Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

# 非言語情報の出現パターンによる会話状況の特徴抽出

中田篤志[†1]　角　康之[†1]　西田豊明[†1]

我々が会話を行う際には，発話や視線，指差しといった行動に一定の構造が存在する．しかし，従来の手法では会話的インタラクションの分析においてこれを網羅的に，かつ数値的な根拠を持って取得することは十分になされていなかった．これを踏まえ，本論文では我々の提案しているインタラクションマイニングの手法を用いて，インタラクションステートとその遷移構造から会話構造をどの程度推定できるかを検証した．

本論文ではインタラクションマイニングの重要要素であるインタラクションステートと $\chi^2$ 検定による特徴的構造検出を説明した．その後 3 人の自由会話に対してインタラクションマイニングを行い，タスクやポスターの異なる実験間でインタラクションステートの頻度分布を比較した．また特徴的構造抽出によって得られた構造に対し検証を行った．その結果，ポスターの種類やタスクの差異によって生じる会話の差異がインタラクションステートの頻度分布差として現れることを示した．また，指差しと発話の共起性，3 者対面会話における視線と発話の関係など，従来より報告されてきた構造が特徴的構造として機械的に抽出できることを確認した．

# Feature abstraction of situation in conversations by pattern of nonverbal interaction

Atsushi Nakata,[†1] Yasuyuki Sumi[†1]
and Toyoaki Nishida[†1]

When we talk, some protocols exist in speech, gaze, and pointing. However, There are few theories before to acquire this protocol by usual methods of conversation analysis totally and numerically. So we verified we could presume conversation structure from Interaction State and the transition of it by Interaction Mining we proposed before.

In this paper, we explained Interaction State and detection of distinctive structure by $\chi^2$ test that was an important element of Interaction Mining. Then we apply Interaction Mining to free conversation of three people, and We compared the distribution of the frequencies of the Interaction State between the experiments with a different task and poster. After that, we verified the structures we got by the detection of distinctive structure. As a result, we suggested the difference in conversation structures from the difference of posters and tasks caused the distribution of the frequencies of the Interaction State. And we found automatically the structure to suggest past theory in the distinctive structure we got,for example the fact that pointing and gaze appears at the same time and relation of gaze and speech in the conversation three people have with seeing the others.

## 1. はじめに

我々が会話を行う際には，発話や視線，指差しといった行動によって会話を制御している．このときの行動は，言語において単語間で節や文法といった構造が存在するのと同様に，一定した構造を持って行われている．例えば「返答を期待するときは，その相手の顔を見ながら話す」「重要な発話のまえには，視線配分やジェスチャを行って他の人の興味を引く」といったものである．

このような会話におけるマルチモーダルな行動の構造を明らかにすることができれば，それを基にエージェントやロボットなどの人工物が今よりも自然な形で人とインタラクションを行うことが可能となる．例えばロボットが伝えたい情報がある場合に，それを唐突に伝えるのではなく，うなずきや視線配分を利用して発話の権利を得てから情報を伝えるといったことが期待される．

そこで本研究では，データマイニングの手法をインタラクション分析に取り入れることによって会話構造を推定することを試みる．これをインタラクションマイニングと呼ぶ．

本論文では，まずインタラクションマイニングの重要要素であるインタラクションステート，および特徴的構造の検出法について説明する．その後 3 人の被験者による自由会話 3 回分のデータに対してインタラクションマイニングを行い，得られたインタラクションステートと特徴的構造に関して以下の観点から検証する．

- 実験別にインタラクションステートの出現頻度を比較し，実験ごとに設けたポスター・タスクの差異が出現頻度び分布にどのような形で現れているかを検証する．
- 全データを統合して特徴的構造の検出を行い，抽出された構造について従来の会話分析で得られている知見を基に検証を行う．

†1 京都大学
Kyoto University

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

## 2. 関連研究と本研究の位置づけ

会話的インタラクションにおける構造を明らかにしようとする試みは数多く行われている[1)2)3)4)5)]．その際の手法としては 2 種類ある．1 つはある仮説を検証するための統制された環境で実験を行うという手法であり[2)3)]，もうひとつは会話の中から分析者が一部のエピソードを取り出し，その中における会話構造を詳細に議論する手法である[4)5)]．前者は様々な種類の会話構造に関して網羅的に検証するのは困難であり，後者は得られた知見がどの程度汎用的なのか，どのような状況で発生しやすいのかといったことを議論するのは難しい．

