

## 人物対話行動における応答時間分布抽出

木村 綾平<sup>†</sup> 武村 紀子<sup>†</sup> 岩井 儀雄<sup>††</sup> 佐藤 宏介<sup>†</sup>

<sup>†</sup> 大阪大学 基礎工学研究科 〒 560-8531 大阪府豊中市待兼山町 1-3

<sup>††</sup> 鳥取大学 知能情報工学科 〒 680-8552 鳥取県鳥取市湖山町南 4-101

E-mail: †{kimura,takemura}@sens.sys.es.osaka-u.ac.jp, ††iwai@ike.tottori-u.ac.jp,

†††sato@sys.es.osaka-u.ac.jp

あらまし 近年、様々なセンサ・認識技術の発展に伴い、人との相互対話を通して活動支援を行うシステムに関する研究が活発に行われている。システムが自然な対話行動を行うためには、時間関係を含めた対話構造を明らかにする必要がある。本研究では、人物の姿勢やジェスチャ、音声といったマルチモーダル情報を用いて人物対話行動を構造化することを目指す。複数人物の対話行動から身体各部位の三次元位置データ及び音声データを取得し、対話行動事例データベースを構築する。事例データベースからは三次元位置データのクラスタリング及び発話区間検出により、行動の基本単位である行動素の抽出を行う。さらに、各行動素間の観測時間差を応答時間とし、その分布を抽出することを本研究の目的とする。行動素と応答時間を抽出することが出来れば、それらを用いて対話行動構造化記述を得ることが出来、対話行動の理解に繋がる。本稿では、人物対話行動構造化のための提案モデルを示し、対話行動構造化に用いる応答時間分布の抽出方法について述べ、応答時間分布抽出実験について記す。

キーワード 対話構造化, 行動素, 時間構造

### 1. はじめに

#### 1.1 背景・目的

近年、各種センサや映像・音声認識技術の進歩を背景として、人との相互対話を通して活動支援を行うシステムに関する研究が活発に行われている。たとえば、画像認識と音声認識を用いて、ユーザの指示した物体の運搬を行うロボットが開発されている [1]。ほかにも、ユーザが観光地候補と、その重要度や時間的制約を入力することで、観光スケジュールを算出するシステムや [2]、指ししや手の振りなど人の身体動作を検出して移動するロボットなど [3]、幅広く研究が行われている。しかし、これらのシステムの多くは、ユーザ側の要求が既に明確に存在し、システムはそれに基づいて必要な情報を提供したり、タスクを実行するという受動的なものとなっている。ここで、人と人との対話行動を振り返ってみると、人は対話相手に自らの要求を一方向的に伝えるのみではなく、対話相手からの提案や条件を加味することでも新たな情報を得ている。つまり、より自然な相互対話を行うシステムに求められるものは、ユーザの明確な要求には適切に応答し、それが無いときにはユーザの意図や興味を推定して積極的に情報を提示する能力である。そのため、システムが人の動作や音声を認識するのみではなく、対話がどのように行われているかを理解することが必要である。そこで本研究では、人物対話行動をシステムが自動的に認識することが出来る形で構造化することを目指す。

音声データから得られる様々な特徴量を用いて、人の

意図や感情を推定するための研究がなされている。桂ら [4] は基本周波数  $F_0$  のパターンによって怒りや喜びといった感情を表している。また、藤原ら [5] はタスク志向型の音声対話において、対話構造や対話意図と、発話タイミング、基本周波数  $F_0$ 、相対発話速度といった対話リズムとの関係の分析を行い、対話リズムが発話内容やその重要度、対話構造の状態、対話相手が予想している発話内容との差など、対話者の内面的要素から強い影響を受けていることを明らかにしている。さらに、西村ら [6] は個々人の音声特徴量と意図や感情との関係性のみでなく、対話の盛り上がりや互いの親密さによって、対話者の韻律変化に同調関係の相互作用があることを示している。また、映像データに対しては、米澤ら [7] は画像処理により推定したユーザの視線方向からユーザの興味・関心の対象を検出し、さらにぬいぐるみロボットに同様の注視行動を行わせることで、ぬいぐるみの視線方向へユーザの興味を誘引・促進できることを示している。また、平山ら [8] は顔器官の動きのタイミングや時間長をモデル化することで、自発的な笑いと意識的な笑いの分離を行っている。

