

博士論文

文楽人形遣いの非言語情報を用いた“型”動作の伝達

Bunraku Puppeteer's Nonverbal Information

Transmission of “Kata” Motion

櫻 哲郎

目次

第1章 序論	1
1.1 はじめに	1
1.2 伝統芸能“文楽”とは	3
1.2.1 文楽の概要と歴史	3
1.2.2 文楽の舞台構造と三業	3
1.2.3 文楽人形の構造	5
1.2.4 文楽人形の“ず”による協調操作	7
1.3 文楽の持つ特殊性とその成立過程	9
1.4 文楽を対象とした先行研究と関連する先行研究	11
1.4.1 文楽の操作システムと先行研究	11
1.4.2 文楽とそれ以外の協調タスクの比較	14
1.4.2.1 協調歩行(Synchronization)	18
1.4.2.2 協調運搬	19
1.4.2.3 ジャズセッション	20
1.4.2.4 クラシック音楽	21
1.4.3 まとめ	22
1.5 本論文の目的と構成	22

第2章 左遣いの“ず”判別の仮説	24
2.1 はじめに	24
2.2 文楽における左遣いの役割	24
2.3 左遣いへの“ず”の伝達媒介	25
2.4 左遣いによる動作判別の仮説	28
2.4.1 表現上の制約	28
2.4.2 左遣いの型判別仮説	29
2.4.3 左遣いの型判別に影響を与える要因	31
2.4.3.1 仮説1：型判別の正確性は、熟達の影響を受ける.	31
2.4.3.2 仮説2：型判別の正確性は動作の観察時間と共に上昇する.	32
2.4.3.3 仮説3：型の種類によって判別に必要な時間が異なる.	32
2.5 まとめ.....	32
第3章 左遣いの型判別実験	33
3.1 目的と概要	33
3.2 左遣い型判別実験.....	33
3.2.1 本実験の目的と概要.....	33
3.2.2 実験計画	34
3.2.3 実験手続き	36
3.3 実験結果	38
3.3.1 熟達度，提示時間，型の種類の正解確信度への影響	38
3.3.2 提示時間ごとの確信度のクラスタ分析.....	40

3.3.3 各提示時間における確信度の主成分分析	42
3.4 考察	44
3.4.1 各要因の影響について（一般化線形混合モデルによる分析）	44
3.4.2 動作特徴の利用について（クラスタ分析）	44
3.4.3 重要視される特徴について（主成分分析）	45
3.5 まとめ.....	45
第4章 人形動作からの型判別情報の抽出	47
4.1 使用する隠れマルコフモデルとパラメータ設定	49
4.2 結果	54
4.2.1 学習と適切な状態数	54
4.2.2 抽出された状態遷移と人形動作の関係.....	54
4.2.3 型と状態遷移の関係と連長圧縮	58
4.2.4 抽出された状態遷移と人形動作の関係.....	60
4.3 考察とまとめ	63
第5章 総合考察.....	64
5.1 解明された左遣いの型判別の特徴	64
5.2 型判別の仕組みが持つ利点.....	65
5.3 型判別の仕組みに関する先行研究との比較	66
5.4 本論文の限界と文楽人形の協調操作の全体像.....	68
5.5 今後の展望	72
5.5.1 アドリブと“ず”の対応関係学習と適切な状態数	72
5.5.2 修行の必然性と技能継承への応用可能性	72

第 6 章 結論	74
謝辞	77
参考文献	78
付録 A 文楽人形の詳細.....	81
A.1 各パーツ詳細	81
A.1.1 かしら	81
A.1.2 左腕.....	82
A.1.3 右腕.....	83
A.1.4 胴.....	83
A.1.5 足.....	83
A.2 各パーツの操作法	87
A.2.1 かしらと胴の操作法	87
A.2.2 左腕の操作法.....	87
A.2.3 右腕の操作法.....	87
A.2.4 足の操作法（立役）	89
A.2.5 足の操作法（女方）	89
A.3 センサ配置詳細	92
A.4 文楽人形のモデル化.....	92
付録 B 実験データの補足	95
B.1 HMM による状態の推移と動作波形の時間変化	95

