり、(b-2)では作業者がそばの道具の状態に気づいていない(その方向を見ていない)ことを支援者が気づいていない。また、道具が倒れた後にも、(b-5)~(b-7)で状況が正確に伝わっていない。作業者が支援者に状況を伝えるためにモダリティを適切に活用できなかったことによる。(c)は必要な指示がタイミング良く与えられなかったために作業が成功しなかった例である。支援者が(c-2), (c-3)で作業者の発話に気をとられ、必要な指示を与えられなかった。(c-2)で作業者が上手に支援者を自分の話題に引きこんでしまったためである。

\*1 対象が作業者の視界に入っているにもかかわらず、対象に関する情報が認知されていることは保証できないが、ここでは理想的な場合を想定している。

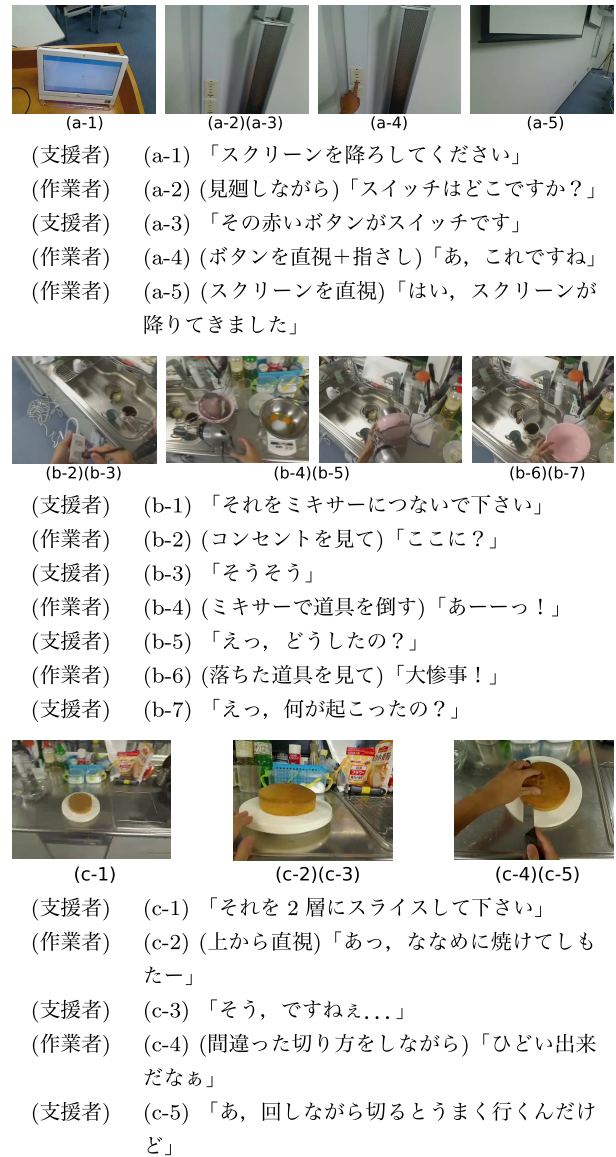


図2 映像対話型行動記録中の典型的なシーン

Fig. 2 Typical scenes of the working support through FPV communication.

### 4.3 頻出パターンと凝集性

図2の例に見られるような、複数のコミュニケーション要素が近接して用いられること(共起)を本研究では「凝集性(cohesiveness)」と呼ぶことにし、(a)~(c)いずれも凝集性が高い状態であると考えられる。(a)においては発話や行動がお互いに呼応し、お互いの内容を参照したり言及したりしているのに対し、(b), (c)では先に表れた発話や行動と関係のない行動や発話がある。本研究では(a)のように、発話や行動が呼応しながら関わり合っている状態を一貫性(coherency)が高い状態であるとし、(b), (c)は一貫性が低いと考える。

このような凝集性とパターンの出現頻度について、本研究では、以下の仮定を置く。

- (i) 共通の対象に注意が向けられ、意志の疎通が良好な状況では、凝集性と一貫性の両方が高くなる。結果とし

て、良好に作業が進んでいる状況ではこのようなパターンが頻出する。

- (ii) 円滑な意志の疎通が行われていない場合には、必要な要素が欠けて（凝集性が低くなって）いたり、一貫性のないコミュニケーションのパターンが現れたりする。このような要素の過不足が頻出することは少ない。

このような仮定に基づき、本研究では、4.1 節にあげたモダリティから自動的に抽出可能なものを特徴とし、その発生系列から頻出するパターンとその性質を解析する。

## 5. 頻出パターンの分析手法

### 5.1 頻出パターンの抽出手法

頻出パターンの抽出には、時系列パターンのマイニングで広く用いられている PrefixSpan [10] を用いた。PrefixSpan は入力として与えられた記号系列の中に一定以上の頻度で出現する記号列を頻出パターンとして高速に抽出する手法である。その際に、PrefixSpan では、個々の頻出パターンが出現している間に無関係の記号の出現が挟まることを許し、複数の記号が同一の順番である（同時に生起する）ことを許す。本研究では、PrefixSpan に対し、個々の特徴を生起時刻順に並べた特徴系列を入力として与え、その中に頻出する特徴の系列を頻出パターンとして抽出する方法をとった。その際に以下のような設定を行った。

- 各特徴の生起時刻は開始時刻とする。
- 同一発話文内の複数の特徴は同時に生起しているとし、発話の開始時刻を生起時刻とする。

これは次のような考察による。特徴は継続区間（開始時刻から終了時刻まで）を持つことができるが、多くの場合、行動や発話が起こる直前に、行為者の意図がほぼ決まっていると考えられる。本研究では、特徴の開始時刻がそれに近いと仮定する。また、発話中の表現や語順は言語の文法構造に依存するため、単語が発話された厳密な時刻を考える意義は少ない。そのため、同一発話文内の特徴はすべて同時刻に生起したものとする。

