Vol.22 No.1, 2020

原著論文

対話型プレゼンテーションにおける非参与者から参与者への状態遷移推定

神原 隆宏 *1　西山 正志 *1　岩井 儀雄 *1

State Transition Estimation from Non-participants to Participants in Free-style Conversation Interaction

Takahiro Kambara*1, Masashi Nishiyama*1, Yoshio Iwai*1

Abstract – **We propose a novel method for inferring state transitions from non-participants to participants in free-style conversational interaction using physical behaviors acquired from cameras and a microphone. The existing methods do not consider non-participants and bystanders who play important roles in the interaction. In the research field of cognitive science, the existing methods consider the psychological aspects of changing from the non-participants to the participant. However, the existing methods cannot be easily implemented because inferring the psychological aspects is a very difficult task. Instead of using the psychological aspects, our method exploits physical behaviors such as standing position, facial direction, and audio direction. We analyzed the parameters of the behaviors to increase the performance of inferring the state transitions using datasets collected in the poster presentations.**

Keywords : Interaction, Non-participants, Bystanders, Participants, State transition

1. はじめに

日々様々な場所で行われている対話型プレゼンテーションは，発表者にとって重要な場面である．展示物の良さをより多くの人物に対して発表者がアピールする意味を持つ．発表者が説明する展示物として，研究成果のポスターや新製品のデモンストレーションなどが挙げられる．

対話型プレゼンテーションにおいて，発表者を除いた人物の状態は大きく分けて以下の 3 つに分類することができる．第 1 の状態として，発表者との会話に参加している参与者がある．この状態の人物は，同時に複数存在する場合や全く存在しない場合があり，入れ替わりも発生する．第 2 の状態として，発表者との会話に参加していないが，プレゼンテーションを見ている傍観者が存在する．展示物や発表者の近くで見ている場合と，展示物を遠巻きに見ている場合がある．この状態の人物は参与者と同様に人数が動的に変化する．第 3 の状態として，展示物や発表者を見ておらず，発表者との会話にも参加していない非参与者がある．この状態も同様に人数が動的に変化する．

会話中に複数の参与者が存在する中で参与構造を分析する研究 [1]~[5] がなされている．これらの研究では人物が着席し会話する状況を想定している．参与構造の分析が主目的であるため，この状況では非参与者と傍観者の推定が主目的にされることは少なかった．一方，対話型プレゼンテーションでは，参与者に加えて，非参与者と傍観者も推定する必要性が高いと想定される．対話型プレゼンテーションにおいて，発表者との会話に参加している参与者を推定する手法 [6]~[12] が提案されている．Hung ら [6] と Alameda-Pneda ら [7] は，会話中の参与者で構成されるグループをその場から検出する手法を提案している．さらに，Vascon ら [8]，Ricci ら [9]，Subramanian ら [10] によっても，同様の目的のための検出手法が提案されている．Tung ら [11] と井上ら [12] の研究では，発表者と参与者を対象とし，ジェスチャや発話の動作から状態を推定する手法が提案されている．このように既存手法は，発表者と参与者との間のインタラクションに主眼を置いていた．

展示物を介して会話を行った参与者の人数を記録することで，発表者自身がプレゼンテーションを振り返る際の助けに利用できると考えられる．さらに，対話型プレゼンテーションの場面では，展示物のアピールという観点からより多くの参与者と会話を行うことが発表者にとって望ましい．このため，参与者に遷移する可能性のある傍観者が重要な存在となる．非参与者から傍観者への状態遷移を推定することができれば，発表者に対して傍観者の存在を知らせることが可能となる．これにより，発表者はより多くの傍観者を発表に呼び込むことに将来的には繋げられると考えられる．

坊農ら [13] は，対話型プレゼンテーションの場に登場する全ての人物にカメラとマイクを取り付け，認知

*1: 鳥取大学大学院 持続性社会創生科学研究科
*1: Graduate School of Sustainability Science, Tottori University

図 1　各状態の人物の身体配置の例.
Fig. 1　Examples of body layout between the participant, bystander, and non-participant.

科学の視点から参与者や傍観者の状態を観察している．ただし，各人物の状態は全て人手で決める制約があり，傍観者と参与者に接触センサを取り付ける制約があった．

そこで本論文では，人物が動的に入れ替わる対話型プレゼンテーションにおいて，非接触のカメラとマイクを用いて非参与者から参与者への状態遷移を推定する手法について述べる．非参与者から参与者へ役割が変化する過程を，動作に基づく状態遷移モデルとして設計する．人物の動作として，カメラで取得された顔向き，身体向き，立ち位置，および，マイクで取得された音声方向を用いる．提案手法の有効性を確認するために評価データセットを収集し，複数の教示者により状態ラベルを与え，状態遷移モデルの有効性を確認した．以下では，2. で非参与者から参与者への状態遷移モデルについて述べ，3. で状態遷移の判定手法について述べる．提案手法の有効性を 4. で確認し，最後に 5. でまとめる．

2.　人物の状態遷移モデル

2.1　考え方

非参与者から参与者への状態遷移について考察するために，既存研究について調査を行った．Goffman [14] は，3 者以上の会話における参与構造のアイデアを提示している．この参与構造では，会話は話し手と受け手に加えて，その受け手として選ばれていない傍参与者で構成されている．会話に参加していない人物として傍観者と盗み聞き者が存在することが述べられている．Clark [15] は，Goffman のアイデアを基に会話の参与構造から図を作成し，会話中における共同注視モデルについて議論している．坊農ら [13] は，これらの既存研究を踏まえた上で，会話場に近づく，発話を受け取る，話しかけるといった動作を経て，状態が遷移するモデルのアイデアを提示している．なお，展示物付近で発表者との会話が行われる空間を本論文では会話場とする．

坊農らのモデル [13] では聞き手と受け手が存在するが，提案手法のモデルでは両者とも参与者として扱うこととする．さらに，坊農らのモデルでは傍参与者も存在する．傍参与者と傍観者の違いは，既存の参与者らに確実に存在を意識されているか否かにあると述べられている．ただし坊農ら [13] は，傍観者から傍参与者へ状態が遷移する時に，人物の動作に変化が表れるとは述べていなかった．このため傍参与者と傍観者の間で，動作の明確な違いを表現することが難しいと考えられる．提案手法では，傍参与者を傍観者の一部として扱うこととする．さらにこの文献では，非参与者から傍観者へ遷移する際，非参与者が会話場に近づいていく動作について言及されていた．提案手法のモデルでは，傍観者をより広く解釈し，状態の遷移時に発表者もしくは展示物を見ている動作を人物が行うこととする．その理由として，発表者もしくは展示物を近くで見ている人物だけでなく，これらを遠巻きに見ている人物も傍観者として扱うためである．

