社会言語科学　第15巻第1号
2012年9月　38-56ページ

The Japanese Journal of Language in Society
Vol. 15, No. 1, September 2012, pp. 38-56

研究論文

時系列データマイニングを援用した会話インタラクションにおけるジェスチャ分析の支援

岡田将吾（東京工業大学大学院），**坊農真弓**（国立情報学研究所），**角　康之**（はこだて未来大学），
高梨克也（科学技術振興機さきがけ／京都大学学術情報メディアセンター）

本研究ではハンドジェスチャの量的な分析を支援するために，データマイニングのアプローチを紹介する．データマイニングの手順として，最初に手の動きをモーションキャプチャでセンシングし，時系列データとして抽出する．次に，時系列データから動作部分・無動作部分を自動分類し，ジェスチャパターンの候補となる頻出する系列パターンを発見・抽出する．データマイニングシステムの有用性を示すために，これを用いた分析事例を紹介する．分析対象には，二人が説明者，一人が聞き手という設定で行うアニメーション説明タスクにおける説明者のジェスチャを選定した．計8セッションの説明タスクから得られるデータに，本手法を適用した．各セッションごとに説明者が用いたジェスチャの頻度，二人の説明者がジェスチャを行ったタイミングに関する分析を行った結果，説明におけるジェスチャの使い方，使う量が場合分けできることを示す．さらに，セッション間で共通して使用された頻出ジェスチャパターンや，アニメーションにおけるキャラクタの小刻みな動きを表象するジェスチャを抽出できることを示す．

キーワード：ジェスチャ分析，定量的分析，非言語インタラクション，自動アノテーション，データマイニング

Gesture Analysis in Conversational Interaction Using a Time-Series Data Mining Approach

Shogo Okada (Tokyo Institute of Technology)
Mayumi Bono (National Institute of Informatics)
Yasuyuki Sumi (Future University Hakodate)
Katsuya Takanashi (PRESTO, Japan Science and Technology Agency/
Academic Center for Computing and Media Studies, Kyoto University)

This paper introduced examples of gesture analysis by using a time-series data mining system. A time-series data mining system is able to reduce time and effort to annotate a large amount of nonverbal behavior data. At first, we obtain the trajectory of hand motion as multi-dimensional time-series data by using motion capture system. Second, each frame samples in the time-series data is classified to no motion patterns (such case that hand is set on home position) and motion patterns including gestures automatically. Third, frequent gesture patterns are extracted in time series by using motif (pattern) discovery algorithm. We applied these data mining methods to analyze gestures which are used in an explanation task. In the task, two participants explain the contents of animation to one participant. We collected datasets of 8 groups which are composed of 24 participants. Results of data mining show that amount of gestures in the task and timing to use gestures are different among groups and we can classify the timing in 8 groups to 3 types. From results of extracting of frequent gesture patterns, we show that 3 kind of gestures are used commonly in some groups.

Key words: gesture analysis, quantitative analysis, nonverbal interaction, autonomous annotation, data mining

1. はじめに

近年，コミュニケーション研究に関するさまざまな分野で，多人数・マルチモーダルインタラクションの分析が盛んに行われている（高梨ほか，2004; 坊農・高梨，2009; Streeh, Goodwin, & LaBaron, 2011; 細馬・片岡・村井・岡田，2011）．人文科学分野では，会話における言語・非言語に関する現象の発見や分析が多方面から行われ，人間同士の対人コミュニケーションにおけるルールや秩序に関する重要な知見が得られてきた（Saks, Schegloff, & Jefferson, 1974; Goodwin, 1981; Patterson, 1983; Heath, 1986; Bull, 1987; Kendon, 1990; Clark, 1996; 大坊，2005; 串田，2006; 西阪，2008; Sheglof, 2007; 坊農，2008; 榎本，2009）．今後，多方面からのコミュニケーションの事例分析研究，質的分析研究と，それらの研究から得られた知見を利用して，対象行動を数量化する量的分析研究がますます盛んに行われると考えられる．

2. 背景と目的

2.1 インタラクションコーパスの分析手順

一般に，談話データに対する探索型の定量的分析から新たな知見・仮説を獲得するためには，(1)分析対象の選定，(2)対象となる談話データの収集，(3)タグの設計（単位の認定，内容記述），(4)タグの付与，(5)タグの集計と統計処理の作業が必要である（伝，2006）．伝(2006)はタグの設定・付与の過程で，当初のタグ設定時には想定しなかった重要な要因が考慮されていない可能性もあるため，分析者は必要に応じて，生データに立ち返り上記の作業を繰り返し行う必要があると述べている．したがって，上記の(1)から(5)の人手による作業を，複数の事例に対して行うには多大な労力を要する．そこで，本研究ではこの労力を軽減し，分析を支援するために，センシングデバイス・情報処理技術によりコミュニケーション分析を支援する方法論を提案する．

2.2 センサによる大規模会話データの収録と分析

一般に，各種センサによる大規模会話データの収録・分析，会話行動パターン・ルールの獲得を実現するためには以下の二つの技術を実現する必要がある．

- 対面会話状況で交わされる身体配置や視線方向といった身体表現と音声言語表現とを統合的に収録する環境の構築
- 上記の環境からデータ取得される非言語パターンを分析し，頻出する会話行動パターンを抽出するためのデータマイニングツール

前者に関して，会話インタラクションのコーパスを作成することを目的とした研究プロジェクトが存在し，分析対象に沿ったデータ収録環境を構築し，データの蓄積を行っている(Carletta et al., 2005; Waibel & Stiefelhagen, 2009; Chen et al., 2006; 角・矢野・西田，2011）．本研究では後者である，視線，ジェスチャ，発話といった複数のモダリティで表出される，会話者の非言語の行動分析を支援するための情報技術に焦点を当てる．

上記の情報技術に関して，著者はジェスチャをセンサで観測し，得られた時系列センサデータにおける頻出パターンの抽出や類似するジェスチャパターンの高精度なグループ化を行う技術の開発を進めてきた（岡田・西田，2010; Okada, Ishibashi, & Nishida, 2010）．これらの情報技術は機械学習・データマイニングと呼ばれている．

機械学習とは，データの背後に潜むルールをモデルとして獲得したり，データを何らかのモデルで記述し，そのモデルを利用して未知のデータを予測する技術である．

データマイニングとは，統計学，パターン認識，機械学習などのデータ解析技法を大量のデータに適用することで新しい知識を取り出す技術である．また取り出した知識は，有識者により選別されモデル

にフィードバックされることで，モデルを徐々に精緻化していく．機械学習に関してはBishop (2006)を，データマイニングに関しては元田・津本・山口・沼尾(2006)を参考にされたい．

2.3　目的

機械学習・データマイニングの技術は人工知能分野の応用だけにとどまらず，医療・経済・環境などさまざまな分野で用いられている．しかしながら，対人コミュニケーションの分析のためのデータマイニングや機械学習の研究事例はほとんどない．これらの技術を利用すれば，大規模な会話データから非言語パターンを抽出・列挙できるため，分析者のアノテーション（タグ付け）作業の負担軽減が見込める．また，また大量のデータにアノテーションを自動で行うことができれば，分析者がデータ中のアノテーションされた非言語パターンを検索・閲覧することが可能となり，データを分析する際の効率が飛躍的に向上すると考えられる．さらに大量データから抽出された非言語パターンは，従来得られていた言語学・ジェスチャに関する研究の知見の正当性を実証したり，客観的な指標の獲得に役立つ可能性もある．

最終的に著者は，大規模な会話データにおける頻出パターンからその背後にある出現ルールに関する仮説を立てる役割を機械側が担い，分析者が発見されたパターン・ルールから仮説を立てて，機械側にフィードバックをするという一連の流れを繰り返すことで，人文科学分野でも得られていない新たな知見を大規模データから発見することを目指している．この目標を達成するためには，人分科学分野の分析者にデータマイニングの手法に関する理解を深めてもらい，利用してもらうことが必要不可欠である．このような背景から，本研究では会話中のジェスチャ分析支援のためのデータマイニング手法の方法論を提案し，分析事例を紹介する．ハンドジェスチャを対象としたのは，近年その分析が盛んになってきているものの(McNeill, 1992, 2005; Kendon, 2004)，言語による発話などとは異なり，ジェスチャ単位の認定などのアノテーション作業には膨大な労力がかかり，また，分析対象としたいジェスチャを含む断片を膨大なビデオデータの中から発見するのが容易でないため，これらの検索やアノテーションを支援することの価値が特に大きいと考えられるためである．

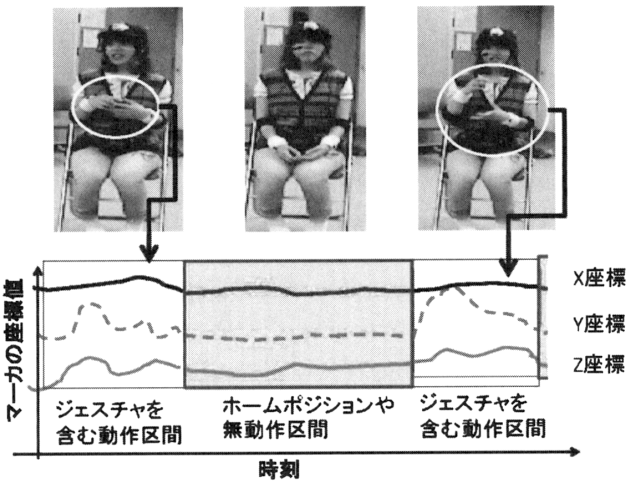


図1　モーションセンサから取得される手の動きを示す時系列データから動作区間を検出している例

注）*x*, *y*, *z*座標の座標値に大きな変動が見られる箇所をジェスチャを含む区間，座標値にほとんど変化が見られない箇所をホムポジションや無動作区間とする．