また共通する課題として，これらの研究の多くは人手でつけた情報を基礎として研究を行っているため，実際にセンサなどで認識するのが困難なものや，会話の意味に踏み込んで初めて議論が可能となるものが含まれている．このような情報は，実際に人工物に応用していく上で困難な面がある．

これらのことを踏まえ，我々はセンサによって取得された計測データに基づいて解釈を極力自動化することを目指している[6)]．本研究ではその中でも，大量のデータを構造化し分類することで，複数のモダリティ・複数のプロセスを経た会話構造について数値データを基に検証することを試みる．

次に，多数のカメラやセンサを用いて大量に会話データを収録し，インタラクションの分析を試みている研究として NIST[7)]，AMI[8)]，VACE[9)] といったものがある．これらの研究における自動解釈はあくまで文のセグメンテーションや非言語行動の検出にとどまっており，会話の意味構造といった抽象度の高いものについては従来の会話分析の手法が用いられている[4)]．

また，計測データの自動解釈に関する先行研究としては，森田らの研究[10)] が挙げられる．この研究では，ウェアラブルセンサにより自動付加されたラベルからのパターン抽出を試みている．この研究では構造の時間変化については検討されておらず，構造の頻出順のみを正規化し示している．

最後に，本研究で提案しているインタラクションステートの概念，およびインタラクションマイニングの手法は福間らの研究[11)] において最初に提案されたものである．この研究では利用している実験データが 1 回分と少なく，またセンサの認識結果をそのまま利用して構造化・分析を行ったためにセンサの誤認識の影響が大きいという課題があった．本研究ではそれを踏まえ，構造化に利用する実験データを増やし，かつセンサの認識結果をある程度手作業で修正している．

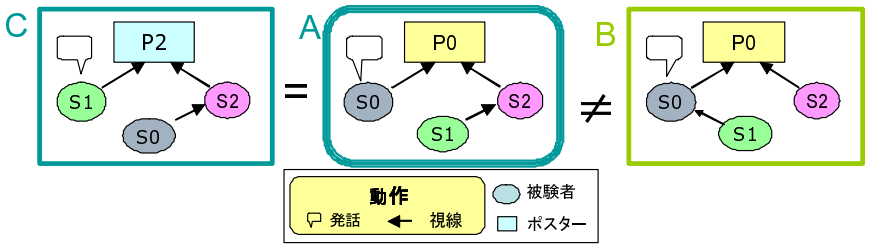


**図 1** インタラクションステートの例
Fig. 1 The examples of Interaction State

## 3. インタラクションマイニングの構成要素

ここでは，本研究で用いているインタラクションステートの概念，および特徴的構造の検出法について説明する．

### 3.1 インタラクションステート

本研究では会話の構造化のための重要概念としてインタラクションステートを用いている．

会話分析において，非言語情報に関する開始点・終了点およびその対象の情報はラベルと呼ばれ，収録した会話に対しラベルを作成することはラベリングと呼ばれる．インタラクションステートは，複数の被験者に関する非言語情報のラベルから，会話の状況を

- 全員でひとつのボードを見ていて，1 人がボードを指差している
- 3 人中 2 人が互いを見ていて，もう 1 人はそのうちの片方を見ている

といった形で抽象化するためのモデル化手法である．この概念の導入により，個々の実験やその中における参加者，非言語行動の対象などを共通化して会話の状態を表現することが可能になる．

インタラクションステートの例を図 1 に挙げる．ここで，ステート A とステート B は S0 と S2 の状態は同じであるものの，S1 が発話者である S0 を見ているか，非発話者である S2 を見ているかという点が異なるため，異なったインタラクションステートである．一方，ステート A とステート C は，S0 と S2 を入れ替え，P0 を P2 と入れ替えることで同じ状態となるため，同じインタラクションステートである．

