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

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 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」と呼び，その映像やその他のデータを記録したものを「映像対話型行動記録」と呼ぶ．我々は，このような場におけるコミュニケーションを統計的に解析する手法を提案してきたが，本稿では，熟練度等の条件の違いによる作業者や支援者の態度の違いをこのような手法によって分析する方法とその結果について紹介する．さらに，この結果と作業の円滑さを吟味し，遠隔対話型行動支援の改善や補助を行なうための良い基礎データが得られることを示す．

キーワード　映像対話型行動記録，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，個人視点映像

Behavior analysis of worker and supporter in working support through FPV communication

Takahiro KOIZUMI†, Kanako OBATA†, Yasuhiko WATANABE††, Kazuaki KONDO†, and Yuichi NAKAMURA†

† 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. FPVC record is a recorded video of an actual FPV in this situation. In this paper, we report the important characteristics concerning their skills and attitudes can be quantified by the statistical methods that we have proposed for analyzing the behaviors of workers and supporters behaviors. Then, we examine the relationship between the results of the above analyses and the failure or accidents, and consider the possibility of supporting a work with FPVC.

Key words　First Person Vision Communication, Interaction analysis, Multimodal comunication, First person vision video

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，指示等を行う．遠隔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，例えば，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等に活用できる．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一層進み，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，本研究では，作業者と支援者の振る舞いのパターンと，作業者と支援者の心理的な態度，意思疎通の円滑さ，また，それらと作業の成否等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．そのためのノンバーバル情報としては，作業動作や見る・捜す (探す) 行動などを重視し，頻出パターンの抽出，情報量などを用いて，定量的な解析を行った．その結果，作業者や支援者の態度，意思疎通の状

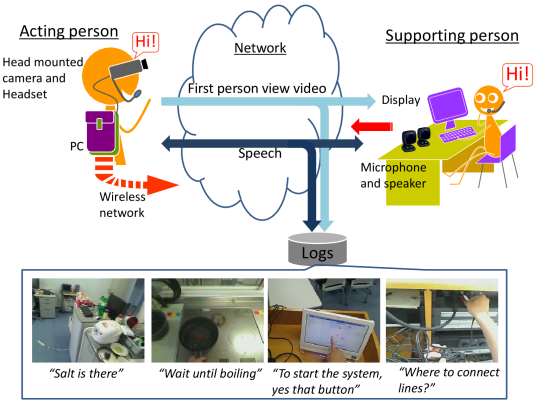


図 1　Overview of working log with FPV communication

況などの差が量として現れることを確認した．

以下本稿では，映像対話型行動支援におけるマルチモーダルな振る舞いの定量的な解析の考え方とそこから得られた知見について紹介する．

2.　問題設定: 何を分析するのか?

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を解析して，以下の項目に関する知見を得ることを目的とする．

(1)　作業者と支援者の振る舞いや態度の記述と定量化・類型化

(2)　作業者と支援者の振る舞いや態度と作業の円滑さや失敗との関係の解析

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る舞いが重要な意味を持つ．つまり，複数のモダリティを活用して意志の疎通が行われるため，上記 (1) として，マルチモーダルな面からの調査が重要な課題となる．従来から，例えば会話分析 [3] [4] 等でも，言語外情報の記述が行われているが，それに比べ，映像対話型行動支援には以下のような特徴がある．

映像対話型遠隔行動支援では，同じ空間に作業者と支援者がいる場合に比べて，コニュニケーションのチャネルが制限され，意思疎通が難しくなる．具体的には，嗅覚や力覚が使えないのに加え，映像通信においても，視点位置の制約，視野の狭さ，環境音の取得など，種々の問題が生じる．逆に，コミュニケーションを解析する場合には，チャネルが制限されているため，それを集中的に調べれば良いという利点がある．

そのため，本研究で重視したことは，コミュニケーションの意味的構造，時間的構造を調査するための**定量的な解析**を行うことである．コミュニケーションの典型的なパターンの洗い出しには時系列パターンのマイニング手法 (PrefixSpan) [5] を用い，さらにそれらのパターン中の各要素の共起性や相互情報量などを時間を基準にして調べる [6]．このような手法を用いることにより，特徴の網羅的な調査や種々の条件の比較が可能になることが期待できる．