2.2　対話型プレゼンテーションの予備調査

対話型プレゼンテーションは発表者と参与者の会話を中心に行われる．会話に参加している人物は F 陣形を形成する可能性が文献 [16], [17] で示唆されている．会話中の人物の立ち位置の内側に現れる空間が O 空間と呼ばれている．O 空間を維持するために F 陣形が形成されている．また，O 空間の外縁的な輪の空間が P 空間と呼ばれている．F 陣形を形成する人物は身体を P 空間に配置する．McNeill [18] は，会話中の参与者が使用するターゲットオブジェクトが存在する場合，そのターゲットオブジェクトによっても O 空間が変形すると述べている．本論文では，対話型プレゼンテーションにおける展示物をターゲットオブジェクトとして取り扱う．

ポスターセッションを観察することで，各人物がどのような動作をとるか予備調査を実施した．対象としたポスターセッションは，国内で開催されたシンポジウムの一部であり，47 件のポスターが展示されていた．なお参加登録者数は 200 名程度であった．筆者らのポスター 1 件について 2 時間ほど観察した．そのポ

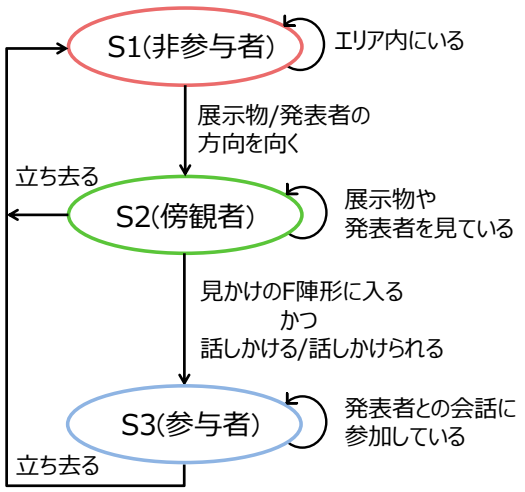


図 2　状態遷移モデル.
Fig. 2　State transition model.

スターに対して，参与者は 1 名から 5 名の間で推移しており，多くの時間帯で 1 名から 2 名の参与者が見られた．予備調査の結果，坊農の観察結果 [17] と同様に，発表者と参与者の会話において，展示物が O 空間の内部に存在する様子が確認できた．さらに，会話を行っている人数が増えるにつれて O 空間は大きくなる傾向にあることが確認できた．ここで注意すべきは，参与者の周囲に存在する傍観者が P 空間に侵入しているように見える場合が存在する点にある．その例における各状態の人物の身体配置を図 1 に示す．なお，図中では，個人情報保護の観点から，ポスターセッションを模擬した環境で撮影された写真を用いた．図中 (a) では傍観者は P 空間の外から展示物を見ていた．図中 (b) では傍観者は展示物に近づき発表者の説明を聞いていた．その後，図中 (c) では傍観者が発表者に話しかけることにより P 空間内で参与者に遷移していた．事後に傍観者へのヒアリングを行ったところ，傍観者が会話を聞きながら発表者と話す機会を伺っていたということであった．その他に，発表者は傍観者の存在に気付いていなかった，傍観者は会話を行いたいとは思わないが展示物の内容を読むために近づいたという例も見られた．このとき傍観者は，発表者から会話への参加が承認されていないにも関わらず P 空間内に侵入していた．傍観者を含む立ち位置の関係は F 陣形の一般の定義を外れるため，以下では見かけの F 陣形と呼ぶ．見かけの F 陣形は，参与者が意識的に形成するものではないが，状態遷移モデルを設計する上で重要になる．観察結果に基づいて設計した状態遷移モデルについて述べる．

2.3　状態遷移モデルの設計

論文中のモデルを図 2 に示し，以下で詳細を説明する．文献 [13]~[18] のアイデアを，工学的に取り扱えるように読み替えた上で，状態遷移モデルを設計する．なお図中のエリアとは，ある展示物に設置されたカメラで観測可能な範囲を指し，その展示物の発表者に対する非参与者，傍観者，参与者を状態推定の対象とする．

2.3.1　各状態における動作

対話型プレゼンテーションが行われているエリアにおいて，状態が非参与者，傍観者，参与者の人物が行う動作を以下に挙げる．

S1(非参与者)：　エリア内にいる
S2(傍観者)：　展示物や発表者を見ている
S3(参与者)：　発表者との会話に参加している

非参与者は，エリア内に身体配置をしているが，展示物や発表者を見ておらず，発表者との会話にも参加していない動作を続けている．傍観者は，発表者との会話に参加していないが，展示物や発表者を見ている動作を続けている．ただし，傍観者は見かけの F 陣形を形成していると取れる場合もあり，身体配置のみから傍観者と参与者を区別することは難しく，会話の開始と終了に気を付ける必要がある．参与者は，発表者と F 陣形を形成し会話行っている動作を続けている．次節では，状態遷移のトリガーとなる動作について詳細を述べる．

2.3.2　状態が遷移する際の動作

対話型プレゼンテーションが行われているエリアにおいて，状態が遷移する際に人物が行う動作を以下に挙げる．

S1 → S2：　展示物や発表者の方を向く
S2 → S3：　見かけの F 陣形を形成し，発表者に話しかける，または，発表者から話しかけられる

S1 から S2 へ遷移する際，対象人物は展示物や発表者の方を向く動作を行う．S2 から S3 へ遷移する際，対象人物は見かけの F 陣形を形成し，かつ，発表者に話しかける，もしくは，発表者が対象人物に話しかける動作を行う．状態 S2，S3 から S1 へ遷移する際，対象人物が 2.3.1 の S2 と S3 の動作を行っておらず，立ち去る動作を行う．最後に対象人物はエリア外へ出ていくこととする．

3.　状態遷移の判定手法

設計したモデルにおいて，人物の動作を用いて状態遷移を判定する手法を構築する．各時刻 t におけるエリア内の人数を I^t とする．なお，変数の右肩に付けられた t は時刻を表す．人物を表す指標を i とし，$i = 1$ を発表者，$i = 2, ..., I^t$ を状態遷移の推定対象の人物とする．展示物の中心位置を原点 **O** とする．各人物

