非言語マルチモーダル情報を利用したグループ対話におけるジェスチャの機能認識

岡田　将吾[†a)]　坊農　真弓[††]　高梨　克也[†††]　角　康之[††††]
新田　克己[†]

Gesture Function Recognition Using Multimodal Features in Small Group Narrative Interaction

Shogo OKADA[†a)], Mayumi BONO[††], Katsuya TAKANASHI[†††], Yasuyuki SUMI[††††], and Katsumi NITTA[†]

あらまし　会話エージェント，会話記録支援システムなどの高度なヒューマンコミュニケーションシステムの実現には，対話中に交わされる言語情報だけでなく視線・ジェスチャといった非言語情報の認識技術が必要である．本研究ではグループ対話中に表出するハンドジェスチャの機能を認識する枠組みを提案する．対話中に用いられる情景記述，発言の強調・調整などのジェスチャの機能を認識するために，(1) Kendon により提案されたジェスチャフェイズに関する特徴量，(2) ジェスチャと共起した発話・視線などのマルチモーダル特徴量を抽出する．3 者の対話タスクを行い，対話中に観測されるジェスチャ・発話・視線情報にアノテーションを付与した後に，(1)(2) に関して複数の特徴量を定義し，抽出した．機械学習を用いてジェスチャの機能の認識実験を行った結果，提案する特徴量を用いることにより認識精度を示す F 値が，手の動作特徴だけを用いた場合よりも 0.28 ポイント向上することが確認された．

キーワード　ジェスチャ認識，社会的信号処理，マルチモーダル情報処理，会話分析

1. まえがき

対面会話は情報交換・意思決定・合意形成を行うための基本的なインタラクション行為である．対面会話では，交わされる言語情報だけでなく，韻律・視線・ジェスチャといった非言語情報を通じてコミュニケーションが行われている [1]．非言語情報の中でもジェスチャは，発話内容に関連する情報を有したり，何かを伝えようという意図のもとに表出されたり，対話の流れを調節するために表出されたりするため，参加者の意図・態度や対話インタラクションの構造を理解するために重要な手掛かりの一つであると考えられる．ジェスチャの対話インタラクションにおける機能を機械的に認識することができれば，話者が情報伝達を行っている，発言を強調しているといった話者の状態推定や，インタラクション構造の自動分析に役立つと考えられる．

本研究では対話中に観測されるハンドジェスチャの機能を認識するために有効な特徴量を抽出・分析し，ハンドジェスチャの機能認識モデルの構築・評価を行う．一般にジェスチャは個人が任意の状況下で用いることから，ジェスチャが使われるタイミング，手の動かし方は人によって異なるため，ハンドジェスチャの機能を識別するための特徴量を抽出することは容易ではない．従来提案されている多くのハンドジェスチャ認識システム [2] では，手・指の形状や動きの特徴量が認識に利用されているが，自然と表出するジェスチャの機能認識にはこれらの特徴量が必ずしも有効ではない．

本研究ではジェスチャ機能の認識のために，社会学・

† 東京工業大学大学院総合理工学研究科知能システム科学専攻，横浜市
Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama-shi, 226–8503 Japan
†† 国立情報学研究所，東京都
National Institute of Informatics, Tokyo, 101–8430 Japan
††† 京都大学，京都市
Kyoto University, Kyoto-shi, 606–8501 Japan
†††† はこだて未来大学，函館市
Future University Hakodate, Hakodate-shi, 041–8655 Japan
a) E-mail: okada@dis.titech.ac.jp

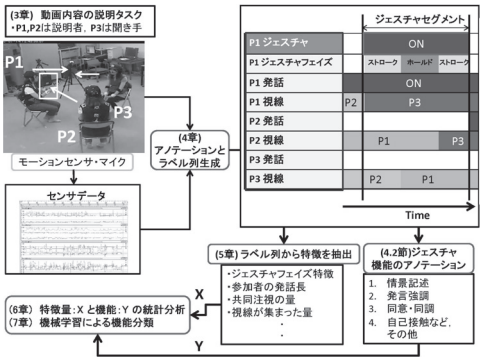


図 1 対話中に表出するジェスチャの機能認識の枠組み
Fig. 1 Proposed framework for conversational gesture function recognition.

言語学の知見を利用して 2 種類の特徴量を提案する．

（1） **ジェスチャフェーズ [3] の特徴量:** ジェスチャフェーズ（gesture phase）はジェスチャを分析するために提案された汎用的な動作の記述方法である．ハンドジェスチャをジェスチャフェーズに分割し，各フェーズの時間長や頻度といった統計量を特徴量に用いる．

（2） **対話参加者のマルチモーダル非言語特徴量:** [4] ではジェスチャの機能を分析する際，ジェスチャに伴う発話情報を同時に観察する重要性が指摘されており，[5] では情報伝達に使われるジェスチャは意図的に他者に向けられており，他者指向性の機能をもつことが指摘されている．これらの知見を利用し，手の動作特徴だけでなく，ジェスチャに付随した発話・ジェスチャを行った人の視線方向，更には他の参加者の視線状態・発話状態に関するマルチモーダル非言語特徴量を利用する．この特徴量を以降ではマルチモーダル特徴量と呼称する．

上記の特徴量が機能認識に有用であることを検証するために，3 人による状況説明対話タスクを計 10 セッション収録し，そのタスクで交わされた手の動作，発話，視線を含めたマルチモーダルデータセットを構築する．対話中に観測されるジェスチャとそのジェスチャの機能，ジェスチャフェイズ，発話，視線に関してアノテーションを行い，アノテーションしたラベル系列から特徴量を抽出する．各特徴量と各機能について統計分析を行い，機能の識別に有用な特徴量を分析した後，機械学習を用いてジェスチャの認識モデルを構築し，認識精度を評価する．本研究の枠組みを図 1 に示す．

2. 関 連 研 究

本研究は社会学・言語学で行われてきたジェスチャ研究の知見を利用して，表出するジェスチャの機能を機械的に認識する枠組みを提案する．**2.1** ではジェスチャの理論研究と本研究の位置づけを述べ，**2.2** ではジェスチャ認識システムを始めとした，工学・情報学における関連研究と本研究の位置づけを述べる．

2.1 本研究で援用するジェスチャ理論研究の知見

表出するジェスチャの機能の分析・理論構築は [3], [6] を始めとして盛んに行われてきた．[3] で提案されたジェスチャフェイズはジェスチャの記述方法の一つであり，対話タスク・個人によらず汎用的に利用可能である．本研究では，対話中のジェスチャの特徴量にジェスチャフェイズの情報を用いる．

[3], [6] ではジェスチャと発話状態は共起関係にあり，発話とジェスチャを統合的に分析することの重要性が述べられており，ジェスチャと同時に観測される発話の長さ，ジェスチャとの時間共起割合などの発話状態はジェスチャの機能認識に有効であると考えられる．