### 3.2 ステート列の構築

ここでは，収録された会話データからインタラクションステートの列を構築する手順を説明する．

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

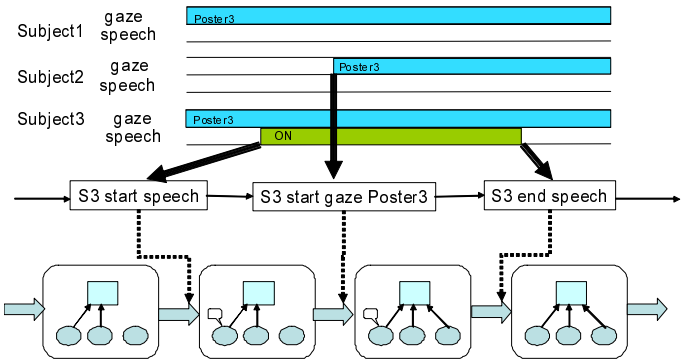


**図 2** ラベルのインタラクションステート列への変換過程
Fig. 2 The process to change labels to a line of Interaction State

まず，着目する会話中の行動に対してラベリングを行う．ここで，本手法では 2 章で述べた背景から，ラベリングはセンサデータを基に自動化することを想定している．ただし自動ラベリングの精度が十分でない場合，ラベリングの際に一定条件で出現するノイズが特徴的構造として抽出される場合があることに注意しなければならない．また，何秒以上の発声を発話ととらえるか，視線の対象となりうるものをどう定義するかなどといったラベリングの規則によって，後述するステート列や特徴的構造は変化する．

その後，作成されたラベルの全ての変化点をインタラクションステートの遷移点として捉え，ステートの列を自動的に構築する．この際，各ステートの時間幅は考慮しないものとする．

図 2 は 3 人の被験者に関する視線・発話のラベルをインタラクションステート列に変換する過程を示したものである．

### 3.3 $\chi^2$ 検定による特徴的構造抽出

ここではインタラクションステートの列を基に，会話状態遷移における特徴的な構造を抽出する手法について説明する．

まず 3.1 の変換で得られたインタラクションステート列を一定の深さを持つ木構造によって構造化する．これによって，それぞれのインタラクションステートごとに木構造およびそれぞれの構造の出現回数を得ることができる．その後,「インタラクションステートの遷移において，被験者間の偏りは存在しない」という帰無仮説を基に $\chi^2$ 検定を行い，仮説が棄却される木構造を抽出する．

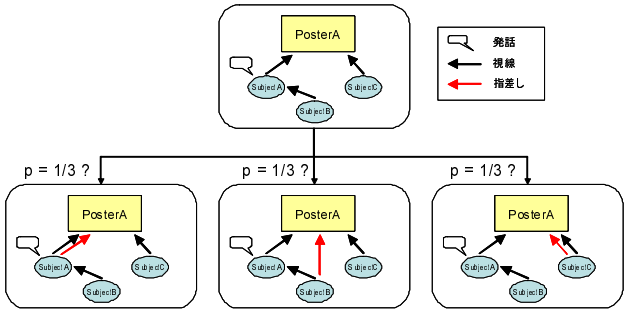


**図 3** $\chi^2$ 検定による構造抽出
Fig. 3 Detecting structure by $\chi^2$test

具体例を図 3 に示す．最初の状態ではどの被験者も指差しを行っていない．ここで前記の帰無仮説を採用すると，どの被験者も一様に行動の開始・終了を決定することになる．すると A，B，C の誰もが指さしを行っていないため，次の行動として指さしを開始する可能性があり，その確率は 1/3 で等しい．すなわち，図 3 の 2 段目のインタラクションステートへの遷移が同じ回数ずつ起きるということになる．しかし，インタラクションの持つ文法を考えた場合，ポスターを見ていない B が指差しを行うことは非常に考えづらい．また，非発話者である C が指差しを行うという遷移の発生回数も A が行う遷移よりも少ないと考えられる．

この抽出プロセスでは，このような偏りを $\chi^2$ 検定で抽出する．これにより，変換プロセスで生成された非常に大きなツリーから，インタラクションの構造によるものであると考えられる部分を検出し抽出できる．

## 4. 3 人自由会話の収録とラベリング

ここではインタラクションマイニングの基礎データとなる会話の収録とラベリングについて説明する．

### 4.1 データの収録

ここでは評価に利用した多人数インタラクションデータの収録について説明する．

まず，被験者の数は 3 人とした．この理由は 2 つある．まず 3 人以上になると，発話交替における立場の変化，指さしに対する注視・非注視，立ち位置の変化といった，会話に伴う興味深い社会現象が多く発生するといわれている[1)2)]．また 4 人以上の人数を対象とした

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

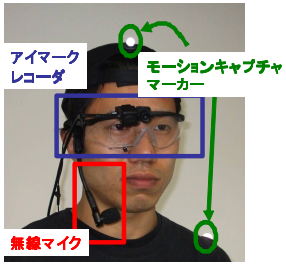


**図 4** センサ装着時の被験者
Fig. 4 The subject with sensor devices

**図 5** 会話環境
Fig. 5 Conversation field

場合，モーションキャプチャのオクリュージョンなどの要因でデータにノイズや欠損が多く含まれることになり，以後の処理に悪影響を与えることになる．

非言語行動を自動処理によってラベリングするため，被験者には頭・腕・背中にモーションキャプチャのマーカーを取り付け，また視線計測用のアイマークレコーダと無線マイクを取り付けた．センサを取り付けた被験者の様子を図 4 に示す．また実験環境には多数のカメラを設置し，マイクによる音声データと合わせて実験の様子を記録した．