の立ち位置をベクトル $\mathbf{p}_i^t$ で表し，各人物の顔向きを単位ベクトル $\mathbf{h}_i^t$ で表す．なお $\mathbf{h}_i^t$ は，立ち位置 $\mathbf{p}_i^t$ を始点として顔が向いている方向を指す．各人物の身体向きを単位ベクトルを $\mathbf{b}_i^t$ で表す．なお $\mathbf{b}_i^t$ は，立ち位置 $\mathbf{p}_i^t$ を始点として身体が向いている方向を指す．展示物から測定される音声方向を単位ベクトル $\mathbf{v}^t$ で表す．なお $\mathbf{v}^t$ は，展示物の原点 $\mathbf{O}$ を始点とし，発話した人物が存在する方向を指す．各人物の $\mathbf{p}_i^t$，$\mathbf{h}_i^t$，$\mathbf{b}_i^t$ はカメラで取得され，$\mathbf{v}^t$ はマイクで取得する．人物 i がエリア内に出現した時点で S1 が与えられる．以下では，遷移条件の手法について述べる．

3.1　S1 から S2 への遷移

非参与者が発表者もしくは展示物を見たかどうかを判定する．人物 i が発表者へどの程度向いているかについて，互いの位置を結ぶ線分 $\mathbf{p}_1^t - \mathbf{p}_i^t$ と顔向き $\mathbf{h}_i^t$ を用いて式 (1) で求める．

$$\theta_i^t = \cos^{-1} \frac{(\mathbf{p}_1^t - \mathbf{p}_i^t, \mathbf{h}_i^t)}{||\mathbf{p}_1^t - \mathbf{p}_i^t||} \tag{1}$$

なお，$\mathbf{p}_1^t$ は発表者の立ち位置とする．図 3 (a) にこれらのパラメータを示す．一方，人物 i が展示物へどの程度向いているかを，展示物から立ち位置への線分 $-\mathbf{p}_i^t$ と顔向き $\mathbf{h}_i^t$ を用いて式 (2) で求める．

$$\phi_i^t = \cos^{-1} \frac{(-\mathbf{p}_i^t, \mathbf{h}_i^t)}{||\mathbf{p}_i^t||} \tag{2}$$

図 3 (b) にこれらのパラメータを示す．展示物もしくは発表者を見たかどうかを式 (3) で決定する．

$$\mathrm{check}_1(\theta_i^t, \phi_i^t) = \begin{cases} 1 & \min(|\theta_i^t|, |\phi_i^t|) < \eta^1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

ここで，min は入力値のうちで小さい方の値を返す関数，η^1 は閾値を表す．関数 check_1 の出力が 1 となった人物 i は，状態が S2 になったと判定する．出力が 0 の人物 i は，状態が S1 のままであると判定する．

3.2　S2 から S3 への遷移

傍観者が見かけの F 陣形を形成しており，かつ，発表者へ話しかけたとき，または，話しかけられたときに参与者へ状態が遷移する．まず，傍観者が見かけの F 陣形を形成しているかどうかを確認する．展示物を中心とした半円で見かけの F 陣形を近似し，人物 i の立ち位置を用いて式 (4) で判定する．

$$\mathrm{check}_2(\mathbf{p}_i^t) = \begin{cases} 1 & ||\mathbf{p}_i^t|| < \eta_1^2 + \eta_2^2 I_f \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

I_f^t は見かけの F 陣形を形成する人数，η_1^2，η_2^2 は閾値を表す．なお，$0 \leqq I_f^t < I^t$ とする．図 3 (c) にパラメータ $\mathbf{p}_i^t$ を示す．

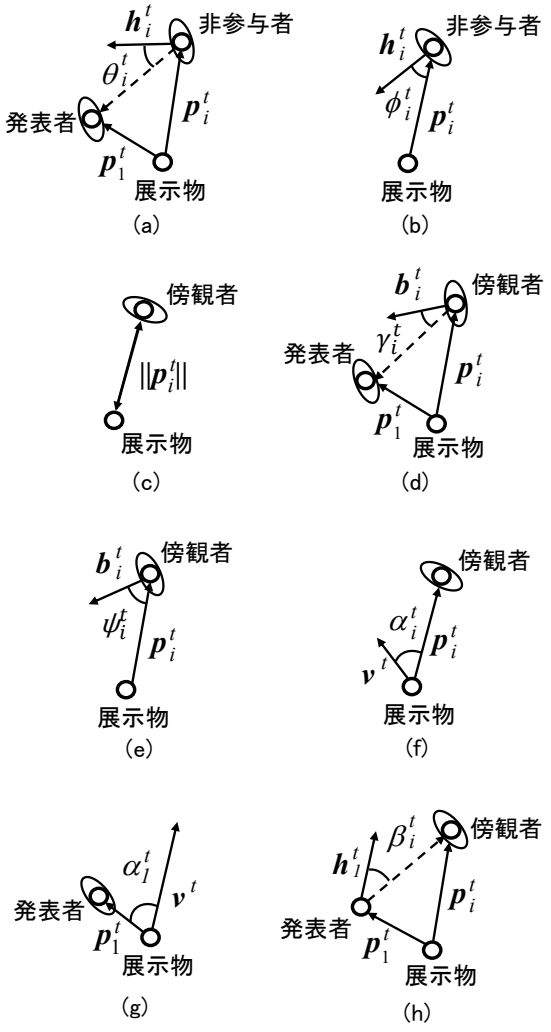


図 3　状態遷移の判定に用いるパラメータ．
Fig. 3　Parameters for inferring state transitions.

F 陣形を形成しているかどうかを判定するには立ち位置だけでなく身体向きも重要である．そこで，傍観者が発表者の方向，もしくは，展示物の方向に身体を向けている条件を設ける．人物 i が発表者へどの程度身体を向けているかについて，互いの位置を結ぶ線分 $\mathbf{p}_1^t - \mathbf{p}_i^t$ と身体向き $\mathbf{b}_i^t$ を用いて式 (5) で求める．図 3(d) にこれらのパラメータを示す．

$$\gamma_i^t = \cos^{-1} \frac{(\mathbf{p}_1^t - \mathbf{p}_i^t, \mathbf{b}_i^t)}{||\mathbf{p}_1^t - \mathbf{p}_i^t||} \tag{5}$$

一方，人物 i が展示物へどの程度身体を向けているかを，展示物から立ち位置への線分 $-\mathbf{p}_i^t$ と身体向きを用いて式 (6) で求める．

$$\psi_i^t = \cos^{-1} \frac{(-\mathbf{p}_i^t, \mathbf{b}_i^t)}{||\mathbf{p}_i^t||} \tag{6}$$

図 3 (e) にこれらのパラメータを示す．展示物もしくは発表者に身体が向いているかどうかを式 (7) で決

定する．

$$\mathrm{check}_3(\gamma_i^t, \psi_i^t) = \begin{cases} 1 & \min(|\gamma_i^t|, |\psi_i^t|) < \eta_3^2 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