[7] では聞き手が発話とジェスチャの双方から同時に情報を取得し，それらを統合して理解していることが示された．この知見は言い換えれば，情報伝達を行うためにジェスチャが使われた場合，聞き手はそのジェスチャに視線を向けて，ジェスチャを観察している可能性が高いことを示している．Özyürek は参加者の配置によって，ジェスチャを行う手や，ジェスチャを向ける方向が変化することを述べている [8]．この知見から，話し手側も，自身のジェスチャを聞き手側が観測しやすいように振る舞っていることが示された．これらの分析結果から，ジェスチャを行う人・聞き手の視線状態を考慮することは機能認識に有効であると考えられる．

2.2 ジェスチャ認識・動作認識に関する研究

近年，ジェスチャ認識システム・アルゴリズムに関する研究が盛んに行われている [2], [9]．これらの研究の多くは，手・指の形状，それらの動作軌跡の類似性に基づきジェスチャのカテゴリーを認識し，このカテゴリーに対応するコマンドを送ることができる，インタフェースの実現に焦点を当てており，対話中に表出するジェスチャの認識・理解に焦点をあてたものは少ない．

一方で [10] はステレオビジョンから，対話中の上半身（手と頭部）の動作軌跡をロバストに抽出できる枠

組みを提案している．評価実験ではインタビュー場面において話者の，自己接触・ビートジェスチャといったジェスチャの自動認識を行っている．この研究は 2 者による対面対話におけるジェスチャの種類の認識に焦点を当てており，本研究のようにグループ対話におけるジェスチャの機能認識に焦点を当てていない．

[11] では手の動き，発話の切れ目などのマルチモーダル特徴量を利用して対話中に利用されるジェスチャの認識精度を向上させる枠組みを提案している．[12] では対話中に用いられるビートジェスチャの分析に焦点を当てている．Xiong らは韻律情報と手の振動の特徴を表す周波数特徴の間の関係を分析している．[11], [12] の研究ではいずれもジェスチャを行った人の音声情報を利用し，マルチモーダル特徴量を統合することで対話中のジェスチャを認識する点で本研究の枠組みと類似するが，本研究はジェスチャを表出した本人だけでなく対話参加者全員の発話・視線状態に関する特徴量を抽出し，ジェスチャの機能認識を行う点が異なっている．

本研究の先行研究 [13] では，参加者の発話・視線・頭部ジェスチャを利用し，説明中に用いられたジェスチャか否かの 2 クラス分類が行われた．本研究では [13] では定義されていなかった共同注視，視線を受けた量，発話交代などの非言語特徴を新たに定義する他，統計分析により機能の識別に寄与する特徴量を明らかにし，4 クラスの機能認識を行う点で [13] の枠組みが拡張されている．

3. ジェスチャの機能認識概要

対話中に観測されるジェスチャの機能を認識するモデルを構築するために，ジェスチャが頻繁に観測される対話タスクを設定し，多様な機能をもつジェスチャデータを収集する必要がある．

3.1 ジェスチャの機能の定義

本研究では [5] で述べられている，ジェスチャがコミュニケーションにおいて果たす機能を「ジェスチャの機能」と定義する．ジェスチャの機能は主に「伝達内容の表現」,「コミュニケーションのメタ調節」に大別される [5].「伝達内容の表現」は情報伝達したい内容そのもの，または発話内容に関連した情報をジェスチャが含んでる場合を示し,「コミュニケーションのメタ調節」は話す順番（ターン）の保持・譲渡，相手の発言への呼応といった対話の調整機能としてジェスチャが使用される場合を示している．本研究では **3.2** で述べる対話タスクで参加者から観測されるジェスチャに対して，上記二つの機能に基づきアノテーションを行い，アノテーションされた機能ラベルの認識を試みる．

3.2 対話タスク：アニメーション説明タスク

3 人のグループ対話タスクとしてアニメーションの内容・各シーンの状況を説明する課題を設定し，ジェスチャデータセットを収集した．本研究では McNeill [6] により考案された，動画を事前に観察した参加者（以後，説明者と呼称）がその動画を見ていない参加者（以後，聞き手と呼称）に動画内容を説明するタスクを選定した．動画コンテンツについても [6] で用いられたワーナーブラザーズ社製の “Canary Row” というアニメーションを使用する．

この説明タスクでは，アニメーションの情景，猫・鳥などの登場人物の動作を表現するためのハンドジェスチャが発言に伴って観測される．McNeill は 1 人の説明者が 1 人の聞き手に説明を行う 2 者対面対話タスクを設定したが，本研究では説明者を 2 人に増やし，動画内容を知る 2 人の説明者が 1 人の聞き手に説明するタスクを設定した．説明者を増やすことで，一方の動画内容の記憶が曖昧な場合に，他方の説明者がそれを補ったり，一方の発言に対して，同意・同調するようなジェスチャが出るなど 2 人対話よりも多様な対話構造が観測できる上，コミュニケーションのメタ調節機能に関するジェスチャが観測できる．

3 人は着座状態で対話を行う（図 1 の左上画像）．人材派遣会社を通じて計 30 名の実験協力者を募集した．募集した 30 名はいずれも初対面の 20 代前半の女性であり，3 人同士を 1 グループとして説明タスクを 10 セッション行い，対話データを収集した．この内，充分に説明を行わなかった 1 グループと，センサデータの欠損が著しかった 1 グループのデータを除外し，計 8 セッションのデータを本研究に使用した．データとして使用した各セッションの平均対話時間は約 11 分（合計で約 700 分）であった．

同世代の女性同士のグループ対話を実験対象とした理由は，女性が男性に比べて多くの表象的ジェスチャを行うことが，同タスクのジェスチャ分析に関する研究 [14] で報告されており，更に「同性同士，同世代」のグループが初対面で一番対話をしやすいためである．本研究では対話中に表出するジェスチャデータを多数収集する必要があったため，本設定でデータ収集を行った．この設定でデータを収集することで，**3.4** で示すように 6 種類の機能をもつジェスチャパターンを多数取得した．

3.3 非言語データの取得環境

ジェスチャの候補となる手の動作・発話状態・視線状態に手動でアノテーションを行う．動画の閲覧を通じたアノテーションの負担を軽減するために，各種センサを用いて，手の動作・発話・頭部動作のセンシングを行い，得られた各信号データを動画と同期させて収録する．各信号データと動画データを同時に閲覧することで，手の動作の開始・終了点，視線方向の変化点などを特定することが，動画のみを閲覧する場合よりも容易となる．

対話者の発話状態を取得するために，指向性無線マイクと録音機材を，顔向け方向，ハンドジェスチャをセンシングするために，モーションアナリシス社製の光学式モーションキャプチャシステム：Mac3D，をそれぞれ用いる．動画・センサデータを用いたアノテーションの手順については **4.** で述べる．

3.4 データから観測されたジェスチャの機能

8 セッション，計 16 人の説明者からジェスチャの候補となる，500 以上の手の動作パターンを観測し，ジェスチャの機能を以下のように列挙した．

（1） 動画の情景描写・登場人物の動作を表現するために使われるジェスチャ
（2） 発話の調子を整えたり，発言を強調するようなジェスチャ
（3） 他人の発話に同調・同意・呼応して出るジェスチャ
（4） 相手に呼びかけを行うジェスチャ
（5） 言いよどみ時に出るジェスチャ
（6） 頬を触ったり，指で数を数えたりする動作や，機能やその意味を特定できない動作