具体的には，まずハンドジェスチャ分析の第一歩として，センサで取得したデータから手の動作を自動で検出する．自動検出の概要を図1に示す．次に手の動作から小刻みな動きを表象するジェスチャや，複数人が同じシーンを説明した時に何度も使われたジェスチャ，類似する手の動かし方などを自動で抽出可能になる．ここで抽出されたパターンには，分析対象のジェスチャと無意味な手の動きとの両方が含まれてしまうため，分析者はパターン群から分析対象となるパターンを精査する必要がある．この工程は頻出するジェスチャ候補パターンの自動抽出と，分析者による精査とを含むため半自動処理となる．

本研究で提案するデータマイニング手法により分析支援可能な項目は以下の二つである．

- 着座状態における会話者のハンドジェスチャを含む手の動作区間と，ホールド状態やホームポジションに手がある場合の無動作区間の抽出(5.2)
- 会話データからの頻出ジェスチャパターンの発見・抽出(5.3)

実験ではセンサルームで三者間におけるアニメーション説明課題タスクを行い，この実験で得られたセンサデータに本手法を適用した．この結果，ジェスチャ区間の抽出・頻出パターンの獲得が可能であることを示す．またこの結果を用いて，個人間におけるジェスチャの総量の比較分析(6.1)，各セッションにおける二人の説明スタイルの違いに関する分析(6.2)，また特定のシーンに見られるジェスチャパターンの分析(6.3)などが可能になることを示す．

3. 関連研究

3.では本研究に関連する情報技術のアプローチを紹介し，本研究の位置づけを明らかにする．

3.1 マルチモーダルデータの収録と蓄積を目的としたプロジェクト

インタラクションのコーパスを作成することを目的として，複数のメディア・センサによりさまざまな形態のインタラクションデータの取得が行われてきている．ここでは主要なプロジェクトとして，Augmented Multi party Interaction (AMI) (Carletta et al., 2005), Computers in the Human Interaction Loop (CHIL) (Waibel & Stiefelhagen, 2009), Video Analysis and Content Exploitation (VACE) (Chen et al., 2006), Interaction Measurement, Analysis, and Design Environment (IMADE) プロジェクト (Nishida, 2007; 角ほか, 2011) を取り上げる．これらのプロジェクトの概要とその一環として得られた研究成果について表1にまとめる．以降では表1に沿って各プロジェクトの研究成果を述べる．

AMIは，グループミーティングのインタラクションを収録し，またそのインタラクション・コーパスを構築した．音声からの会話書き起こし，頭部方向，手のジェスチャ，視線方向といった非言語行動のアノテーションが行われた．CHILはオフィスや教室におけるグループの会話インタラクションを中心に収録している．また音声・映像データからの表情・感情理解，機械学習手法による人の動作自動検出や状況認識の技術開発が行われている．VACEも主に着座式のミーティングに焦点を当て，インタラクションを収録した．このプロジェクトでは画像処理技術を用いた会話構造の特定が行われた．IMADEプロジェクトはインタラクションのコーパスを作成することを目的として，視線計測装置，動作計測装置，音声録音設備を備えた実インタラクション収録環境を構築している．これらの環境は分析のためのインタラクションコーパスを作るためだけでなく，会話に参加できるロボットや，会話コンテンツの蓄積・活用，会議記録システムや体験記録システムなど，会話を軸とした工学的研究を行うために活用されている(Nishida, 2007).

複数センサ環境で獲得されたセンサ時系列データを可視化し，アノテーションするためのソフトとして，iCorpusStudio（來嶋・坊農・角・西田，2007）が開発された．ほかにも，分析者のアノテーションツールとしてAnvil[1]，ELAN[2]などが開発されている．

コミュニケーションをマルチモーダルデータとして観測できるデバイスと獲得されたデータをアノテーションするツールが開発されたことで，コミュニケーションデータの分析を行える環境が以前より整いつつある．今後は，センサデータの機械学習を通じて自動的なアノテーションを行うことで，アノテーション作業を軽減したり，大量データを閲覧・検索しやすくしたりする技術開発が進めば，人文科

表1 センサデータの収録とそのデータを用いたインタラクション分析・パターン認識に関するプロジェクト

プロジェクト名	情報取得環境	収録会話タイプ	コーパスを用いた研究事例
AMI	音声・動画像・MD	着座式会話	同意・非同意パターンの自動抽出
CHIL	音声・動画像	着座式会話	動画像を利用した音声認識
VACE	音声・動画像，MD	着座式会話	発言権遷移の自動抽出
IMADE	視線，音声・動画像・MD	着座式会話，立ち会話	多人数会話の非言語構造抽出

注）情報取得環境のMDとはモーションセンサにより取得した動作データを示す．

学系の分析者にとっても，コーパス作成を行う分析者にとっても有用である．

3.2　機械学習によるコミュニケーション行動の自動認識に関する研究

コミュニケーションシーンを対象としてビデオ・マイク・各種センサから獲得されるデータをシステムに解釈させるための研究は盛んに行われている．

Germesin & Wilson (2009) はAMIで構築されたデータセットを用いて聞き手の同意・非同意パターンの認識を行った．Lucey, Potamianos, & Sridharan (2007) はCHILで構築されたデータセットを用いて，複数の視点から撮影された顔の映像と音声データより，高精度に音声認識を行う手法を提案した．Chen & Harper (2009) はVACEで構築されたデータセットの，マルチモーダル情報を用いて発言権の遷移の自動抽出を試みた．Germesin & Wilson (2009)，Lucey et al. (2007)，Chen & Harper (2009) をはじめとしたAMI, VACE, CHILのプロジェクトで行われた研究に共通するのはあらかじめ認識対象の非言語行動・マルチモーダル行為を定義しておいて，モデルを構築するところである．

本研究の枠組みとAMI, VACE, CHILで行われた従来研究の枠組みとの相違点を図2に示す．図2の左はAMI, VACE, CHILのプロジェクトで行われた研究の枠組みを示す．これらの研究ではマルチモーダルデータ収録環境で収録された会話データから，機械学習によりさまざまな会話場面におけるコミュニケーション行動の自動認識器が構築された．その中では，機械学習技術は分析支援のためではなく，発話権の切り替わりなど，あらかじめ定義された事象を計算機で自動認識するために用いられる．

図2の右は本研究の枠組みを示す．本研究で紹介する機械学習・データマイニングの技術は計算機による自動認識のためではなく，人文科学分野の分析者の分析プロセスの支援のために用いられる．具体的には，大量データからジェスチャを含む動作区間の抽出を行うことにより，ジェスチャ分析者に各話者ごとのジェスチャを使う頻度に関する統計情報を提供し，ジェスチャが頻繁に使われている分析対象のデータを選択するための手がかりを与える．また

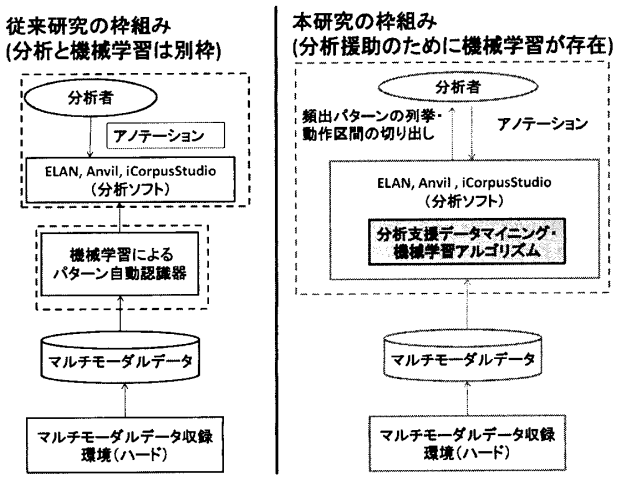


図2　我々のアプローチと従来研究の枠組みの違いを示す概念図

頻出パターンの抽出を行うことで，分析者のアノテーションを支援したり，分析者の分析対象がデータに含まれているかを確認させることができる．

3.3　センサデバイスを用いたコミュニケーション行動の分析・発見に関する研究

中田・角・西田(2011) はIMADEプロジェクトの一環として，多人数インタラクションにおける頻出する会話シーンの抽出を行い，ポスター環境発表会話とポスター環境自由会話における会話構造の比較を行った．ポスター環境発表会話はポスター発表場面の会話であり，ポスターを前にして説明者が二人の聞き手に説明をする場面を指す．ポスター環境自由会話は複数人が移動しながら複数のポスター展示物を閲覧し，自由に会話する場面を指す．発表会話では明確に説明者・聞き手の役割が決まっているのに対し，自由会話では任意のタイミングで話し手・聞き手の役割が変化する．この研究では発話・あいづち・指さし・視線・頷きをモーションセンサ・アイトラッカ・マイクより取得し，非言語行動パターンに手動でアノテーションを行った後，行動パターン列をインタラクションイベントとして抽出した．

中田ほか(2011) の研究は，3.2の自動認識を目指した研究と違い，会話データから会話構造を発見するためのデータマイニングを行っている．ここで行われたデータマイニングでは頻出パターンの抽出を行うことにより，アノテーションを支援していると

みることもできるため，本研究の目的と類似している．Bono, Suzuki, & Katagiri (2003) も，視線計測装置・マイク・環境カメラを設置して，ポスター発表会場での非言語行動・会話構造を分析している．ポスターを使った会話ではポインティングジェスチャが多数観測されることから，これらの研究ではジェスチャに関してポインティングのみを対象としており，ポインティング以外のジェスチャの頻度や使われ方を分析対象としていない上，センサデータからの非言語パターンのアノテーションは手動で行われている．