次に人物 i が発表者へ話しかけたか，もしくは，発表者から話しかけられたかを判定する．ここでは 2 名以上の人物が同時に発話しないと仮定する．提案手法では発話した人物を特定するために音声方向 $\mathbf{v}^t$ を用いる．発話した人物は展示物を中心とする見かけの F 陣形を形成しているため，$\mathbf{v}^t$ は展示物の位置 $\mathbf{O}$ から観測する．まず傍観者が発表者へ話しかけた場合について述べる．発話が起こった方向と人物 i が存在する方向との間の角度を，立ち位置 $\mathbf{p}_i^t$ と音声方向 $\mathbf{v}^t$ を用いて式 (8) で求める．

$$\alpha_i^t = \cos^{-1} \frac{(\mathbf{p}_i^t, \mathbf{v}^t)}{||\mathbf{p}_i^t||} \quad (8)$$

図 3 (f) にこれらのパラメータを示す．発表者へ話しかけたかどうかを式 (9) で決定する．

$$\mathrm{check}_4(\alpha_i^t) = \begin{cases} 1 & |\alpha_i^t| < \eta_1^3 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

ここで，η_1^3 は閾値を表す．次に傍観者が発表者から話しかけられた場合について述べる．発表者が傍観者へ話しかけた時は相手の顔を見ていると仮定し，発表者の顔向き $\mathbf{h}_1^t$ と，互いの位置を結ぶ線分 $\mathbf{p}_i^t - \mathbf{p}_1^t$ を用いて式 (10) で求める．

$$\beta_i^t = \cos^{-1} \frac{(\mathbf{p}_i^t - \mathbf{p}_1^t, \mathbf{h}_1^t)}{||\mathbf{p}_i^t - \mathbf{p}_1^t||} \quad (10)$$

図 3 (g)，(h) にこれらのパラメータを示す．発表者から話しかけられたかどうかを式 (11) で決定する．

$$\mathrm{check}_5(\alpha_1^t, \beta_i^t) = \begin{cases} 1 & |\alpha_1^t| < \eta_2^3 \wedge |\beta_i^t| < \eta_3^3 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

ここで，α_1^t は発表者について式 (8) で求めた角度，η_2^3 および η_3^3 は閾値を表す．関数 check_2 の出力が 1 であり，かつ，関数 check_4 もしくは関数 check_5 の出力が 1 となった人物 i は，状態が S3 になったと判定する．出力が 0 の人物 i は，状態が S2 のままであると判定する．

3.3 各状態から S1 への遷移

各状態の人物 i が立ち去ったかどうかについて，展示物を見ておらず，かつ，展示物から離れているかの条件を用いて判定する．ここでは，関数 check_1 と関数 check_2 の両方の出力が 0 となった人物 i は，状態が S1 になったと判定する．少なくとも一方の出力が 1 の人物 i は，状態が元のままであると判定する．

4. 実験

4.1 評価データセット

4.1.1 セッティング

提案手法の有効性を確認するために，ポスターセッションを模擬したエリアにカメラとマイクを設置し，対話型プレゼンテーションの評価データセットを収集した．展示物として 26 インチディスプレイと紙製タイトルを設置した．ディスプレイ上にポスタースライドを表示し，発表者が参与者に向けて説明を行った．カメラとマイクが埋め込まれた Microsoft Kinect v2 を 6 台用い，それらの配置は図 4 とした．非参与者と傍観者の動作を取得するために，エリア内の遠距離から中距離をカバーする K_1 から K_3 の 3 台を使用した．発表者の動作を頭上から取得するために K_4 の 1 台を使用した．また傍観者と参与者の動作を取得するために，近距離をカバーする K_5 と K_6 の 2 台を使用した．カメラは 30 フレーム毎秒とし，マイクのサンプリングレートは 16 キロヘルツとした．カメラの配置と台数を決定する際，エリア内に同時に存在する 6 名の人物の顔を，隠れなく観測できることとした．ただし，エリア内の人物がさらに増加した場合には隠れの問題が発生すると考えられる．今後の課題としてカメラの台数や配置を最適化する手法の開発が必要である．

展示物の中心が $\mathbf{O}$ となるように世界座標系を定義し，各 Kinect の回転行列と並進ベクトルをキャリブレーション手法 [19] で求めた．人物の立ち位置 $\mathbf{p}_i^t$ を，各 Kinect から出力された胸部位置から平均を求めることで決定した．統合する際，Kinect 間の人物対応付けのみ人手で行った．顔向き $\mathbf{h}_i^t$ は，同一人物に対して Kinect から出力された角度の平均とした．身体向き $\mathbf{b}_i^t$ は，Kinect から出力された左肩の位置と右肩の位置を結ぶ線分に対する法線方向とした．音声方向 $\mathbf{v}^t$ は，K_5 の内部マイクから出力された角度とした．

エリア内に人物が 1 名存在する場合について，各測定値の誤差を評価した．エリア内の 44 点 (0.5 メートル間隔) で計測したところ，立ち位置 $\mathbf{p}_i^t$ の誤差は 0.1 ± 0.1 メートル，顔向き $\mathbf{h}_i^t$ の誤差は 20.7 ± 14.4 度であった．次に，録音した音声を展示物の中心から 2 メートル離れた 7 点 (15 度間隔) で 1 点ずつ再生したところ，$\mathbf{v}^t$ の誤差は周囲からの雑音が無い場合で 2.2 ± 1.3 度になった．周囲から雑音が入る場合を評価するために，展示物から 2 メートルと 3.5 メートルの位置にスピーカをそれぞれ設置し，録音した音声を同音量で 30 秒間再生した．その結果，2 メートルの位置にあるスピーカの音声方向を上記の精度で 30 秒中 23.9 秒獲得できていた．

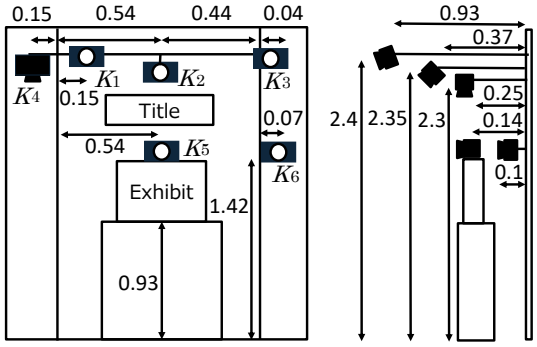


図 4　カメラとマイクが埋め込まれた Kinect の配置 [メートル].
Fig. 4　Setup of Kinect devices of the cameras and the microphone [m].