一方で、これらの研究は音声、映像といった個別の研究に対して行われている。しかし、人と人との対話行動においては、人は対話者に意図を伝える際、種々の情報を総合的に利用している。たとえば、身振りやうなずきは発話権の受け渡しや対話リズムの調整を行っているし、発話内容が聞き取りづらいときは発話者の唇の動きからその内容を理解しようとする。そこで本研究では、人物対話行動の理解のために視覚・聴覚情報を統合的に

利用する．複数人物の対話行動から，人物の姿勢やジェスチャ，音声といったマルチモーダルデータを取得し，人物対話行動事例データベースを作成する．事例データベースから，対話行動を記述するのに必要な情報を自動的に抽出し，対話行動構造化記述を得ることを本研究の目的とする．

## 1.2 関連研究

川嶋ら [9] は円滑な会話を進める上で，視覚的に観察可能な身体動作により発話権の受け渡しが行われていることについて，落語における頭部動作の開始と発話開始・終了の時間的構造の解析を行っている．そして，動作による発話権取得タイミングの解析や，役柄交代開始までの区間長の抽出を行っているが，それらの状態の遷移などを構造化するには至っていない．

本研究と同様に，各種センサから得られたマルチモーダルデータを用いて対話行動構造化を行ったものとして，福岡ら [10] の研究が挙げられる．この研究では，ポスター会話における三者間対話行動を発話情報，視線方向，指差し方向の3つのアノテーションによって記述している．対話状態遷移は，一時刻前の状態にのみ影響を受けると仮定し，N-gram モデルによってツリー構造化することで人物対話行動を記述している．しかし，アノテーションのように人によって解釈可能なものの生成には，必ず人手による情報の取捨選択が必要であり，コンピュータによって完全に自動化することはできない．また，N-gram モデルによって対話状態遷移を表すことで，対話行動において重要な働きを持つ対話者間の時間関係が，状態の前後関係のみにまで縮退してしまっており，その働きを十分に表せているとはいえない．これに対し，本研究では対話行動を人物行動の基本単位である行動素を用いて表す．行動素は，人物行動をさらに小さく分割することで得られ，人手による意味付けの必要がない．そのため，身体各部位の三次元位置データや音声データといったセンサ情報から自動的に抽出することが出来る．また，対話者間の時間関係は，観測された行動素間の時間差を応答時間として抽出し，これを対話状態の遷移確率の学習に用いることで表現する．これによって単にタスクとして与えられた対話行動を構造化するのではなく，対話者の意図や興味といった暗示的な情報を構造化に組み込むことが出来る．

## 1.3 本論文の構成

本論文では，人物対話行動構造化の概要と，対話行動構造化に用いる応答時間分布を抽出する手法について記す．本論文の構成は以下のようになっている．第2章では，対話行動構造化モデルについて述べる．第3章では，応答時間分布抽出手法の概要と各処理の詳細について述べ，第4章では，応答時間分布抽出実験について述べる．最後に，第5章で結論と今後の課題についてまとめる．