これらの手法を用いて，本研究では，上記 (2) であげたように，作業者と支援者の行動やコミュニケーションのパターンと作業の状況やその失敗との関係を調査した．意志の疎通が作業の円滑さや成否に密接に関わることは明白であるが，実際に上記 (1) で調べられるパターンやその特徴にどのように現れるかを定性的な特徴だけでなく，定量的な特徴に基づいて解析することが目的である．ただし，意志の疎通がうまく行かない場合には，支援者から与えられる指示が不足している場合，作業者の注意が不足している場合等，様々な状況が考えられる．また，作業者，支援者それぞれの経験・知識や個人的な傾向も深く関わる．このように，コミュニケーションの分析には多くの難しい問題が含まれるが，本研究では，意味的に深く解析することよりも，表面的に観測できる作業者・支援者の振る舞いと作業の円滑さとの関係を調査することに重点を置いた．

このような性質が明らかになれば，その知見を，映像対話型行動支援におけるコミュニケーションスキルを評価する基礎として用いることができる．例えば，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することが考えられる．また，コミュニケーションの評価をリアルタイムに行なって，評価をフィードバックすることにより，作業支援の補助をすることも考えられる．さらに，支援者の代わりに，エージェントが支援するような設定における，エージェントの振る舞いを設計する指針にもなる．

3.　行動とコミュニケーションの記述

3.1　モダリティ (特徴) 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 (特徴) 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 (手の動き等)，移動，その他

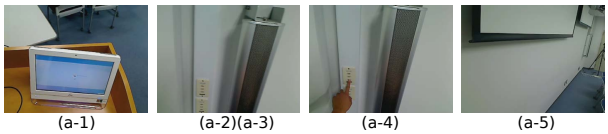
物体・環境:　物体や環境の外観，外観の変化，位置関係，その他

「発話」以外の 3 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，敢えて別モダリティとしている．また,「環境の状態」が対人コミュニケーションにおけるモダリティとして考えられることは稀であるが，映像対話型の環境では，それが作業者と支援者の間の重要な情報共有・情報交換の対象となっていることから，ここでは一つのモダリティとした．

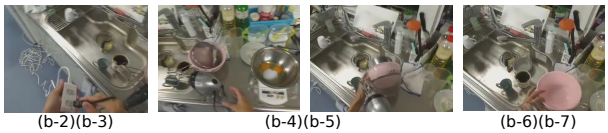
3.2　典型的なシーンとインタラクションのパターン

映像対話型行動支援ではこれらの特徴が複合して現れる．そ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) は円滑にコミュニケーションと作業が進んでいる理想的な例である．作業者と支援者の間で映像と発話をうまく使うことにより，円滑に情報交換が進められた結果である．


(a-1) (a-2)(a-3) (a-4) (a-5)

(支援者) (a-1) 「スクリーンを降ろしてください」
(作業者) (a-2) (見廻しながら)「スイッチはどこですか？」
(支援者) (a-3) 「その赤いボタンがスイッチです」
(作業者) (a-4) (ボタンを直視＋指さし)「あ，これですね」
(作業者) (a-5) (スクリーンを直視)「はい，スクリーンが降りてきました」


(b-2)(b-3) (b-4)(b-5) (b-6)(b-7)

(支援者) (b-1) 「それをミキサーにつないで下さい」
(作業者) (b-2) (コンセントを見て)「ここに？」
(支援者) (b-3) 「そうそう」
(作業者) (b-4) (ミキサーで道具を倒す)「あーーっ！」
(支援者) (b-5) 「えっ，どうしたの？」
(作業者) (b-6) (落ちた道具を見て)「大惨事！」
(支援者) (b-7) 「えっ，何が起こったの？」


(c-1) (c-2)(c-3) (c-4)(c-5)

(支援者) (c-1) 「それを 2 層にスライスして下さい」
(作業者) (c-2) (上から直視)「あっ，ななめに焼けてしもたー」
(支援者) (c-3) 「そう，ですねぇ...」
(作業者) (c-4) (間違った切り方をしながら)「ひどい出来だなぁ」
(支援者) (c-5) 「あ，回しながら切るとうまく行くんだけど」

図 2　映像対話型行動記録中の典型的なシーン

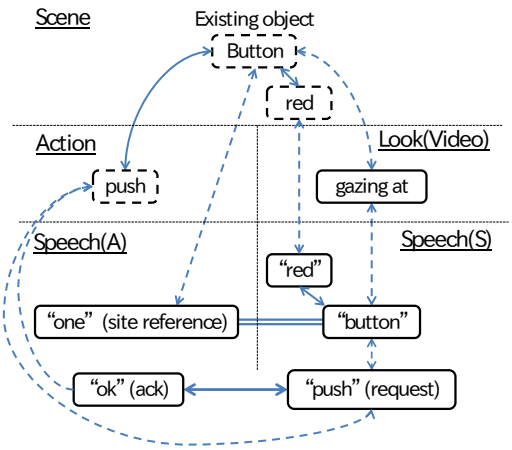


図 3　Intermodal relationships

図 3 に図 2 の (a) における要素間の相互関係を示す．例えば，(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[(注1)]．これは，映像を介して支援者にも伝わる．また，行った指さし (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，複数のモダリティが密接な対応関係や依存関係を持ち，それぞれの発話や行為が適切なタイミングで起こっていること，それにより，物，行為，意図等に関する情報が正確に伝わっていることがわかる．