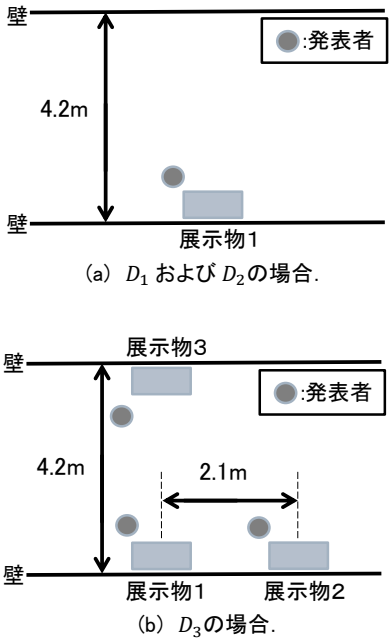


図 5　展示物の配置図.
Fig. 5　Layouts of the exhibits.

図 6　評価データセットの動画の例.
Fig. 6　Examples of the video sequences of our dataset.

4.1.2　撮影時に実験協力者へ与えた条件

実験協力者の中から発表者をランダムに選んだ．展示物 1 か所につき発表者を 1 名配置した．その他の実験協力者には，展示物を見学し興味があれば質問を行うように指示した．発表者を除いた全ての実験協力者がエリア外にいる状態から撮影を開始した．収集された評価データセットを以下に挙げる．

- D_1:　実験協力者を 11 名，展示物を 1 か所として撮影した．展示物の配置を図 5 (a) に示す．全員がエリア外に出るまで撮影し，その動画の長さは 140 秒となった．動画の一部を図 6 (a) に示す．
- D_2:　実験協力者を 11 名，展示物を 1 か所として撮影した．展示物の配置は D_1 と同じとした．全員がエリア外に出るまで撮影し，その動画の長さは 415 秒となった．動画の一部を図 6 (b) に示す．
- D_3:　実験協力者を 9 名，展示物を 3 か所として撮影した．展示物の配置を図 5 (b) に示す．6 台の Kinect を図中の展示物 1 に設置した．撮影された動画の長さは 600 秒とした．実験協力者へ，時間中に 3 か所の展示物を自由に見て回り議論へ参加するように指示を追加した．動画の一部を図 6 (c) に示す．

展示物で取り扱った話題について述べる．D_1 では情報処理技術の学術的な話題を扱った．D_2 では多くの人が会話に参加できるようスマートフォンの話題を扱った．実際のポスターセッションに近づけるために，D_3 では展示物の設置箇所を増やし学術的な話題を扱った．

4.2　複数の教示者による状態ラベルの付与

撮影された評価データセットに対して状態を人手で教示した．教示者は 5 名とし，ELAN [20] を用いた．ELAN とは，教示者が動画と音声を視聴しながら，各時間の各人物に状態ラベルを手作業で与えるソフトウェアである．各データセットの各人物に対して，付与者が状態ラベルを以下の手順で与えた．まず始めに，推定対象となる各人物に対して，状態ラベルを与える対象時刻の始端と終端を決定した．いずれかのカメラで対象人物が観測された時刻を始端とし，全てのカメラで対象人物が観測できなくなった時刻を終端とした．次に，始端と終端の間で，各時刻における状態ラベルを付与者が与えた．ある人物 i のある時刻 t において，付与者が最も当てはまると判断した 1 つの状態ラベルを与えるよう指示した．付与者はこの作業を始端と終端の間の全時刻で繰り返した．なお作業前には，付与者に 2.3.1 で述べた状態 S1，S2，S3 の動作のみを説明し，2.3.2 の状態遷移の判定に用いる動作について

表 1　教示された状態ラベルの Kappa 係数.
Table 1　Kappa coefficients of the state labels given by the workers.

	Kappa 係数
D_1	0.84±0.07
D_2	0.87±0.07
D_3	0.96±0.03

表 2　評価データセットにおける状態遷移の回数.
Table 2　The number of state transitions for each dataset.

	S1 → S2	S2 → S3	S2 → S1	S3 → S1
D_1	10	3	7	3
D_2	10	5	5	5
D_3	13	3	7	2

は説明しなかった．また，付与者にエリアの定義を説明し，発表者がどこに存在するかを説明した．なお，教示者はこれらの評価データセットを作業時に初めて見た．

教示者 5 名が与えた状態ラベルを比較し，教示者間でどの程度一致していたかを評価した．付与された状態ラベルから Fleiss の Kappa 係数 [22] を求め，教示者間のばらつきを評価した．その結果を表 1 に示す．Landis らの文献 [23] では Kappa 係数が 0.81 以上であれば，互いにほぼ一致していると述べられている．このことから，5 名の教示者間で状態ラベルはほぼ一致していると言える．次節の定量評価では，各時刻における状態ラベルの正解を，教示者 5 名の多数決の結果とした．発生した状態遷移の回数を表 2 に示す．各時刻における状態ごとの人数の変化を図 7 に示す．

4.3　提案手法を用いた状態遷移推定の評価

提案手法の閾値 η^1，η_1^2，η_2^2，η_3^2，η_1^3，η_2^3，η_3^3 を変化させたときの推定精度の違いを評価データセットで検証した．各時刻において，教示された各人物の状態と，提案手法で推定された状態とを比較した．評価指標として，適合率と再現率の調和平均である F 値を用いた．F 値が 1.0 に近づくにつれ，教示された状態ラベルと提案手法で推定された状態が一致していると言える．

ある閾値を固定しそれ以外の閾値を変化させたときの平均 F 値を図 8 に示す．多重比較検定のボンフェローニ法を適用した．η^1 では，5，25，45 度が 65 度よりも高く，かつ，25，45 度が 5 度よりも高く，有意差が見られた．η_1^2 では，1.35，1.45，1.55 メートルが 1.25 メートルよりも高く，かつ，1.35，1.45 メートルが 1.55 メートルよりも高く，有意差が見られた．η_2^2 では，0.1，0.2 が 0，0.3 よりも高く，有意差が見られた．η_3^2 では，15，25，35 度が 5 度よりも高く，有意差が見られた．η_1^3 では，1，11，21 度が 31 度よりも高く，かつ，11 度が 1 度よりも高く，有意差が見られた．η_2^3 では，15，25，35 度が 5 度よりも高く，有意差が見られた．η_3^3 では，5，15，25 度が 35 度よりも高く，かつ，15，25 度が 5 度よりも高く，有意差が見られた．よって，精度が高くなる閾値は，η^1 が 25 度から 45 度の間，η_1^2 が 1.35 メートルから 1.45 メートルの間，η_2^2 が 0.1 から 0.2 の間，η_3^2 が 15 度以上，η_1^3 が 11 度から 21 度の間，η_2^3 が 15 度以上，η_3^3 が 15 度から 25 度の間であった．η_2^2 が 0 の時は精度が低下することから，見かけの F 陣形を形成する人数に応じてその陣形を制御することの有効性が確認された．以上の結果より，状態遷移の推定精度に対する提案手法の閾値の影響を確認した．