会話環境としては，被験者が図 5 のように配置された 6 枚のポスターを見ながら自由に会話を行うという形式をとった．これは，我々が興味の対象としている，話題の発生や発話者の遷移，さらにその前後における話者や聞き手の非言語行動を多く観測するための設定である．また，同時に自動検出を容易にすることも目的としている．

本研究では，提案手法が多くの種類の会話に対し適用できることを示すために，収録ごとにポスターの内容や教示内容を少しずつ変えながら行った．行った収録は下記の 3 回である．

**session1:** 室町時代の京都市内を描いた絵画[⋆1]をポスターとし，描かれた人物や歴史的背景について雑談する

**session2:** 京都市内の航空写真をポスターとし，地図上の建物などについて雑談する

**session3:** 京都市内の航空写真をポスターとし，地図上の寺社仏閣を探すタスクを行う

### 4.2 ラベリング

今回のインタラクションマイニングでは，分析の対象として発話の有無，指差しの対象，視線の対象の 3 種類を選択した．理由は，これらの非言語行動は会話の中で重要な役割を果たすことが多くの研究で述べられていると共に，相槌や複雑なジェスチャなどと比較して自動ラベリングが容易なためである．

本研究では，ラベリングを極力自動処理で行い，かつインタラクションマイニングにおけるノイズを排除するため，一定のアルゴリズムによる自動ラベリングの後手作業で修正を行うという手法をとった．以下で，視線・指差し・発話のそれぞれについて自動認識の手法と手作業修正の基準について述べる．なお，文中における時間幅などのパラメータについては，実験の前に行われた数度の予備実験で得られた，自動処理で最も精度よく取れる値を利用している．

#### 4.2.1 視　　線

視線ラベルの生成に当たっては，まずアイマークレコーダとモーションキャプチャから視線ベクトルを計算し，その後他の人の頭部をモデル化した球体と，ポスターをモデル化した長方形との衝突判定を行った．さらにセンサからデータが取得できない場合があることを考慮し，250ms 以下の短いラベルに対して補間を行った．

その後，人手での訂正が必要な箇所については，それぞれのアイマークレコーダの映像を参考にして訂正を行った．

ただし，3 回目の実験における被験者のうち 1 人の視線データは，アイマークレコーダのキャリブレーション用データに重大な欠損があったため，完全手作業でラベルを作成した．

#### 4.2.2 指　差　し

指差しラベルの生成のため，まず指差しベクトルを計算した．このベクトルの始点は頭部のモーションキャプチャのマーカー位置から推定した眼の位置であり，その方向は手首につけられたモーションキャプチャのマーカーの位置である．次に，このベクトルとポスターをモデル化した長方形との衝突判定を行った．その後衝突判定から得られたラベルのうち間が 250ms 以下のものを補間し，さらに発生区間が 150ms より短いラベルについては削除を行った．

また，人手での訂正は，基本的に誤認識したラベルを削除することで行ったが，モーションキャプチャのマーカーが隠れている等の理由でラベルが断続的に生成されている場合は補間を行った．また，ラベルの開始点や終了点に問題があり，ずれている場合は，指先もしくは手のひらがボードに向いているかどうかを判断基準として調整を行った．

#### 4.2.3 発　　話

発話ラベルの生成は，まず各被験者が装着したマイクから得られた音声波形を 50msec ご

⋆1 上杉本洛中洛外図屏風を用いた．

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

とに分割し，FFT を用い音量を計算したうえで適当な閾値で 2 値化を行った．閾値は実験の最初の部分に人手でラベルを付加したうえで，それらの部分で再現率 90% を超え適合率を最大とする値とした．次に，隣接するラベル同士の間が 250ms 以下のものを補間した．最後に，発生区間が 150ms より短いラベルを削除した．

また，人手での訂正では，基本的にラベルの追加は行わず誤認識したラベルを削除することで行った．ラベルの開始点や終了点の調整が必要な場合には，それらの点は音の立ち上がり・立下りの点にあわせた．

## 5. インタラクションステートの頻度分布と実験間の会話の差異

ここでは 3 回の実験におけるインタラクションステートの頻度に着目して，それぞれの実験における会話の違いについて考察する．

それぞれの実験におけるインタラクションステートのうち，総ステート数の 2% 以上の出現頻度を持つインタラクションステートを取り出したところ表 1 のようになった．ここで，