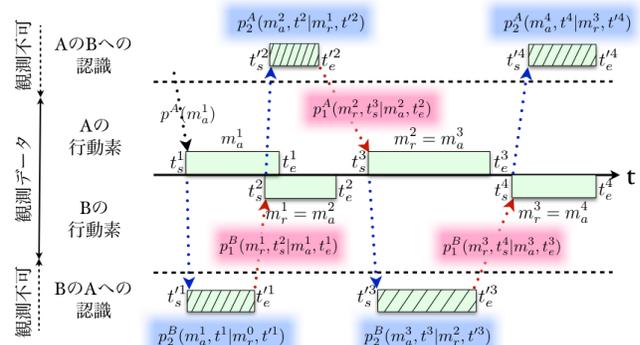


図1 対話行動構造化モデル

## 2. 対話行動認識モデル

本章では，本研究で提案する人物対話行動認識モデルについて説明する．

### 2.1 対話行動構造化

本研究で用いる人物対話行動構造化モデルを図1に示す．人物対話行動における対話者をA, Bとする．本研究では対話行動を記述する要素  $m$  として行動素を用いる．行動素とはマルチモーダルデータから自動的に抽出される行動の最小単位である．行動素の抽出方法の詳細は第3章に記す．行動素が観測された順に  $m_1, m_2, \dots, m_N$  とし， $m_i$  の開始・終了時刻を  $t_s^i, t_e^i$  とする．行動素を用いて，対話行動を (Aの先行行動素 Bの認識 Bの後続行動素 Aの認識) といった基本対話行動  $I_S$  の遷移として扱う． $i$  番目の基本対話行動  $I_S^i$  の後続行動素  $m_r^i$  は， $i+1$  番目の基本対話行動  $I_S^{i+1}$  の先行行動素  $m_a^{i+1}$  となる．これらを用いて人物対話行動を3つの確率  $p_0, p_1, p_2$  の連鎖モデルとして記述する．以下説明のため，基本対話行動  $I_S^i$  において，Aによる先行行動素  $m_a^i$  とBによる後続行動素  $m_r^i$  が観測されるとする．

- 初期行動素確率  $p_0$

A, B が対話状態にないときに，Aによる行動素  $m_a^1$  が観測されて対話行動が開始する確率であり，以下のように記述される．

$$p_0^A(m_a^1) \quad (1)$$

- 行動素遷移確率  $p_1$

基本対話行動  $I_S^i$  において，時刻  $t_e^i$  にAによる先行行動素  $m_a^i$  の観測が終了したときに，ある時刻  $t_s^{i+1}$  においてBによる後続行動素  $m_r^i$  の観測が開始される確率を．以下の条件付き確率で表す．

$$p_1^B(m_r^i, t_s^{i+1} | m_a^i, t_e^i) \quad (2)$$

これによって，対話行動における対話者間の時間関係を，

先行行動素と後続行動素の観測時間の差と置き換えて表す．この時間を応答時間とする．応答時間抽出の詳細は第3章で述べる．行動素遷移確率  $p_1$  は、各行動素間の時間的相関を示しており、センサ情報から自動的に抽出することが出来る．

- 対話継続確率  $p_2$

基本対話行動  $I_S^i$  が基本対話行動  $I_S^{i+1}$  に遷移する際、状態  $I_S^i$  での B による後続行動素  $m_r^i$  が新たな状態  $I_S^{i+1}$  での先行行動素  $m_a^{i+1}$  となるが、この遷移は常に起こる訳ではない．A が行動素  $m_r^i$  を受けたときに、 $m_r^i$  を自分が反応を返す必要がある行動素、つまり、 $m_r^i = m_a^{i+1}$  であると感じたときにのみ基本対話行動は遷移する．この確率を、A の行動素  $m_a^{i+1}$  に対する認識時間  $t^i$  を用いて以下の条件付き確率で表す．

$$p_2^A(m_a^{i+1}, t^{i+1} | m_r^i, t^{i+1}) \quad (3)$$

A の認識と B の行動素のずれが大きいつき、B が対話行動の継続を望んでいても対話行動は継続しない．この確率は、A, B 互いの心理状況や意図、目的、社会的立場、周囲の状況など直接的に観測出来ない要素からの影響を受けている．

これら行動素遷移・対話継続確率によって、観測される対話行動は以下の確率連鎖モデルに変換される．

$$p^A(m_a^1) p_2^B(m_a^1, t^1 | m_r^0, t^1) p_1^B(m_r^1, t_s^2 | m_a^1, t_e^1) \\ \times p_2^A(m_a^2, t^2 | m_r^1, t^2) \cdots p_2^B(m_r^N, t^N | m_a^{N-1}, t^N)$$

以下、本稿では提案モデルのうち行動素遷移確率  $p_1$  を抽出することを目的とする．その際、タスク試行対話を対話行動事例データとして用いることで、対話継続確率  $p_2$  はタスク終了時を除いて常に 1 に近似出来る．

## 2.2 認識

対話行動毎に作成した対話行動事例データベース内の学習データを全て式 (2.4) で表される状態に変換する．そして、対話行動  $IM$  に関する学習データを用いて以下のような  $\{t^i\}$  を抽出する．

$$D^{IM} = \arg \max_{\{t^i\}} \prod_{i=1}^{N/2} p_2^B(m_a^{2i-1}, t^{2i-1} | m_r^{2(i-1)}, t^{2(i-1)}) \\ \times p_1^B(m_r^{2i-1}, t_s^{2i} | m_a^{2i-1}, t_e^{2i-1}) \\ \times p_2^A(m_a^{2i}, t^{2i} | m_r^{2i-1}, t^{2i}) \\ \times p_1^A(m_r^{2i}, t_s^{2i+1} | m_a^{2i}, t_e^{2i})$$