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) で状況がうまく伝わっていない．作業者が支援者に状況を伝えるためにモダリティをうまく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よく与えられなかったために作業がうまく行かなかった例である．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．

3.3　凝集性と頻出パターン

物，行為，意図等に関する情報が複数のモダリティに跨って現れたり，相互に参照されることがコミュニケーションの重要な構造となっている．

その最も基本的なパターンとしては，あるモダリティにおける要素が他のモダリティ中の要素を参照する場合や，一つの現象や行動が複数のモダリティに現れる場合等がある．依頼と結果等の因果関係もよく現れる．典型的な例としては，支援者が作業を指示・依頼することによるその後の作業者の行動や状況の変化，また，完了の報告などの一連の関係があげられる．図 2(a) の例では，(a-1) の依頼によって (a-2) 以降の発話や行動が引き起こされ，(a-4) の報告まで一連の関係が続いている．さらに上記 2 種類の関係に比べて出現が少なく，その定義も曖昧なものとなるが，質問とそれに対する説明の関係も現れる．典型的な例としては，支援者からの質問に対して作業者が周囲の状況を確認して返答する場面があげられる．図 2(a) の例を考えると，(a-2) で作業者が質問をするが，その前後にスイッチを探す動作が現れ，(a-3) で支援者から返答を受けた時に注目行動や指示動作が現れる．

図 2(a) の例だけでなく，(b)，(c) の例でも，複数のモダリティが並行して用いられ，それが作業者と支援者の相互理解や作業の成否と強く影響している．このような共起を本研究では「モダリティ間の**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る．

ただし，(a) においては相互関係や行動が一貫性 (coherency) を持っているのに対して，(b), (c) では一貫していない (incoherent な) 部分がある．そのため，(a) は Norris の定義する modal density [7] [8] が高い状態であると言えるが，(b), (c) に

(注1)：対象が作業者の視界に入っていても，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．

ついてはそれが成り立たない(注2).

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，以下のような仮定を置く.

- 凝集性も一貫性も高いパターンは意志の疎通が良好な状況で起きる．通常は良好な状況が大半を占めるため，このようなパターンが頻出する.
- 凝集性が高く一貫性が低い場合には，両者の注意が異なる対象に向けられている．そのため，意志の疎通が不足していたり，齟齬が起こりやすい.

このようなことから，凝集性の高い部分に注目すること，および，頻出するパターンを抽出することが重要な分析手段となる.

4. 頻出パターンの分析

4.1 頻出パターンの抽出

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良く用いられる PrefixSpan [5] を用いたが，そのために以下のような設定を用いた.

- 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全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 (例えば，$S = \{F_1, F_2, ..., (F_i, ..., F_m), ..., F_z\}$) とする．ただし，本研究では，各アイテム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が，比較的簡単にその意味がわかる．例えば，$L_1 =<$ (発話：依頼, 発話：行動), 移動, 作業 $>$ は，支援者からの発話による行動依頼があって，作業者が移動し，作業が行われたことを意味する(注3).

4.2 時間的性質

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ている．より詳細な時間的構造をとらえるため，本研究では，生起時刻の差をパラメータとした特徴の共起性，および，擬似的な相互情報量を求めることにした.

共 起 性

あるカテゴリ F_a の特徴 f_i が生起したときに，それに対して 3.3 で述べたような対応関係のいずれかを持つ 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.

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

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.

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

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

$\triangle t = 0$ の場合には，以下の値とする.

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

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 $[t_i^s, t_i^e]$ との重なりの長さを表す.

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.

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

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.

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

このようにして得られる $\tilde{C}_k(F_a, F_b, \triangle t)$ は，F_a に属する特徴の生起区間に $\triangle t$ 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(注4)．ここで，F_a，F_b の組が 3.3 で述べたパターンの一つであれば，それらが 3. で述べた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

共起性が高くても，それぞれの特徴の生起確率が常に大きい場合には，特徴的なパターンとは言えない．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言える.