$freq = \frac{Numberofthestates}{Numberofallstates} \times 100(\%)$

である．出現頻度が 2% 以下であるものはまとめて“ else ”とした．また，表の state に対応するインタラクションステートを図 6 に示す．後の考察を分かりやすくするため，ステートにポスターが含まれないものを A，含まれるものを B として区別してある．

これらのデータからいくつかの考察が得られた．

### 5.1 共通のインタラクションステート

表 1 から，B1 や B2 は 3 回の実験で共通して頻度の高いインタラクションステートであることがわかる．したがって，これらのインタラクションステートはポスターの種類やタスクの有無に関わらず，会話環境に依存して出現するステートだと考えられる．

### 5.2 ポスターの種類による差異

ポスターとして京都を描いた絵画を用いた session1 と，京都の航空写真を用いた session2, session3 を比較すると，頻度の高いインタラクションステートの多くが異なっていることが分かる．session1 では人と人との対面会話に関するインタラクションステートが多く出現し，session2・session3 ではポスターに向かい合った会話のインタラクションステートが多く出現している．

このような結果になった理由はポスターの種類によってもたらされたのではないかと考えられる．実際の映像と音声で確認したところ，session2・session3 では基本的に話題がポスター上の建物や通りに関するものであったのに対し，session1 では絵画そのものの時代背景

**表 1** 各実験のステート頻度分布
Table 1 State frequency distribution of experiments

| session1 (picture, no task) | | session2 (Aerophotograph, no task) | | session3 (Aerophotograph, task) | |
|---|---|---|---|---|---|
| state | freq | state | freq | state | freq |
| A1 | 4.027 | B1 | 6.176 | B1 | 10.251 |
| A2 | 3.930 | B2 | 4.950 | B2 | 7.572 |
| A3 | 3.572 | B5 | 3.724 | B6 | 5.252 |
| A4 | 3.310 | B6 | 2.832 | B9 | 3.902 |
| A5 | 2.910 | B7 | 2.676 | B5 | 3.860 |
| B1 | 2.855 | B8 | 2.163 | B7 | 3.649 |
| A6 | 2.744 | else | 77.479 | B10 | 3.227 |
| B2 | 2.386 | | | B11 | 2.784 |
| A7 | 2.372 | | | B12 | 2.320 |
| B3 | 2.317 | | | B13 | 2.215 |
| A8 | 2.082 | | | else | 45.032 |
| B4 | 2.069 | | | | |
| else | 65.426 | | | | |

やそこから発展した被験者の歴史知識・雑学といった，ポスターの内容とは離れた話題が多く含まれていた．そのため session1 では他の会話に比べて他の被験者に興味の対象が当たることが多く，このようなステートの頻度差が生まれたと考えられる．

### 5.3 タスクの有無による差異

ポスターとして航空写真を使った session2 と session3 を比較したところ，頻繁に出現しているインタラクションステートは共通しているのに対し，その出現頻度の分布には大きな差が見られた．具体的には，session2 においては 2% 以上の出現頻度を持つインタラクションステートにステート全体の約 22.6% が含まれるのに対し，session3 においては全体の約 55.1% が含まれていた．

このような結果になった理由はタスクの有無によってもたらされたのではないかと考えられる．実際の映像と音声で確認したところ，session2 では会話の進行に合わせて

