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の定量化*

小泉　敬寛[†a)]　小幡佳奈子[††b)]　渡辺　靖彦[†††c)]　近藤　一晃[††d)]
中村　裕一[††e)]

Coherency Quantification of Interactions in Working Support through FPV Communication*

Takahiro KOIZUMI[†a)], Kanako OBATA[††b)], Yasuhiko WATANABE[†††c)], Kazuaki KONDO[††d)], and Yuichi NAKAMURA[††e)]

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．具体的には，映像対話に頻出するコミュニケーションのパターンから，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，両者に密接な関係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

1. まえがき

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．そこでは，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う場合等に活用できる [1]〜[3]．今後，種々の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 [4]〜[6]．

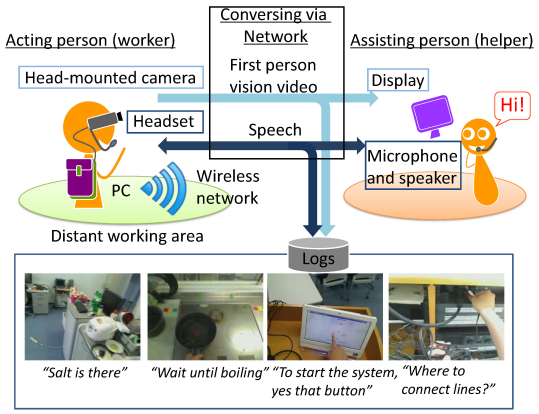


図 1　映像対話型行動支援の概要
Fig. 1　Overview of the working support through FPV communication.

このような背景から，我々は，映像対話型行動支援におけ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

† 京都大学工学研究科，京都市
Dept. of Engineering, Kyoto University, Kyoto-shi, 606–8501 Japan
†† 京都大学学術情報メディアセンター，京都市
ACCMS, Kyoto University, Kyoto-shi, 606–8501 Japan
††† 龍谷大学理工学部，大津市
Faculty of Science and Technology, Ryukoku University, Otsu-shi, 520–2194 Japan
a) E-mail: koizumi@media.kyoto-u.ac.jp
b) E-mail: obata@media.kyoto-u.ac.jp
c) E-mail: watanabe@rins.ryukoku.ac.jp
d) E-mail: kondo@media.kyoto-u.ac.jp
e) E-mail: yuichi@media.kyoto-u.ac.jp
* 本論文は 2014 年度 HCG シンポジウム推薦論文である．
DOI:10.14923/transinfj.2015HAT0003

た [7]．具体的には，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，作業者・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てきた．

これらの結果を踏まえ，本研究では更に，個々の状況における作業者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として，コミュニケーションの「一貫性」を考え，コミュニケーションの基本的なパターンやそこからの逸脱によって一貫性の程度を推定することを試みた．その結果，一貫性と基本的なパターンの時間的性質に密接な関係があることがわかった．

以下本論文では，**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．更に，**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 一貫性の考え方と関連研究

作業が円滑に進むためには，作業者と支援者が同一の対象に注意を向け，それぞれの意図が正確に伝達されることが望ましい．コミュニケーションがこのように円滑に行われていることを客観的に評価できれば，作業が円滑に行われるように作業者と支援者に対してコミュニケーション方法の訓練を行ったり，失敗や事故の原因を探ったり，またその予防を行うことに大いに役立つ．更に，それらが自動化・実時間化されれば，実際の作業の場で，状況の良さを作業者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．このような問題に対し，我々はこれまでの研究で，作業者と支援者の間のインタラクションの統計的性質に両者の態度や失敗の起こりやすさなどがよく反映されることを示した [7]．

本研究では，これまでの研究に加え，個々の場面，つまり比較的短時間のコミュニケーションに対する良さを評価することを目的とする．そのために，「作業者と支援者の注意が同一の対象に向けられ，その対象に関連する言動がタイミング良く現れていること」を「一貫性 (coherence)」が高い状態として考え，それを評価することにする．

一貫性の高い状態では，隣接ペア (adjacency pair) [8] と呼ばれる先行発話と後続発話の関係が良く現れる．先行発話によって，後続の発話が期待され (この期待を「投射 (projection)」と表現している)，その二者の関係が満たされているものを対話行為の交換の最小単位と考えるものである．更に，隣接ペアを身体動作にまで拡張したものが「企図ペア (projective pair)」として Clark により提案されている [9]．これらは一貫性の良い手がかりとなるが，良好な意思疎通が起こっている場合はこれらに限らない．つまり，後続の発話や身体動作により先行発話や動作の話題や焦点が継続される場合も含める必要がある．例えば，先行発話の内容に対して後続発話が新しい情報を付け加えていく状態などである．他の例として，支援者がある対象について言及したときに，作業者がその対象を手にとって眺める行為を考えてみる．必ずしも支援者は作業者の行為を期待 (投射) していたわけではないが，支援者の発話と作業者の行為には明確な因果関係が認められるため，良好な意思疎通が行われていることを示唆している．