頻出パターンの抽出処理は以下のように表すことができる。同時に生起している特徴を () でくくり、それ以外は生起時刻に順に並べたものをトランザクション（たとえば、 $S = \{F_1, F_2, \dots, (F_i, \dots, F_m), \dots, F_z\}$ ）とする。ここで、 $S$  は意味的なタスクや作業で区切られたトランザクション、 $F$  は特徴である。また、インデックス番号  $t$  は  $F$  に付与されるユニークな番号である。ただし、本研究では、トランザクション中の各要素は特徴のカテゴリとし、それ以上の区別をしない。

マイニングによって得られる頻出パターンも特徴カテゴリの系列となる。たとえば、<(支援者:依頼, 見直し), 移動, 作業> は、支援者からの依頼と同時に作業者の見回しがあって、その後作業者が移動し、作業が行われたことを意味する。

### 5.2 時間的性質の分析手法

前節の頻出パターンの抽出では各特徴の生起順序のみを考慮した。以下では、より詳細な時間的性質をとらえるため、各特徴の持続区間を考慮した特徴間の共起性、および、擬似的な相互情報量を求める。

#### 共起性

4.3 節で述べたように凝集性と一貫性の高いパターンではある要素に対し様々な特徴が呼応して現れる。このように、カテゴリ  $F_a$  の特徴  $f_i$  が生起したときに、それに対応するいずれかの特徴  $f_j$  が存在する確率を考えよう。特徴の生起時刻や持続区間を考慮に入れない場合には、これは以下のように表せる。

$$P(R_k(f_i, f_j) | f_i \in F_a) = \frac{N(R_k(f_i, f_j) \text{ となる } f_j \text{ が存在する } f_i)}{N(f_i \in F_a)} \quad (1)$$

ただし、 $f_i$  と  $f_j$  に対応関係  $R_k$  があることを  $R_k(f_i, f_j)$  と表し、 $N(\cdot)$  は条件を満たす事象の生起数を表す。

各々の特徴  $f_i$  が持続区間  $[t_i^s, t_i^e]$  を持つ、つまり、開始時刻  $t_i^s$  と異なる終了時刻  $t_i^e$  を持つことができる場合には、その対応関係  $C_k$  を以下のように定める。

$$C_k(f_i, f_j, \Delta t) = \begin{cases} 1 & (\Delta t \text{ のオフセットで共起している}) \\ 0 & (\text{それ以外}) \end{cases} \quad (2)$$

ここで、 $\Delta t$  のオフセットで共起しているとは、 $t_i^e + \Delta t$  ( $\Delta t > 0$  の場合)、または  $t_i^s + \Delta t$  ( $\Delta t < 0$  の場合) が区間  $[t_j^s, t_j^e]$  に含まれることと定義する。

$\Delta t = 0$  の場合には、以下の値とする。

$$C_k(f_i, f_j, 0) = \frac{\text{overlap}}{t_i^e - t_i^s} \quad (3)$$

ここで overlap は  $f_i$  の持続区間  $[t_i^s, t_i^e]$  と  $f_j$  の持続区間  $[t_j^s, t_j^e]$  との重なるの長さを表す。

次に、ある特徴カテゴリ  $F_a$  に属する特徴  $f_i$  が特徴カテゴリ  $F_b$  のいずれかの特徴  $f_j$  と対応関係を持つことを次のように表す。

$$\hat{C}_k(f_i, F_b, \Delta t) = \max_{f_j \in F_b} C_k(f_i, f_j, \Delta t) \quad (4)$$

これを以下のように  $F_a$  に属する特徴について平均する。

$$\tilde{C}_k(F_a, F_b, \Delta t) = \frac{\sum_{f_i \in F_a} \hat{C}_k(f_i, F_b, \Delta t)}{N(f_i \in F_a)} \quad (5)$$

このようにして得られる  $\tilde{C}_k(F_a, F_b, \Delta t)$  は、 $F_a$  に属する特徴の持続区間に  $\Delta t$  のオフセットを与えた時刻において、対応する特徴  $f_j \in F_b$  が存在する割合を表すものである\*2。ここで、 $F_a, F_b$  の組が 4.3 節で述べたパターンに含

\*2 本質的には式 (1) と同様の考え方であるが、ここで用いている値は確率とはいえない。対応する特徴が複数ある場合に式 (4) のように max をとる定義としたためである。

まれていれば、それらが分析対象とすべき重要なシーンを構成していることになる。

時間的性質の指標

共起性が高くて、それぞれの特徴の生起確率がつねに大きい場合には、特徴的なパターンとはいえない。また、逆に生起確率が低い場合に共起性が低いことに対しても同様のことがいえる。このようなことから、擬似的な相互情報量を頻出パターンの時間的性質の指標として用いることにした。具体的には、式 (1)~(5) で示した  $\tilde{C}_k(f_i, F_j, \Delta t)$  を用いて、以下の  $\tilde{I}_{\Delta t}(F_b; F_a)$  を計算する。

$$\tilde{I}_{\Delta t}(F_b; F_a) = \sum_{f_j \in F_a} \sum_{f_i \in F_b} \tilde{C}_k(f_i, f_j, \Delta t) \log \frac{\tilde{C}_k(f_i, f_j, \Delta t)}{P(f_i)P(f_j)} \quad (6)$$

ここで、 $P(f_i)$  は特徴単独の生起確率を表す。

$\tilde{I}_{\Delta t}(F_b; F_a)$  は、相互情報量の定義における  $P(f_i, f_j)$  の代わりに  $\tilde{C}_k(f_i, F_j, \Delta t)$  を用いたものであり、相互情報量の近似的な指標となっている。また、以下で述べる実験では、 $F_a = \{f_a, \bar{f}_a\}$ ,  $F_b = \{f_b, \bar{f}_b\}$  のように、特定の特徴が生起している場合とそうでない場合の二択を主に用いた。

6. データ収集と頻出パターンの検出

映像対話型の行動支援を用いて調理作業を行い、得られた記録から頻出パターンの検出を行った。

6.1 データ収集システムとタスク

システムは表 1 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。これは、支援者側の様子を映像で伝えることの重要性が低いためである\*3。

作業としては調理タスクを選んだ。作業者は筆者らの研究室に設置したシステムキッチンで調理を行い、支援者は他の部屋で映像を見ながらアドバイスを与える。作業にか

表 1 実験システムの設定

Table 1 Configuration of the experimental system.