- 全員で議論する場面
- 分かれて建物を捜索する場面
- 2 人が熱心に議論している一方で 1 人が離れている場面

といった様々な会話場面が頻繁に入れ替わっていた．それに対し，session3 では大まかに分けて

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

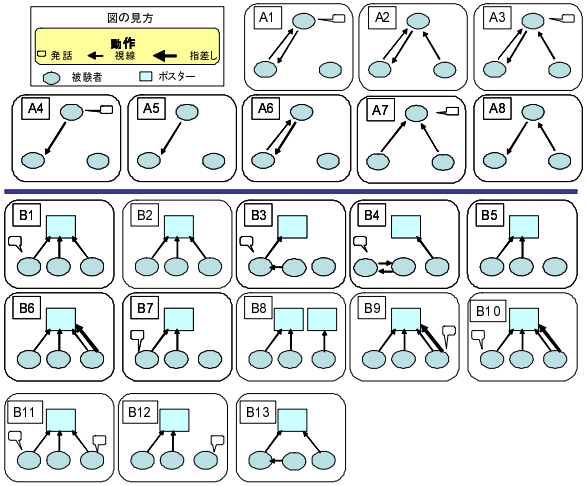


**図 6** 表 1 のインタラクションステート
Fig. 6 Interaction State in table 1

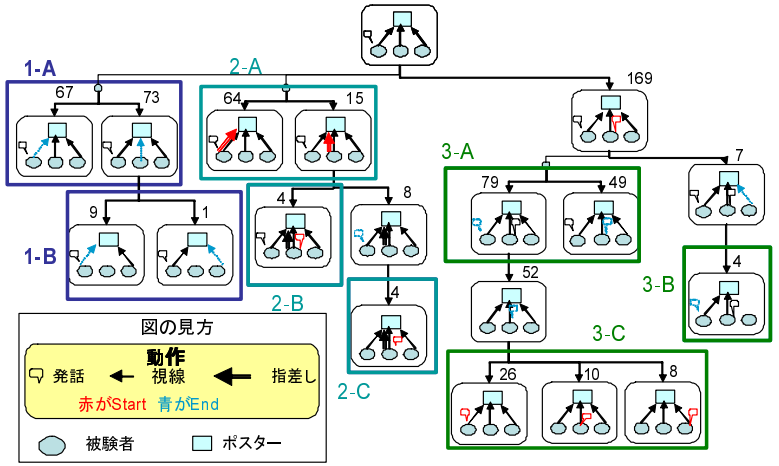


**図 7** 最頻インタラクションステートを基点とした構造
Fig. 7 The structure based on the interaction state appearing most often

(1) あるポスターで寺社仏閣を探す
(2) 全員で次のポスターへ移動

という 2 種類を繰り返す流れが存在した．このようなタスクの有無による会話場面の遷移の頻度がインタラクションステートの出現頻度分布に現れたのではないかと考えられる．

## 6. 特徴的インタラクション構造の検証

3 回分の実験データに関して全ての木構造と出現頻度を合算し，特徴的な構造の抽出を行った．このとき，$\chi 2$ 乗検定の有意水準は 5% とした．ここでは，これによって抽出された特徴的構造を下記の 2 通りの手法で検証する．

(1) 最も多く出現したインタラクションステートを初めとした構造に関して，4 段階の遷移までを検証する
(2) 1 段階の遷移で有意差が見られた構造のうち，総ステート数が多いもの上位 5 つに関して検証する．

### 6.1 最頻インタラクションステートを基点とした構造

図 7 が，最も多く出現した（出現回数 972 回）インタラクションステートを頂点とする構造である．このインタラクションステートは,「3 人がボードを注視し，かつそのうちの 1 人が発話中である」という状態である．図 7 では発現頻度に有意差が見られた構造，およびそれに関連する構造のみを記述し，他は省略してある．矢印の側の数字はそれぞれの遷移の回数を表している．

以下は得られた構造を 3 つの着眼点で検証する．

#### 6.1.1 視線の遷移に関する構造

1-A，1-B は，いずれも被験者の視線の遷移に関する特徴的な構造を示したものである．

1-A に関してだが，初期の「全員がボードを注視」という状態から，発話中の人物が視線を外す回数が 67 回，発話していない人物が 73 回となっている．仮に定まった会話構造が存在しない場合，発話中の人物と発話中でない人物が視線を外す確率はそれぞれ 1/3，2/3 となるはずであり，ここには一定の会話構造が存在していると考えられる．

これらの構造が発生している場面を実験の映像から確認したところ，発話者・非発話者に関わらず，他の発話者に視線を送る場合や別のポスターに視線を向けて会話場を変化させる場面が多く見られた．

以上のようなことから，発話者は非発話者よりも頻繁に視線を配分し，会話をコントロールしているのではないかと考えることができる．このような現象は従来研究でも論じられて

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

いる[1)2)].

また，非発話者が視線を外した場合でも，その後発話者が視線を外す回数はもう 1 人の非発話者が視線を外す場合に比べて有意に多い（1-B）．この事実も，1-A で見られた会話構造を補強するものだと考えることができる．

### 6.1.2 指差しに関わる構造

2-A，2-B，2-C は，いずれも発話と指差しの関係についての構造である．

2-A に関してだが，初期のステートから注視対象を指差す回数は，発話者が 65 回，非発話者が 15 回となっており，発話者の指差し回数は非常に多いといえる．

また，非発話者が指差しを行った場合に注目すると，指差しを行った被験者が直後に発話する場面は見られたが，行っていない被験者が発話する場面は見られなかった（2-B）．さらに，初期状態で発話していた被験者が発話を終了した後，指差しを行った被験者が発話する場面はみられたが他の被験者が発話する場面は見られなかった (2-C)．