逆に，両者のいずれかが直前の言動と関連性のない発話をしたり，関連性のない行為を行っていることを一貫性が低い状態であると考える．作業者と支援者が異なる空間にいる状況では，同じ空間を共有する場合に比べ，2 者が異なる刺激や情報を受け取ることが多くなる．そのため，直前の言動と関連性のない言動が現れる，例えば，直前の話題と無関係な話題を切り出したり，相手が言及していないものを触ったりする言動などがある．これらの結果，話題が変わってしまう場合も多くなる．しかし，話者の片方が直前のトピックを継続させようと努力する (直前と同じ発話を繰り返す) 場合もあり，必ずしも話題が変わるわけではない．

この一貫性と類似する概念として「結束性 (cohesion)」や「凝集性 (cohesiveness)」も良く用いられる．ただし，結束性は修辞的な文脈構造を含んで議論されることが多いため，ここでは一貫性を用いることにした．また，凝集性はコミュニケーションの要素間の関連性だけでなく，要素の数が多いことも考慮するため，少し異なる意味を含む．

一貫性の考え方は，Clark らによる “grounding in communication” の考え方とも整合する．[10] では，共通の対象に関して対話を進めるための段階や方法が述べられており，これは，映像対話型のコミュニケーションの一貫性を評価したりそれを保つ方法としても参考になる．ただし，Clark らは言語的行為を対象としているのに対し，本研究では，以下で述べるような，マルチモーダルなコミュニケーションを考えることに

新規性がある．更に，表層的な特徴から機械的に一貫性を推定する指標を設定することも本研究の新しい特徴である．

マルチモーダルなコミュニケーションを考える際には，Norris による “modal density” の考え方が参考となる [11], [12]．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これまでに議論されていない．

会話の分析に関する先行研究でも，視線や相槌といった行為と話者交代などとの関連性が調査されてきており，複数のモダリティを扱うことの有用性が示されてきた [13], [14]．このような知見もマルチモーダルな指標を設定することの有効性を示唆している．ただし，映像対話型行動支援では実際の作業を対象としていることや，作業に関連する動作が重要な意味をもつことから，新しい手法が必要となっている．

2.2 典型的なコミュニケーション例と一貫性

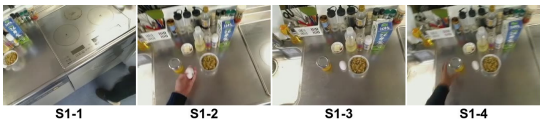
映像対話型行動支援におけるコミュニケーションの性質を考えるために，典型的な状況を考えてみよう．表 1 の (a) から (d) は，キッチンで調理作業を行ったときの映像対話型行支援の記録で見られる典型的なコミュニケーションの例であり，(a) と (b) は円滑な意志の疎通が行われている例である．(a) では，両者の注意が一つの対象に対して向いている，つまり，注意 (焦点) を共有している状態である．良好に作業が進行している場合にはこの状態が多く見られる．発話だけでなく，見回す行動などもコミュニケーションとして重要な役割を果たしている．(b) では，関連性のない二つの焦点に対して注意が向けられているが，作業者・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かかわらず，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．

作業者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，例 (c), (d) のように，言動の量が多いだけでは十分な条件とは言えない．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．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．

表 1 典型的な映像対話の例
Table 1 Example of the typical FPV communications.

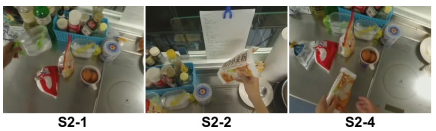
(a) 焦点一つ，または，複数が密に関係

(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 個若しくは 0.5 個分，はどこですか？」
(S1-3) A: 「そこの瓶に 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



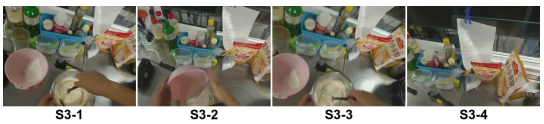
(b) 複数焦点．お互いに関連なし

(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: 「オーブ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．ところで，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！」

2.3 一貫性の要件

これまでの議論から，一貫性に対して，(I) 話題や注意 (焦点) の連続性が認められること，直前の言動との因果関係が認められること，典型的には隣接ペア，企図ペアを成していることを正の条件 (成立条件) とし，(II) 直前の言動と関連性や因果関係が認められない言動が起こっていることを，負の条件 (非成立条件) と考える．

例えば，前節の例 (a)，(b) は一貫性の高い場面である．(a) では (S1-1) の「○○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．また，これらの言動は注意や焦点を共有している．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．

これらとは対照的に，前節の例 (c), (d) は一貫性の低い場面である．これらの場面ではお互いに呼応していない言動が目立つ．(c) では，(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．