上記で列挙した全ての動作・ジェスチャの機能を認識できることが望ましいが，(5)(6) のジェスチャが観測されたのは各 10 回未満と少なかったため，(4) に統一して「その他」のジェスチャとした．また (4) に含まれる「頬を触る」といった行為はその人の意図を表している重要なジェスチャである可能性も考えられるが，意図的であるかどうかの判断が困難であるため，このようなジェスチャについても「その他」のジェスチャとしてまとめた．したがって本研究では (1)～(4) の機能認識に取り組む．(1) のジェスチャは「伝達内容の表現」の機能の一部であり，(2)～(3) は「コミュニケーションのメタ調節」の機能の一部であるとみなせる．

3.5 機能認識モデルの構築手順

ジェスチャの機能認識モデルの構築を以下の手順で行う．

（1） 説明タスクで取得したセンサデータ・動画データを利用して参加者の手の動作・発話・視線・ジェスチャの機能のアノテーションを行う（**4.**）．
（2） アノテーションされたマルチモーダルラベル系列から，ジェスチャの機能認識のための特徴量を抽出する（**5.**）．
（3） マルチモーダル特徴量で構成されるジェスチャデータセットを用いて，機械学習により機能認識モデルを構築・評価する（**7.**）．

4. 非言語パターンのアノテーション

本論文は，ジェスチャフェイズに関する特徴量と対話参加者全員のマルチモーダル特徴量のジェスチャの機能認識における有用性を検証することに焦点を当てるため，ジェスチャセグメント（手の動作区間）・ジェスチャフェイズ・視線（顔向け）状態のアノテーションは人手で行い正確なラベルデータを生成し，アノテーションされたラベル系列からの特徴抽出・認識モデル構築を自動処理・機械学習により行う．

4.1 ジェスチャセグメント・フェイズのアノテーション

両腕に装着したマーカの三次元座標値と動画を観察し，手の動作区間であるジェスチャセグメントとジェスチャフェイズのアノテーションを行う．

今回の収録環境では肘掛なしの椅子に参加者は座って対話をするため，参加者は最初膝の上に手を置いていた．ここで膝の上に手を置いている状態の間を無動作区間と定義した．膝上以外の場所に手が置かれている場合，また手が動いている区間をジェスチャセグメントと定義した．

次にジェスチャセグメント内に含まれるジェスチャフェイズのアノテーションを行う．Kendon の定義によると，ジェスチャフェイズは「準備」,「ストローク」,「ホールド」,「復帰」の 4 種類に大別される．「準備」は動作区間の始まりの初期動作とし，「復帰」は動作区間の終わりから無動作区間への移行動作と定義した．「ホールド」は膝上の位置以外で手が止まったままキープされている状態とし，「ストローク」は動作区間で手が移動し続ける状態と定義した．

本研究で定義したジェスチャセグメントは複数のジェスチャフェイズで構成される．図 1 の右上の非言語パターンのアノテーションの例では，対話参加者 P1

の手の動作区間（P1 ジェスチャ）としてジェスチャセグメント（ON の区間）が観測されている．このジェスチャセグメントは順に「ストローク」,「ホールド」,「ストローク」の 3 区間のジェスチャフェイズにより構成されている．

アノテーションは 3 人のコーダによって以下の手順で行われた．

Step1. ジェスチャセグメントと無動作区間のアノテーションを行い，各区間をジェスチャフェイズのセグメントと定義する．

Step2. 「準備」,「ストローク」,「ホールド」,「復帰」区間のアノテーションを行う．

Step3. 右手・左手のアノテーションラベルを統合する．両手のアノテーションラベルが等しければ，そのラベルが最終的なラベルと決定される．片手の状態が無動作区間で，もう片方がジェスチャセグメントである場合，ジェスチャセグメントのラベルを付与する．両手のラベルがストローク・ホールドで異なった場合，ストロークを優先しストロークのアノテーションを付与する．

4.2 ジェスチャ機能のアノテーション

4.1 でアノテーションした一つのジェスチャセグメントを一つのジェスチャパターンの単位と定義し，**3.4** で述べた 4 種類のジェスチャの機能：(1) 情景記述，(2) 発言強調・調整，(3) 他者発言への同意・同調，(4) 自己接触，意味の付与が困難な動作を含む，(1)〜(3) に属さない，その他のジェスチャ（以降では，「その他」と呼称）のいずれか一つのラベルをパターンに付与する．

(1) について，発話内容と照らし合わせて，動画中の情景，キャラクタ，その動作などを形作っている動作軌跡が 1 か所でもジェスチャセグメント中に含まれていれば，情景描写と定義する．2 人の説明者同士で動画内容を確認し合う場面が観測されたが，そのときに上記の動作軌跡が含まれている場合も，情景描写と定義した．(2) について，発言の最中に手が動いているが，発話内容とその動作に一致する点が見られないもの，発言の調子を整えるように手を振動させるようなビートジェスチャが観測できた場合を発話強調・調整と定義した．

(3) については，以下の例で説明する．時刻 t より前の参加者 j の発話・ジェスチャに対して，参加者 i が呼応して発言したときに付随するジェスチャが観測されたり，参加者 j のジェスチャを模倣して同じ形のジェスチャを参加者 i が行った場合，その機能を (3) に分類した．この作業は 1 人のコーダが行った．

4.3 視線方向のアノテーション

本タスク環境において三者の距離は十分離れており，頭部方向のセンシングで，誰が誰を見ていたかを特定できることを確認したため，頭部マーカの位置座標の変化から視線方向を近似する．具体的には頭部に装着した二点のマーカの位置座標の変化と動画を観察し，参加者が，その参加者から見て左または右の参加者のどちらに顔を向けているかをアノテーションした．

4.4 発話区間のアノテーション

環境音などが問題にならない環境で実験を行っており，参加者の発話は接話マイクを用いて収録しているため，収録した音声データに含まれるノイズは無視できるとみなした．このため，音声区間の検出は自動で行った．

音声区間検出には Julius [15] を用いた．このツールでは零点交差法により音声区間の候補を抽出し，事前に音声区間・無音区間を学習しておいた混合ガウシアンモデルを用いて音声区間を検出している．[16] にならい，700ms 以下の短い音声断片を削除した結果を発話区間としてアノテーションした．

5. 機能認識に用いる特徴量の抽出

アノテーションされたラベル系列と，モーションセンサから取得される手の動作量から，ジェスチャの機能認識に用いる特徴量を抽出する．図 2 には各特徴量の計算例を示す．以降ではアノテーションされた一つのラベルパターン（例えば図 2 上の gs）をセグメントと呼称する．

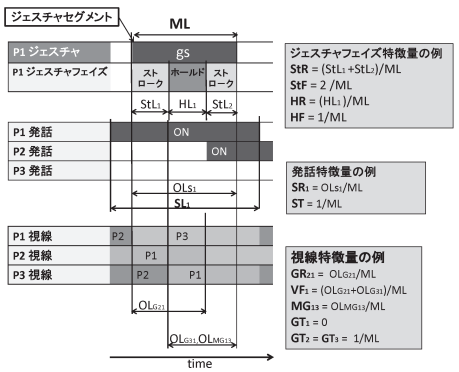


図 2 抽出されたマルチモーダル特徴量の例
Fig. 2 Examples of extracted multimodal nonverbal features.