第 1 章 序論

1.1 はじめに

文楽とは、三味線による演奏、太夫による語り、人形による芝居の 3 つの要素で構成される日本の伝統芸能である（図 1-1）。芝居に用いられる人形は、1 体につき 3 人の人形遣いにより分割操作されるにもかかわらず、多彩かつ情緒深い動作が実現される（図 1-2）[1], [2]。この人形操作技術は、人形浄瑠璃が流行した江戸時代に、人形構造の複雑化と共に徐々に洗練されていった暗黙的な技であり、当事者以外には、連携のために用いられる“ず”と呼ばれる合図を認識することすら困難である。また、この協調操作の仕組みは明文化されておらず、人形遣い達は 10 年を超える修行によって習得すると言われている[3]。文楽人形操作は、他の人形劇はおろか、協調タスクとしてみても他に例のない特徴を備えた特殊な協調操作システムを有する。その詳細を 1.4 節で述べる通り、本協調操作システムは、マスター—スレーブ関係のはっきりした協調タスクに分類され、通常は打ち合わせや文脈情報もしくは発話による情報伝達が必要とされるタスク中の動作内容の変更や挿入を、マスターが分担操作する人形の動作自体を媒介とした合図で可能にしている。これは長年の修行で磨かれた技の判別技術と非言語情報（人形動作）のみではほぼ実現されていると考えられる。もし、その通りであれば、本システムでは、他の協調タスクでは観察が困難な、打ち合わせなどが無い時の協調している者の振る舞いを観測することができるため、主従関係のはっきりした協調タスクにおける判別プロセスのモデルケースとしての知見を得られる可能性がある。

本論文では、他に類をみないこの協調操作の仕組みを研究対象とする。本章では、まず 1.2 節で文楽の概要とその背景となる歴史的な成立過程について説明し、現在の人形の構造とその操作方法の概略について解説する。そして 1.3 節で、文楽人形の操作がなぜ世界に類をみない仕組みを生み出すにいたったかについて、文楽全体がもつ特殊性を

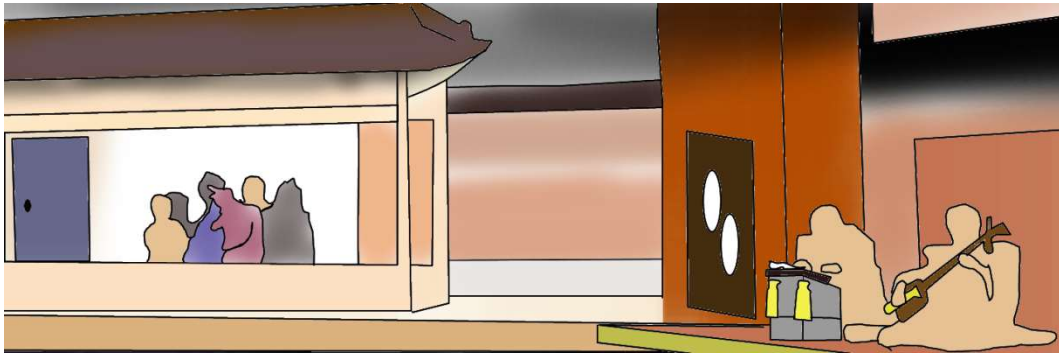


図 1-1 伝統芸能文楽外観 ([1]をもとに筆者が作成)

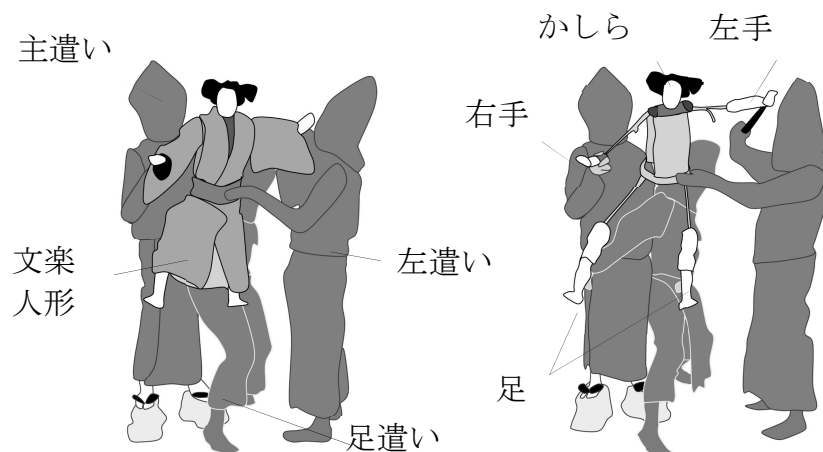


図 1-2 文楽人形の外観と操作分担 ([1]をもとに筆者が作成)