(注2)：modal density はモダリティ間で関連しあっていること (cohesion) と一貫していること (coherence) を含む．本研究では，集中していることに注目し，これを「凝集性」と表現することにする

(注3)：なお，PrefixSpan では，系列の中に他のアイテム (本研究では特徴) が含まれているものも同一の頻出パターンの出現数として数え上げられるため，例えば，$L_2 =<$ (発話：依頼, 発話：行動), 発話：応答, 移動, 作業 $>$ は，L_1 の一つの派生パターンとなる.

(注4)：本質的には (1) 式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言えない．対応する特徴が複数ある場合に (4) 式のように max をとる定義としたためである.

表 1　頻出パターンと意志の疎通

	頻出パターン	コミュニケーションの要因
(1)	< 発話 (質問)，手を使った作業 >	作業と関連する質問
(2)	< 発話 (説明)，映像 (見まわし)>	説明と見回し (確認)
(3)	< 発話 (指示)，映像 (移動)>	作業指示と場所移動
(4)	< 発話 (物体名)，映像 (移動)>	具体物の名と場所移動

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，(1)〜(5) 式で示した $\tilde{C}_k(f_i, F_j, \triangle t)$ を用いて，以下の $\tilde{I}_{\triangle t}(F_b; F_a)$ を計算する．

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

ここで，$P(f_i)$ は特徴単独の生起確率を表す．

$\tilde{I}_{\triangle t}(F_b; F_a)$ は，相互情報量の定義における $P(f_i, f_j)$ の代わりに $\tilde{C}_k(f_i, F_j, \triangle t)$ 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，$F_a = \{f_a, \overline{f_a}\}$，$F_b = \{f_b, \overline{f_b}\}$ 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．

4.3　頻出パターンと作業者・支援者の態度

作業者と支援者の意志の疎通が円滑に行われている場合には作業も円滑に進むことが期待できる．このような意志の疎通には様々な要因が考えられるが，重要なものには，作業に必要な情報を作業者が質問しているか，支援に必要な情報を支援者が獲得できているか，お互いに共通の対象に注意を払っているか，作業の状況や結果を相手に伝えているか等があげられる．

上記の頻出パターンからこれらの要因に強く関連するものを選ぶことができる．数例を表 1 に示す．例えば，表の (1) は，作業とそれに関連する質問の対である．このパターンにより，作業者が作業の際にどのような頻度で質問するか，また，作業に先立って質問する傾向があるか，作業を始めてから質問する傾向があるか等を確かめることができる．同様に，(2) は作業者が説明を受けた場合に見回し確認を行うパターンであり，どの程度，また，どのタイミングで説明された対象に注意を向けたかを確認できる．

このように，頻出パターンの頻度，また，その時間的構成を調べることにより，種々の傾向がわかる．ただし，このような傾向の絶対的な評価を行うための知見はまだ存在しないため，本研究では，条件の異なる試行を行なって，その傾向を比較することにより，コミュニケーションと作業の円滑さとの関係を探ることにした．これらの条件には，作業者・支援者の作業スキル，性格などの個人差，作業環境等が考えられる．また，作業が円滑に進んでいる場合とそうでない場合，意志の疎通がうまくいっている場合と齟齬がある場合などを比較することにより，コミュニケーションのパターンの違いが及ぼす影響などを調べることができる．

5.　実験: データ収集と解析

映像対話型行動支援システムを用いて調理作業を行ったデータを収集し，種々の性質を確認した．

表 2　実験システムの設定

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 10fps)

表 3　画像からの特徴抽出

特徴	備考 (検出条件)	略号
物体	物体領域	V:O
手を用いた作業	手 (肌色領域) が動き続けている	V:W
指示	手 (肌色領域) の動きが止っている	V:p
視線停留	カメラモーションが小さな状態が続く	V:h
見回し	カメラモーションが見回す動きを示す	V:l
移動	カメラモーションが前進していることを示す	V:m

5.1　データ収集

システムとタスク

システムは表 2 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した．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 (注5)．

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．作業にかかる時間はおよそ 30 分である．

被験者の設定

被験者 (作業者および支援者) は 20 代〜30 代の 5 名 (男性 4 名，女性 1 名) で，普段から良く調理を行うもの 2 名，時々調理を行うもの 1 名，ほとんど行わないもの 2 名である．また，その社会的立場は，教員 (2 名)，研究員 (2 名)，学生 (1 名) である．以下，それぞれを，KO(中上級者，研究員)，KK(中上級者，教員)，MY(中級者，研究員)，TK(初心者，教員)，KM(初心者，学生) と記号で呼ぶことにする．今回の実験で収録した (作業者，支援者) の組み合わせは，(KK, KO), (TK, KO), (KM, KO), (MY, TK) である．まだ網羅的にデータを収録できていないが，収録されたデータは，それぞれ作業スキルや立場が異なった組み合わせとなっている．