表 1　ジェスチャの機能認識に用いた特徴量のまとめ
Table 1　Feature set using for classification of gesture roles.

特徴量 ID	変数名	特徴量の説明	特徴の次元数
手の動作に関する特徴量（計 9 次元）			
1	MM, MV	ジェスチャセグメントの動作変化量・分散	2
2	ML	ジェスチャセグメントの時間長	1
3	StM, StV	ストロークセグメントの動作変化量・分散	2
4	StR	ストロークセグメントの時間長割合	1
5	StF	単位時間当たりのストローク回数	1
6	HR	ホールドセグメントの時間長割合	1
7	HF	単位時間当たりのホールド回数	1
発話に関する特徴量（計 9 次元）			
8	SR_i	参加者 i の発話の内ジェスチャと時間共起した割合	3
9	SL_i	ジェスチャと時間共起した参加者 i の発話断片の総時間長	3
10	$SR_{/1}, SL_{/1}$	ジェスチャの受け手の特徴量 $SR_{2,3}, SL_{2,3}$ の各合計	2
11	ST	ジェスチャが行われている際，話者が変わった回数	1
視線に関する特徴量（計 18 次元）			
12	GR_{ij}	参加者 i が参加者 j に視線を向けた割合	6
13	VF_i	参加者 i が他の参加者から視線を向けられた割合の総和	3
14	MG_{ij}	参加者 i と参加者 j が相互注視した割合	3
15	GT_i	単位時間当たりの参加者 i の視線先が変化した回数	3
16	$VF_{/1}, GT_{/1}$	ジェスチャの受け手の特徴量 $VF_{2,3}, GT_{2,3}$ の各合計	2
17	MG_{1*}	ジェスチャを行っている人が他の参加者 j と相互注視した割合の合計	1
			合計 36

アノテーションされたジェスチャセグメントの集合を以下のように定義する．$\mathcal{GS} = \{gs_1, \ldots, gs_x, \ldots, gs_{N_g}\}$，$gs_x = \{st_x, et_x\}$．$N_g$ はジェスチャセグメントの総数であり，st_x, et_x は gs_x の始点・終点のフレーム番号を示す．モーションセンサのサンプリングレートが 120 フレーム/秒であるため，ジェスチャ・音声・視線ラベル系列の開始・終了時刻をフレーム表記に変換した．本研究では，聞き手のジェスチャはほとんど観測できなかったことから，二人の説明者から観測されたジェスチャを対象とした．ジェスチャgs_x が観測されたとき，そのジェスチャを行った説明者を ID = 1，もう一人の説明者の ID を 2，聞き手の ID を 3 とする．本章より以降では，ジェスチャを行った参加者以外の参加者（ID = 2, 3）をジェスチャの受け手と呼称する．以下で説明する特徴量を表 1 にまとめる．

5.1　手の動作に関する特徴量

4.1 の処理で検出されるジェスチャセグメントは準備，ストローク，ホールド，復帰といったジェスチャフェイズのセグメントから構成される．準備と復帰のセグメントはストローク・ホールドに比べて非常に短く，今回定義した機能の認識に寄与しなかったため，特徴量として扱わなかった．したがって，ストローク・ホールドの時間長や頻度といったジェスチャフェイズに関する特徴量と，ジェスチャセグメント内の腕のマーカの三次元位置座標の系列から特徴量を定義する．

ジェスチャセグメントの時間長: st_x, et_x 間のフレーム数: ML

手の動作特徴量: モーションキャプチャにより取得される両手首に装着したマーカの三次元座標（計 6 次元ベクトル: m）の系列 $M = \{m_{st_x}, \ldots, m_{et_x}\}$ から統計量を計算する．M の各次元 d ごとに時間方向の差分ベクトルを計算し，そのベクトル長で正規化したノルム V_d の次元間での最大値を，ジェスチャセグメントの変化量と定義する．この値は，ジェスチャセグメント内で手の位置が変化した量を示している．また V_d の次元間での分散を求める．計算された変化量を MM，分散を MV と定義する．

ストローク・ホールドセグメントの頻度: 状況説明下で表出するジェスチャセグメント中にはストローク・ホールドのセグメントが複数回観測されるため，それらの頻度情報は有用な特徴である．gs_x 中に含まれるストローク・ホールドのセグメントの回数を MT で割った値を頻度 StF, HT とそれぞれ定義する．StF, HT の計算例を図 2 の上段に示す．

ストローク・ホールドの占める時間割合: ホールドセグメントがジェスチャセグメント内で占める時間長割合を以下の式で計算する．

$$HL = \sum_{i=1}^{HT} HL_i / ML \tag{1}$$

HL_i は i 番目のホールドセグメントのフレーム長である．同様にストロークセグメントの時間割合 StL も

計算する．ストロークセグメント区間内での手の動作特徴量を MM, MV と同様に計算し，StM, StV と定義する．

5.2 ジェスチャと共起するマルチモーダル特徴量

ジェスチャセグメントと時間共起する発話状態，視線状態などマルチモーダル特徴量を抽出する．gs_x と時間的に共起したセグメントを探索する．発話・視線のセグメントパターン p_y と gs_x とが時間的に共起したフレーム長を以下の式により計算する．

$$OL_{p_y} = \max(0, (\min(et_x, et_{p_y}) - \max(st_x, st_{p_y}))) \tag{2}$$

式 (2) において st_{p_y}, et_{p_y} は p_y の始点・終点，$\max(a,b)$, $\min(a,b)$ は数値 a と b の最大値・最小値を返す関数である．

5.2.1 発話に関する特徴量

3 人の対話参加者の発話セグメントセットを $\mathcal{S}_i$ $(i = 1,2,3)$ とする．i は参加者の ID である．

発話長・発話共起割合: $\mathcal{S}_i$ に含まれるパターン S_i の内，$OL_{S_i} > 0$ である，S_i^* のフレーム長を発話長 SL_i とする．S_i^* が複数存在する場合，それらの総和を SL_i とする．発話共起割合 SR_i を式 (1) と同様に計算する．図 2 の中段に計算例を示す．