このような例からわかるように，目的のはっきりした作業におけるコミュニケーションでは，人間が呼応関係や焦点の共有に基づいて一貫性の高低を判断することは比較的容易である．本研究の目的は，このような一貫性を機械的，自動的に評価できるような指標を設定することである．その際に，深い意味の自動解析は難しい問題であるため，本研究では表層的な特徴を用いて上記 (I) を定量化することを目指した．

3. 映像対話における一貫性の指標

3.1 基本的な考え方

一貫性の評価には「呼応パターン」を用いる．呼応パターンとは前節の (I) が満たされている可能性が高い言動のペアのことであり，詳しくは **3.3** で説明する．手法の概要は以下のようになる．

（1） 映像対話中に頻繁に出現するインタラクションのパターン (以下「頻出パターン」と呼ぶ) を抽出する．

表 2 画像からの抽出する特徴
Table 2 Feature detection from images.

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	カメラモーションの前進を示す	V:m

表 3 発話から抽出する特徴
Table 3 Feature detection from speeches.

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

（2） 頻出パターンの部分パターンの中から呼応パターンを選ぶ．

（3） 各時刻でのインタラクションが各呼応パターンの標準的な特徴量に適合する度合いを数値化し，総合したものを評価値とする．

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記録中に頻出している．
- 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があるが，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．

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 [7] を参照されたい．

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．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．画像特徴については表 2 に示したものを用いた．視線停留，見回し・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．ここで頻出パターンは，トランザクションデータ中に一定以上の頻度で出現する n 個の特徴からなる系列とする．このとき，頻出パターンの出現区間ではパターンと無関係の特徴が挟まることと，複数の特徴が同時に生起する

ことを許す．また，頻出パターンはトランザクションデータ内での一致区間が最小になるものが選ばれるものとする．抽出アルゴリズムとして，時系列パターンのマイニングで広く用いられている PrefixSpan [15] を用いた．PrefixSpan は複数の系列データを入力とし，そこから抽出される数がしきい値以上になる部分系列パターンを出力とする．後述する呼応パターンの選出の際には $n > 3$ となる頻出パターンが抽出され，そこから頻出パターンが一致する区間内に出現する特徴数と，その一致区間長で昇順にソートした上位 20 の頻出パターンを用いた．

以降，頻出パターンや後述する呼応パターンでは，{作業者発話，支援者発話，作業者行動 (画像より得られる)} をそれぞれ {“W”，“S”，“V”} とし，“:” を挟んで表 2，3 に挙げた特徴と並べた形で表記する．例えば，“(V:l→W:D→S:D)” は，作業者が周囲を見回した後，作業者が何らかの説明をし，支援者が更に何らかの説明をすることを表す．

映像対話型行動支援では，カメラに写っていない部分が視覚的に伝わらないこと，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与えられるという特徴がある．そのため，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り，記録データの解析によって多くの重要な性質を捉えられることが期待できる．

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コミュニケーションの一貫性と強い関係が期待されるものを以下の指針を用いて呼応パターンとして選ぶ．

- 3 特徴以上の頻出パターンに含まれる 2 特徴の組を呼応パターンとする
- 頻出パターンの生起している間に生起する他の特徴の平均が少ないほど優先する(注1)

それぞれの指針が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．前者が必要となるのは，2 特徴しか含まない頻出パターンには，個々の特徴の出現頻度が高いために，実際の呼応とは関係なく頻度が高くなるものがあるためである．例えば，(S:D→W:D) のような頻出パターンは個々の特徴の発生数が多いため，実際に意味のある呼応関係にあるかどうかとは関係なく様々な箇所で頻出パターンとして検出されてしまう．後者は，頻出パターンの生起区間に，頻出パターンに含まれている特徴以外のものを多く含んでいる場合，頻出パターンに含まれている特徴同士が実際に呼応していない場合が多くなることによる．例えば，(V:m→S:D→V:m) というパターンでは，平均して 17.8 個の特徴が発生している．つまり，作業者の移動後に支援者の説明が発生するまでと，支援者の説明後に多数の特徴が現れているため，頻出パターン中の特徴同士の呼応関係が意味のあるものだとは考え辛くなる．

(注1)：PrefixSpan の考え方として，頻出パターンの要素の生起は連続していなくても良い．つまり，他の要素の生起が挟まっていても良い．

表 4 各呼応パターンの生起時刻のずれの分布の Shapiro-Wilk の正規性検定による p 値と歪度，尖度の値
Table 4 p-value of the Shapiro-Wilk nomality test and kurtosis, skewness from the distributions of the Δt at each adjacency patterns.