これがテストデータとの比較の対象となる、対話行動  $IM$  を表す認識モデルである．入力として与えられた人物対話行動を、学習データと同様に行動素・応答時間抽出によって (2.4) のような構造化記述への変換を行う．

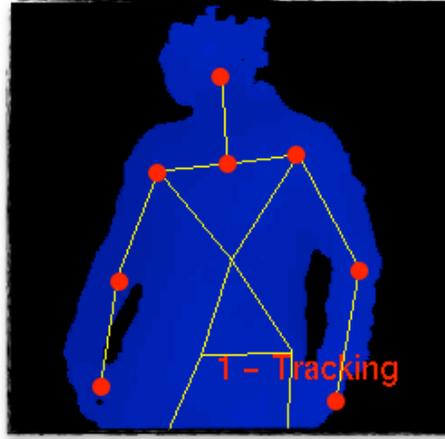


図 2 三次元位置データ取得部位

$p_1^A, p_2^A, p_1^B, p_2^B$  の分布形状はそれぞれの対話行動モデルにより異なるため、

$$\arg \max_{IM} D^{IM}$$

となるような対話モデル  $IM$  を選択することにより、対話行動を認識出来る．

## 3. 応答時間分布抽出

本章では、本研究で対話行動遷移確率の学習に用いる行動間の応答時間分布抽出について述べる．

### 3.1 学習プロセス概要

学習用データである対話行動事例を各種センサを用いて身体各部位の三次元位置データとして取得する．行動認識部で行動素遷移を検出した結果を用いて、対話行動における先行行動とそれに対応する後続行動間の応答時間を抽出する．得られた応答時間は、各認識部での結果と合わせて対話行動事例データベースとして保存し、未知データの対話行動遷移確率を学習するために用いる．

### 3.2 姿勢推定

対話参加者の身体各部位の三次元位置データを、距離画像センサと可視光カメラを内蔵する Microsoft 社 Kinect を利用して取得する．そして、得られた三次元位置データについて前処理、姿勢データのクラス分類を行うことで行動素抽出を行う．

#### 前処理

Microsoft 社 Kinect を用いて、毎秒 30 フレームで三次元座標位置  $x = (x, y, z)^t$  を計測する．取得する身体部位は図 2 に示すように、頭、胸、両肩、両肘、両手の 8 カ所である．取得した三次元位置データに補正処理を行い、すべての身体動作、人物において同条件で姿勢データのクラス分類が出来る状態へと変換する．処理内容として、欠損データの補間、座標系の変換を行う．

## 欠損データの補間

各部位について，Kinect で三次元位置座標が取得出来なかった時刻に対して，前後の時刻から線形補間することにより位置を推定する．線形補間を用いるのは，Kinect によって座標が取得出来なかった時刻が比較的短かったことと，身体動作における各部位の座標点は，短い時間区間においては線形に推移すると仮定したためである．

## 座標系の変換

Kinect によって得られた三次元位置データは，実験室内の一点を原点とする座標系によって定義されており，取得された情報には，Kinect に対して部屋の奥側にいるであるとか，部屋の手前側にいるといった情報を含んでいる．しかし，室内のどの場所においても同一の姿勢は同じ行動として認識される必要がある．つまり，行動素は人物の位置に依存せずに抽出されることが求められるため，座標系の変換処理を行う必要がある．そこで，座標原点を人物の胸の位置へと変換する．これにより，変換後の座標系では，人物の胸の位置を中心とした相対運動を観測することができ，人物の位置によらずに行動素が抽出できる状態に変換される．

## 姿勢データのクラス分類

これまでの処理によって得られた対話行動の三次元データから行動素を抽出する．まず，三次元データから図3で示すような上半身の関節角度  $\Theta = \{\theta_i | i = 1, 2, \dots, 8\}$  を姿勢特徴量として抽出する．関節角度を用いる利点として，