発話者の交代回数: st_x, et_x 間に，話者が交代した回数を計数する．全参加者（$i = 1,2,3$）の発話 $\mathcal{S}_i$ について $OL_{S_i} > 0$ である，S_i^* を列挙した後，発話開始時間 $st_{S_i^*}$ で昇順にソートする．ソート後，t 番目の発話断片を S_{Pt^*}，それを発した参加者を P_t^* とし，st_x, et_x 内で $S_{P_t^*}$ が終了した後，次の発話 $S_{P_{t+1}^*}$ が観測され，$P_t^* \neq P_{t+1}^*$ である場合，1 回の発話交代と認定する．上記に従い計数した値を，発話者の交代の回数 ST と定義する．この回数が少ない場合，そのジェスチャ中に同じ話者が発話をキープしていることを示しており，多い場合，複数の話者の発話断片が交互に観測される．図 2 の中段の例では，P1 から P2 に話者が 1 回交代したため $ST = 1/MT$ となる．

ジェスチャの受け手の発話長・割合: ジェスチャの受け手 $i = 2,3$ の SR_i, SL_i を足し合わせた値 $SR_{/1}, SL_{/1}$ を特徴量として定義する．

5.2.2 視線に関する特徴量

各参加者の視線状態セグメントセットを $\mathcal{G}_{ij}$ $(i, j = 1,2,3,\ i \neq j)$ と定義する．G_{ij} は参加者 i が参加者 j に視線（顔）を向けた視線状態セグメントを示す．

視線の時間共起割合: 視線状態セグメント G_{ij} とジェスチャセグメント gs_x の時間共起割合 GR_{ij} を SR_i と同様に，式 (1) に従い計算する．

参加者 i が視線を受けた割合: 各参加者 i が自分以外の参加者に視線を向けられている割合を $VF_i = \sum_j G_{ji}$ として計算する．

参加者 i が共同注視した割合: 参加者 i と参加者 j が共同注視した割合 MG_{ij} を計算する．式 (2) を拡張し，三つのセグメント gs_x，G_{ij}，G_{ji} 間の共起する時間長 $OL_{MG_{ij}}$ を定義し，$MG_{ij} = OL_{MG_{ij}}/ML$ として計算する（図 2 の下段）．

視線方向の変化回数: ジェスチャセグメントの開始・終了時間までの間に，参加者 i の視線方向が G_{ij} から G_{ik} $(j \neq k)$ に変化した回数を計数し，正規化した値を視線方向の変化回数 GT_i と定義する．

ジェスチャの受け手の視線に関する特徴量: ジェスチャの受け手が視線を受けた割合を $VF_{/1} = VF_2 + VF_3$ とする．ジェスチャを行った参加者が受け手と相互注視した割合の総和を $MG_{1*} = MG_{12} + MG_{13}$ とする．ジェスチャの受け手の視線方向の変化回数を $GT_{/1} = GT_2 + GT_3$ とする．

6. 機能認識に有用な特徴量の分析

特徴量の頻度がジェスチャの機能ごとに異なるかを分析する本研究では，ジェスチャフェイズの特徴，ジェスチャと共起した他の非言語特徴がジェスチャの機能認識に有効な特徴であると仮定した．本章ではこの仮定が正しいことを検証するため，「情景記述」，「発言強調・調整」，「他者発言への同意・同調」，「その他」の四つの機能カテゴリー間で各特徴量の頻度平均値に差があるかを検定し，各カテゴリー間で有意に平均値が異なる特徴量を明らかにする．

8 セッション，計 16 人の説明者から 547 個のジェスチャセグメントが抽出された．この内，モーションキャプチャのデータが欠損しているデータを除き，計 473 個のジェスチャセグメントを本章の分析・次章の機械学習に用いる．各カテゴリーのデータの数は，「情景記述」が 229，「発言強調・調整」が 32，「他者発言への同意・同調」が 28，「その他」は 184 となっている．

6.1 分析手順

機能の 4 カテゴリーを水準，特徴量の変数 X を一つの要因とした 1 元配置分散分析を行った後，多重比較検定を行うことで，各カテゴリー間で特徴量の平均値に有意差があるかを検証する．検定手順を以下に述

べる．

（1） 全ジェスチャデータの特徴量 X の値を平均 0，分散 1 になるように正規化した後，最小値が 0 になるよう値を変換する．

（2） 特徴量 X の値を四つのカテゴリーに従い分割する．

（3） 特徴量の変数 X を要因，カテゴリー数を水準とした 1 元配置分散分析を行った後，テューキーの検定を用いて，互いに有意差があるカテゴリーペアを列挙する．有意水準は $p = 0.05$ とした．

（4） (1)〜(3) の手順を表 1 の全特徴量に関して行う．

6.2 特徴量の分析結果

表 1 の特徴量ごとに，(1) 手の動作に関する特徴，(2) 発話に関する特徴，(3) 視線に関する特徴に分けて，検定結果を示し，考察する．(1)，(2)，(3) に関して有意差の認められた各変数の平均値を図 3，図 4，図 5 にそれぞれ示す．各図の横軸は変数名を示し，縦軸は頻度（frequency）を示す．図中の棒グラフ間に ⊓ が記載されている場合，その棒グラフに対応するカテゴリーの平均値の間には $p < 0.05$ で有意差が認められたことを示す．

6.2.1 手の動作特徴量の検定結果

図 3 より，ジェスチャセグメント長 ML，ストロークに関する特徴 StL, StT，ホールドに関する特徴 HT, HL 共に，「情景記述」ジェスチャの値が他のカテゴリーより大きいことが示された．ジェスチャを用いて情景記述を行う場合，そのジェスチャ長は長く，そのジェスチャの中で，ストローク，ホールドジェスチャが多用されることが示された．一方で，他の 3 カテゴリー間に有意差が認められるペアは存在しないことがわかった．

6.2.2 発話に関する特徴量の検定結果

図 4 より全参加者の共起発話長 SL_i，発話の共起割合 SR_i について，幾つかのカテゴリー間で有意な差が確認された．ジェスチャを行った人の発話特徴 SL_1, SR_1 について，「情景記述」，「発話強調」のジェスチャと共起した発話長・共起割合は他のカテゴリーに比べて，値が大きいことがわかる．逆に「他者発言への同意・同調」，「その他」のカテゴリーのジェスチャを行っているときに伴う発話の長さは短いことを示している．「情景記述」，「発話強調」時にはジェスチャを行っている人が発話権を保持している可能性が高いため自然な結果である．併せて「発話強調」時に発話交代 ST が低頻度になっていることからも，ジェスチャを行っている人が発話権を保持し続けていることが確認できる．

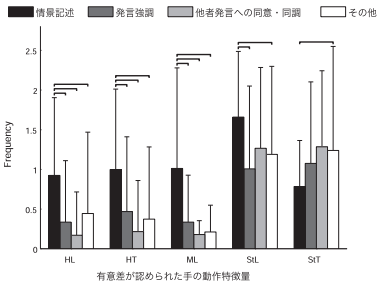


図 3 手の動作特徴量の多重比較分析結果
Fig. 3 Result of multiple comparison tests for hand motion features.