Pattern	p 値	kurtosis	skewness
W:D→S:D	1.8357e-09	0.28817	0.057257
S:D→W:P	1.496e-08	-0.037804	0.10995
S:R→W:D	0.015937	-0.92047	0.0014528
W:U→S:D	0.0012327	1.7607	0.77729
S:D→W:D	1.2805e-07	0.1035	0.081357
V:h→S:D	0.0035644	-0.53922	0.12116
S:R→V:h	0.025941	-0.97225	-0.060641
W:D→S:R	0.00081645	-0.87447	0.0066899
W:D→S:P	0.0014884	-0.41494	0.23897
S:D→V:h	3.4302e-08	-0.90746	-0.10602
V:l→S:D	0.0094854	-0.51693	0.049902
S:D→V:m	2.0317e-05	-0.85965	-0.054036

次に，呼応パターンの時間的性質を考える．十分大きな時刻差を想定すれば，二つの言動が呼応している確率は，二つの言動が時間的に離れる (生起の時刻差が大きくなる) につれて小さくなる．そのため，一様分布などではなく，時間的に局在した分布を考えるのが妥当である．また，図 2 より，それぞれが中央に偏った分布になっている傾向がわかる．そのため，本研究ではその頻度を正規分布で近似する方法をとる．起点の特徴を F_a，2 番目の特徴を F_b とし，F_a と F_b に属する特徴の間 (起点の特徴が終わってから 2 番目の特徴が始まるまで) が Δt である確率分布を $P(F_b|F_a;\Delta t)$ としたとき，それを平均 $\mu_{\Delta t}$，分散 $\sigma_{\Delta t}$ の正規分布 $N(\mu_{\Delta t},\sigma_{\Delta t})$ で近似した．実際の確率分布 (頻度によるもの) と正規分布で近似した値の差の平均は 0.010125，分散 0.0011107 となる．Shapiro-Wilk の正規性検定でも，5%の有意水準で正規分布に対する帰無仮説が棄却されていない．そのため，統計的にも妥当な近似となっていることが示唆されている．

図 2 より，大半の呼応パターンでは 2 秒以内にピー

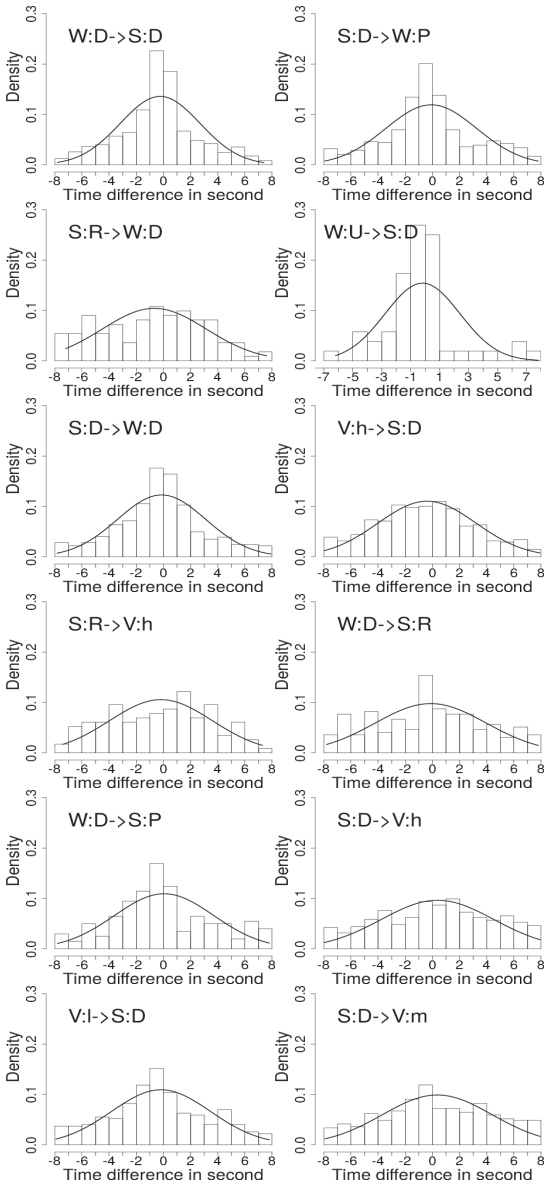


図 2　各呼応パターンの生起時刻ずれの分布と正規分布による近似カーブ

Fig. 2　Approximation curve and histogram of the distribution of the Δt at each adjacency patterns.