本研究で紹介する技術はセンサデータから動作区間の抽出と，頻出パターンの抽出・列挙を行うことで，ハンドジェスチャの分析の支援を試みている．またハンドジェスチャでなくても，センサ波形データであれば適用することは可能であり，Bono et al. (2003)，中田ほか (2011) の行った会話構造抽出におけるジェスチャ以外の行動の前処理にも援用することができる．

4. 分析対象のデータセット

本研究ではMcNeill (1992) が実験で用いた，“Canary Row”というアニメーションの内容・情景を説明するタスクを行う際のインタラクションを分析対象とした．このタスクでは，アニメーションの情景を説明するために，ハンドジェスチャを伴う発言が多く含まれる傾向にある．このためジェスチャの量的な分析の対象として，適切な課題の一つと考えられる．以後，このタスクを説明タスクと呼称する．

McNeill (1992) は一人の説明者が一人の聞き手に説明を行う場面を分析した．これに対し，本研究では将来的に多人数会話の分析や，共語り手によるジェスチャの同期現象の観察を目的としているため，人が協調して人の聞き手に説明を行う場面を分析対象とする．ここで共語り手とは複数人の説明者が聞き手に説明行為を行うことである (Lerner, 2002)．ジェスチャの同期現象とは複数の説明者間で同じジェスチャが引き継がれる行為である（城・細馬，2009）．

図3に示すように，三人は着座状態で会話を行う．

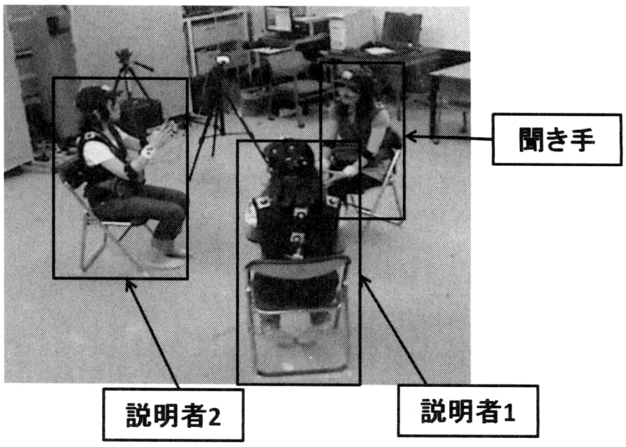


図3　三者のアニメーションタスク

上記の説明タスクを，異なる実験協力者のグループで複数回行い，各セッション（三人1グループで行う1回の説明タスク）における説明者の腕の動作に関する時系列データを分析する．人材派遣会社を通じて計30人の実験協力者を募集した．募集した30人はいずれも20～24歳の女性である．三人1グループで計10セッションの説明タスクを行い，センサデータを収集した．1セッション目のデータは多人数・マルチモーダルインタラクション研究のためのプラットフォーム構築に向けて収録されており，公開されている[3]（坊農・角・高梨・岡田・菊地・東山，2011）．

10セッションのうち，収録エラーなどのあった2セッションを除外して，計8セッションのデータをデータマイニングによる分析対象とした．

5. データマイニングシステムの狙いと工夫

本研究で提案するデータマイニングシステムを図4に示す．提案システムは動作区間の抽出アルゴリズムと頻出するジェスチャパターンの発見アルゴリズムにより構成されている．以下にシステム構成を述べる．手順の各ステップ番号は図4の番号と対応している．

1. 会話データにおける説明者の手の動作データの取得 (5.1)
2. 手の動作データから説明者の動作区間を抽出 (5.2)
3. 動作区間のデータから頻出するジェスチャパ

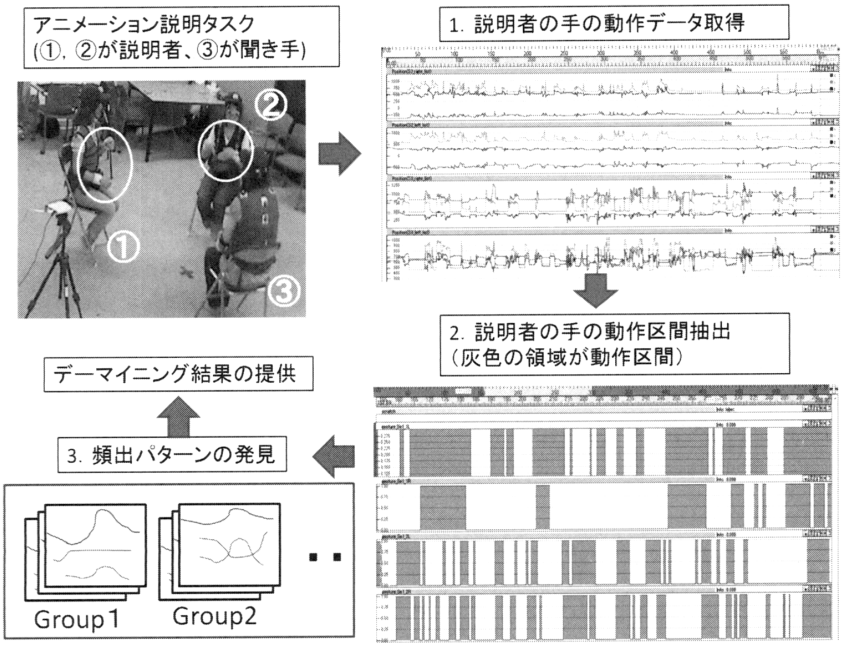


図4　提案するジェスチャ分析のためのデータマイニングシステムの概要

ターンを発見(5.3)

5.1　ジェスチャデータの取得環境

本研究ではマルチモーダルデータ収録環境(IMADEルーム；角ほか，2011）でジェスチャデータの取得を行った．IMADEルームでは，動作センサデータ，視線，生理指標，音声，動画像が収録可能である．各種センサを実験協力者に装着している例を図5に示す．データ収録機材のうち，ハンドジェスチャを計測するために，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を用いた．

Mac3Dを用いる場合，会話者は関節などに図5に示す複数のマーカを装着する．複数のラプターカメラでマーカに，カメラの方向からライトを当てると，反射して反射光がカメラに戻る．この反射を利用して，関節の位置座標を示す各マーカの三次元座標位置が計測できる．

モーションキャプチャにより得られる各マーカの三次元座標は部屋の隅を原点とする座標系の値である．本研究では会話者の手の動きを測定するため，各会話者の中心位置を原点とした座標系に変換する．以下に座標変換方法を説明する．会話者の中心位置，座標変換に利用した肩のマーカの位置と両手

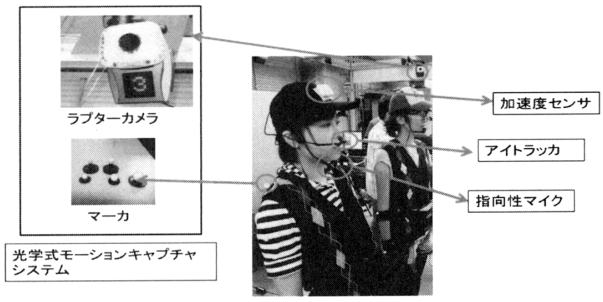
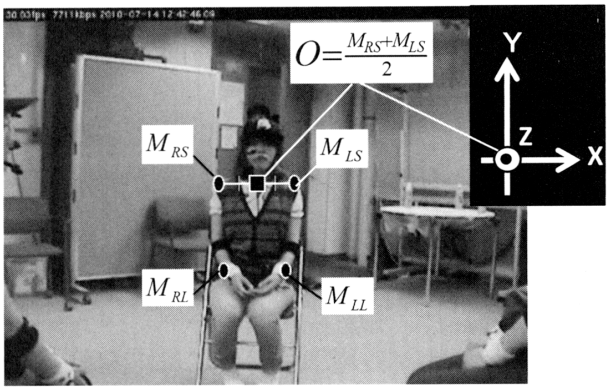


図5　データ収録で用いた各種センサ

注）ハンドジェスチャのデータマイニングには腕に設置したマーカの三次元座標値を用いる．



M_*, Oは3次元の位置座標

図6　会話者の中心位置，座標変換に利用した肩のマーカの位置と両手首のマーカの位置

首のマーカの位置を図6に示す．まず原点とする会話者の中心座標を両肩に付けたマーカの中心座標Oと定義する．

$$O=\frac{M_{RS}+M_{LS}}{2} \tag{1}$$

上式においてM_{LS}, M_{RS}は左肩，右肩のマーカの三次元座標をそれぞれ示す．左，右手首に装着したマーカの三次元座標をM_{LL}, M_{RL}とすると，会話者の中心座標を基準とした両手首のマーカの三次元座標は以下のように計算される．

$$\begin{aligned}\hat{M}_{LL}&=M_{LL}-O\\ \hat{M}_{RL}&=M_{RL}-O\end{aligned} \tag{2}$$

会話者の中心座標に対し，水平方向の動きをx軸，垂直方向の動きをy軸，奥行き方向の動きをz軸とする

以下の動作区間抽出，ジェスチャパターン発見の入力には取得される両手首のマーカ$\hat{M}_{LL}, \hat{M}_{RL}$の座標値$(x, y, z)$の3次元データを用いる．両手首にマーカを装着し，両腕のジェスチャを観測するため，合計6次元の時系列データを取得する．データは120フレーム毎秒で取得される．会話中に会話者の身体が動く可能性があるため，毎フレームごとに肩のマーカの位置座標M_{LS}, M_{RS}を計測し，式(1)，式(2)に従い$\hat{M}_{LL}, \hat{M}_{RL}$の計算を行う．

5.2　動作区間の抽出

一般的に会話中では話者は意識的であれ，無意識的であれ手を動かす場面と，手がホールド状態やホームポジションにある，手が動かない場面が存在する（細馬，2009）．まず大量の時系列データから手の動きを検出できれば，ジェスチャを行っている可能性のある区間を特定することができる．この結果，分析者は手の動作区間とジェスチャが行われていない無動作区間を分離して観察することができるため，分析の負荷を軽減できるというメリットがある．