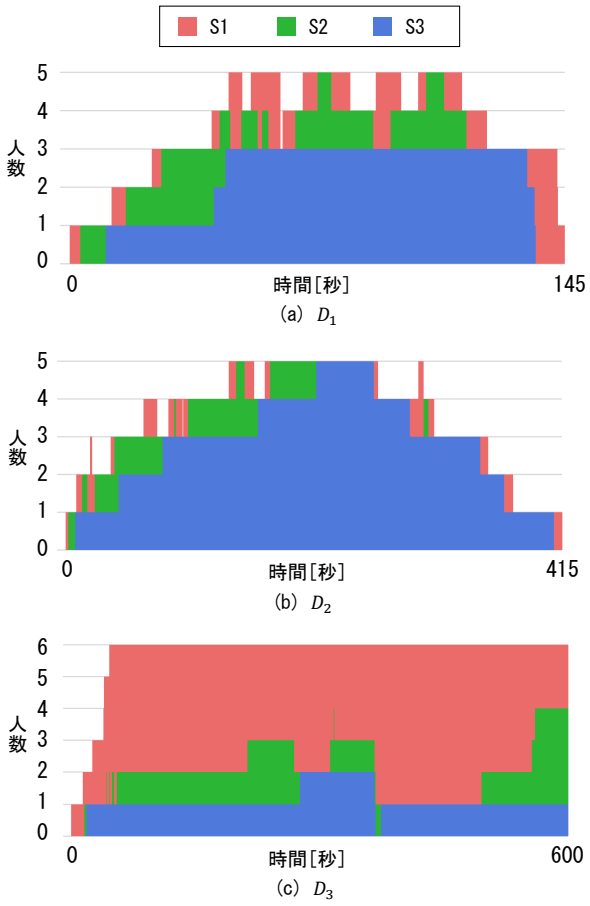


図 7　評価データセットにおける各時刻の各状態の人数の変化.
Fig. 7　The number of participants temporally changing the states in our dataset.

全データセットでの平均 F 値の最大は 0.94 であり，データセットごとの F 値は D_1 で 0.86，D_2 で 0.96，D_3 で 0.92 であった．D_2 および D_3 では F 値が 1.0 に近く，提案手法の状態ラベルと付与者の状態ラベルは近いと言える．D_1 の F 値が D_2，D_3 の F 値と比べて低い理由は 4.4.1 で考察する．なお，提案手法の閾値は全データセットで共通とし，η^1 が 25，η_1^2 が 1.55，η_2^2

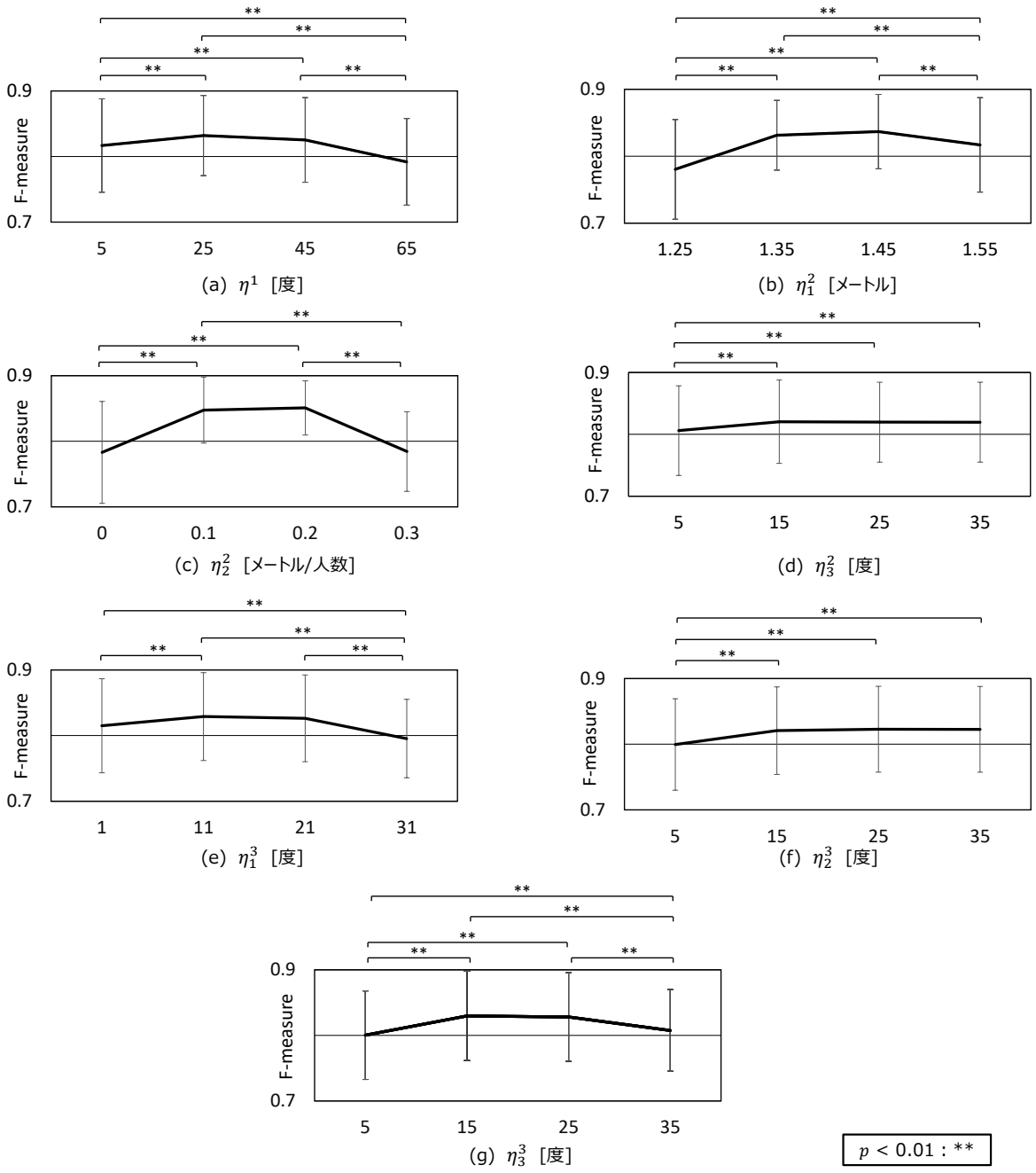


図 8　ある閾値を固定しその他の閾値を変化させたときの平均 F 値.
Fig. 8　Average F-measures when a certain threshold is fixed and the other thresholds are changed.