クが現れ，上述した呼応関係があることが示唆されている．ただし，S:R→W:D や S:D→V:h のように明確なピークがないものもあった．これは，作業の指示に対する作業状況の説明や視線の停留(注2)ように作業の実行を伴うことが必要な場合には，呼応に時間がかかることが多いためだと考えられる．

(注2)：多くの場合，作業の実行に起因する．

4. 一貫性の計算

呼応パターンの特徴間の間隔から，それぞれの時刻における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とおりである．

まず，それぞれの呼応パターンの起点の特徴の生起ごとに，その終了時刻に最も近い時刻に生起した 2 番目の特徴を求める．その間隔によってスコア付けする．そのために，呼応関係の良さを表す正スコア ($s_p(t)$ とする) に前節の $P(F_b|F_a;\Delta t)$ の値を与える．実際の計算では，図 2 で表したような正規分布 $N(\mu_{\Delta t},\sigma_{\Delta t})$ で近似した値を用いる．この $s_p(t)$ は，良いコミュニケーションが大半を占めているという考え，つまり，生起確率の高いパターンほど良い呼応関係である可能性が高いという考えに基づく．

更に，呼応関係の悪さを表す負スコア (以下 $s_n(t)$ とする) を考え，その値は $P(F_b|F_a;\Delta t) - \max_{\Delta t}(P(F_b|F_a;\Delta t))$ とする．ここで，$\max_{\Delta t}(P(F_b|F_a;\Delta t))$ では，同一の呼応パターンの確率分布の値が最大となる Δt を求める．つまり，この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補助的なスコアとなる．実際の計算には，正スコアと同様に，$P(F_b|F_a;\Delta t)$ の正規分布を用いる．

次に，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．本研究では，以下で述べるように S_m，S_s，S_n の三つの異なる総合方法で求められる指標を検討した．

（1） コミュニケーションの受け手が送り手にタイミング良く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．そのため，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，$S_m(t) = \max_i s_p(t,i)$ 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” 等の観点から良いコミュニケーションが行われている状態であると考える．そのため，$S_s(t) = \sum_i s_p(t,i)$ が良さを表す指標となる．

（3） コミュニケーションの受け手が適切に反応していないことは，最も良く呼応しているパターンが不十分なものとなっていることによってわかる．そのた

め，$S_n(t) = \max_i s_n(t,i)$ がその指標となる．

以上の三つそれぞれが異なる観点からの指標であるため，本研究では指標を一つに絞り込むことをせず，それぞれの指標を実際のデータに対して求め，その適切さについて検証する．

5. 実　　験

5.1 データ収集

システムは図 3 のような構成とした．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，ヘッドセットにより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 (図 3 左下) 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．

作業者と支援者は表 5 のように異なる立場の被験者の組み合わせとした．これは，作業者と支援者の立場やタスクへの熟練度の違うデータを幅広く収集するためである．

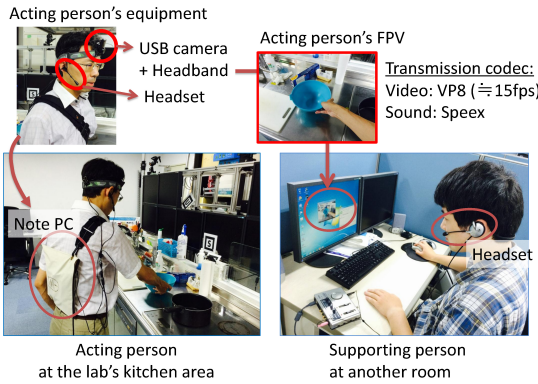


図 3　実験環境：作業者がキッチンで調理をしている場面と支援者が別の部屋で支援を行っている様子

Fig. 3　Experimental environment: Acting person at kitchen area with equipments and Supporting person at another room.

表 5　被験者の組み合わせ

Table 5　Pairs of research participants.

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)
d	スタッフ (熟練者)	学生 (初心者)

5.2　個々の場面における一貫性の指標

まず個々の典型的な場面に対する一貫性の指標の有効性について述べる．図 4 と図 5 で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，図 6 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った例，図 7 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場面の例であり，それぞれのグラフが各場面での指標値を示している．

図 4 の例では，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 (A-a2) で移動を始め，それを (A-a3) で発話による応答として知らせている．(A-a7)〜(A-a10) でも支援者の指示に対して動作と発話による応答を与えている．このように，状況の変化を映像と発話で伝えることができている場面で S_s が特に高い値になった．同様に，図 5 の例では，(B-b2) のように，会話の中で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．

一貫性の低い例である図 6 では，(C-c3) で作業者

id	Time (Sec)	特徴	内容
A-a1	202.5	S:D	「黄身と白身をむらなく良く混ぜる」
A-a2	204.2	V:m	移動
A-a3	205.8	W:P	「はい」
A-a4	205.8	V:h	停留
A-a5	207.9	S:D	「で，だし汁，はそのボウルに入っているやつ」
A-a6	209.5	V:m	移動
A-a7	211.2	S:D	「ちょっともう見えんけど，それ，それ」
A-a8	211.2	V:h	停留
A-a9	213.3	S:D	「もうちょっと頭さげて，あ，それ，だし汁」
A-a10	215.4	W:P	「はい」