5.2.1　動作区間抽出アルゴリズムの概要

モーションセンサから得られる手の動きの時系列データの例が2.の図1である．図から明らかなように，手が動いていない領域のグラフの灰色で示した無動作区間では時系列データの変化量が極端に少ないことがわかる．

そこで，本研究では動作区間抽出にHidden Marko Model (HMM)を用いる．HMMを用いた無動作区間抽出の概念図を図7に示す．HMMとは，各状態（図7の左におけるSl, S2）が確率的に遷移することを仮定した確率状態遷移モデルの一つである．各状態は，時系列中の三次元座標データの確率分布を有している．三次元座標データの時系列変化を状態の遷移によって表現していると解釈できる．図7右の波形データにおける無動作区間の特徴量（微小に変化する，または変化のないデータ）が各状態内の確率分布として表現される．

HMMのアルゴリズムを実装するためには，時系列データが入力された際のHMMの確率計算と，HMMのパラメータ推定方法を理解する必要がある．詳細はRabiner (1989)を参照されたい．

無動作区間では手の微小な揺れなどの理由で，少量の時系列変化が生じる．この変化はジェスチャに関わる動作ではないノイズと見なせるが，この微小な変動を吸収するために確率モデルであるHMMを用いる．ここで腕のマーカーのy座標がある閾値以上大きく変化したら動作区間と判定するようなシンプルな方法では少量の時系列変化に伴うノイズに反応してしまうため，検出精度が低下してしまう．この結果から，HMMが動作の学習に適していることがわかる．

HMMへの入力には腕のマーカの三次元位置座標の時間差分を用いた．無動作区間の判定基準となる学習データを選定する．今回は筆者がビデオと時系

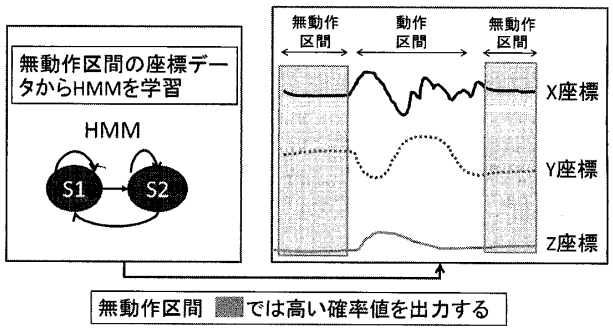


図7　HMMを用いた無動作区間の検出

列波形を観察した結果，セッションの開始後3秒まで会話者の手がホームポジションにあることを確認したため，開始後3秒間のデータを学習データとした．

HMMの学習後には手を動かしていない場合の系列データが入力されると，HMMからは無動作区間が高い確率で出力され，手の動作が始まると，低い確率が出力される．この性質を利用し，学習したHMMの出力確率が大きい時間区間を無動作状態（ホームポジションまたはホールド状態）と判定した[4]．

5.2.2　HMMの動作例

図8は説明シーンにおいてある話者の左手のセンサから得られた3次元の位置座標を80秒間取得した時系列データに対してHMMによる無動作区間抽出を行った結果を示している．左手が動くとセンサは反応するが，左手が無動作の状態ではマーカの観測ノイズを除いて変化はない．図において灰色領域が無動作と判定された区間である．無動作区間と判定されたフレームを1，動作と判定したフレームを0としている．マーカの観測ノイズを含む無動作区間が正しく抽出できていることがわかる．

5.3　インタラクションデータからの頻出ジェスチャ候補パターンの発見

動作区間を抽出した後，動作区間に対応する時系列データから頻出する時系列パターンを抽出する．頻出する時系列パターンとは時系列データの中に含まれる類似する信号パターンである．類似するパターンは類似する手の動かし方を示しており，ジェスチャパターンの候補であると考えられるため，このパターンを頻出パターンとして抽出することを試みる．

5.3.1　時系列データからのパターン発見アルゴリズムの概要

本研究では長時間の手の動きのデータから頻出するパターンの抽出を行える，SAX+Random Projection (Chiu, Keogh, & Lonardi, 2003)を用いる．SAX+Random Projectionの詳細や，実装方法についてはChiu et al. (2003)を参照されたい．

SAX+Random Projectionは一次元の時系列データに適用可能な方法であるため，本研究では両手首のマーカから得られた6次元の時系列データから各次元ごとに計算を行う．この結果として，右手だけを垂直に真上に挙げるようなジェスチャであれば，右手首マーカのy座標（上方向成分）の時系列データから頻出するパターンが抽出されることになる．最後に各次元で抽出されたパターンを統合する．SAX+Random Projectionのアルゴリズムの挙動例を図9に示す．

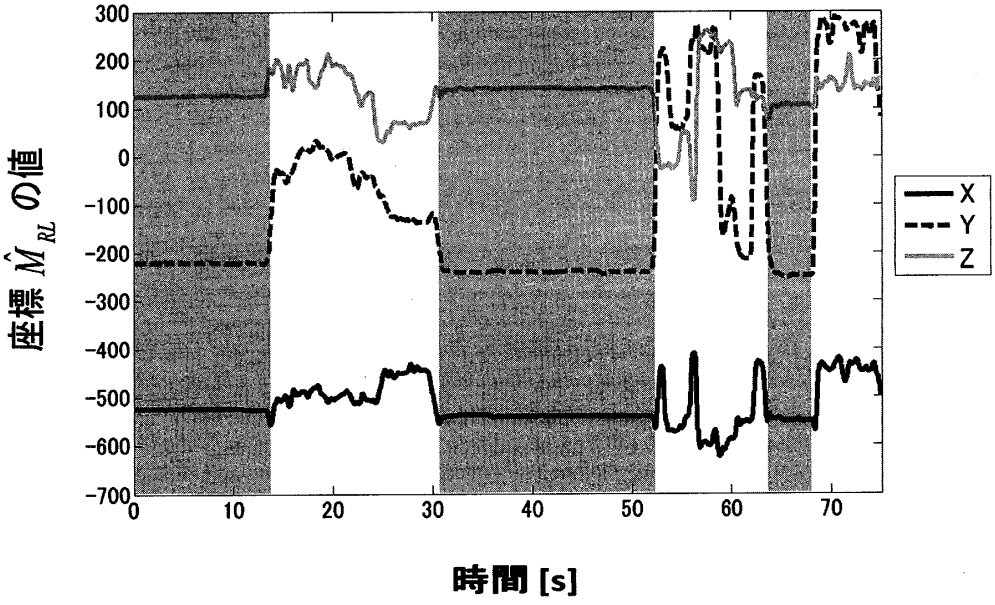


図8　無動作区間および動作区間の抽出例

注）無動作と判定されれば1，動作と判定されれば0を出力する．灰色領域は無動作と判定された区間である．

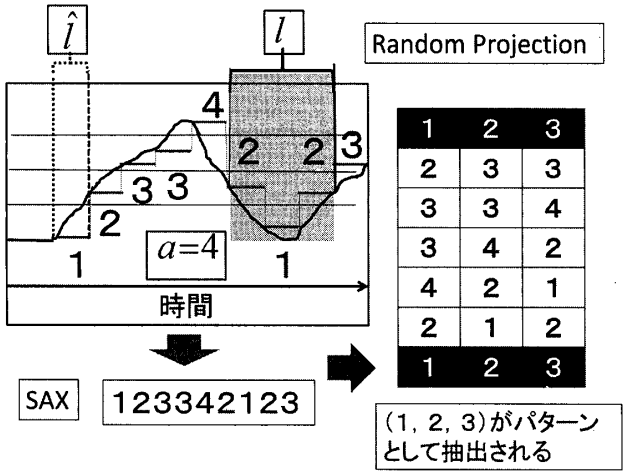


図9　SAX+Random Projectionアルゴリズムの概念図

最初に連続量の時系列データの各フレームのサンプル値を離散化するためにSAX (Symbolic Aggregate approXimation)を用いる．離散化とは，図9の左図の時系列データ波形を自然数列に変換する処理である．時系列データはSAXにより自然数列に変換され，Random Projectionアルゴリズムにより，数字列のマッチングが行われ，ここで類似する自然数列（例えば(1, 2, 3)と(1, 2, 2)など）はジェスチャパターンの候補として抽出される．また伸縮を有する時系列データ，例えば(1, 2, 2, 2, 2, 3, 3)のような自然数列は(1, 2, 3)と変換して扱う．この操作により，時系列データの各部位が伸縮するパターンを同じグループとして抽出できる．SAX+Random Projectionを動作させるために，a, $\hat{l}$, lのパラメータを設定しておく必要がある．

aは離散化レベルである．図9ではa=4となっており時系列データは1から4までの数字のいずれかに割り当てられ，自然数列として出力される．センサの座標の数値同士は細かすぎるため，ほとんどマッチしない．このため，一定の規則に基づき離散化する必要がある．逆に，離散化が粗すぎる（aが小さい）と，異なる手の動きを同一の離散レベルに分類してしまう危険がある．離散化の度合いはデータごとに異なるため，この部分はパラメータ化して，対象データに合わせて調整できるようにしておく必要がある．離散化は座標データの大きさ（図9の左図のy軸方向)に関するものであるのに対して，同様の調整を時間軸（図9の左図x軸方向）について行うためのパラメータが$\hat{l}$である．つまり，$\hat{l}$はどのくらいの長さの時系列データを一つの数字に置き換えればいいかを調整するためのパラメータである．最後にlは抽出したいジェスチャ候補パターンの系列長を示す．図では長さ$\hat{l}$の領域を灰色で示し，lの領域を点線区間で示している．また図ではジェスチャ候補パターンの系列長lを$3\hat{l}$に設定している．