5.2　利 用 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 に示したものを用いた．手による作

(注5)：このような非対称性については Billinghurst [9] らが解析している．支援者側から作業者側への資料提示等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．

表 5　Number of occrrences (feature)

特徴	略号	生起数
発話	-	694
物体名	S:c	100
現場参照	S:s	94
要求	S:R	17
説明	S:D	561
応答	S:T	70
質問	S:Q	41

特徴	略号	生起数
物体領域	V:O	257
手を用いた作業	V:W	159
視線停留	V:h	79
見回し	V:l	47
移動	V:m	46

表 6　Frequent patterns (result of PrefixSpan)

Order	Pattern	Number
1	(S:D)	387
2	(S:D)(S:D)	383
3	(S:D)(S:AD)	379
4	(S:D)(S:D)(S:D)	379
	...	
231	(S:D)(V:m)	318
232	(S:D)(V:m)(S:D)	318
	...	
368	(S:D)(V:l)	278
369	(S:D)(V:l)(V:h)	278
	...	
1062	(V:h)	80
1063	(V:h)(S:D)	80
	...	
1438	(S:c)(V:m)	67
1439	(S:D S:c)(S:D)(S:c)(S:AD)	67
	...	
1701	(V:l)(S:D)	50

業や現場指示の検出のためには手領域の位置と動きを特徴とする．視線停留，見回し・視線移動，移動を検出するための手がかりとしては，カメラの動きが利用できる．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ことを利用する．なお，物体領域に関しては，発話で参照されているものだけを対象とする．これは，画像中に膨大な数の物体領域 (注6) があり，それらを網羅的に抽出することが困難なためである．

発 話 特 徴

発話特徴として，表 4 に示した単語や名詞句等の特徴，文のタイプ，役割等を用いた．これらは，発話文を書き起こしたものから，音声認識，発話文の形態素解析，係り受け解析，シソーラスの参照等によって抽出し，それを人手で修正したものである [10]〜[12].

5.3　実 験 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 5 分 24 秒，合計で約 21 分となった．出現した特徴の数を表 5 に示す．694 個の発話，588 個の画像特徴などが全体として現れたが，一分あたりでは，20〜30 個程度となる．

頻出パターン

5.2 に挙げた特徴を対象に，それらが近接して出現する頻出パターンを 4.2 で述べた方法により求めた結果を表 6 に示す (注7)．ただし，この検出では PrefixSpan の minimum support を 0.3 に設定した．図 4 に頻出パターンの順位と出現数のグラフを示す．横軸が出現回数の多いものから並べた順位を表し，縦軸が出現数を表す．多くの頻出パターンが検出されているが，全体で 50 回以上現れたのは約 1700 パターンであった．

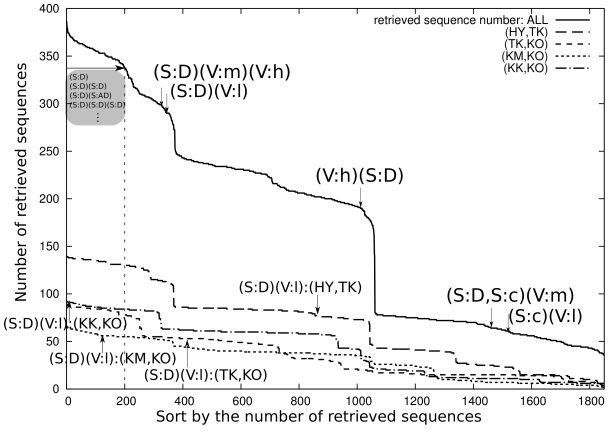


図 4　Number of occurrences (frequent patterns)

このうち，200 番台ぐらいまでの頻出パターンは，< S:D, S:D, S:D > のように，発言がいくつか続いたというだけのパターンがほとんどであり，特に重要な意味のあるものではない．300 番台ぐらいから画像特徴を含んだ興味深いパターンが多く現れる．例えば，348 番の < S:D, V:m, V:h > は説明の発話の後に移動と移動先での視線の停留が起きている場合，1483 番に現れる < (S:D,S:c),V:m > は物体名を参照することと，それについての説明が一発話の中で起こり，その後場所の移動が現れる場合，1510 番の < S:c, V:l > は物体名を参照し，その後見回し行動を行う場合である．