が 0.1，η_3^2 が 15，η_1^3 が 11，η_2^3 が 5，η_3^3 が 15 とした．

提案手法の状態ラベルと付与者 5 名が与えた状態ラベルとの間で Fleiss の Kappa 係数を求めた．データセットごとの Kappa 係数は D_1 で 0.71，D_2 で 0.81，D_3 で 0.91 であった．Landis ら [23] は Kappa 係数が 0.81 以上であれば Almost perfect agreement，0.61 から 0.80 の間であれば Substantial agreement であると述べている．このことから，提案手法と付与者の間で D_2，D_3 についてはほぼ一致しており，D_1 についても概ね一致していると言える．

4.4　考察

4.4.1　提案手法における推定誤りの原因

提案手法の推定誤りを引き起こす原因を特定するために，データセット D_2 や D_3 と比べて F 値の低い D_1 において S1，S2，S3 の状態ごとに評価を行った．その結果，状態ごとの F 値は S1 で 0.78，S2 で 0.66，S3 で 0.93 であり，特に状態 S1 および S2 の推定精度が低かった．その原因として次の二つが挙げられる．

・S1 から S2 へ遷移する時の推定誤り
・S2 から S3 へ遷移する時の推定誤り

図 9　S2 から S3 への遷移の推定誤り例.
Fig. 9　The example of estimation error of transition from S2 to S3.

まず前者の原因を述べる．対象人物の位置は Kinect から得られているものの，推定された顔向きの誤差が大きく，状態遷移を正しく判定できなかった．今後の課題として，顔向き推定を安定に行える技術の開発が必要である．次に後者の原因を述べる．ある人物が発表者に向けて発話した際，その人物付近に位置する人物が発話したと誤って検知されたため，状態遷移を正しく判定できなかった．後者の具体例を図 9 に示す．付与者の状態ラベルでは，人物 1 は時刻 t_1，t_2 で傍観者であり，t_3 で非参与者であった．人物 2 は t_1，t_2，t_3 ともに参与者であった．提案手法を適用したところ，人物 1 の t_2 で，傍観者ではなく参与者であると誤って推定された．これは，人物 2 が t_2 で発話した際に獲得された音声方向の延長線上に人物 1 が存在したためである．今後の課題として，精度よく話者を特定できる技術の開発が必要がある．

4.4.2　人物対応付けの自動化に向けた課題

本実験で使用した Kinect に付属する人物追跡ライブラリでは，自動出力される追跡結果に誤りが含まれる場合が存在した．具体的には，ある 1 台の Kinect で 2 名の人物がすれ違う様子を観測した際に，一方の人物に隠れが発生し，他方の人物がもう一方の人物と誤って対応付けられる事例が確認された．ある 1 台の Kinect でこの事例が発生すると，他の Kinect との追跡結果を統合する際，複数の人物が同一人物となる問題が生じた．このため，本実験ではカメラ間の人物対応付けを人手で行った．人物対応付けを自動化するためには，すれ違った後に同じ人物を正しく追跡できているかどうかを検知する技術が必要である．その際に，Person Re-Identification の技術 [24], [25] の導入が今後の課題として考えられる．

4.4.3　会話における同時発話

提案手法において，2 名以上が同時に発話しないと仮定した理由について述べる．S2 の傍観者から S3 の参与者への遷移を判定する際，式 (9) の関数 check_4 と，式 (11) の関数 check_5 で音声方向を使用する．これらの関数では，S2 から S3 への状態遷移の時刻において同時発話の発生確率が非常に低いことを想定している．本実験で用いた評価データセットを確認したところ，S2 から S3 へ遷移する時刻付近での同時発話は見られなかった．なお，S3 に遷移した後の人物については同時発話が多数見られた．提案手法では，それらの人物に対して状態遷移が既に完了しているため，推定精度に影響が見られなかった．ただし，エリア内の人数が非常に多い場合には質問が飛び交い，同時発話により S2 から S3 への状態遷移が正しく推定できない恐れがある．今後の課題として，同時発話への対応は必須であると考えられる．

4.4.4　展示物の大きさ

本実験では，顔を動かさずとも展示物全体を見ることができる大きさのディスプレイを使用した．ディスプレイ中心に対する人物の顔向きを，提案手法の判定条件に用いたことは妥当であると考える．しかし，展示物が大型化した場合，顔を様々な方向に動かさなければ全体を見ることができない場合が発生する．この場合は，展示物の大きさを考慮した判定条件が必要になると考えられる．今後の課題として，展示物の寸法パラメータを持たせるなど展示物の大きさに関する判定条件を検討する必要がある．

4.4.5　参与者から傍観者への遷移

実際のポスターセッションでは，参与者から傍観者へ遷移する可能性があると考えられる．その際，2 つのパターンが存在すると想定される．第 1 のパターンとして，F 陣形を形成する O 空間から参与者が離れて，その展示物を傍観する場合が考えられる．この場合，非参与者へ遷移した後に傍観者へ遷移したとして提案手法は状態を推定できる．第 2 のパターンとして，参与者が立ち位置を変えずに会話を終了し傍観者に遷移する場合が考えられる．その場合，提案手法では状態遷移を正しく推定できない．この問題を解決するためには，会話構造を推定する技術の導入が今後の課題として必要である．

4.4.6　参与者同士の会話

発表者と複数の参与者との間の会話において，参与者が一時的に他の参与者へ話しかける場合が存在する．提案手法は参与者同士の会話を考慮していないが，参与者として推定されていれば，他の参与者と会話を行った場合でも，展示物から離れて非参与者へ状態が遷移しない限り，そのまま参与者として推定され続ける．会話構造の詳細をさらに推定するためには，参与者同士の会話を検知する技術が今後の課題として必要である．

4.4.7　展示物の密集

展示物が密集する場合について考える．この場合，ある展示物に対する F 陣形の O 空間が，他の展示物に対する発表者と参与者が形成する F 陣形の O 空間を内包する可能性がある．提案手法では複数の O 空間の重なりを考慮していないため，状態推定を正しく行うことが出来ない．今後の課題として，参与者がど

の O 空間に属しているかを判定する手法など，さらなる研究開発が必要である．

5. むすび

本論文では，対話型プレゼンテーションにおいて，非参与者から参与者への状態遷移を推定する手法について述べた．認知科学分野の既存モデル [13] とポスターセッションを観察した結果を用いることで，人物動作を用いて状態が遷移するモデルを設計した．非接触のカメラとマイクを展示物周辺に複数配置し，人物の立ち位置，顔向き，音声方向を取得することで，状態遷移を判定した．対話型プレゼンテーションの評価データセットを収集し，提案手法の有効性を実験により確認した．

本論文では決定論的な手法を採用したが，大塚らの研究 [21] のような確率論的な手法も有用であると考えられる．今後の課題として，人数がさらに増えた場合の評価，参与者間の関係性の考慮，実際のポスターセッションでの実証実験などが挙げられる．

参考文献

[1] L. Chen, R.T. Rose, Y. Qiao, I. Kimbara, F. Parrill, H. Welji, T.X. Han, J. Tu, Z. Huang, M. Harper, F. Quek, Y. Xiong, D. McNeill, R. Tuttle, and T. Huang. Vace multimodal meeting corpus. In *Machine Learning for Multimodal Interaction*, pp. 40–51, 2006.