- 使用する特徴量の数が少ない
- 特徴量同士の相関が小さい
- 個人差を吸収できる

などが挙げられる．得られた姿勢特徴量をクラスタリングし，生成されたクラスタを行動素とする．クラスタリングのアルゴリズムとして，*k*-means クラスタリングを利用している．このとき，行動素の分類結果数 *M* はクラスタ数 *k* と一致し，同一クラスタに属する姿勢データは，同一の行動素として分類される．また，適切なクラスタ数 *k* については，実験を通して検討する．

以上の処理によって，学習データ毎に対話参加者それぞれについて，関節角度を用いた姿勢遷移の時系列データを得る．

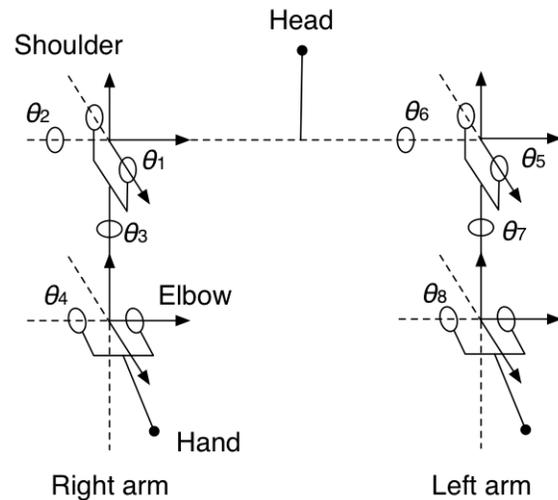


図3 上半身の姿勢パラメータ

## 3.3 応答時間分布抽出

姿勢推定の結果として，行動素の遷移データが得られる．対話参加者の全ての行動素遷移データを統合することで，対話行動遷移データを得る．これを用いて第2章で提案した対話行動構造化モデルの行動素遷移確率  $p_1$  を，各行動素のペア毎に，先行行動・後続行動間の応答時間を取得する．ここで，抽出する応答時間は，先行行動終了時刻と後続行動開始時刻の差である．応答時間抽出を多くの学習データに対して行い，応答時間確率分布の形を推測するためにヒストグラムを作成する．応答時間確率分布の形状が推測出来れば，学習を繰り返すことで確率分布のパラメータの推定精度を高めることができる．

## 4. 応答時間分布抽出実験

### 4.1 実験設定

人物対話行動構造化に用いる応答時間分布抽出実験について述べる．実験では結婚式場での受付対話を想定し，一連の対話行動における応答時間の抽出を試みた．2名の被験者によって26回分の対話行動を Kinect, B 環境カメラを用いて収録し，得られた身体各部位の三次元位置データを基に，対話行動事例データベースを作成した．一連の対話行動はおよそ1分程度の長さであり，下記の大まかな流れを除いては，被験者に自由に対話を行わせた．図4に実験環境と，以下に対話行動の一例を示す．

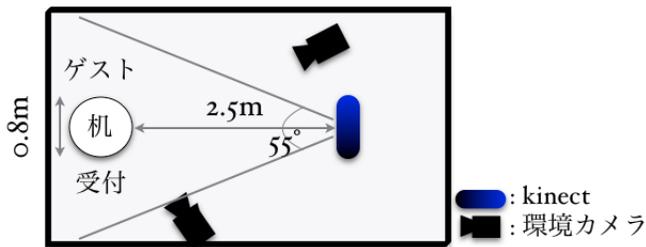


図4 実験環境図面

・ゲスト来場

受付係はゲストへ向かって一礼し、互いに挨拶を交わす。



・ご祝儀の手渡し

ゲストはお祝いの言葉を述べつつ、受付係へご祝儀袋を手渡す。受付係はお礼を述べつつ両手で受け取り、お盆の上へ。



・記帳のお願い

受付係は、ゲストへ芳名帳への記帳のお願いをする。ゲストは芳名帳へ、住所と名前を記入する。



・ゲスト退室

受付係は、ゲストへ席次表などを手渡し、控え室の案内をする。ゲストは配布物を受け取り、控え室へ向かう。受付係は一礼して見送る。



4.2 実験結果および考察

作成した事例データベースを用いて、身体動作に関する行動素と、各行動素間の応答時間分布の抽出を行った。行動素数を6, 8, 10として応答時間分布を抽出した結果を図6, 図7, 図8に示す。各グラフの見方について、行動素数8, 先行行動素  $m^3$ , 後続行動素  $m^7$  のグラフ(図5)を基に説明する。このグラフでは、横軸に先行行