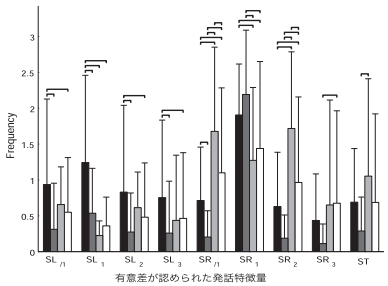


図 4 発話の特徴量の多重比較分析結果
Fig. 4 Result of multiple comparison tests for speech features.

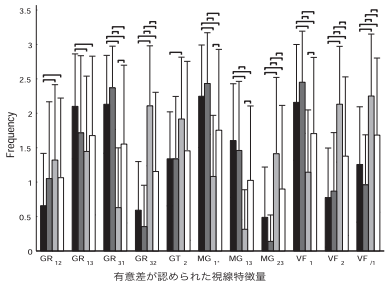


図 5 視線の特徴量の多重比較分析結果
Fig. 5 Result of multiple comparison tests for gaze features.

ジェスチャの受け手の発話に関する変数 $SR_{2,3,/1}$ について，「他者発言への同意・同調」，「その他」のカテゴリーの値は，他の 2 カテゴリーに比べて大きいことが示された．これは受け手の発言に同調するときに手が動いたり，受け手の発話中に自己接触・無意味な動作を行っているためであると考えられる．

6.2.3 視線に関する特徴量の検定結果

図 5 より $GR_{13,31}, MG_{1*,13}, VF_1$ について，「情景記述」，「発話強調」の平均値が有意に大きいことから，このカテゴリーのジェスチャを行っている最中は，ジェスチャの受け手から視線を多く集めていることが分かり，更にジェスチャの受け手と共同注視する頻度も高いことがわかる．

一方で，$GR_{12,32}, GT_2, MG_{23}, VF_{2,/1}$ において，「他者発言への同意・同調」，「その他」の平均値が有意に大きいことから，このカテゴリーのジェスチャを行っている最中は，ジェスチャの受け手に視線が集まっていることが多かったり，ジェスチャの受け手である二人が共同注視を行ったりしており，ジェスチャの受け手である参加者の視線方向も高頻度に変化することがわかる．

GR_{13} について，前者では「情景記述」の平均が「発話強調」より高く，GR_{12} においてはその逆である．参加者（$i = 3$）は物語を知らない聞き手であるため，ジェスチャを用いて情景記述を行う際には聞き手に視線を向ける頻度が大きかったためと考えられる．

6.2.4 特徴量の検定結果まとめ

上記の分析結果をまとめて以下の知見が得られた．

- ジェスチャフェーズを含む手の動作特徴は，「情景記述」を他のカテゴリーと識別するために有用である．
- 「情景記述」のジェスチャに伴うその人の発話長は長い．「説明強調」のジェスチャとその人の発話は共起する割合が大きい．
- 「他者発言への同意・同調」，「その他」のジェスチャと共起するジェスチャの受け手の発話は長く，割合も大きい．
- 「情景記述」，「説明強調」のジェスチャが観測される間，ジェスチャを行う人が視線を集めやすく，「発言の同意・同調」，「その他」のジェスチャが観測される際，ジェスチャの受け手が視線を集めやすい．

上記の結果より，本研究で提案するジェスチャフェイズ特徴量とジェスチャと共起した発話・視線の非言語特徴量は，機能の認識に有用であることが示唆された．

7. 機械学習によるジェスチャ機能認識

機械学習アルゴリズムを用いて，ジェスチャの機能認識モデルを構築し，評価することで本研究で提案した特徴量がジェスチャの認識に寄与するかを検証する．

7.1 実験設定

認識モデルにはガウシアンカーネルを用いた非線形 SVM と AdaBoost [17] アルゴリズムを用いた．ガウシアンカーネルのバンド幅のパラメータは $\gamma = 0.1$ と設定し，AdaBoost の木構造の分割数は $T = 3$ と設定した．各特徴量は平均 0，分散 1 になるように正規化した．

SVM，AdaBoost について各カテゴリーごとに分類器を準備し，対象カテゴリーに属する訓練データを正クラス，それ以外のカテゴリーに属する訓練データを負クラス，として訓練することで，多クラス分類を行う．テスト時には各カテゴリーの分類器からの出力値をスコアとして比較し，最大スコアを出力する分類器のカテゴリーを認識結果とする．

評価実験は 10 分割交差検定により行う．各カテゴリーごとにデータを 10 分割し，9 割を訓練，1 割をテストに用いる．分類器を訓練する上で，「発話強調」，「他者発言への同意・同調」のデータ量は少ないため，この場合，負例データが正例データより多くなりすぎる傾向にある．この 2 カテゴリーについては正例・負例データのバランスを取るため，負例からランダムに N_m 個をサンプリングする．予備実験より $N_m = 100$ として評価実験を行った．サンプリングのランダム性を考慮して，1 回の実験に付き 10 回のランダムサンプリングを行った．すなわち，10 分割交差検定の各実験において 10 回のランダムサンプリングを行い負例データセットを構築したため，合計 100 回の実験を行いテストデータの精度を評価する．

7.2 認識実験結果

各特徴量の認識精度への寄与を検証するために，6 種類の特徴量セットを以下のように構築した．

$F1$: 手の動作特徴量
（特徴量 ID：1〜2）

$F2$: $F1$ ＋ ジェスチャフェイズ特徴量
（特徴量 ID：1〜7）

$F3$: $F2$ ＋ 参加者全員の発話・視線特徴量（全特徴量）
（特徴量 ID：1〜17）

$F4$: $F2$ ＋ ジェスチャを行った人（当事者）の発話・視線特徴量
（特徴量 ID：1〜7）と（特徴量 ID；8, 9, 12, 15 の $i = 1$）

$F5$: $F2$ ＋ 全員の発話特徴量
（特徴量 ID：1〜11）

表 2　各特徴セット（$F1$〜$F6$）を用いた場合のジェスチャの機能認識精度の比較
Table 2　Classification accuracy of gestural functions by SVM and AdaBoost.

F 値（再現率/適合率）	$F1$ 手の動作量	$F2$ $F1$ + ジェスチャフェイズ	$F3$ 全特徴量
SVM	0.31 (0.33/0.33)	0.36 (0.35/0.38)	**0.59 (0.64/0.56)**
AdaBoost	0.38 (0.40/0.41)	0.42 (0.43/0.44)	0.47 (0.47/0.50)

F 値（再現率/適合率）	$F4$ $F2$ + 当事者の発話・視線	$F5$ $F2$ + 全員発話	$F6$ $F2$ + 全員視線
SVM	0.43 (0.48/0.43)	0.51 (0.54/0.49)	0.53 (0.54/0.52)
AdaBoost	0.45 (0.45/0.48)	0.46 (0.46/0.50)	0.47 (0.47/0.51)