作業者用システム	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	skype (約 10 fps)

\*3 このような非対称性については Billinghamurst ら [11] が解析している。支援者側から作業者側への資料提示などが有用なタスクである場合には、ヘッドマウントディスプレイなどへの表示を検討する必要がある。これは今後の課題としたい。

かる時間はおよそ 30 分である。

事前に作業者・支援者に作業内容に関するレシピを渡し、両者はそれを参照しながら作業を進めた。レシピの内容は作業者と支援者で異なっており、作業者には一般的なレシピと同様、簡便に作業手順のみを記したものを、支援者にはそれに加えて詳細な作業手順やコツが付加されたものを与えた。これにより、作業者が初心者である場合には、作業に関する情報が不足し、支援者には指示や助言などの情報を多く持たせることを狙った。

作業者と支援者の設定

作業スキルや立場の違いによる、コミュニケーションの違いを調べるために、表 2 のように、作業スキルと立場の異なる作業者と支援者のペアによる実験を行った。被験者ペアの A から D にはケーキの土台作りをタスクとした。被験者ペア E にはだし巻き卵の調理をタスクとした。

本実験では、様々な振舞いの違いが現れることが望ましいため、立場やスキルが異なる 6 人の被験者を様々な組み合わせることにより、表 2 のような特徴的なペアを構成した。そのため、同一人物が複数回の試行に関わっているが、その影響については以下のように考えられる。同一人物が作業者として複数回の試行に関わる場合には、必ず初見のタスクを実行しているため、記憶を頼りに作業を進めることができず、それぞれの試行において、初心者としての振舞いや情報要求がコミュニケーションに現れている。また、同一人物が支援者として複数回の試行に関わる場合には、支援者として上達した可能性があるが、知見をコミュニケーションを通してしか活用することしかできないため、違いがコミュニケーションの記録に現れることになる。ただし、支援者と作業者のペアはつねに異なる人物同

表 2 被験者の組合せ

Table 2 Skill and status of the each pairs.

ペア	作業者	支援者	見られた特徴
A	N1 (初心者) (学生)	M1 (熟練者) (スタッフ)	N1 が M1 の説明をよく聞き、さらに注意深く確認をしながら作業を進める
B	N2 (初心者) (教員)	M1 (熟練者) (スタッフ)	N2 が M1 の説明を聞きつつも、自分の判断を加えて作業を進める
C	M2 (熟練者) (スタッフ)	N2 (初心者) (教員)	M2 はあまり N2 の説明を聞かずに自分の判断で作業を進め、確認も求めない
D	M3 (熟練者) (教員)	M1 (熟練者) (スタッフ)	M3 は M1 に説明を求めないが、与えられた説明は聞き、時々確認を求めながら作業を進める
E	N3 (初心者) (学生)	N1 (初心者) (学生)	N1 と N3 で相談し、注意深く確認をしながら作業を進める

表 4 発話から抽出する特徴

Table 4 Features extracted from the utterances.

特徴	検査対象	条件	値	略号
具体物	名詞句	シソーラス [12] で具体物に分類されている	2 値 (偽, 真)	c
現場参照	指示詞※	現場参照を表す (※ 連体修飾を含む)	2 値 (真, 偽)	s
発話的役割	文	発話における役割	指示, 疑問, 説明, 応答	R, U, D, P

表 3 画像からの特徴抽出

Table 3 Features extracted from the videos.

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

士であるため、お互いに相手に合わせる必要があり、それによる振舞いの差が大きく、個別の被験者の習熟が議論できるほどデータは整理できていない。

被験者 (作業者および支援者) は 20~30 代の 6 名 (男性 5 名, 女性 1 名) である。ふだんからよく調理を行うもの 2 名と時々調理を行うもの 1 名を熟練者, ほとんど行わないもの 3 名を初心者とする。各ペアの振舞いには表の右項にあげたような特徴があったが, これは筆者らが要求したのではなく, 両名のスキルや年齢, 立場の関係から自然に現れたものである。

ペア A は支援者が作業者の目上であり, 支援者がそのタスクに習熟しているため, 作業者が支援者からの指示を待ってから行動する傾向が強く見られた。ペア B とペア C, ペア D は年齢が近い同僚であるため, 支援者が優位に立つ形態は見られず, スキルや性格の違いによる振舞いの違いが目立つ。ペア B では作業者は指示を待つ姿勢を持つが, 自身で判断できる場合は指示を待たずに行動することも多かった。ペア C では支援者が初心者であるため, 作業者が自身の判断で行動する余地が大きくなり, 指示を待たずに行動することが多くなった。また, どちらも作業に習熟しているペア D では, 直接作業と関係のない雑談が多く見られた。ペア E は同学年の学生同士で両者が初心者であったため, 気兼ねなく相談しながら作業を行っていた。ただし, 4.2 節で述べたように, 支援者が作業者の話題や作業状況に引きこまれて, 支援が必要なタイミングで行えないシーンが見られた。