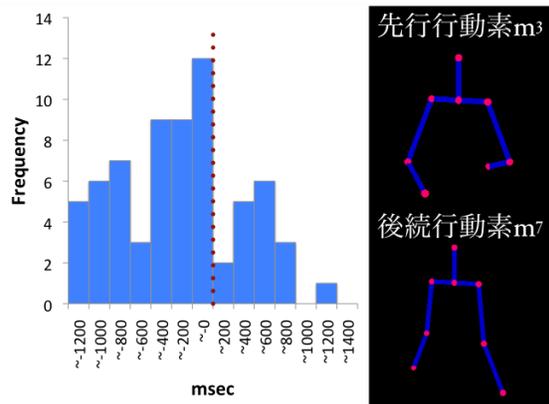


図5 応答時間ヒストグラム(行動素数8, 先行行動素  $m^3$ , 後続行動素  $m^7$ )

動素が終了してからの時刻を200msec毎に示しており、縦軸に各時刻間で後続行動素が開始した回数を記し、ヒストグラム化している。マイナスの時刻は、動作のオーバーラップを示している。このようなヒストグラムを各行動素ペア毎に作成し、まとめて表したものが図6, 図7, 図8である。

行動素数についての検討

3つのグラフから分かるように、行動素の数を増やすにつれて同一の行動素に含まれていた姿勢が分離され、それぞれ別の行動素に属するものとして扱われるようになる。あまりに行動素の数が少ないと、十分な姿勢近似を行うことが出来ていない。たとえば、行動素数8のときの行動素  $m^6$ (あるいは行動素数10のときの行動素  $m^2$ ) のような姿勢は、人の目から見ると特徴的な姿勢であるが、観測回数が少ないこともあり、行動素数6のときには抽出することが出来ていない。逆に、行動素の数が多きほど動作が細かく分類され姿勢近似精度は向上するが、行動素の遷移パターンが複雑になるために、行動素数8のときの  $m^4, m^5$ (あるいは行動素数10のときの行動素  $m^5, m^7, m^8$ ) のように同じような行動素や、行動素数8のときの  $m^2$ (あるいは行動素数10のときの行動素  $m^9$ ) のように出現頻度の少ない行動素など、重要でないと考えられるものが抽出される可能性が高まる。そのため、対話行動を表す本質的な行動素の抽出回数が少なくなってしまう、行動素遷移確率の学習能力が低下してしまうと考えられる。特に本実験のように、一般的な受付対話行動を、被験者に自由な裁量を与えて行ってもらった場合、対話行動における人物姿勢のバリエーションが少なくなってしまうことも考えられる。そのため、行動素数は構築する対話行動構造化モデルに合わせて、無駄な行動素を省いて応答時間分布の抽出精度を高めつつ、各姿勢を近似するのに十分な数を保つように設定する必要があると考えられる。