頻出パターンの時間的性質

4.3 節で述べたように，作業の円滑さと意志の疎通に関わる頻出パターンを作業者・支援者が異なる条件で比較した．図 5〜図 7 のグラフは，頻出パターン $I_{\triangle t}(F_b; F_a)$ をプロットしたものである．この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，右は生起後に F_b が生起したことを表す．

図 5 は表 1(1) にあげた < S:Q, V:W > (作業と関連する質問) の時間的性質をプロットしたものであり，4 組の作業者・支援者の間の違いが顕著に見られるのがわかる．(TK, KO), (KK, KO) の組は質問しながら作業する傾向が強く，また支援者に依存する傾向は (TK, KO) の方が，時間的にも生起確率的にも高いことが現われている．それに対して，質問する傾向が低いのが (HY, TK) の組である．事前にはあまり質問せず，事後の質問も平均すると高くない．最も注意深く作業を進めている (KM, KO) 組は作業の後に確認の質問をする傾向が強く見られる．

(注6)：さらに，一つの領域でも定義によって多数の見方がある

(注7)：表 6 中の < S:AD > は < S:D > と同じ説明の発話であるが，何か行動を示す役割を持つ発話であることを表している．頻出パターンの計算で用いる際には，これらは同一の “説明” の発話特徴としている．

表 4　発話から抽出する特徴

特徴	検査対象	条件	値	略号
具体物	名詞句	シソーラス [10] で具体物に分類されている	2 値 (偽，真)	S:c
現場参照	指示詞※	現場参照を表わす (※ 連体修飾を含む)	2 値 (真，偽)	S:s
発話的役割	文	発話における役割	指示，質問，説明，応答	S:R, S:Q, S:D, S:T