図9中の左の波形データは実験協力者の右腕部の装着したマーカの三次元座標の内のy座標であり，垂直に真上に挙げる動作(1, 2, 3)がジェスチャ候補として抽出されていることを示している．

5.3.2　SAX+Random Projectionの動作例と各次元における結果の統合方法

実際の腕のマーカから取得された時系列データに本アルゴリズムを適用した例を図10に示す．黒太線の四角で囲まれた二つの領域に含まれる波形は同じ軌跡を持つと推定されるジェスチャを示している．上の図が入力時系列で，下の図が入力時系列にSAXを適用して離散化した後の結果である．この離散シンボル列同士は非常に近い値になっていることがわかる．これらのシンボル列はRandom Projectionにより抽出される．

最後に各次元で獲得された頻出するパターンを統合する．右手または左手の動きから得られるマーカの三次元(x, y, z)の座標の時系列データから抽出されたパターンを統合する例を図11に示す．まずx座標の時系列からはp1, p2, y座標の時系列からはp3, z座標の時系列からはp4のパターンが抽出されている．これを各次元間で統合する場合，例えばp2とp3は同じ時刻で抽出されていることから，p2, p3を統合して抽出する．この統合パターンを2とする．同様に同時刻で抽出されたパターンを統合していくと，図11の例ではc1, c2, c3という種類のジェスチャ候補パターンが最終的に発見される．

6.　提案データマイニング手法を援用した分析事例

紹介したデータマイニング技術をコミュニケー

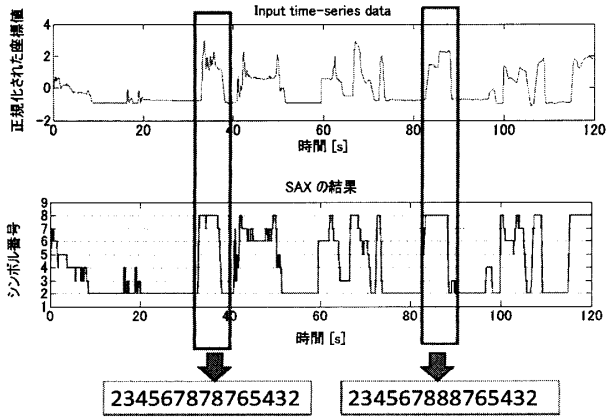


図10　SAXアルゴリズムにより時系列データがシンボル列に変換される様子

注）類似するパターンは類似するシンボル列に変換されていることがわかる．

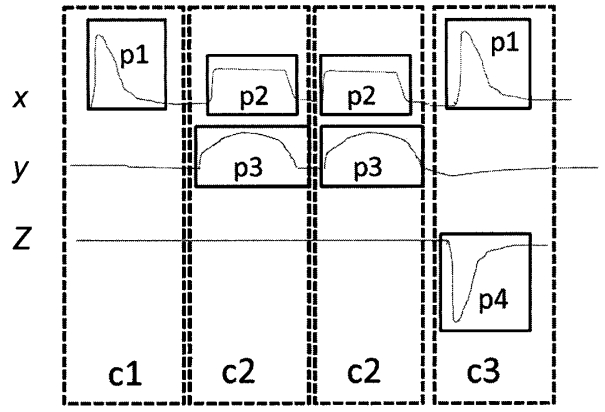


図11　各次元で抽出されたパターンを統合する手法の概念図

ション分析に援用することで，以下の機能が実現される．

- 大規模データからジェスチャを含む動作区間を抽出し，各セッション・各会話者ごとの動作頻度を自動的に算出する．
- 大規模データから頻出パターンを列挙し，分析者に提示する．

4.のデータセットの分析に，上記の二つの機能を援用した場合に可能となる分析支援の内容を以下に挙げる．

1. セッション間・シーン間でのジェスチャの総量の比較分析(6.1)
   各セッションの説明者が行ったハンドジェスチャの分量を比較することにより，ハンドジェスチャを多用して説明しているセッションや，各セッションの中でハンドジェスチャを多用しているシーンを特定できる．
2. 二人の説明者のジェスチャ頻度比較による説明方法の分析(6.2)
   検出された動作区間にラベルを付与した後に，ラベル系列を観察することで，各時刻における二人の説明者のジェスチャを使った箇所とその頻度を分析できる．この分析からは，二人が交互に説明をしているのか，一方の説明者が一方的に説明を行っているのかを検証することが可能となる．
3. 各セッションで共通に用いられるジェスチャパターンの抽出・分析(6.3)
   5.3で述べた手法を用いたアニメーション課題データ中に現れるジェスチャパターンの抽出・分析事例を紹介する．

6.1　分析事例1：セッション間・シーン間でのジェスチャの総量の比較分析

6.1.1　各説明者の動作区間抽出

HMMに基づく手法を用いて，各セッションにおける二人の説明者の動作区間を抽出した．ここで各セッションにおいて説明タスクにかかる時間が異なるため，説明タスクにかかった時間を用いて正規化し，以下の式で動作区間の割合を計算する．

$$\text{動作区間の割合} = \frac{\text{動作区間と判定されたフレーム総数}}{\text{説明タスクにかかったフレーム総数}}$$

各セッションにおける動作区間の割合の比較を表2に示す．表2の一番右の列は各説明者の右手・左手の動作量の差の絶対値を示す．S*の*はセッション番号を，Sp*の*は説明者番号をそれぞれ示す．なお説明者番号の1, 2は図3の説明者1, 2に対応する．実験結果より，各セッションで動作区間の割合が異なることがわかる．セッション4, 5, 6のジェスチャは他のセッションに比べ，動作区間の割合が小さい．比較的動作区間の割合が大きかったセッション1, 2, 3, 7, 8では以下の傾向がみられた．セッション1と8では説明者の両方とも動作区間の割合が大きく，二人の説明者ともに説明中にたくさんの動作

表2　各セッションにおけるジェスチャの割合

話者ID	左手動作割合：r_1	右手動作割合：r_2	左右動作割合差：$\lvert r_1-r_2\rvert$
S1-Sp1	0.80	0.69	0.11
S1-Sp2	0.67	0.74	0.07
S2-Sp1	0.54	0.56	0.02
S2-Sp2	0.36	0.43	0.07
S3-Sp1	0.61	0.77	0.16
S3-Sp2	0.56	0.42	0.14
S4-Sp1	0.12	0.15	0.03
S4-Sp2	0.16	0.21	0.05
S5-Sp1	0.23	0.38	0.15
S5-Sp2	0.29	0.25	0.04
S6-Sp1	0.45	0.49	0.04
S6-Sp2	0.47	0.35	0.12
S7-Sp1	0.61	0.45	0.16
S7-Sp2	0.22	0.32	0.10
S8-Sp1	0.73	0.85	0.12
S8-Sp2	0.66	0.85	0.19

を行っているため，ジェスチャを多用している可能性が高い．一方でセッション2, 3, 7では一方の説明者の動作割合は一方の説明者のそれに比べて多い．この結果より，これらのセッションでは一人の説明者が説明を主体的に行っている可能性が高い．セッション1, 2, 3, 7, 8のビデオ映像を検証したところ，上記の自動抽出結果による説明者のジェスチャ頻度に関する仮説が正しいことがわかった．

次に，各話者の右手・左手の動作量に着目して分析を行った．8セッションの計16人の説明者の内，15人が右利き，一人が左利きであった事実から，右手の動作量の平均値が多いであろうという仮説を立てた．この仮説を確かめるために，t検定の片側検定を行った．また同一人物の右手・左手の動きは関連するため，対応のある検定を行った．図3の各値は動作区間の割合（比率）であるため，角変換を行った後．両手の動作量に差がないという帰無仮説に従い，検定を行った．検定の結果，有意水準5%で帰無仮説は採択された($t=-1.10, p=0.145$)．この結果より，右手・左手の動作量の平均に差があるということは認められず，右手の動作量が多いという仮説は成り立たないことが確認された．この原因を調べるため，各セッションの説明者ごとに動作量の違いを分類した．

セッション1, 7では説明者1の左手，説明者2の右手の動作量が多く，セッション3, 5, 6では逆に説明者2の左手，説明者1の右手の動作量が多い．各セッションの説明者はセッション6の説明者1を除いて右利きであることから，セッション1, 3, 5, 6, 7の6名の説明者は利き手でないほうの手が頻繁に動いていることがわかる．

上記の理由を以下のように考察した．Özyürek (2002)は，話し手から見た聞き手の座席位置の違いがジェスチャの産出に影響を与えていることを指摘している．本研究のデータでも，二人の話し手のそれぞれから見て，聞き手役の参与者が自分のどちら側にいるかが異なっている．そのため，自分と聞き手との位置関係によっては，利き手ではない左手でのジェスチャが多用されることになるのではないかという仮説を導くことができる．

6.1.2　各シーンにおけるジェスチャ量の比較分析

次に説明タスクにおける話題ごとに，動作区間の割合の比較分析を行う．この分析を行える条件は，話題の変化が明確なタスクであることと，事前にその話題の変化点をアノテーションできていることである．“Canary Row”のアニメーションでは明示的にシーンが分かれているうえ，そのシーンを説明する時に特徴のあるジェスチャが使用される．すべてのセッションに共通して説明された内容を，“Canary Row”におけるシーンに従って時系列順に列挙する．

0. アニメーションの登場キャラクターと設定
1. 猫がマンションの正面から入るシーン
2. 猫が排水管の外側を伝ってヒヨコを捕まえようとするシーン
3. 猫が排水管の中を伝ってヒヨコをつかまえようとするシーン
4. 猫が猿になりすましマンションに侵入するシーン
5. 猫がドアボーイになりすますシーン
6. 猫がシーソーを使って飛び上がるシーン