含めて説明する．そのうえで 1.4 節にて，文楽とその協調操作について現在までに明らかになっていることや関連する協調タスク研究を整理した上で，1.5 節で本研究の位置付けと目的を述べる．

1.2 伝統芸能“文楽”とは

本節では，伝統芸能文楽の概要とその舞台構成について述べる．

1.2.1 文楽の概要と歴史

文楽は，前述した通り，太夫・三味線・人形の3要素（三業と呼ばれる）によって構成される総合芸術である．その成り立ちは江戸時代初期にさかのぼり，古くはあやつり人形，後に人形浄瑠璃と呼ばれるようになった[4]．貞享元年（1684年）頃，竹本義太夫（後に筑後掾）が道頓堀に竹本座を開設して義太夫節を樹ててより後は，竹本義太夫による義太夫節と近松門左衛門の作品とが結び付くことで全盛期を迎えた．その後，様々な人形浄瑠璃座（浄瑠璃を興行する一座およびその芝居小屋）が盛衰を繰り返し，1872年に淡路仮屋の初世の植村文楽軒が「西の浜の高津新地の席」という小屋を大坂高津橋南詰で建てて文楽座を設立し，人形浄瑠璃を再興させた．その後は，再び衰退した時期もあったが，1963年に財団法人文楽協会がつくられ，現在では文楽が人形浄瑠璃の中心的な存在とみなされるようになった．また，文楽座が一定規模以上の人形浄瑠璃の公演を行う唯一の団体であるため，現在では「文楽」の名称が人形浄瑠璃と同義に用いられる場合もある．

（人形浄瑠璃文楽座の座員によって演じられる）文楽は1955年に文化財保護法に基づく重要無形文化財に指定された．先述した財団法人文楽協会は，国，大阪府，大阪市，日本放送協会の助成金により運営されており，国が保証する伝統芸能の形が整えられている[5]．2003年にユネスコにより「人類の口承及び無形遺産に関する傑作」として宣言され，2008年に「人類の無形文化遺産の代表的な一覧表」に記載された，日本を代表する伝統芸能の一つである．

1.2.2 文楽の舞台構造と三業

文楽の公演は現在主に大阪（国立文楽劇場）と東京（国立劇場の小劇場）の2つの舞台で行われている．舞台の構造は図1-3のようになっており，舞台中央に人形が劇をする空間がとられ，右手（上手）側に，三味線と太夫が現れる回転式の舞台（床）が据えられている[1]．

次に文楽を構成する三業の役割について説明する．太夫はナレーションと登場人物の

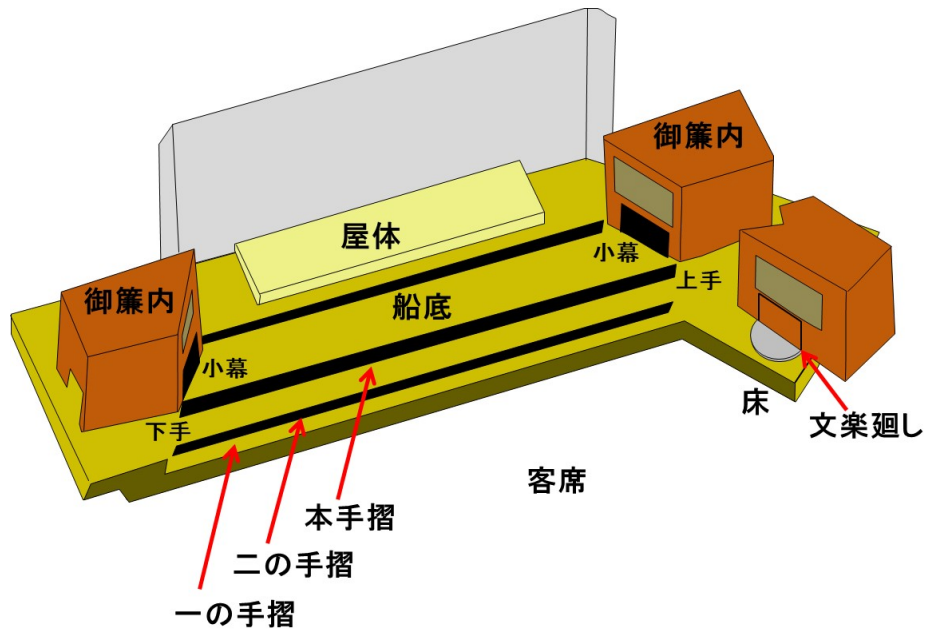


図 1-3 文楽の舞台外観 ([1]をもとに筆者が作成)