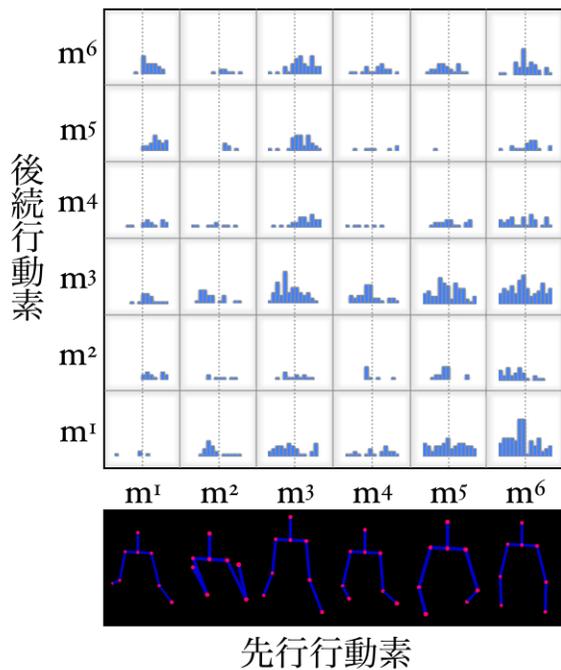


図 6 行動素数 6

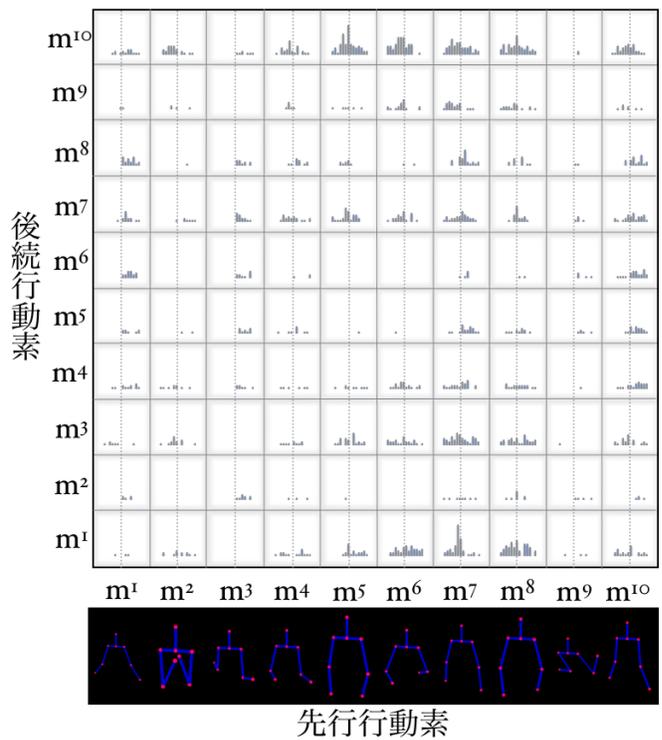


図 8 行動素数 10

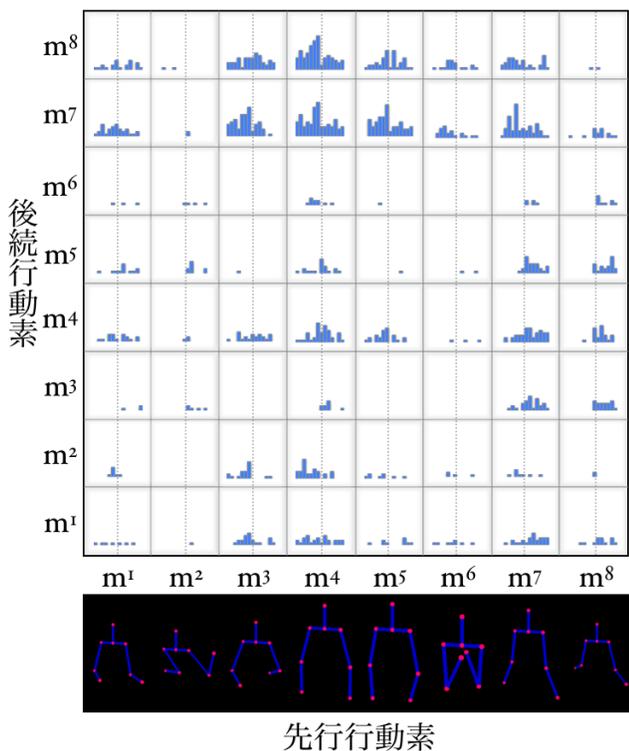


図 7 行動素数 8

#### 応答時間分布についての検討

まず応答時間分抽出の際、各行動素間の応答時間が1200msecより小さかったもの、あるいは1400msecより大きかったものは、対話構造において、行動素間の相関が小さいと考え、グラフから省略した、行動素数を6とした図6のグラフについて考察する。応答時間分布の形状は大別すると、

- (i) 先行対話終了直後～1000msec間にピークを持つもの ( $(m^3, m^6)$  など)
- (ii) 先行対話終了前 -600msec～終了直後にピークを持つもの ( $(m^5, m^3)$ ,  $(m^6, m^1)$  など)
- (iii) 明らかなピークが見られず、全体に広がりを持っているもの ( $(m^5, m^1)$  など)
- (iv) 行動素間にほぼ相関が見られない ( $(m^2, m^4)$ ,  $(m^4, m^5)$  など)