## 6.2 分析対象とする特徴

本研究では, 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。ただし, 現在の段階では誤りのない自動抽出が望めないため, 人手で正解データを用意した。

### 画像特徴

画像特徴については表 3 に示したものをを用いた。視線停留

留, 見直し・視線移動, 移動は, カメラの動きから判断した。本研究では頭部装着カメラを用いているため, 見るための頭の動きがカメラの動きとして現れるためである。

### 発話特徴

発話特徴として, 表 4 に示した単語や名詞句などの特徴, 役割などを用いた。これらは, 発話文を書き起こしたものから, 発話文の形態素解析, 係り受け解析, シソーラスの参照などによって抽出し, それを人手で修正したものである [12], [13], [14]。人手での修正は, 主に定型的な応答を説明として識別された誤りに対して行った。これは, 話し言葉では“はい”というようにはっきりとした応答だけでなく, “あー”や“ええ”といった曖昧な応答が多く含まれる, また, それらが応答だけでなく, 次の発話へつなぐための間として使われる場合があり, 現在の処理では正しく識別がされなかったためである。そこで, 本実験では自動処理では正しく識別されなかったもののうち, 明確に応答であると判断可能なものについて識別結果を修正した。修正数は全データに対して 41 カ所であった。以上のように, 今回の実験ではより正確な解析結果を得る目的で人手による修正を行ったが, 修正内容は単一の識別結果の誤りに対してのみである。実用的なシステムを構築する際の自動処理の有効性は十分に期待できる。

以降では, これらの特徴に発話者を示す略号 W (Worker) と S (Supporter) を先頭に加えて, W:D (作業者:説明) や S:P (支援者:応答) のように表記する。

## 6.3 頻出パターンの抽出結果

収録されたデータのうち手動でタグを付与された映像データは, 平均 22 分 30 秒, 合計で約 112 分となった。出現した特徴の数を表 5 に示す。757 個の発話, 226 個の画像特徴が全体として現れた。

6.2 節にあげた特徴を対象に, それらが近接して出現する頻出パターンを 5.1 節で述べた PrefixSpan により求めた。ここで, PrefixSpan の minimum support を 0.1 に設定している。minimum support は, 系列パターンが頻出しているかどうかを決定するためのしきい値である。入力された系列数が  $N$  であり, そこからある系列パターン  $p$  が出現した数を  $N_p$  とする。このとき  $N_p/N$  が minimum support として設定した値以上であれば, その系列パターン  $p$  は頻出パターンとして抽出される。

表 5 特徴の出現数

Table 5 Number of features.

特徴	略号	生起数			特徴	略号	生起数
		合計	作業者	支援者			
物体名	c	108	66	42	視線停留	h	104
現場参照	s	79	61	18	見回し	l	65
指示	R	27	19	8	移動	m	57
質問	U	33	22	11			
説明	D	530	327	203			
応答	P	167	59	108			
発話数	-	757	427	330			

表 6 コミュニケーションと頻出パターンの対応関係例

Table 6 Example of the correspondence relationship between the communications and the frequent patterns.

Order	パターン	対応する場面	出現数
79	<S:D, W:P, S:D>	支援者：長めの説明や指示 作業者：相槌を打って説明の続きを促す	92
109	<V:h, S:D, W:D>	作業者：作業状況の報告 支援者：映像中の状況に対して指示や説明, 注意	56
119	<W:D, V:h, S:D, W:D>	作業者：作業状況の報告 支援者：映像中の状況に対して指示や説明, 注意	54
146	<S:D, V:l>	作業者：支援者からの説明を聞いて対象の確認 支援者：様々な指示や説明	46
172	<V:l, S:D, W:D, S:D>	支援者：映像中の状況について指示や説明 作業者：作業について質問や確認	42
177	<V:l, W:D, S:D, W:D>	作業者：映像と発話による作業状況の伝達 支援者：了解の応答, 注意	40
214	<S:D, W:D, V:m>	支援者：必要な道具の指示や説明 作業者：場所の確認, その他質問など	33

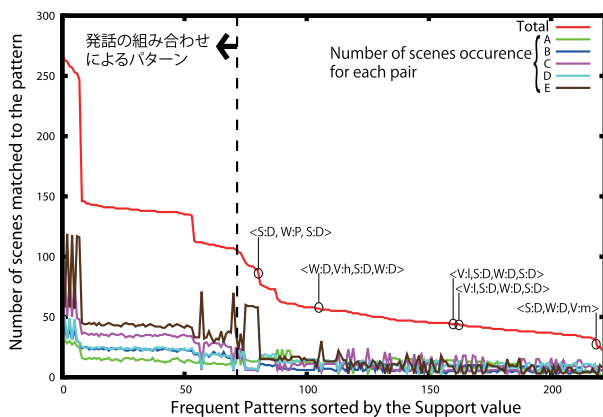


図 3 頻出パターンの出現数

Fig. 3 Number of frequent patterns.

検出された頻出パターンは 222 個あり、出現数が最小のもので 22 回、最大のものは 263 回出現していた。図 3 は抽出された頻出パターンに対応する場面の出現数を、出現数が多いものから並べたグラフを示す。出現数の多い順に 72 番目までのパターンは作業者と支援者の発話、<W:D> や <S:D> やそれらが複合したパターンのみであった。これは、他の特徴に比べ発話特徴の数が多いためである。76

表 7 パターン長ごとの頻出パターンの出現数

Table 7 Number of frequent patterns for each pattern length.