[2] F. Pianesi, M. Zancanaro, B. Lepri, and A. Cappelletti. A multimodal annotated corpus of consensus decision making meetings. *Language Resources and Evaluation*, Vol. 41, No. 3, pp. 409–429, 2007.

[3] M. Poel, R. Poppe, and A. Nijholt. Meeting behavior detection in smart environments: Nonverbal cues that help to obtain natural interaction. In *Proceedings of 8th IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–6, 2008.

[4] Y. Sumi, M. Yano, and T. Nishida. Analysis environment of conversational structure with nonverbal multimodal data. In *Proceedings of International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, pp. 44:1–44:4, 2010.

[5] 石井亮, 大塚和弘, 熊野史朗, 松田昌史, 大和淳司. 複数人対話における注視遷移パターンに基づく次話者と発話開始タイミングの予測. 電子情報通信学会論文誌 A, Vol. J97-A, No. 6, pp. 453–468, 2014.

[6] H. Hung and B. Kröse. Detecting f-formations as dominant sets. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, pp. 231–238, 2011.

[7] X. Alameda-Pineda, Y. Yan, E. Ricci, O. Lanz, and N. Sebe. Analyzing free-standing conversational groups: A multimodal approach. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pp. 5–14, 2015.

[8] S. Vascon, E.Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, and V. Murino. Detecting conversational groups in images and sequences: A robust game-theoretic approach. *Computer Vision and Image Understanding*, Vol. 143, pp. 11–24, 2016.

[9] E. Ricci, J. Varadarajan, R. Subramanian, S. R. Bulo, N. Ahuja, and O. Lanz. Uncovering interactions and interactors: Joint estimation of head, body orientation and f-formations from surveillance videos. In *Proceedings of International Conference on Computer Vision*, pp. 4660–4668, 2015.

[10] R. Subramanian, J. Varadarajan, E. Ricci, O. Lanz, and S. Winkler. Jointly estimating interactions and head, body pose of interactors from distant social scenes. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pp. 835–838, 2015.

[11] T. Tung, R. Gomez, T. Kawahara, and T. Matsuyama. Multiparty interaction understanding using smart multimodal digital signage. *IEEE Transactions on Human-Machine Systems*, Vol. 44, No. 5, pp. 625–637, 2014.

[12] K. Inoue, Y. Wakabayashi, H. Yoshimoto, and T. Kawahara. Speaker diarization using eye-gaze information in multi-party conversations. In *Proceedings of Interspeech*, pp. 562–566, 2014.

[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. 認知科学, Vol. 11, No. 3, pp. 214–227, 2004.

[14] E. Goffman. *Forms of Talk*. University of Pennsylvania Press, 1981.

[15] H.H. Clark. *Using Language*. Cambridge University Press, 1996.

[16] A. Kendon. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, 1990.

[17] 坊農真弓. 会話構造理解のための分析単位：F 陣形. 人工知能学会誌, Vol. 23, No. 4, pp. 545-551, 2008.

[18] D. McNeill. Gesture, gaze, and ground. In *Machine Learning for Multimodal Interaction*, pp. 1–14, 2006.

[19] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330–1334, 2000.

[20] H. Lausberg and H. Sloetjes. Coding gestural behavior with the neuroges-elan system. *Behavior Research Methods*, Vol. 41, No. 3, pp. 841–849, 2009.

[21] 大塚和弘, 竹前嘉修, 大和淳司, 村瀬洋. 複数人物の対面会話を対象としたマルコフ切替えモデルに基づく会話構造の確率的推論. 情報処理学会論文誌, Vol. 47, No. 7, pp. 2317–2334, 2006.

[22] J. L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.

[23] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, Vol. 33, No. 1, pp. 159–174, 1977.

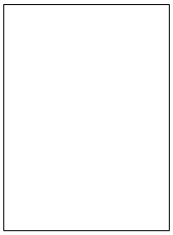
[24] S. Gong, M. Cristani, S. Yan and C. C. Loy. Person Re-Identification. *Springer*, 2014.

[25] L. Zheng, Y. Yang, A. G. Hauptmann. Person Re-identification: Past, Present and Future. arXiv:1610.02984, 2016.

（2019 年 5 月 8 日受付，8 月 30 日再受付）

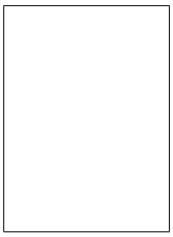
著者紹介

神原　隆宏



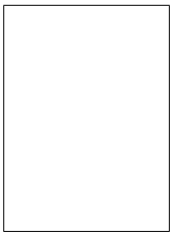
2018 年鳥取大学工学部知能情報工学科卒業．現在，鳥取大学大学院持続性社会創生科学研究科博士前期課程工学専攻情報エレクトロニクスコースに在学中．

西山　正志



2002 年 岡山大学大学院博士前期課程了．同年 株式会社東芝研究開発センターに勤務．2011 年 東京大学大学院学際情報学府にて博士（学際情報学）を取得．2015 年より現在 鳥取大学大学院工学研究科准教授．画像認識，インタラクションの研究に従事．2009 山下記念研究賞などを受賞．電子情報通信学会，情報処理学会の会員．

岩井　儀雄　（正会員）



1992 年 (平成 4 年) 大阪大学基礎工学部情報工学科卒業．1997 年 (平成 9 年) 大阪大学大学院基礎工学研究科博士課程後期修了．同年同大学院助手．2003 年 (平成 15 年) 同大学院助教授．2004 年 (平成 16 年)5 月〜2005 年 (平成 17 年)3 月英国ケンブリッジ大学客員研究員．2007 年 (平成 19 年) 同大学院准教授．2011 年 (平成 23 年）鳥取大学大学院工学研究科教授．コンピュータビジョン，パターン認識の研究に従事．博士（工学）