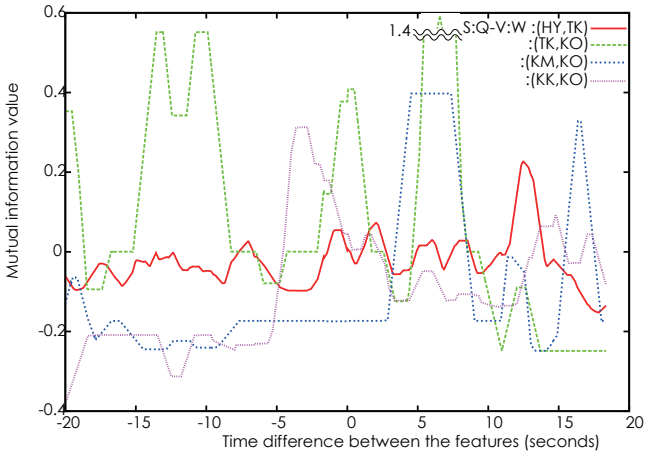


図 5　Mutual information of < S:Q, V:W >

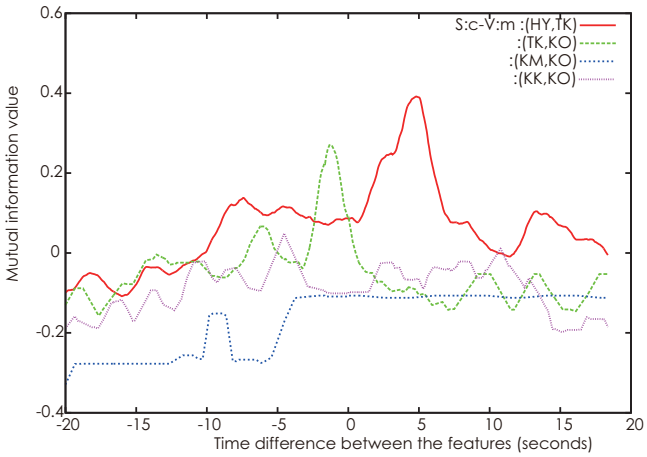


図 6　Mutual information of < S:c, V:m >

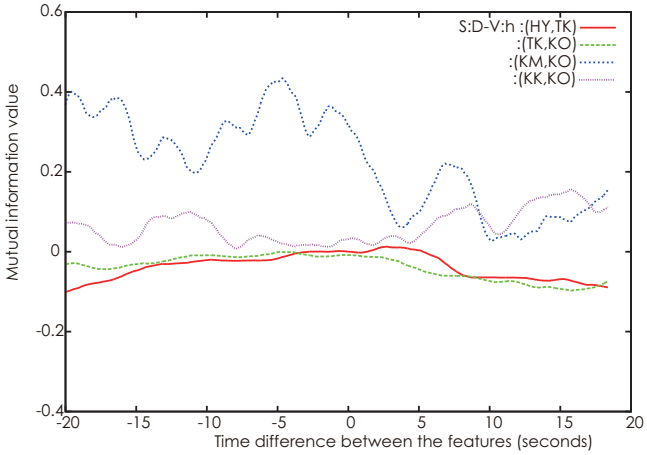


図 7　Mutual information of < S:D, V:h >

図 6 に表 1 (4) の < S:c, V:m >(具体物の名と移動) を示す．これは可視物の具体名が与えられてから移動を行うパターンであり，多くの場合は対象を探したり，取りに行ったりする行動である．この例では，(HY, TK) の組, (TK, KO) の組で，具体物が参照された際に移動していることが顕著に現れている．特に，(HY, TK) の組で，参照中や参照された後から取りに行く割合が目立ったが，これは段取りが不十分であることを示唆している．その他の 2 組には，大きな変化はなかったが，(KK, KO) より (KM, KO) の方が絶対的な値が低い傾向があり，段取りが良く，落ち着いた対応をしていることが示唆されている．

表 7　失敗が起こりやすい状況

注意・指摘の有無	記号	状況
注意・指摘なし	S1	(支援者が) 状況に気づいていない
	S2	(作業者が) 情報を与えない/要求しない
注意・指摘あり	S3	(作業者が) 注意を払わない，話を聞かない
	S4	(支援者の) 注意・指摘が間に合わない

表 8　失　敗　例

状況	関係するパターン	失敗例
S1	< S:D, V:h >	間違ったスプーンを使おうとしている
S2	< S:Q, V:W > < S:c, V:m >	指示の前に間違った場所に物を取りに行ってしまった
S3	< S:Q, V:W > < S:D, V:h >	粉を混ぜる際に注意を聞かずにダマを作ってしまった
S4	< S:D, V:h >	間違った粉の混ぜ方を続けてしまった

図 7 に < S:D, V:h >(説明と視線の停留) を示す．(KM, KO) の組では高い値を示している．つまり，注視対象 (視線が停留している部分) に対して説明を与えている傾向が顕著に見られるが，これは，作業者が情報を必要としていることを支援者が気づいて説明を与えているためである．その後，値が落ちるのは，作業や次のアクションを始めていることを示す．(HY, TK) の組, (TK, KO) の組では視線を停留させずに説明を聞いていることがわかる．これらに対し，(KK, KO) の組は説明があった後視線を停留させる傾向が若干出ており，説明事項を確認している様子が示唆されている．

これらの結果において重要な点は，モダリティの凝集性の高いパターンを機械的に抽出できること，それらのパターンを定量的に比較したり分析できることである．上記では，4 つの作業者・支援者の組を比較すれば，その振る舞いの差が定量的にはっきりと出ていることがわかる．

作業状況とコミュニケーションパターン

失敗，あるいは失敗しそうになった状況と作業・コミュニケーションのパターンを調べた．ただし，一試行あたりの失敗数は数回〜10 回程度であり，失敗のパターン自身を統計的に評価できるほど数は多くない．そのため，ここでは，失敗の原因となり得る振る舞いと頻出パターンの関係を議論する．

まず，失敗や失敗しそうな状況の支援者と作業者の振る舞いの典型的なものを表 7 にあげる．このような，支援者側からの注意や指摘の有無やタイミング，作業者側の情報要求や与えら

れた情報に対する対応は，頻出パターンとその時間的性質に顕著に現れる．

状況 (S2) では < S:c, V:m > と < S:D, V:h > のパターンが関係する．作業者がひとりで作業を始めたり，作業の確認を取らないことが原因となる．図 6 に示されているように (TK,KO), (HY,TK) の組では移動を始めてから物について指示や注意を受けているため，表 8 の (S2) にあげたような失敗例が起こりやすいと考えられる．状況 (S1) や (S4) では < S:D, V:h > のパターンが強く関係する．作業者が支援者の説明が終わる前に作業を行うこと，作業者が支援者に状況を伝えない (支援者が確認できない) こと等が原因となる．図 7 に示されているように (TK,KO)，(HY,TK) の組では説明に対して状況を確認するための視線の停留がほとんどないため，支援者に適切に状況を伝えられていないことがわかる．それが，表 8 の (S1) や (S4) の失敗例の大きな要因になっていると考えられる．