Length	Number
1	15
2	41
3	70
4	49
5	36
6	11

番目から表 6 にあげたような特徴的なパターンが見られる。画像と発話の両方の特徴を含むパターンが現れるのは 82 番目からであった。

検出されたパターンの系列長 (L) は L = 2 から L = 6 まであり、それぞれの検出数は表 7 に示すとおりである。系列長が L = 5, 6 のパターンは作業者と支援者の発話が連続するパターンのみであった。系列長が L = 3, 4 のものには、表 6 に示すような、作業支援の状況をよく表すパターンが見られた。たとえば、1 行目の <S:D, W:P, S:D> は支援者の説明を作業者が相槌を打ちながら聞く状況を表す。2 行目以降は映像通信の特徴が顕著に現れた例であり、対象

表 8 失敗を引き起こしやすい状況  
Table 8 Situations that easily cause the failure.

要因	状況例	状況が起こった理由
コミュニケーションにおける伝達情報の不足	作業者が道具の間違った使い方をした	支援者からの注意，作業者からの質問の両方がなかった
	作業者が道具を見つけられなかった	支援者が現場の状況を把握できず，適切な指示を与えられなかった
コミュニケーションのタイミングの悪さ	作業の途中でやり方を変える必要があった	支援者からの指示が遅れた
	作業者が質問できずに，間違った方法を選んだ	支援者は作業者が手順を聞きかたがっていることを認識できず，関係のない話を続けてしまった
単純な失敗	作業者が後ろの食器を倒した	周囲に対する注意不足など（コミュニケーションの問題には含まれない）

を見ながら説明や報告を行ったり，説明や指示の後に場所の移動が起こったりするパターンである．遠隔作業支援のためには，作業に必要な情報を作業者が得る（質問する）こと，支援に必要な情報を支援者が獲得すること，お互いに共通の対象に注意を払うこと，作業の状況や結果を相手に伝えることなどが重要となるが，これらのコミュニケーションが実際の頻出パターンとして検出されたことが分かる．

6.4 失敗を引き起こすコミュニケーションの例とその要因

実験中に失敗したり，失敗しかけた状況\*4は 41 例あった．その典型的な例を表 8 に示す．コミュニケーションの悪さに起因するものには，「伝達情報の不足」と「タイミングの悪さ」があった．「伝達情報の不足」に関しては，表 8 の 1 例目にあるように，支援者からの注意が明らかに不足していた場合もあるが，指示に対する相槌を省略したり，作業状況を確認するための発話がなかったりすることによるコミュニケーション不足が原因となった場合も多い．「タイミングの悪さ」に関しては，表 8 の 3 例目のように，単純に指示が遅れたものの他に，支援者の発話が続いたために作業者がタイミング良く質問できなかった例もある．

7. コミュニケーションの特徴と頻出パターンの性質

実験で得られた頻出パターンの性質と作業者・支援者の振舞いの関係を調査した．その基本的考え方は 4.3 節に述べたが，より詳細には以下のとおりである．

- (a) 頻出パターンが，必要な情報を伝達するために常用されるコミュニケーションパターン（以下本論文では，「基本パターン」と呼ぶ）となっている．
- (b) 「伝達情報の不足」は，(i) 基本パターンが成立しているにもかかわらず，それが伝える内容が不足している場合と，(ii) 基本パターンが成立していない，つまり，基本パターンから何らかの要素が欠けている場合がある．
- (c) 「タイミングの悪さ」には，(i) 基本パターンが成立し

ているが，一部の要素が時間的にずれている場合と，(ii) 要素の順番が変化して基本パターンが成立していない場合がある．

本研究で目的とするのは，深い意味的解釈を必要としない表層的なパターンによるコミュニケーション解析であるため，(b)(i)，(c)(ii) を直接扱うことは困難である．そのため，以下では，(b)(ii)，(c)(i) を対象とする．

7.1 コミュニケーションの特徴と頻出パターンの出現数

前節の (b)(ii) を以下のように評価する．まず，頻出パターン  $F_i$  から  $j$  番目の要素を取り除いたものを  $F_i^j$  とし， $F_i$  を上位のパターン， $F_i^j$  を下位のパターンとする． $N(F)$  をパターン  $F$  に対応する場面の出現数としたとき，上位と下位の出現数の比  $r(F_i, j) = N(F_i)/N(F_i^j)$  を求める．全ペアの総和についてこれを求めたものを  $r_{all}(F_i, j)$ ，それぞれのペア (A~E) について求めたものを  $r_A(F_i, j) \sim r_E(F_i, j)$  とする．なお，PrefixSpan の性質として， $F_i$  が抽出される部分に対して  $F_i^j$  も抽出されること，そのため  $r \leq 1$  となることに注意されたい．

$r$  の例を表 9 にあげる．ここで， $r_{all}$  と各ペアごとの比率  $r_A \sim r_E$  間の有意水準 5% で検定を行い\*5，有意差が認められたものには下線を引いてある．

表 9 の 1 行目は，支援者の説明の間に作業者の相槌 (W:P) が挟まれる場合に， $\langle S:D, W:P, S:D \rangle$  が  $\langle S:D, S:D \rangle$  と省略される場合の生起数の比を表している．ペア C で比の値が優位に小さく（相槌が省略され），逆にペア E では値が優位に大きくなっている（相槌が省略されない）．これは，表 2 で述べたペア C における M2 の振舞いや，ペア E で N1 と N3 がよく相談しながら作業を進めていることなどに符合する．

表 9 の 2 行目は，表 6 にあげたパターン ( $\langle V:l, W:D, S:D, W:D \rangle$ ) に関するものであり，作業者が見回した後に，作業側からの説明・報告が起こらなかった割合を示している．ペア間で有意差は認められないが，ペア A とペア E で他よりも小さな値となる傾向が見られた．これは，作業者

\*4 そのまま作業を続ければ確実に失敗を起こす状況であったが，作業者や支援者の機転により失敗を未然に防げた場合．

\*5 「各ペアの母比率に差はない」という帰無仮説において，正規近似により得られた有意確率に有意水準 5% で検定を行った．



表 9 全データとペアごとの  $r$  の値  
Table 9  $r$  values for the all data and the each pairs.