$F6$：$F2$ + 全員の視線特徴量
（特徴量 ID：1〜7, 12〜17）
$F1, F2$ は手の動作特徴とジェスチャフェイズを含めた手の特徴量をそれぞれ示し，$F3$ は提案する参加者全員のマルチモーダル特徴量である．$F4$〜$F6$ はマルチモーダル特徴量の内，どの特徴量が有効であったかを検証するための特徴量セットであり，$F4$ はジェスチャを行った人の発話・視線，$F5$ は全員の発話，$F6$ は全員の視線を $F2$ に加えた特徴量セットである．ここで $F4$ は，$F2$ と $SR_1, SL_1, G_{12}, G_{13}, GT_1$ を使った計 14 次元の特徴で構成されている．

表 2 は各特徴量セットを用いた場合の認識結果を示している．各実験における認識精度は，各カテゴリーごとに再現率，適合率，F 値（再現率と適合率の調和平均）を算出し，その値の平均値とした．100 回の実験における認識精度の平均値を表 2 に記載している．ここでは小数点第 3 位を四捨五入した値を表記している．

表 2 の上段より，$F2$ の精度を $F1$ の精度と比較すると，F 値が SVM では 0.05 ポイント，Adaboost で 0.04 ポイント向上していることがわかる．この結果は対話中のジェスチャを認識する際，ジェスチャフェイズの特徴量が有用であることを示している．$F3$ の精度を $F2$ の精度と比較すると，F 値が SVM では 0.23 ポイント，Adaboost で 0.05 ポイント向上していることがわかる．この結果は対話中のジェスチャを認識する際，全員の視線・発話といったマルチモーダル特徴量も有用であることを示している．

表 2 の上下段より $F1$〜$F6$ を用いた SVM，Adaboost の結果を比較して，全特徴量を用いた場合（$F3$），SVM の F 値は最大値 0.59 をとり，$F1$ から 0.28 ポイント向上した．一方 Adaboost では SVM ほどの精度向上は見られなかったが，$F1$ から 0.09 ポイント向上した．

次に，各モダリティの特徴量の認識精度への寄与を検証する．最初に参加者全員の特徴量がどの程度，認識精度に寄与したかを検証する．表 2 の下段よりジェスチャを行った当事者のマルチモーダル特徴量を用いた場合（$F4$），SVM の F 値は 0.43 であり，全員のマルチモーダル特徴量を用いた場合（$F3$）のそれより 0.16 ポイント下回っている．この結果から話者交代回数 ST，共同注視 MG や視線を受けた比率 VF，視線変化 GT といった参加者全員を含むグループで定義される発話状態・視線状態に関する特徴量が認識に有用であることを示している．

次にマルチモーダル特徴量の内，発話・視線のどちらの特徴が認識精度に寄与したかを検証する．SVM による認識結果より手の動作特徴（$F2$）に発話特徴を加えた場合（$F5$）F 値が 0.15 ポイント，手の動作特徴（$F2$）に視線特徴を加えた場合（$F6$）F 値が 0.17 ポイントそれぞれ向上しているため，視線特徴による精度への寄与が 0.02 ポイント大きいことがわかる．これらの結果から，提案する枠組みにより対話中のジェスチャの機能認識精度が向上することが確認できた．

8. 考　　察

3 人のグループ対話で観測されるジェスチャからジェスチャフェイズに関する特徴量と，各参加者のマルチモーダル特徴量を観測することで，ジェスチャのコミュニケーションにおける機能を認識する枠組みを提案・評価し，有効性を示した．提案する枠組みの汎用性を向上させるためには実験設定の一般化（**8. 1**），多様な対話タスクで観測されるジェスチャ機能認識への応用（**8. 2**），非言語情報認識の自動化（**8. 3**）の三つの課題が存在する．本章では，実験結果を基に現時点での提案手法の応用範囲を考察し，上記の課題をまとめる．

8. 1　実験設定の一般化に関する考察

本研究では 20 代前半の同世代の女性 3 人によるグループ対話を対象とし，説明課題を対話タスクに設定した．今後，男性や異なる世代同士の対話や，4 人以

上の多人数対話に対象を広げる必要がある．

8.1.1 世代・性別を含む個人差への対応

今回の実験から得られた知見が，20 代女性以外の参加者の対話に適用できる保証はない．しかし，「情景記述」，「説明強調」のジェスチャは，発話と共起しやすいなどの特徴（**6.2.4**）は世代・性別に限らず共通に観測される特徴であると考えられるため，20 代前半の女性同士の対話という限られた対話設定であったものの一般的な会話構造・特徴が抽出されており，本研究の実験設定は一定の妥当性をもつと考えられる．換言すると，対話参加者全員の視線状態・発話状態をジェスチャと同時に観測する枠組みは，他の世代・性別の参加者同士の対話におけるジェスチャ機能認識にも適用できることを示唆した．

ただしジェスチャの頻度や，使用される機能の頻度などは，世代・性別を含め個人によって異なるため，他の世代・性別の協力者による対話や異性同士の対話データの収集を行い，個人差によるジェスチャの機能認識精度への影響を検証する必要がある．

8.1.2 多人数対話への対応

本研究では 3 人のグループ対話のタスクに対して，提案する枠組みが有効であることを示した．今後 4 人以上の多人数対話においても，提案する枠組みが適用可能かどうかを検証する必要がある．

6.2.4 の知見である「情景記述」，「説明強調」のジェスチャを使って参加者が説明を行っている場合，その参加者に視線が集まり，「他者発言への同意・同調」，「その他」のジェスチャを行っている参加者は，発話者でない可能性が高いので視線を向けられにくいという特徴は一般的であり，3 人以上の対話でも観測されると考えられるため，提案する枠組みが適用できる可能性がある．

ただし，多人数対話において，ある参加者が発話をしている最中，別の参加者同士が話を始めたり，複数の会話場が存在する場合，視線状態の扱いが複雑になるため本枠組みが適用できない可能性が高い．したがって，多人数対話であっても一つの会話場で構成されており，発話者の発言を他の参加者が傾聴する姿勢を保っている状況であれば本枠組みが適用可能であると考えられる．

8.2 ジェスチャの機能に関する考察

対話タスクの種類によって，用いられるジェスチャの機能は異なると考えられる．「コミュニケーションのメタ調節」の中に，相手の発話を牽制したり，発話権を奪取・譲渡するジェスチャが含まれる [5]．また，相手の意見に対して同意・非同意の態度を示すためのジェスチャが観測される [18]．上記の牽制，非同意といった機能のジェスチャは，ディベートや交渉のように相手と対立する状況で観測される．本研究では説明を行うことを目的とした協調的な対話であったため，上記の対立した態度を示すようなジェスチャは観測されなかった．

また説明タスクでも対話を行う環境によって頻出するジェスチャは異なると考えられる．本実験の対話環境では，説明者は説明に用いる資料などを一切もっていないため，身振りで聞き手に説明する必要があり「情景記述」のようなジェスチャを多く観測することができたが，ポスターなど対象物の前で説明を行うタスクでは，ジェスチャよりもポインティングでその対象を指し示す動作が多用されるであろう．一方で，本実験データから観測された発話の調整・強調，他者発言への同意・同調などのジェスチャは，多くの対話タスクで見られると考えられる．したがって，対話タスクの種類において頻出するジェスチャとその機能を分析し，類型化することも重要な課題である．