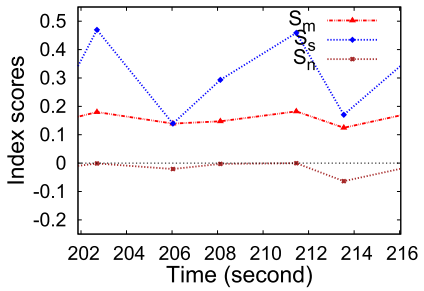


図 4　一貫性の高いインタラクションの例とそのときの各指標値の変化を示すグラフ (A)

Fig. 4　Example of high coherency interaction and the graph of each indices values (A).

id	Time (Sec)	特徴	内容
B-b1	122.8	W:D	「えー，回転台…」
B-b2	123.8	V:l	見回し
B-b3	124.5	S:D	「要はその上に乗せて冷やす」
B-b3	125.1	W:D	「ありますね」
B-b4	125.3	S:D	「冷やせと」
B-b5	126.8	W:P	「はい」

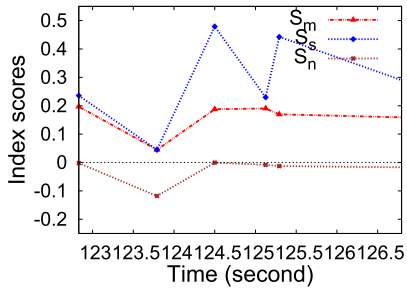


図 5　一貫性の高いインタラクションの例とそのときの各指標値の変化を示すグラフ (B)

Fig. 5　Example of high coherency interaction and the graph of each indices values (B).

id	Time (Sec)	特徴	内容
C-c1	128.0	W:D	「うーん」
C-c2	129.0	V:l	見回し
C-c3	130.3	W:D	「おさら」
C-c4	130.3	V:m	移動 (食器棚へ)
C-c5	133.3	V:l	見回し
C-c6	134.7	V:m	移動 (シンクへ戻る)
C-c7	136.7	V:h	停留
C-c8	137.8	V:l	見回し
C-c9	138.7	V:m	移動 (食器棚へ)
C-c10	141.0	S:D	「ん，おさら？」
C-c11	141.0	V:h	停留
C-c12	145.3	S:D	「おちるおちる」

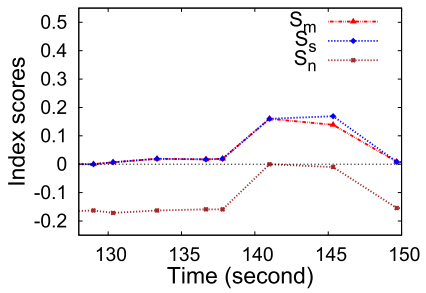


図 6　一貫性の低いインタラクションの例とそのときの各指標値の変化を示すグラフ (C)

Fig. 6　Example of low coherency interaction and the graph of each indices values (C).

id	Time (Sec)	特徴	内容
D-b1	68.3	V:m	移動
D-b2	70.2	W:D	「ミキサー」
D-b3	71.0	S:D	「あれ？」
D-b4	71.8	W:D	「ミキサー」
D-b5	72.5	S:D	「その上に置いてありませんでしたっけ」
D-b6	75.0	W:D	「ミキサー」
D-b7	75.8	S:D	「棚の所にはありませんか？」

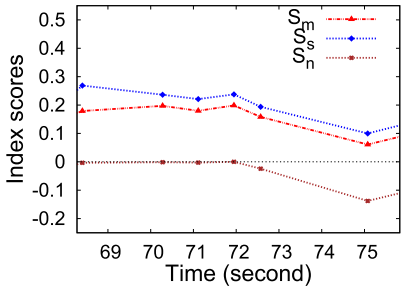


図 7　一貫性の低いインタラクションの例とそのときの各指標値の変化を示すグラフ (D)

Fig. 7　Example of low coherency interaction and the graph of each indices values (D).

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-c10) に対して作業者も返事をしていない．そのため，S_s, S_m, S_n の値が低くなっている．特に作業者の発話である (C-c3) に対する応答がないことがの指標の低下の主な要因である．(C-c7) と (C-c8) の視線移動に反応して (C-c10) があることで指標はいったん改善しているが，その後作業者の応答がないことで，再び指標は悪化している．

同様に，図 7 の例では，作業者の発話 (D-b1)～(D-b4) のそれぞれに支援者が応答している箇所では 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，(D-b5) 以降の応答で各々の指標が低くなっている．

このように，典型的な例に対して一貫性の指標がコミュニケーションの状態を良く表していることが確認できた．また，複数のモダリティを有効に利用している部分では S_s が高くなっていることから，“modal density” の観点からの一貫性を議論するために S_s が有効であることも示唆された．

5.3　一貫性の指標の全般的な有効性

収録されたデータ中の一貫性の高い部分と一貫性の低い部分を人手でタグづけした．タグは筆者自身が両者の発話音声を含む映像を見ながら，発生している隣接ペア，企図ペアが **2.3** で述べた負の条件 (非成立条