セリフを担当し、三味線は楽曲と効果音、そして人形遣いがそれらに合わせて芝居を行うという形態をとる．このうち、太夫と三味線は互いに対等な立場にあり、お互いが間をはかりながら、“床”と呼ばれる楽曲要素を作り上げていく．人形遣いは、作り出された床の場面描写に合わせつつ、情緒深い演技を行っていく．ただし、毎回全く同じことを行うわけではなく、三味線や太夫がその都度タイミングをアドリブ（即興的に）で調整したり、人形もそれに合わせて動作を挿入したり変化させたりする．これらのタイミングの関係性については、渋谷らによって、太夫の語りと三味線の撥音、そして呼吸の同期性の関係の可視化の研究が行われている[6]．

この“床”については、太夫は床本と呼ばれる台本に沿って進行し、三味線と太夫は事前に入念な打ち合わせや練習を行う．人間国宝である七代目竹本住太夫は、太夫を野球のピッチャーに、三味線引きをキャッチャーにたとえ、「一度バッテリーを組んだら、息が通じるまで何度でも稽古します」と述べている[7]．一方で人形芝居については、通し稽古は公演前の1日のみしか行わず、三味線と太夫のような密接な打ち合わせは行われない[3]．

さらに、3人の人形遣いの組み合わせは常に同じではなく、主遣いと手伝い（てったいと読む、ここでは左遣いと足遣いのこと）については、「主遣いにとってはどんな手

伝いが来ても自分の意思を伝えられなければならないし、若い人形遣い達はどの主遣いについても主遣いの指示通りの確に左や足が遣えなければならない」と言われている[3]。しかし、1.2.4 節で後述する高度な操作技術によって、あらゆる人形遣いの組み合わせであっても、即座にアドリブ（即興要素）を含む様々な動作を可能としていると考えられる。

1.2.3 文楽人形の構造

文楽では、その楽曲や動作がある程度伝統芸能としての型にはまってはいるものの、その中で可能な動作のアレンジや、アドリブによる動作の変更、挿入なども様々に行うことが可能である。そのため、文楽における人形操作は、あらかじめ決まった動作を決められたタイミングで繰り返すだけでなく、様々な動作やアレンジに臨機応変に対応できる高度な協調タスクだと言える。本節では、人形の構造と操作技術について説明する。

文楽人形は、頭、胴、腕、脚の各部位が分割されているという特有の構造を有する。文楽人形の大まかな構成と構造は図 1-2 と図 1-4 の通りである。部位は、“頭（かしら）”（頭部）、手、足、胴の 4 つに分かれており、手および足は胴体に紐によって繋がれている。足があるのは“立役（たちやく）”（男性の人形）だけであり、“女形（おんながた）”（女性の人形）に足はない。これらの部位を 3 人の人形遣いが分担し、操作する。3 人の人形遣いはそれぞれ、“主遣い（おもづかい）”、“左遣い（ひだりづかい）”、“足遣い（あしづかい）”と呼ばれており、主遣いが左遣いと足遣いに指示を出すことによって、人形を操作している。主遣いが人形の頭と右手、左遣いが人形の左手、足遣いが人形の足を、それぞれ操る。女形の場合には、足遣いは人形の着物の裏から手を入れ、自らの手を使って足を表現する。

人形の各部位は、紐やくじらのヒゲや猫の皮によるバネ要素で可動する仕組みになっているが、部位間の接続は図 1-4 に示すように緩く、着物の下には空洞部分が多い。見かけ上の手足の長さやプロポーションは人形遣いが意識しなければ簡単に崩れるので、操作時に注意が必要となる。したがって、人形のパーツは紐でつながってはいるが、これを強く引っ張るような操作は行うことができない。これら人形の構造の詳細は付録 A.1 に記載する。