7. 猫がロープを使いターザンのまねをするシーン
8. 猫が電線をつたい，感電するシーン

ここでは事前に話題の変化点を音声データから特定しておき，そのシーンごとに動作区間の量を比較する．動作区間の長かったセッション1, 3, 8について，各話題ごとの動作区間の時間を比較した結果を図12に示す．この3セッションでは動作区間の割合が非常に大きいため，各話題の説明における動作区間の総時間を単位に用いている．図において横軸は各シーンの番号であり，縦軸は各セッションにおける説明者のIDである．各マスの色は該当するシーンにおける動作区間の総時間を示している．

図12より60s以上，手を動かしながら説明しているシーンについて考察する．3セッションで共通しているのは，シーン0についての説明にジェスチャを多く用いるとともに，丁寧に説明していることがわかる．セッション1, 8の説明者はシーン4, 5についての説明をジェスチャを多用しながら説明していることがわかる．セッション8は特に多くのシーンでジェスチャを多用していることがわかる．

各シーンに着目してジェスチャを分析したい分析者にとって，この分析結果を利用して着目するシーンで動作を多く行っているセッションから分析を始めることが可能となる．例えばマクニールはシーン3について実験協力者が説明するシーンを例にとり，キャッチメントの存在を説明している(McNeill, 2005)．このキャッチメントの存在を本実験で確かめたければ，シーン3で多くのジェスチャを行っているセッション8から分析すると効率的である．

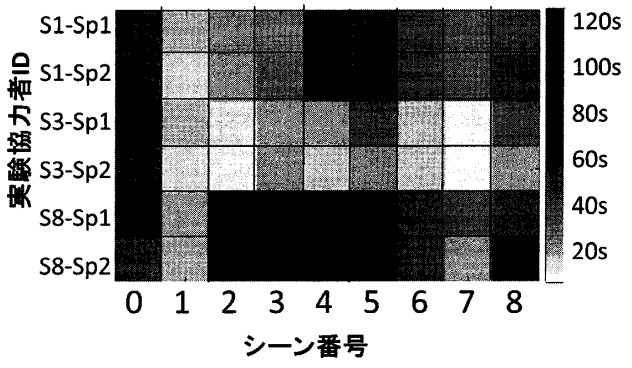


図12　各シーンにおけるジェスチャ量の比較

注）横軸は各シーンの番号であり，縦軸は各セッションにおける説明者のIDである．各マスの色の濃淡は該当するシーンにおける動作区間の総時間を示している．各マスの色の濃淡は右軸のグラデュエーションの色と対応しており，濃いほど総時間が長く，淡いほど短い．

6.2　分析事例2：二人の説明者のジェスチャ頻度比較による説明方法の分析

6.2では，動作区間を時系列順に並べて可視化し，各セッションにおける動作区間の時系列変化を分析する．各セッションにおける二人の説明者が同期的に手を動かしているのか，それぞれが交互に手を動かすのかを調べることで，二人の説明方法の特徴を分析する．ここでは6.1.1で，動作区間の総量が多かったセッション1, 2, 3, 7, 8について動作区間の時系列変位について分析した．その結果これら五つのセッションにおいて，二人の協調的説明には大きく分けて三つのパターンが存在することが示された．典型的な三つのパターンを図13に示す．図13にはセッション1，セッション8，セッション3において，7秒から200秒までに抽出された動作区間を時系列に並べた様子を示している．ここでは左手・右手の動作があった区間をマージして，各説明者につき1本の時系列データを記載している．

まず一番上のセッション1では，50秒から65秒区間を除き，一方の説明者が動作をしたのちに，もう一方の説明者が動作を行っている様子がわかる．このセッションではどちらかが説明を始めると，そちらに委ね，交互に動作を行いながら説明している様子が見て取れる．次にセッション8では，二人の動作区間がほぼ同期的に現れ，二人とも同時に手が動いている場面が多いことがわかる．最後にセッション3では，50から60秒までは説明者S1の手は動いていないが，他の場面ではS2が手を動かしているいかんにかかわらず，S1が手を動かしている様子がわかる．実際にセッション3ではS1が発話の主導権を取っていることが多く，表2からもわかるように動作の量も偏っている．

以上をまとめるならば，5.2で述べた動作区間の自動抽出の結果を利用することで，6.1.1，6.1.2，6.2の分析・また分析補助を行うことができることを示

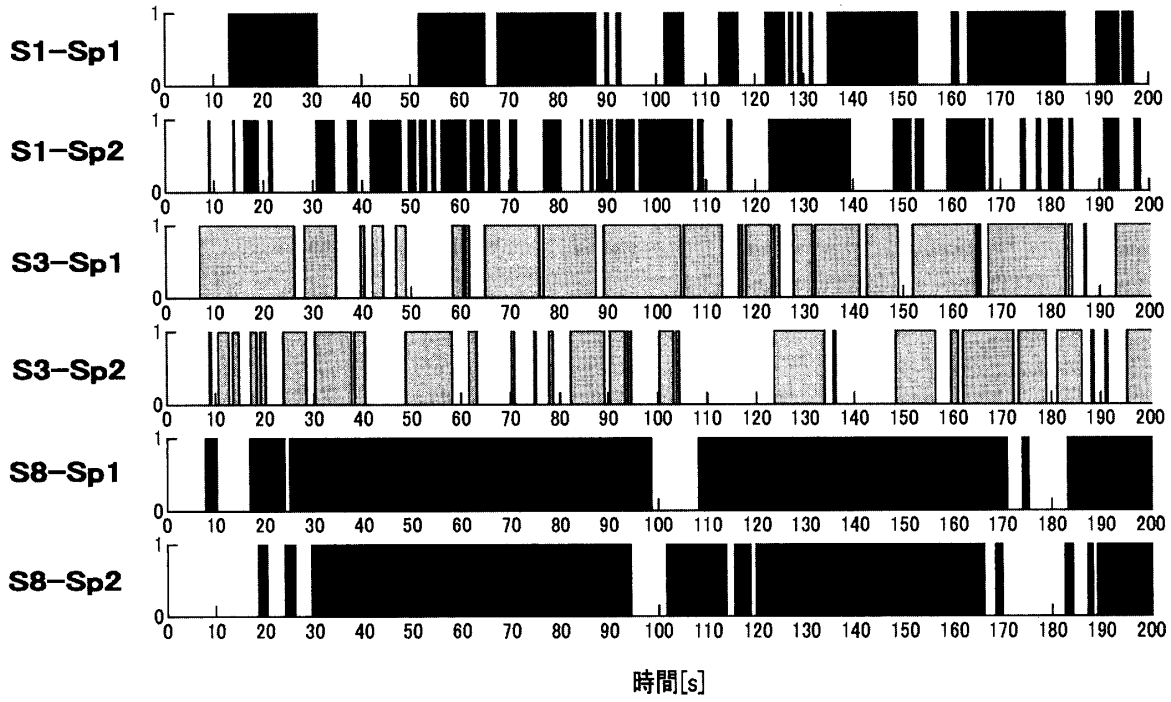


図13　各セッションにおける二人の説明者の動作区間の時系列変化

した．上記のところで行った分析の結果，説明タスクにおいて，説明者二人がジェスチャを多用し，協調的に説明を行っているセッションは1, 8であることがわかった．

6.3　分析事例3：各セッションで共通に用いられるジェスチャパターンの抽出・分析

6.3では5.3で述べたパターン発見アルゴリズムを用いて，頻出するジェスチャ候補パターンの抽出を行う．6.1.2でジェスチャを多用していた，セッション1, 3, 8における説明者の腕のマーカの時系列データから頻出パターンの抽出を行う．

まず以下の事前実験を行い，アルゴリズムのパラメータを決定する(5.3)．セッション1の実験協力者の右手の動きから取得した3分間（2160フレーム）の時系列データに，手をホームポジションから上に上げる動作が3回含まれおり，この3回の動作が正確に抽出できるパラメータを探索した．この結果パラメータをa=8, $\hat{l}$=1と決定した．抽出したいジェスチャパターンのフレーム数lはあらかじめ未知なので，パラメータを固定せず短いパターンから長いパターンまで探索を行う．具体的にはl=[150, 200, 250, 300, 400]の場合につきパターンの発見を行う．120フレーム毎秒でデータが取得できるため，フレーム長150から400のパターン発見は1.25秒から3.33秒までの長さのジェスチャ候補パターンの発見を行うことに相当する．

実験結果を表3に示す．表中の「パターン」とは抽出された「箇所」の総数，「クラスタ」はパターンの「種類」の数を示す．パターンのフレーム長150 (1.25秒)，200 (1.68秒) に設定した場合，短い系列パターンを抽出するため，大量の候補パターンが得られていることがわかる．これらのパターンは非常に短く，例えば手を少し上げたり，下げたりするパターンに対応している場合が全てである．このことから，ハンドジェスチャを分析する場合には，できる限り長いパターンを分析する必要があることがわかる．パターン長250, 300, 400に設定した場合に得られるパターンの多くは，ホールド状態に近い状態で手が動いている動作が多く抽出されたものの，意味をもつジェスチャが3種類抽出された．この3種類は，「双眼鏡の形を作るジェスチャ」，「傘を振り下ろすしぐさを表すジェスチャ」，「手回しオルガンを弾く小刻みな動きを表象するジェスチャ」であった．抽出された三種類のジェスチャを図

表3　頻出ジェスチャ候補パターン発見の結果

パターンのフレーム長l	パターン数	クラスタ数
150	2074	215
200	825	120
250	421	95
300	196	70
400	71	31

14～16にそれぞれ示す．各図において，上図はジェスチャをしている説明者をキャプチャした動画から作成したイラスト，下図は右手の動き（$\hat{M}_{RL}$の座標値の系列データ）から抽出されたジェスチャ候補パターンの波形を示す．以下では，これらの3種類のジェスチャについて，データ上の特徴と，ジェスチャ単位内のフェーズ（細馬，2009）の観点からの考察を述べる．