これらは生起回数が少ないので必ずしも定まった会話構造をあらわしているとは言い難いが，これが一定の会話構造である場合は「指差しと発話は強い共起関係がある」と考えることができる．事実，これらのステートの発生箇所を確認したところ，被験者が発話をしながら発話内容に関わる指差しを行っている場面が多かった．

### 6.1.3 同時発話時の発話権の遷移に関する構造

3-A，3-B，3-C は，いずれも 2 人が同時発話を行ったときの発話の遷移に関わる構造である．

3-A に関してだが，初期のステートから「2 人が同時発話」の状態に移った後，先に話していたほうが発話をやめる回数が 79 回，後から話し始めた方が発話をやめる回数が 49 回であった．一定の会話構造がない場合の発話終了確率はどちらも 1/2 であり，ここには一定の会話構造が存在すると考えられる．

また回数は少ないものの，3-B の構造から，この会話構造は間に何らかのイベントを挟んでも成り立つのではないかと考えられる．

さらに特徴的な場面として 3-C が挙げられる．この構造は，発話が重なって双方が発話をやめた後，初めに発話をしていた被験者が発話をする回数が，他の被験者よりも有意に多いことを示している．これは，「一般的な会話では，片方の話者が一方的に話すよりもある程度の発話のやりとりをしながら話すほうが自然である」ということを表していると考えられる．

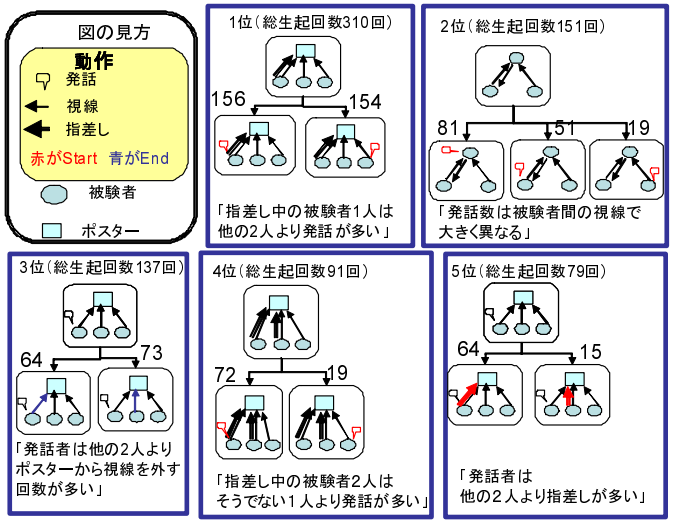


**図 8** 1 段階遷移における発現尤度の高い構造
Fig. 8 Structure with high departure present likelihood in one stage transition

### 6.2 1 段階遷移における特徴的構造

図 8 の 5 つの構造は，1 段階のみの遷移において特徴的な構造を持つもののうち，総生起回数が多い順に 5 つを取り出したものである．

これらの構造のうち，1 位・4 位・5 位は，指差しと発話の共起関係に関する構造であると考えられる．まず，1 位と 4 位の構造から，「指差しを行っている被験者は，その直後に発話を行う場合が他の被験者に比べて多い」ということがわかる．さらに，5 位では「発話者は非発話者に比べて指差しをする場合が多い」ということが分かる．

以上のことから，「発話と指差しには強い共起関係がある」という会話構造が読み取れる．これは普段の日常会話でも納得のできるものであり，また発話とジェスチャの共起関係については先行研究でも議論がなされている[1)]．

次に 2 位の構造に着目する．この構造は 3 人が互いの顔を見ながら話している場面で，発話回数が下記の順に多くなっていることを示している．3 人会話において視線が非常に大きな役割を果たしていることは榎本らによって研究がなされている[2)]．

(1) 他の被験者の 1 人と視線を交し合っており，かつもう 1 人の被験者から視線を受け

Vol.2009-HCI-135 No.13
Vol.2009-UBI-24 No.13
2009/11/12

ている人物

(2) (1) と視線を交し合っている人物

(3) (1) を注視しているが (1)(2) のどちらからも注視されていない人物

上記で述べた従来研究で議論されている構造が自動的に発見できたことは，インタラクションマイニングの有用性を示す結果であると考えられる．

最後に 3 位の構造に着目する．この構造から,「発話者は全員が見ているポスターの対象から視線を外す場合が非発話者よりも多い」ということがわかる．この構造から，発話者は他の被験者を見て様子を伺う，他のポスターを見て会話場の移動を促すといった，会話をコントロールする行動を多くとっているのではないかと考えられる．