実際の失敗との対応例を表 8 に示す．状況 (S3) では，頻出パターンの < S:Q, V:W > や < S:D, V:h > が強く関係する．(HY,TK) の作業者は，初心者である支援者の説明を聞かずに作業を進める傾向が強いことは前節の図 5 や図 7 のグラフに示されている．その結果，例にあげたような失敗が起こっている．

全般的に，(KM,KO) の組み合わせでは，種々の頻出パターンが示すように，作業者が支援者の指示を聞いて，確認を行ってから作業を行い，その結果を報告・質問することが多く，それが失敗を少なくしている．

また，直接失敗の原因になるものではないが，熟練度の差によるコミュニケーションの違いもはっきり現れている．例えば，(KK,KO) の組では次の作業についての質問，つまり段取りを確認するための質問が多かった．これは図 5 で質問の発話前後-5〜5 秒の間作業を行なっている割合が高くなっていることに現れている．一方で，(TK,KO) の組では頻繁に「これでよかったっけ？」といった質問が多く現れていたり，説明を受けてもすぐに対象に注目したり注意を払うことがない．前者は図 5 の質問と作業の相互情報量が常に大きいことに，後者は図 7 で視線停留が説明の発話があった後も大きくならないことに現れている．

以上のような定量的な性質をもとに作業者・支援者の振る舞いを評価し，それをフィードバックすることにより，良い映像対話型行動支援を行うための訓練を行うことが期待できる．例えば，さらに，このような振る舞いを自動認識できれば，振る舞いの良さや現在の状況の良さを自動推定し，作業者・支援者を補助・支援することが可能になると考えられる．

6. おわりに

本稿では，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，作業者・支援者の振る舞いの傾向を良く捉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なっているが，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．そのため，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今後の興味深い課題となっている．

文　献

[1] P. Garner, M. Collins, S Webster, D. Rose, “The application of telepresense in medicine”, BT Technolo J, Vol.15, No.4, pp.181–187, 1997

[2] “UMET”, http://www.tanizawa.co.jp/umet/, 2013 年 10 月 1 日 参照

[3] 泉子・K・メイナード: 会話分析, くろしお出版, (1992).

[4] 坊農真弓, 高梨克也, “多人数インタラクションの分析手法 (知の科学)”, オーム社, 2009.

[5] 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.

[6] 小泉敬寛, 中村裕一, 近藤一晃, 小幡佳奈子, 渡辺靖彦, “映像対話型行動記録におけるモダリティ間の関係と凝集性”, 信学技報, Vol. 112, No. 176, pp.1–76, 2012.

[7] S. Norris: “Analyzing multimodal interaction”, Routledge, 2004

[8] S. Norris: “Identity in (Inter)action”, De Gruyter Mouton, 2011

[9] M. Billinghurst, S. Bee, J. Bowskill, H. Kato, “Asymmetries in Collaborative Wearable Interface”, The Third International Symposium on Wearable Computers, pp.133–140, 1999

[10] 国立国語研究所 (編): 分類語彙表 増補改訂版, 大日本図書, (2004).

[11] 黒橋禎夫, 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).

[12] 黒橋禎夫, 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).

[13] J. Siegel, R. Kraut, B. John, K. Carley, “An Empirical Study of Collaborative Wearable Computer Systems”, Proc of the ACM Conference on Computer Supported Cooperative Work (CSCW1995), pp.312–313, 1995

[14] S. Mann, “Smart Clothing: Wearable Multimedia Computing and Personal Imaging to Restore the Technological Balance Between People and Their Environments”, Proc. of the ACM Conference on Multimedia, pp. 163–174, 1996

[15] 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

[16] D. Gergle, R. Kraut, S. Fussell, “Action as language in a shared visual space”, Proc. of the ACM Conference on Computer Supported CooperativeWork (CSCW 2004), 2004

[17] H.Clark, D. Wilkes-Gibbs, “Referring as a collaborative process”, Cognition, Vol.22, pp.1–39, 1986

[18] J. Gemmell, G. Bell and R. Lueder: “MyLifeBits: a personal database for everything”， Communications of the ACM, vol. 49, Issue 1 (Jan 2006), pp. 88–95. 2006.

[19] Y. Nakamura, J. Ohde, and Y. Ohta. Structuring personal activity records based on attention - analyzing videos from head-mounted camera. In *15th Int’l Conference on Pattern Recognition Track4*, pp. 220–223, 2000.

[20] 清水彰一，西尾和晃，木村誠，藤吉弘亘，“First Person Vision のための Inside-Out カメラの提案”, 電子情報通信学会論文誌 D Vol.J94-D No.11 pp.1909-1918, 2011