に分けられる。(i),(ii)に関しては、先行対話に対しその終了直後、あるいはオーバーラップ気味に後続行動の開始頻度が高まるという知見は、川嶋ら[9]による会話における発話移行区間の抽出による知見と同様のものあり、人物対話行動においても一定のピークを持った応答時間分布モデルが存在すると思われる。(iii)に関しては、本実験においては特定のピークをもつ分布として扱うことは出来ないが、学習データの不足や、行動素・応答時間抽出精度に問題があるのか、行動素間に時間的相関関係がないのかは現時点では判断出来ない。(iv)に関しては行動素間の相関が非常に小さく、このような対話遷移をする可能性は極めて低いと考えられる。

## 実験方法に関して改善出来る点

実験の際，kinect の配置を被験者のほぼ真横に設置したために，受付は左半身の，ゲストは右半身の三次元座標値を取得出来ない時刻が多く存在した．座標値が取得出来なかった時刻に対しては，取得出来た逆半身の座標値から見て，胸の位置に対して対称の位置に存在すると考え補完処理を行ったが，そのために片手のみで行われた動作を行動素として抽出することが難しくなりました．これは各被験者にそれぞれ一つの Kinect を割当て，それぞれの正面からみた身体座標値を計測するなどして改善出来ると考えられる．また，本実験では上半身の関節角度のみを姿勢パラメータとして用いたが，下半身を含む全身の関節角度を用いたり，身体各部位の速度を補助的に用いることでも行動素抽出精度が向上すると考えられる．

## 5. 結 論

本研究では行動素と応答時間分布を用いた対話行動の構造化モデルを提案した．実験により，複数人物の対話行動における身体各部位の三次元位置データと音声データを取得し，受付対話行動事例データベースを作成した．作成した事例データベースを基に，三次元位置データから行動素の抽出を行った．また，提案モデルの行動素遷移確率の学習に用いるために，各行動素間の応答時間分布を抽出した．今後の課題として，提案モデルへの音声データの利用と，提案モデルで用いた対話者の意図や興味といった直接観測出来ない情報に影響を受ける対話継続確率の学習方法を考案し，それらを用いた人物対話行動認識システムの構築を目指す．

## 文 献

- [1] 榎原靖, 白井良明, 島田伸敬: “対話を用いた物体認識のための候補選択戦略,” 電子情報通信学会技術研究報告, vol.104, No.668, pp.49-54, 2005.
- [2] 栗山恭嘉, 花野博司, 澤悠太, 村田佳洋, 柴田直樹, 安本慶一, 伊藤実, 藤原礼征: “多数の観光候補地から効率良い観光スケジュールを自動的に作成・提案するシステム P-Tour の Google Maps を利用した設計と実装,” 情報処理学会研究報告, vol.2008, No.21, pp.261-266, 2008.
- [3] 深谷健一: “手ジェスチャによる移動ロボットの操作,” 北海学園大学工学部研究報告, vol.37, pp.59-68, 2010.
- [4] 桂聡哉, 広瀬啓吉, 峯松信明: “感情音声合成のための生成過程モデルに基づくコーパスベース韻律生成とその評価,” 電子情報通信学会技術研究報告, SP2002-184, pp.31-36, 2003.
- [5] 藤原敬記, 伊藤敏彦, 荒木健治: “タスク指向対話における相互の対話意図を考慮した対話リズムの分析,” 言語・音声理解と対話処理研究会, vol.50, pp.45-50, 2007.
- [6] 西村良太, 北岡教英, 中川聖一: “音声対話における韻律変化をもたらす要因分析,” 音声研究, vol.13, No.3, pp.66-84, 2009.
- [7] 米澤朋子, 山添大丈, 内海章, 安部伸治: “視線コミュニケーションのためのぬいぐるみの視線行動の設計と分析,” 電子情報通信学会論文誌, vol.92, No.1, pp.81-92, 2009.
- [8] 平山高嗣, 川嶋宏彰, 西山正紘, 松山隆司: “表情譜: 顔パーツ間のタイミング構造に基づく表情の記述,” ヒュー

- マンインタフェース学会論文誌, vol.9, pp.271-281, 2007.
- [9] 川嶋宏彰, 西川猛司, 松山隆司: “落語の役柄交替における視覚的「間合い」の解析,” 情報処理学会論文誌, vol.48, No.12, pp.3715-3728, 2007.
- [10] 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,” 電子情報通信学会技術研究報告, vol.109, pp.119-126, 2009.