## 7. おわりに

本論文では，マルチモーダルな多人数会話データをデータマイニングの手法で分析するインタラクションマイニングを試みた．まずインタラクションマイニングの重要な要素であるインタラクションステート，および $\chi^2$ 検定による特徴的構造検出について説明した．その後 3 人の被験者による自由会話に対してインタラクションマイニングを行い，得られたステートの頻度分布や特徴的構造について考察した．

結果，インタラクションステートの分布に関する考察では，共通のステートやポスターの種類による差異・タスクによる差異などがステートの出現頻度分布に現れることを示した．また特徴的構造の検証では，指差しと発話の共起性，3 者会話における視線と発話の関係など，従来研究でも議論されてきたインタラクションの構造が抽出できることを明らかにした．また，発話のオーバーラップ時の遷移に関わる構造という複雑なインタラクションの構造を抽出することができた．

今後の展望としては，まず対象とする非言語行動を増やしていくことが考えられる．今回は発話・視線・指差しという 3 つの行動のみに注目して分析を行ったが，このほかにうなずきや相槌などといった会話における重要な行動を追加することで，より複雑な構造の発見が期待できる．

また，今回は抽出された構造の中でも頻度の多いものに関して議論を行ったため，従来から議論されているような分かりやすい会話構造を抽出するのみにとどまった．しかし，より多くのデータを基にして特徴的構造の抽出を行えば，中程度の頻度を持つ部分ではこれまで議論がされていなかった新たな会話構造が見つかる可能性がある．

最後に，完全自動で作成されたラベルからの分析が課題として挙げられる．前述した取り組みを行っていくためには大量のインタラクションデータが不可欠だが，それらのデータに対しラベルの作成や修正を手作業で行うには限界がある．今後はより自動認識の精度を高める実験デザイン・認識手法を用いて，自動ラベリングによる構造化・構造抽出に取り組んでいきたい．

**謝辞** 本研究は，文部科学省科学研究費補助金「情報爆発時代に向けた新しい IT 基盤技術の研究」の一環で実施されました．また，本研究における実験協力や論文へのアドバイスなど多大な助力を頂いた，西田・角研究室の皆様に感謝いたします．

## 参 考 文 献

1) 坊農真弓：日本語会話における言語・非言語表現の動的構造に関する研究，ひつじ書房 (2008).
2) 榎本美香，伝 康晴：3 人会話における参与役割の交替に関わる非言語的行動の分析，人工知能学会， Vol.SIG-SLUD-A301, pp.25 – 30 (2003).
3) Chen, L., Harper, M., Franklin, A., R.Rose, T., Kimbara, I., Huang, Z. and Quek, F.: A Multimodal Analysis of Floor Control in Meetings, *MLMI*, Vol.3869, pp.36–49 (2006).
4) McNeill, D.: Gesture, gaze, and ground, *MLMI*, pp.1–14 (2005).
5) 細馬宏通，石津香菜，繁松麻衣子，中村智代，矢野雅人：身体を示し合う会話－自分の身体で相手の身体を語ること－，社会言語科学会第 14 回大会，pp.67 – 70 (2004).
6) 角 康之，西田豊明，坊農真弓，來嶋宏幸：IMADE : 会話の構造理解とコンテンツ化のための実世界インタラクション研究基盤，情報処理学会学会誌， Vol.49, No.8, pp. 945–949 (2008).
7) Michel, M., Ajot, J. and Fiscus, J.G.: The NIST Meeting Room Corpus 2 Phase 1, *MLMI*, pp.13–23 (2006).
8) Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., Lathoud, G., Lincoln, M., Lisowska, A., McCowan, I., Post, W., Reidsma, D. and Wellner, P.: The AMI Meeting Corpus: A Pre-announcement, *MLMI*, pp.28–39 (2005).
9) Chen, L., Rose, R., Qiao, Y., Kimbara, I., Parrill, F., Welji, H., Han, T.X., Tu, J., Huang, Z., Harper, M.P., Quek, F. K.H., Xiong, Y., McNeill, D., Tuttle, R. and Huang, T.S.: VACE Multimodal Meeting Corpus, *MLMI*, pp.40–51 (2005).
10) 森田友幸，平野 靖，角 康之，梶田将司，間瀬健二，萩田紀博：マルチモーダルインタラクション記録からのパターン発見手法，情報処理学会論文誌， Vol.47, No.1, pp. 121–130 (2006).
11) 福間良平，角 康之，西田豊明：人のインタラクションに関するマルチモーダルデータからの時間構造発見，情報処理学会研究報告，Vol.2009, No.23 (2009).