まず双眼鏡を目の前に持ってくるジェスチャが複数人のデータより発見された（図14）．図14の下図は右手の動きから獲得されたパターンを示すが，左手についても同様に膝付近から，手で双眼鏡を作り，目の前に双眼鏡を持ってくるジェスチャを抽出できた．抽出されたジェスチャはフレーム長400（3.33秒間）であった．このジェスチャでは，双眼鏡を覗き込んでいる間は両手がほとんど上下動しないが，この区間は「双眼鏡を覗く」という意味のある動作を表現する部分であるため，ジェスチャフェーズとしてはストローク(stroke)部分に当たると考えられる．

次に，猫を叩くために傘を振り下ろすジェスチャが抽出された（図15）．このジェスチャは一人の話者の手の動きから複数回連続で抽出された．図15の下図において，フレーム長300（2.50秒間）の黒の太線で四角く囲まれ部分が頻出パターンとして抽出された．腕の急な上昇が準備(preparation)，急な下降がストロークに，それぞれ対応しており，この上下動の波形の回数から，「傘を振り下ろす」という動作が何回行われたかがわかる．

最後に手回しオルガンを弾く小刻みな動きを表象するジェスチャが獲得された（図16）．このジェス

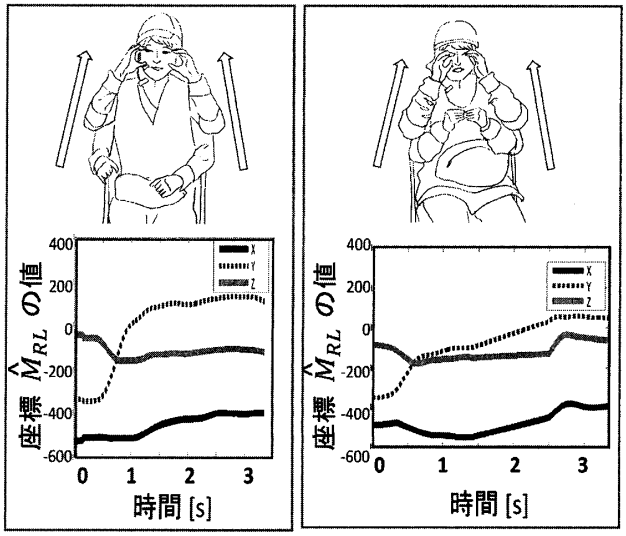


図14　抽出されたジェスチャ1：双眼鏡の形を作るジェスチャ

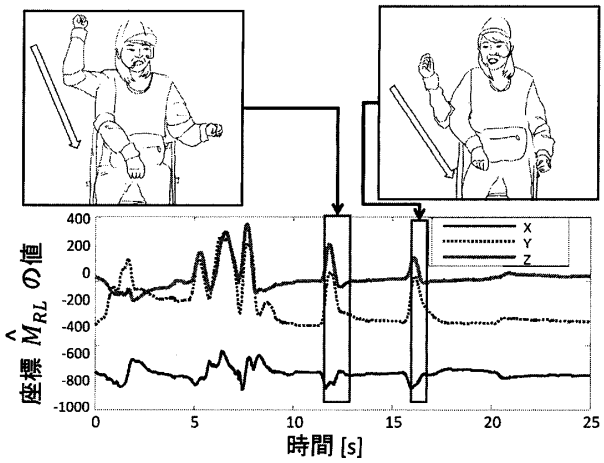


図15　抽出されたジェスチャ2：傘を振り下ろすしぐさを表すジェスチャ

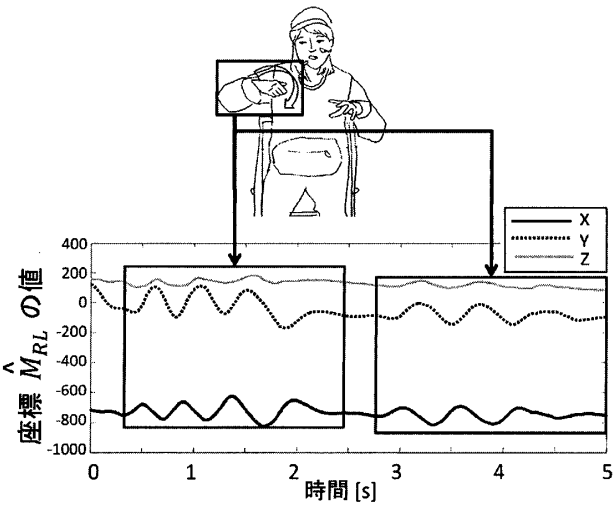


図16　抽出されたジェスチャ3：手回しオルガンを弾く小刻みな動きを表象するジェスチャ

チャも一人の話者の手の動きから複数回抽出された．図16についても，フレーム長250（2.08秒間）の黒の太線で四角く囲まれ部分が頻出パターンとして抽出された区間を示す．上記の結果より，類似する腕の軌跡を持つジェスチャから得られる時系列パターンの形状は類似していることがわかる．ストロークが1回の直線的な動きではなく，小刻みな動きの反復によって表現されるという意味で，特徴的なジェスチャであることがわかる．

7. 分析結果の考察と今後の課題

7.では分析結果から得られた知見と今後の課題をまとめる．

7.1 データマイニングが人手分析に与える効果

分析では，まず説明者のジェスチャを含む動作区間の自動抽出技術を用いた結果，アニメーション説明課題におけるアニメーションの各シーンごとのジェスチャ頻度の分析，語り手が交互に説明するのか，一方の語り手が一方的に説明するのかの説明スタイルの分析が可能となった．また頻出するパターンを抽出し，最終的に3種類のジェスチャを抽出した．

上記の分析結果と同様の結果は，人手分析でも得られると考えられる．したがって今回の分析例では，データマイニングによる分析が人手分析を凌駕するような結果を得られたとは言い難い．しかし，今回用いた8セッションのビデオの合計時間は91分27秒であり，約1時間半のビデオデータからジェスチャ区間を抽出したり，頻出パターンを数え上げる作業は非常に手間がかかる．本研究の技術では，自動でジェスチャ区間を抽出し，頻出パターンを見つけることが可能であることから，アノテーションの手間を省くことができている点で分析を支援していると考えられる．

セッション数が100以上のオーダになった場合，一人の分析者によりアノテーションを行うのは現実的に難しくなり，複数の分析者によりアノテーションを行う必要が生じるため，人的コストがかかる．複数の分析者によりアノテーションを行う場合には，全員で一致したアノテーションルールを作成する手間もかかる．これに対しセンサデータとビデオ観察の結果付与された，高品質のアノテーションデータから，本研究のように無動作区間を学習してジェスチャ区間を検出すれば，人目で観察するよりも正確なジェスチャ区間の切り出しを行えると考えている．学習されたモデルは，複数人の分析者の間で共有できて便利である．

1.で述べたように，本研究の最終目的は，データマイニング手法と分析者の協調分析により，人文科学分野でも得られていない新たな知見を大規模データから発見することである．この目標のためには，抽出されたジェスチャパターン群から分析者が重要なパターンを選択し，この結果をデータマイニング側にフィードバックする．このフィードバック結果を利用して，データマイニング機構のパラメータ調整やアルゴリズム自体の修正や新たな処理の追加などを行ったり，分析者の選択したジェスチャパターンに類似するパターンを抽出したりするために，データマイニング側と分析者側の継続的な相互作用が必要となる．

7.2 パターン発見アルゴリズムの今後の課題

本研究においてパターン発見アルゴリズムは頻出するジェスチャパターンの数え上げを支援するツールである．この機能は坊農・高梨(2009)で定義されたカテゴリカルアプローチの，ジェスチャ分析を支援していると解釈できる．しかしながらノイズも多く，抽出された多くのジェスチャは分析対象ではない，無意味なパターンであった．今後，これらの無意味なパターンを分析者の教示に従い削除するアルゴリズムを実現する必要がある．

今後はジェスチャパターンだけでなく，発話やあいづち，視線，うなずきなどについても同様の手法を用いたデータマイニングが可能なので，マルチモーダルな複数時系列データの組み合わせから，参与者個人での複数の行動の共起や複数参与者間の行動の継起などのマルチモーダルインタラクションパターンの発見も可能であろう．

角ほか(2011)はセンサデータから階層的にインタラクション行動を規定するための，インタラクションの階層的解釈モデルを提案している（角ほか，

2011における図2)．このモデルにおいて，1層目は，センサから得られる座標データから発話の有無や動作の有無などの要素行動パターンを抽出する役割を担う．次に，2層目は要素行動パターンの組み合わせで，ある対象を「指さした」，ある人に向けて手を挙げたなどのインタラクション要素を定義する役割を担う．3層目ではインタラクション要素を組み合わせて，「共同注視」といった複数人によるインタラクションイベントの抽出が行われる．このモデルにおいて本研究の枠組みは，ジェスチャに関するインタラクション要素となるパターンを抽出できる．将来，マルチモーダルインタラクションパターンの発見が可能になれば，複数人によるインタラクションイベントの抽出も可能になると考えられる．

7.3　時系列マイニング手法を普及させるための課題

今後アルゴリズムの拡張・精緻化・高速化を行うとともに，多くのジェスチャ分析者に時系列マイニングの手法を普及させる必要がある．このため，ジェスチャパターン分析のための時系列データマイニング手法のライブラリの開発と配布を今後行う予定である．今後，簡便に分析者の有する環境でジェスチャデータを取得・分析できるよう，安価かつ簡便にデータを取得できる，加速度センサ・地磁気センサ（加速度，角速度，地軸の情報を観測できるセンサ）等のセンシングデバイスを用いた場合にも本手法が適用できることを検証する必要がある．