8.3 非言語情報認識の自動化に関する考察

本研究では，ジェスチャセグメント・ジェスチャフェイズ・視線 (顔向け) 状態のアノテーションは人手で行ったが，これらの非言語行動を自動推定することも今後の課題である．カメラを用いた，対話中のジェスチャの認識・手の動作のトラッキング，頭部方向に基づく視線状態推定は [10], [19] でそれぞれ行われており，これらを参考に各非言語情報の自動認識を試みる予定である．現状の認識精度に関して，「説明強調」，「他者発言への同意・同調」の認識精度（F 値）が最良の場合でも 6 割以下と低いが，これは訓練事例の数が少ないことに起因するため，対話データの量を増やすことで精度向上が見込める．上記の改善を行うことでジェスチャの機能認識を行い，参加者の説明態度・会話参加態度といった高次の状態推定に取り組む予定である．

9. むすび

本論文はグループ対話中に用いられたジェスチャの機能を認識するために，ジェスチャフェイズの特徴量と対話参加者の発話・視線状態に関する特徴量を利用する枠組みを提案した．3 者による説明タスクに本枠組みを適用した結果，ジェスチャ行為者以外の参加者

の非言語情報がジェスチャの機能の認識精度の向上に寄与することを示した．今後，**8.** で述べた実験設定の一般化，非言語情報認識の自動化に取り組み，更に多くのジェスチャの役割を機械に自動判別させることを目指す．

謝辞　本研究は科研費若手研究（B）25730132 の助成による．

文　献

[1] M.L. Knapp, Nonverbal communication in human interaction, Wadsworth Publishing Company Inc., 1972.
[2] S. Mitra and T. Acharya, “Gesture recognition: A survey,” IEEE Trans. Syst. Man Cybern., B, vol.37, no.3, pp.311–324, 2007.
[3] A. Kendon, “Gesticulation and speech: Two aspects of the process of utterance,” The relationship of verbal and nonverbal communication, vol.25, pp.207–227, 1980.
[4] A. Kendon, Gesture: Visible Action as Utterance, Cambridge University Press, 2004.
[5] 齋藤洋典，喜多壮太郎，ジェスチャ・行為・意味，共立出版，2002.
[6] D. McNeill, Hand and Mind: What Gestures Reveal about Thought, University of Chicago Press, 1992.
[7] J. Cassell, D. McNeill, and K.E. McCullough, “Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information,” Pragmatics & Cognition, vol.7, no.1, pp.1–34, 1999.
[8] A. Özyürek, “Do speakers design their cospeech gestures for their addressees?,” J. Memory and Language, pp.688–704, 2002.
[9] Y. Wu and T.S. Huang, “Vision-based gesture recognition: A review,” Proc. International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, pp.103–115, Springer-Verlag, 1999.
[10] A. Marcos-Ramiro, D. Pizarro-Perez, M. Marron-Romera, L.S. Nguyen, and D. Gatica-Perez, “Body communicative cue extraction for conversational analysis,” Proc. IEEE FG, pp.1–8, 2013.
[11] R. Sharma, J. Cai, S. Chakravarthy, I. Poddar, and Y. Sethi, “Exploiting speech/gesture co-occurrence for improving continuous gesture recognition in weather narration,” Proc. IEEE FG, pp.422–427, 2000.
[12] Y. Xiong, F. Quek, and D. McNeill, “Hand motion gestural oscillations and multimodal discourse,” Proc. ACM ICMI, pp.132–139, ACM, 2003.
[13] S. Okada, M. Bono, K. Takanashi, Y. Sumi, and K. Nitta, “Context-based conversational hand gesture classification in narrative interaction,” Proc. ACM ICMI, pp.303–310, 2013.
[14] 荒川　歩，鈴木直人，“ジェスチャーは会話スタイルの一部か?—発話の近言語的特徴とジェスチャー頻度との関係およびその性差—,” 対人社会心理学研究，vol.6, pp.57–64, 2006.
[15] T. Kawahara, A. Lee, K. Takeda, K. Itou, and K. Shikano, “Recent progress of open-source LVCSR engine Julius and Japanese model repository,” Eighth International Conference on Spoken Language Processing, pp.3069–3072, 2004.
[16] L.-P. Morency, I.D. Kok, and J. Gratch, “Context-based recognition during human interactions: Automatic feature selection and encoding dictionary,” Proc. ACM ICMI, pp.181–188, ACM, 2008.
[17] Y. Freund and R.E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” Proc. EuroCOLT ’95, pp.23–37, Springer-Verlag, 1995.
[18] K. Bousmalis, M. Mehu, and M. Pantic, “Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools,” Image and Vision Computing, vol.31, no.2, pp.203–221, 2013.
[19] K. Otsuka, H. Sawada, and J. Yamato, “Automatic inference of cross-modal nonverbal interactions in multiparty conversations: “who responds to whom, when, and how?” from gaze, head gestures, and utterances,” Proc. ACM ICMI, pp.255–262, 2007.

（平成 26 年 4 月 8 日受付，9 月 5 日再受付）

岡田　将吾　（正員）

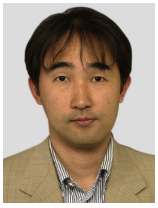
2008 東京工業大学大学院知能システム科学専攻博士課程了．同年京都大学情報学研究科知能情報専攻特定助教，2011 東京工業大学大学院知能システム科学専攻助教．博士 (工学)．人間行動解析，社会的信号処理，パターン認識の研究に従事．

坊農　真弓

2005 神戸大学大学院総合人間科学研究科博士課程了．ATR，京都大学，JSPS 特別研究員 (PD)，UCLA ポスドク研究員，テキサス大学オースティン校客員研究員，国立情報学研究所助教を経て，2014 より同研究所准教授．博士 (学術)．多人数インタラクション研究及び手話研究に従事．

高梨　克也

2000 京都大学大学院人間・環境学研究科博士課程単位取得退学．情報通信研究機構，京都大学，科学技術振興機構さきがけ専従研究者などを経て，現在京都大学学術情報メディアセンター産官学連携研究員．博士（情報学）．コミュニケーションの組織化を支える認知的・社会的プロセスの解明に従事．



角　　康之　（正員：シニア会員）

1995 東京大学大学院工学系研究科情報工学専攻了．（株）国際電気基礎技術研究所（ATR）主任研究員，京都大学大学院情報学研究科准教授を経て，2011 より公立はこだて未来大学教授．博士（工学）．研究の興味は，知識や体験の共有を促す知的システムや，人のインタラクションの理解と支援にかかわるメディア技術．

新田　克己　（正員）

1980 東京工業大学大学院博士課程了．同年電子技術総合研究所に入所．1989 から 1993 まで（財）新世代コンピュータ技術開発機構に出向．1994 東京工業大学大学院総合理工学研究科教授．工学博士．法的推論システム，ヒューマンインタフェース，マルチエージェントシステムなどの研究に従事．