$N(F_i)/N(F_i^j)$	$r_{All}$	Pair				
		$r_A$	$r_B$	$r_C$	$r_D$	$r_E$
$N(<S:D, W:P, S:D>)/N(<S:D, S:D>)$	0.357	0.5	0.222	<u>0.12</u>	0.233	<u>0.496</u>
$N(<V:l, W:D, S:D, W:D>)/N(<V:l, S:D, W:D>)$	0.90	0.83	1.0	1.0	1.0	0.67
$N(<V:l, S:D, W:D>)/N(<S:D, W:D>)$	0.27	<u>0.75</u>	0.174	0.139	0.36	0.214
$N(<S:D, W:D, V:m>)/N(<S:D, W:D>)$	0.225	0.188	0.174	0.278	<u>0.44</u>	0.095

が見回した後、作業員からの説明や報告を待たずに支援者からの指示や説明が与えられている場合が他よりも多いことを表す。ここから、作業員が見回した意図を支援者が汲み取っていることが示唆される。

その他、有意差が認められた例として表 9 の 3, 4 行目がある。これらは、支援者からの指示・説明と作業員の説明・報告が続けて起こっている場合 (<S:D, W:D>) と作業員の見直し (V:l) や移動 (V:m) との共起を表す。ペア A では作業員の見直し後に支援者からの指示・説明と作業員の説明・報告が起こっている割合が有意に高いが、これは上述した 2 行目のパターンと同様の議論となる。ペア D では、支援者からの指示・説明と作業員の説明・報告が与えられた後に移動している割合が有意に高い。作業員と支援者の双方が熟練者であるため、次に使う物や手順についての説明が多く交わされていたためである。

以上のように、“応答の有無やその方法”のペア間の差が頻出パターンの出現数の差として現れていることが分かる。同様に、作業員からの状況説明の過剰が、作業員の行動と共起した会話の発生頻度の違いとして現れている。

### 7.2 コミュニケーションの特徴と頻出パターンの時間的性質

上述した (c)(i) について、5.2 節で定義した擬似相互情報量の時間的変化とコミュニケーションの特徴との関係について述べる。

図 4, 図 5 は、特徴  $F_a, F_b$  間の時間的性質  $I_{\Delta t}(F_b; F_a)$  をプロットしたものである。グラフの中央付近の「0」が  $F_a$  の生起時刻を表し、それより左は  $F_a$  の生起前、右は生起後を表す。縦軸が擬似相互情報量となる\*6。ここでは、表 6 にあげた頻出パターンに含まれる特徴の組合せのなかで、興味深い性質を示した特徴の組合せを選んでいる。

図 4 は支援者の説明の前後に作業員の見回しが起こるパターン (<S:D, V:l> と <V:l, S:D> を合算したもの) である。ここで他と異なる傾向が見られるのはペア A で、説明の 10 秒前および説明の 10 秒後に見回しが起こる確率が高くなっている。これは、作業員 N1 が支援者 M1 の説明を聞いた後、周囲を見回していること、逆に、N1 が周囲を見

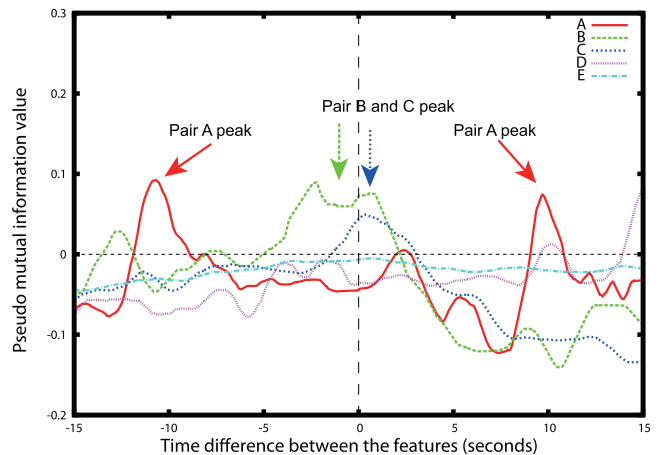


図 4 <S:D, V:l> の擬似相互情報量のグラフ

Fig. 4 Pseudo mutual information value graph of <S:D, V:l>.

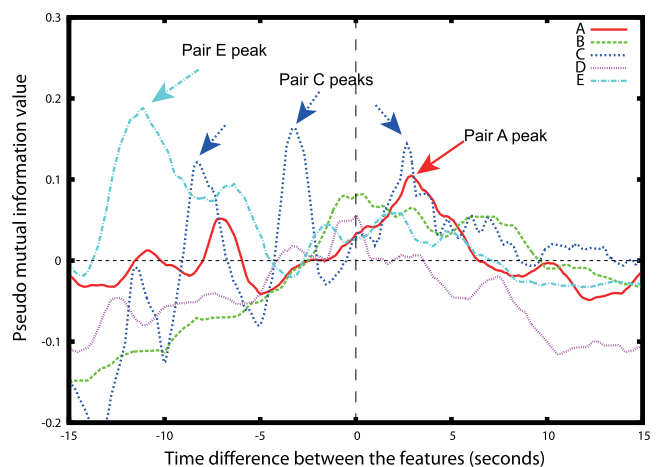


図 5 <V:h, (S:D, W:D)> の擬似相互情報量のグラフ

Fig. 5 Pseudo mutual information value graph of <V:h, (S:D, W:D)>.