件) に合致していると判断できる場面を一貫性の低い場面としてタグ付けした．それ以外の場面については，基本的なコミュニケーションは成立している状態とみなして，一貫性の高い場面に含まれるものとしている．いずれの実験データでも，人間が見て因果関係を判断することが困難な隣接ペア，企図ペアは見合たらなかった．このタグづけによる，一貫性の高い場面は 46 場面 (呼応数 181)，一貫性の低い場面は 48 場面 (呼応数 134) であった．

これを用いて，収録データに対する一貫性の各指標の値と一貫性の高い/低い部分との一致を確認した．その結果を表 6 に示す．表が示すように，S_m，S_s，S_n のいずれの指標も，一貫性の高低を表す指標となっていることがわかる．このことは，有意水準を 1%とした t 検定により確認した．表右端の p 値が示すとおり，はっきりと有意差が認められた．

5.4 コミュニケーションの傾向と一貫性の指標の関係

作業者と支援者のペアによる振る舞いの傾向の違いと，一貫性の指標との関係について述べる．

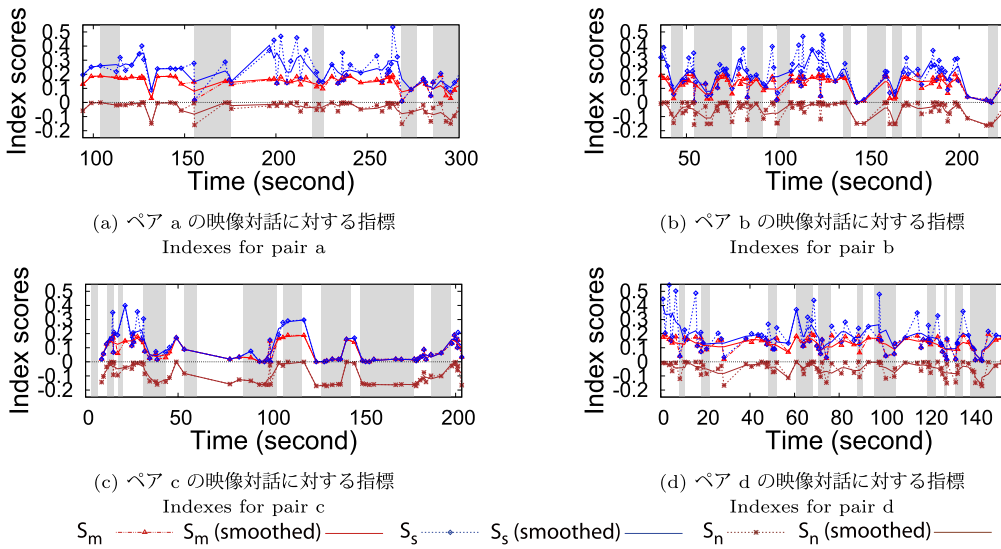
図 8 (a)〜8 (d) は各ペアの映像対話記録から 200 秒程度を抜き出し，一貫性の指標の変化をプロットしたものである．点線は特徴が発生するごとに得られる値をプロットしたもので，実線は 3 秒間の移動平均により平滑化したものである．また，この図で網掛けになっている区間が一貫性の低い箇所であり，それ以外の部分は一貫性の高い部分，または，コミュニケーションが行われていない部分である．これらの結果から，全般的な傾向として，一貫性の低い部分またはその周辺で，S_m，S_s，S_n の一つまたは複数が低くなっていることがわかる．これは前節の結果と整合する．

次に各ペアの振る舞いの傾向と一貫性の指標との関係を概観する．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，過不足なく一貫性の高いコミュニケーションをとりながら慎重に調理を進めた．図 8 (a) より，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．

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一貫性の低い場面が幾つか見られた．図 8 (b) より，S_m が高くなる箇所が多いことがわかるが，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，発話だけで情報を伝えていることを示唆している．

表 6　一貫性の指標の統計的検証
Table 6　Statistical verification of coherency indexes.

	一貫性の高い場面		一貫性の低い場面		
指標	平均	分散	平均	分散	p 値
S_m	0.16060	0.00809	0.10013	0.00940	2.6733e-08
S_s	0.22800	0.02542	0.12666	0.01767	6.1560e-09
S_n	-0.93648	0.03057	-1.05597	0.03474	1.7897e-08



(a) ペア a の映像対話に対する指標
Indexes for pair a

(b) ペア b の映像対話に対する指標
Indexes for pair b

(c) ペア c の映像対話に対する指標
Indexes for pair c

(d) ペア d の映像対話に対する指標
Indexes for pair d

図 8　各ペアの映像対話に対する指標
Fig. 8　Indexes for each pairs.