8.　結　　論

本研究ではセンサで取得した話者のハンドジェスチャを，多変数時系列データとしてとらえ，この系列データからジェスチャを行っている可能性のある動作区間の抽出を行い，この動作区間に対し，頻出する非言語パターンを抽出する手法を提案した．

実験では三者間におけるアニメーション説明課題タスクを行い，この実験で得られたセンサデータに本手法を適用した．この結果，ジェスチャ区間の抽出・頻出パターンの抽出が可能であることを示した．またジェスチャ区間の抽出アルゴリズムを使って，個人間におけるジェスチャの総量の比較分析，セッション間における二人の説明スタイルの違いに関する分析を行えることを示した．

この分析結果を踏まえ，機械学習・データマイニング技術がコミュニケーションにおけるジェスチャの自動抽出・半自動処理を可能にし，ジェスチャ分析に有用であることを示した．

謝　辞

本研究は，国立情報学研究所公募共同研究費，JSPS科研費22700146，JST戦略的創造研究推進事業さきがけ「多人数インタラクション理解のための会話分析手法の開発」，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」の助成を受けたものです．ここに記して感謝いたします．

注

1) http://www.anvil-software.de/
2) http://www.lat-mpi.eu/tools/elan/
3) http://www.nii.ac.jp/cscenter/idr/interaction/interaction.html
4) 本研究ではホームポジションとホールド状態を明示的に分類することを行わなかった．各ジェスチャの詳細な分析を行う場合にはホームポジションとホールド状態を分離したほうが良い場合もあるだろう．その場合は，手のホームポジションまたはホールド状態を別のHMMとして定義・学習し，無動作と判定された区間からホームポジションとホールド状態を分離する必要がある．技術的にはマーカの三次元位置座標の時間差分だけでなく，座標位置そのものをHMMへの入力として学習することで，空中で手が静止するようなホールド状態を抽出することが可能となる．

【参考文献】

Bishop, Christopher M. (2006). *Pattern recognition and machine learning*. New York: Springer-Verlag. (元田浩他訳(2008).パターン認識と機械学習—ベイズ理論による統計的予測—（上・下巻）シュプリンガージャパン)

坊農真弓　(2008).　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

坊農真弓・角康之・高梨克也・岡田将吾・菊地浩平・東山英治　(2011).　多人数・マルチモーダルインタラクション研究のためのプラットフォーム構築　情報処理学会研究報告，Vol. 2011-HCI-145, No. 6, 1-6.

Bono, Mayumi, Suzuki, Noriko, & Katagiri, Yasuhiro (2003). An analysis of participation structure in

conversation based on Interaction Corpus of ubiquitous sensor data, In Matthias Rauterberg, Marino Menozzi, & Janet Wesson (Eds.), *INTERACT 03: Proceedings of the Ninth IEIP TC13 International Conference on Human-Computer Interaction*. pp. 713–716. Zurich, Switzerland.

坊農真弓・高梨克也（編）(2009). 多人数インタラクションの分析手法（人工知能学会編集「知の科学」シリーズ） オーム社

Bull, Peter E. (1987). *Posture and gesture*. Pergamon Press.（市河淳章・高橋超編訳，飯塚雄一・大坊郁夫訳 (2001). 姿勢としぐさの心理学 北大路書房）

Carletta, Jean, Ashby, Simone, Bourban, Sebastien, Flynn, Mike, Guillemot, Mael, Hain, Thomas, Kadlec, Jaroslav, Karaiskos, Vasilis, Kraaij, Wessel, Kronen thal, Melissa, Lathoud, Guillaume, Lincoln, Mike, Lisowska, Agnes, Mc-Cowan, Iain, Post, Wilfried, Reidsma, Dennis, & Wellner, Pierre (2005). The AMI meeting corpus: A pre-announcement. *Second International Workshop on Machine Learning for Multimodal Interaction (MLMI2005), Vol. 3869 of Lecture Notes in Computer Science*. pp. 28–39. New York: Springer.

Chen, Lei, & Harper, Mary P. (2009). Multimodal floor control shift detection. *Proceedings of the 2009 International Conference on Multimodal Interfaes (ICMI-MLMI'09)*. 15–22. ACM, New York.

Chen, Lei, Rose, Travis R., Qiao, Ying, Kimbara, Irene, Parrill, Fey, Welji, Haleema, Han, Tony, Tu, Jilin, Huang, Zhongqiang, Harper, Mary, Quek, Franis, Xiong, Yingen, McNeill, David, Tuttle, Ronald, & Huang, Thomas (2006). VACE multimodal meeting corpus. *Second International Workshop on Machine Learning for Multimodal Interaction (MLMI2005), Vol. 3869 of Lecture Notes in Computer Science*. pp. 40–51. New York: Springer.

Chiu, Bill, Keogh, Eamonn, & Lonardi, Stefano (2003). Probabilistic discovery of time series motifs. *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*. 493–498. ACM, New York.

Clark, Herbert H. (1996). *Using language*. Cambridge: Cambridge University Press.

大坊郁夫（編） (2005). 社会的スキル向上を目指す対人コミュニケーション ナカニシヤ出版

伝康晴 (2006). 談話データの定量的分析—タグの設計と集計— 伝康晴・田中ゆかり（編） 講座社会言語科学6 方法 pp. 208–228. ひつじ書房

榎本美香 (2009). 日本語における聞き手の話者移行適格場の認知メカニズム ひつじ書房

Germesin, Sebastian & Wilson, Theresa (2009). Agreement detection in multiparty conversation. *Proceedings of the 2009 international conference on Multimodal interfaces (ICMI-MLMI'09)*. 7–14. ACM, New York.

Goodwin, Charles (1981). *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press.

Heath, Christian (1986). *Body Movement and Speech in Medical Interaction*. Cambridge University Press.

細馬宏通 (2009). ジェスチャー単位 坊農真弓・高梨克也（編）多人数インタラクションの分析手法 オーム社 pp. 119–136.

細馬宏通・片岡邦好・村井潤一郎・岡田みさを (2011). 特集「相互作用のマルチモーダル分析」 社会言語科学, **14**(1), 1–4.

城綾実・細馬宏通 (2009). 多人数会話における自発的ジェスチャーの同期 認知科学, **16**(1), 103–119.

Kendon, Adam (1990). *Conducting interaction: patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

來嶋宏幸・坊農真弓・角康之・西田豊明 (2007). マルチモーダルインタラクション分析のためのコーパス環境構築 情報処理学会研究報告（ヒューマンコンピュータインタラクション）(99), 63–70.

串田秀也 (2006). 相互行為秩序と会話分析—「話し手」と「共-成員性」をめぐる参加の組織化— 世界思想社

Lawrence, Rabiner, R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 257–286.

Lerner, Gena H. (2002). Turn-sharing: The choral co-production of talk-in-interaction. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The language of turn and sequence*. pp. 225–256. Oxford: Oxford University Press.

Lucey, Patrick, Potamianos, Gerasimos, & Sridharan, Sridha (2007). A unified approach to multi-pose audio-visual ASR. *Proceedings of INTERSPEECH. 2007*, 650–653.

McNeill, David (1992). *Hand and mind*. Chicago: The University of Chiacgo Press.

McNeill, David (2005). *Gesture and thought*. Chicago: The University of Chicago Press

元田浩・津本周作・山口高平・沼尾正行 (2006). データマイニングの基礎 オーム社

中田篤志・角康之・西田豊明 (2011). 非言語行動の出現パターンによる会話構造抽出 電子情報通信学会論文誌, **J94-D** (No. 1), 113–123.

Nishida, Toyoaki (2007). Conversational informatics and human-centered web intelligence. *IEEE Intelligent Informatics Bulletin*, **8**(1), 19–28.

西阪仰 (2008). 分散する身体—エスノメソドロジー的相互行為分析の展開— 勁草書房

Özyürek, Asli (2002). Do speakers design their cospeech gestures for their addressees: The effects of addressee location on representational gestures. *Jounal of Memory and Language*, **46**, 688–704.

Okada, Shogo, Ishibashi, Satoshi, & Nishida, Toyoaki (2010). On-line unsupervised segmentation for multidimensional time-series data and application to spatiotemporal gesture data, *International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE2010)*. 337–347.

岡田将吾・西田豊明　(2010).　自己増殖型ニューラルネットワークを用いた時系列データの追加学習型クラスタリング　日本神経回路学会論文誌, **17**(4), 174–186.

Patterson, Milies L. (1983). *Nonverbal behavior: A functional perspective*. New York: Springer Verlag. (工藤力（監訳）(1995).　非言語コミュニケーションの基礎理論　誠信書房)

Sacks, Harvey, Schegloff, Emanuel A., & Jefferson, Gail (1974). A simplest systematics for organization of turn-taking for conversation. Language, **50**(4), 696–735. (西阪仰訳　(2010).　会話のための順番交替の組織－最も単純な体系的記述－会話分析基本論集：順番交替と修復の組織　世界思想社　pp. 5–153.)

Scheglof, Emanuel A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Vol 1. Cambridge: Cambridge University Press.

Streech, Juergen, Goodwin, Charles, & LeBaron, Curtis (Eds.) (2011). *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press.

角康之・矢野正治・西田豊明　(2011).　マルチモーダルデータに基づいた多人数会話の構造理解　社会言語科学, **14**(1), 82–96.

高梨克也・伝康晴・榎本美香・森本郁代・坊農真弓・細馬宏通・串田秀也　(2004).　ワークショップ：多人数会話における話者交替再考—参与構造とノンバーバル情報を中心に—　社会言語科学会第13回大会発表論文集, 144–153.

Waibel, Alexander & Stiefelhagen, Rainer (Eds.) (2009). *Computers in the human interaction loop*. NY: Springer.

(2011年12月9日受付)
(2012年7月21日修正版受付)
(2012年8月11日掲載決定)