回していることに M1 が気づいて、見回しが収まった後に説明を与えていることの結果である。これは丁寧なコミュニケーションが行われていることを示唆している。それに対し、ペア B では説明と見回しがほぼ同時に起こる傾向が強いことが分かる。作業員 N2 が支援者 M1 の説明が終わるのを待たずに見直し始めたり、M1 が N2 の見直し中に説明を与える\*7など、意志の疎通がちぐはぐな状態があることを示唆している。ペア C でも同様の傾向が少し見られ

\*6 擬似相互情報量は確率の近似的な値を用いているため、負の値をとることがある。

\*7 見回している途中では対象をしっかりと視認できていないことが多く、その結果、説明を十分に理解できていないことがある。

る。それに対し、ペア D、ペア E では目立った傾向は見られない。

図 5 は作業者の視線が停留した前後で、支援者の説明と作業者の報告が行われるパターン (<V:h, (S:D, W:D)> と <(S:D, W:D), V:h> を合算したもの)の時間的性質である。このグラフから、ペア A は視線が停留してから 5 秒程度後に報告や説明が行われる傾向が分かる。逆に、ペア E では、説明や報告があってから視線の停留が起こる傾向が顕著に現れている。ペア A では作業者 N1 の行動に合わせて支援者 M1 が指示を与えているのに対し、ペア E では作業者 N3 と支援者 N1 が相談し、N1 の指示に従って N3 が行動していることがこれらの傾向の要因になっている。それらに対し、C のペアでは視線停留の前に 2 カ所、視線停留の直後に 1 カ所のピークが見られる。これは、作業者 M2 が目にとまったものに関する発話を一方的に始めること、支援者 N2 が M2 の作業状況を正しく把握できず、場当たり的に必要な道具や手順などを話すことに起因する。ペア B、ペア D では目立った特徴は現れていないが、上記の中ではペア A の傾向に近い。

以上のように、同じパターンであってもペアによってその行動と発話の発生タイミングが異なっており、特にピークの有無として明確に表れている。

### 7.3 コミュニケーションの特徴と失敗例の関係

各ペアが失敗または失敗しかけたシーンを人手で分類し、その分類ごとに発生数をまとめたものを表 10 に示す。失敗または失敗しかけた数がペアによって異なることが分かる。

ペア B、ペア C では、伝達情報の不足に起因する失敗または失敗しかけた数が多い。7.1, 7.2 節で述べてきたように、ペア B、ペア C では他と比較して応答(相槌など)の不足、見直しや視線の停留と説明・報告などのタイミングがちぐはくであり、これが頻出パターンの出現数や時間的性質にも定量的に現れている。ただし、ペア B の作業者が初心者であるため、意志の疎通の齟齬がそのまま失敗につながるのに対し、ペア C では必ずしも失敗としては顕在化していない。

ペア D では、タイミングが悪いことによる失敗または失敗しかけた例が目立つ。D では、熟練者が自分のペースで

表 10 失敗や失敗を引き起こしやすいコミュニケーションとなる主な要因

Table 10 Major communication factors that cause or easily raise a failure situation.

失敗の主要因	A	B	C	D	E	合計
伝達情報の不足	2	10	6	1	1	20
タイミングの悪さ	2	2	1	6	1	12
単純な失敗		6	2		1	9
合計	4	18	9	7	3	41

行動するため、支援者もタイミング良く指示や説明を与えることが難しかったことが原因である。このような特徴に対し、図 4、図 5 の両方の例で作業者の説明・報告のピークが現れないことから、「発話」と「見る行為」が同期していないことが示されている。そのため、支援者がタイミング良く必要な情報を与えることが難しくなっていることが示唆されている。

ペア A、ペア E では、7.1, 7.2 節で述べたように、注意深くコミュニケーションを行っていることが定量的な特徴にも現れており、その結果、作業者が初心者であるにもかかわらず、失敗が少ない。失敗したケースとしては、支援者が何を指示すべきかを思いつかなかつたために、適切な指示を与えられなかった例があった。具体的には、支援者が先に大まかな手順を伝えてあり、作業者がそれに従って作業をしている場面で、支援者の作業状況の把握が遅れて事前の説明で省略していた詳細な手順の説明が遅れてしまう例があった。

以上のように、失敗を引き起こしやすい被験者ペアの振舞いが頻出パターンの出現数や時間的性質として現れることを確認できた。ただし、それぞれの指標がどのような失敗の主要因となっているかなど、種々の因果関係についてはまだ十分なデータに基づく解析を行っていない。十分なデータの収集に基づくさらなる解析が今後の課題となっている。

### 7.4 今後の課題

頻出パターンの出現数、時間的性質に各ペアの特徴が現れること、また、それによって失敗を引き起こす原因が示唆されることなどを述べてきた。現在はコミュニケーションで伝わる特徴について人手で正解を与えているが、ここで用いている特徴は、完全な精度を要求しなければ現在でも自動的に抽出可能なものである。今後、このような処理の自動化が期待できる。ただし、本研究ではまだ以下のような問題には取り組めておらず、今後のいっそうの研究を必要としている。

まず、注目すべき頻出パターンや、そこから注目すべき性質を選択する指標が必要である。すなわち、7.1, 7.2 節では、分かりやすい性質を示す頻出パターンを筆者らが選んだが、これを自動化する方法論は明らかではない。頻出パターンの性質を網羅的に自動で調べ、重要な性質や特異な性質が現れたものをユーザに示すなどの方法が考えられるが、重要な性質や特異な性質の定義が明らかでないため、今後、多くの試行錯誤が必要とされる。

また、データ数が十分でない場合には、頻出パターンだけではコミュニケーションの性質を十分に把握できない。たとえば、本研究の実験でも、質問(Q)は出現数が少ないため、それを含むパターンが頻出パターンとして検出されなかった。このようなパターンの扱いは今後の課題で