ペア c で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，作業者側から質問や報告を積極的に行わなかったため，図 8 (c) では一貫性の高い部分が少数しか見受けられない．また，S_s が S_m とがほぼ同じ値をとる部分が多いことは，マルチモーダルな情報をうまく使えていないことと整合する．

ペア d では作業者の熟練度が高いことから，作業者が主導して作業を進め，それを支援者が追認する場面が多かった．全般的に一貫性の指標は高くないが，S_s が S_m よりも大きい部分が見られ，映像をうまく使ってコミュニケーションをとっていることが指標にも表われている．

以上のように，各作業者・支援者ペアの振る舞いに対して一貫性の指標がその性質の一部を良く捉えていることがわかる．ただし，細かい部分では，一貫性が低い部分としてタグ付されているにもかかわらず指標が高い部分やその逆も見られる．例えば，支援者が作業とは無関係の発話を側にいた第三者に対して発した際に，たまたま作業者が独り言をつぶやいたため，それら 2 特徴が呼応しているものとして S_s が高くなった．このように，意味的には相手の発話内容と関係がなくとも，何らかの発話や行動があれば指標が高くなることは避けられない．このような場合は比較的少数ではあるが，これらを対象から除く仕組みについては今後の課題としたい．

6. むすび

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の円滑さを評価するために，インタラクションの一貫性に着目し，表層的に観測可能な特徴から一貫性を定量化することを提案した．そのために，映像対話記録データ中の頻出パターンから得られる呼応パターンの利用とそれによる一貫性の指標化方法を提案した．実際の映像対話記録に対して提案手法を適用した結果，一貫性の指標がおおむね良好な振る舞いを示し，実際のコミュニケーションの一貫性の高低と一貫性の指標に高い相関があることが実証された．

今後の課題として，より豊富なデータに対してシステマティックに本手法の有効性を検討していく必要がある．表層的な特徴のみで一貫性を評価することの問題も明らかであり，意味的な処理を加えることを検討する必要もある．更に，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行動を起こす必要がある．このようなコミュニケーションの良さの包括的な評価も重要な課題である．

文　献

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

[2] M. Billinghurst, S. Bee, J. Bowskill, and H. Kato, “Asymmetries in collaborative wearable interface,” Third International Symposium on Wearable Computers, pp.133–140, 1999.

[3] R. Kraut, M. Miller, and J. Siegel, “Collaboration in performance of physical tasks: Effects on outcomes and communication,” Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 1996), pp.57–66, 1996.

[4] S. Fussell, R. Kraut, and J. Siegel, “Coordination of communication: Effects of shared visual context on collaborative work,” Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 2000), pp.21–30, 2000.

[5] R. Kraut, D. Gergle, and S. Fussell, “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), pp.31–40, 2002.

[6] D. Gergle, R. Kraut, and S. Fussell, “Action as language in a shared visual space,” Proc. ACM Conference on Computer Supported CooperativeWork (CSCW 2004), pp.487–496, 2004.

[7] 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頻出パターンに基づいたコミュニケーションの分析，” 情処学論，vol.56，no.3，pp.1068–1079, 2015.

[8] E. Schegloff and H. Sacks, “Opening up closings,” Semiotica, vol.8, pp.289–327, 1973.

[9] H. Clark and M.A. Krych, “Speaking while monitoring addressees for understanding,” J. Memory and Language, vol.50, no.1, pp.62–81, 2004.

[10] H. Clark and S. Brennan, “Grounding in communication,” in Perspectives on socially shared cognition, Chapter 7, American Psychological Association, 1991.

[11] S. Norris, Analyzing multimodal interaction, Routledge, 2004.

[12] S. Norris, Identity in (Inter)action, De Gruyter Mouton, 2011.

[13] 坊農真弓，鈴木紀子，片桐恭弘，“多人数会話における参与構造分析：インタラクション行動から興味対象を抽出する，” 認知科学，vol.11, no.3, pp.214–227, 2004.

[14] 坊農真弓，高梨克也（共編），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.

[15] J. Pei, J. Han, B. Mortazavi-asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, “PrefixSpan: mining se-

quential patterns efficiently by prefix-projected pattern growth," 17th International Conference on Data Engineering (ICDE '01), pp.215–224, 2001.

（平成 27 年 4 月 6 日受付，8 月 14 日再受付，
10 月 2 日早期公開）

小泉　敬寛　（正員）

2005 年筑波大学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．同年京都大学工学研究科電気工学専攻博士後期課程入学．2007 年同大学退学．同年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などの研究に従事．電子情報通信学会．情報処理学会各会員．

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．

渡辺　靖彦　（正員）

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　（正員）

2004 年同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)．情報処理学会，電子情報通信学会各会員．

中村　裕一　（正員）

1985 京大・工・電気工学第二卒．1990 同大大学院博士課程了．同年京都大学工学部助手．1993 筑波大学電子・情報工学系講師．1999 機能工学系助教授，2004 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 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 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，人工知能学会，ACM，IEEE 各会員．