ある。これには、大量のデータに対して抽出された頻出パターンを記録しておき、少量しかないデータの解析にも用いるなどの手法が考えられる。

さらに、本研究では表層的な特徴のみを用いることを目指したが、深い意味的な特徴によって初めて明らかになる性質に関する調査を行っていない。この検討を行い、本手法の有効性とその限界をより明確にすることも重要な課題である。

## 8. おわりに

本論文では、映像対話型行動支援で発生するコミュニケーションやその場の状況を記録し、それを分析する手法について述べた。映像対話型行動支援ではコミュニケーションに使えるモダリティが制限されているため、少数のモダリティの相互関係を分析するだけで、作業・支援者の振舞いの傾向をよくとらえることができることを実際の例で示した。これらは作業の円滑さや失敗などに深く関わるため、このような分析が作業の分析や訓練に応用されることが期待できる。現在は、正確な解析を行うために、データ入力を人手で行っているが、ある程度の誤りが許される場合には、既存の技術を用いた自動化も可能であり、さらに、その精度も今後改善されていくことが予想される。将来的に、コミュニケーション状況をシステムが自動で認識し、支援者や作業者に補助を行うことも興味深い課題となっている。

## 参考文献

- [1] Garner, P., Collins, M., Webster, S. and Rose, D.: The application of telepresence in medicine, *BT Technolo. J.*, Vol.15, No.4, pp.181–187 (1997).
- [2] UMET, available from (<http://www.tanizawa.co.jp/umet/>) (accessed 2014-06-01).
- [3] Siegel, J., Kraut, R., John, B. and Carley, K.: An Empirical Study of Collaborative Wearable Computer Systems, *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 1995)*, pp.312–313 (1995).
- [4] Mann, S.: Smart Clothing: Wearable Multimedia Computing and Personal Imaging to Restore the Technological Balance Between People and Their Environments, *Proc. ACM Conference on Multimedia*, pp.163–174 (1996).
- [5] Kraut, R., Miller, M. and Siegel, J.: Collaboration in performance of physical tasks: Effects on outcomes and communication, *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 1996)* (1996).
- [6] Fussell, S., Kraut, R. and Siegel, J.: Coordination of communication: Effects of shared visual context on collaborative work, *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 2000)* (2000).
- [7] Kraut, R., Gergle, D. and Fussell, S.: The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence, *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 2002)* (2002).
- [8] Gergle, D., Kraut, R. and Fussell, S.: Action as language

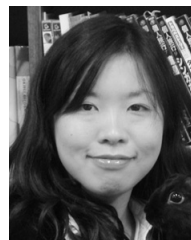
in a shared visual space, *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 2004)* (2004).

- [9] 畑村創造工学研究所, 失敗知識データベース, 入手先 (<http://www.sozogaku.com/fkd/index.html>) (参照 2014-03-28).
- [10] Pei, J., Han, J., Mortazavi-asl, B., Pinto, H., Chen, Q., Dayal, U. and Hsu, M.-C.: PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth, *17th International Conference on Data Engineering (ICDE '01)*, pp.215–224 (2001).
- [11] Billinghamurst, M., Bee, S., Bowskill, J. and Kato, H.: Asymmetries in Collaborative Wearable Interface, *The 3rd International Symposium on Wearable Computers*, pp.133–140 (1999).
- [12] 国立国語研究所 (編): 分類語彙表増補改訂版, 大日本図書 (2004).
- [13] 黒橋慎夫, 河原大輔: 日本語形態素解析システム JUMAN version 5.1 使用説明書, 京都大学 (2005).
- [14] 黒橋慎夫, 河原大輔: 日本語構文解析システム KNP version 2.0 使用説明書, 京都大学 (2005).



小泉 敬寛 (正会員)

2005年3月筑波大学大学院工学研究科知能機能システム専攻博士課程修士号取得・退学。同年4月京都大学大学院工学研究科電気工学専攻博士後期課程入学。2007年4月同大学院退学。同年5月京都大学工学部助教となり現在に至る。修士(工学)。画像・映像処理, ライフログ映像検索等の研究に従事。電子情報通信学会会員。



小幡 佳奈子

2004年大阪府立大学経済学部経営学科卒業。同年8月より京都大学学術情報メディアセンター教務補佐員となり現在に至る。ヒューマンコミュニケーションと料理に興味を持つ。



渡辺 靖彦 (正会員)

1991年京都大学工学部電気工学第二学科卒業。1995年同大学大学院博士課程退学。博士(情報学)。龍谷大学理工学部助手を経て, 2002年より龍谷大学理工学部情報メディア学科専任講師, 現在に至る。自然言語処理, 知識情報処理の研究に従事。



近藤 一晃 (正会員)

2004年大阪大学大学院基礎工学研究科システム人間系専攻博士前期課程修了。2007年同大学院情報科学研究科コンピュータサイエンス専攻博士後期課程修了。同年同大学産業科学研究所特任研究員。2009年京都大学学術情報メディアセンター助教就任後現在に至る。反射屈折光学系，知能ロボット，マンマシンインタラクション，知的行動支援に関する研究に従事。博士（情報科学）。電子情報通信学会会員。



中村 裕一 (正会員)

1985年京都大学工学部電気工学第二学科卒業。1990年同大学大学院博士課程修了。同年京都大学工学部助手。1993年筑波大学電子・情報工学系講師。1999年機能工学系助教授，2004年京都大学学術情報メディアセンター教授，現在に至る。博士（工学）。画像処理・認識，映像処理，ヒューマンコンピュータインタラクション，自然言語処理等の研究に従事。1996年カーネギーメロン大学ロボティクス研究所客員研究員。1998～2001年科学技術振興事業団さきがけ21研究「情報と知」領域研究員（兼任）。電子情報通信学会フェロー，人工知能学会，ACM，IEEE各会員。