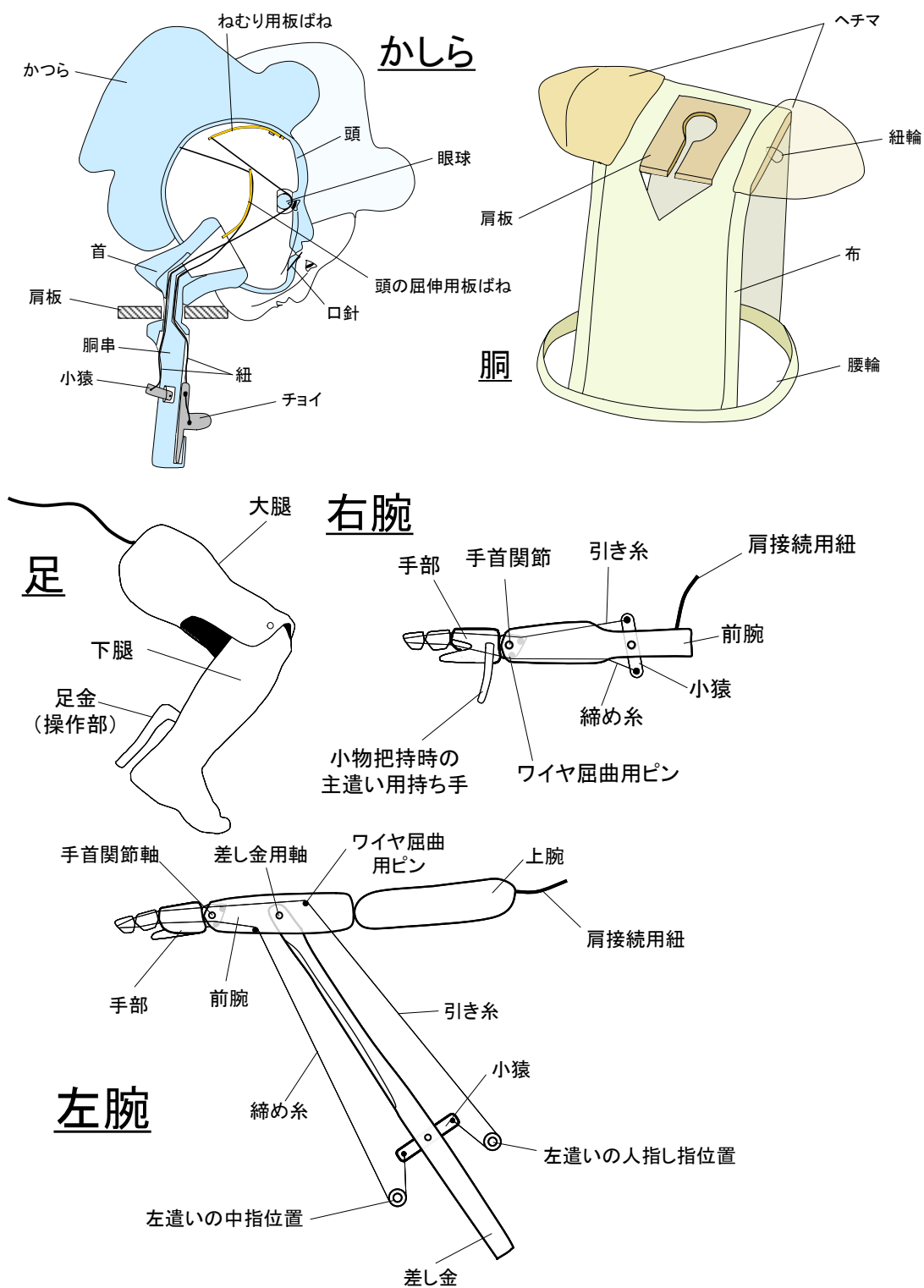


図 1-4 文楽人形の各部構造[8]

文楽人形は胴に縫い付けられた着物と、複数の剛体部位で構成される。文楽人形の構造の詳細は付録 A.1 を、操作方法は A.2 を参照。

1.2.4 文楽人形の“ず”による協調操作

1.2.3 節では人形の構造と各人形遣いが担当する部位を説明した。次に、文楽人形の協調操作方法の概要として、文献[3], [9]から言及されている内容を以下にまとめる。

- (a) 人形は様々な動作を自由に行える。その中でも頻繁に行われる日常動作（座る動作やお辞儀など）や象徴的な動作（ねじや後ろぶりなど）は“型”と呼ばれ、一連の動作手順がある程度定まっている。
- (b) 3人で操作部位を分担して操るが、全体の動作の意思決定を行なうのは、頭部と右手を操作する主遣いと呼ばれる人形遣いである。残りの2人（左手を操作する左遣いと脚部を操作する足遣い）は主遣いからの非明示的な合図“ず”を受け取り、“イキのあった”動作を実現する。
- (c) 主遣いは役の心情や性格、年齢、身分を自ら解釈して動作に反映させるため、同じ“型”でも人形の動きは人形遣いや場面によって変化する。
- (d) 舞台上での人形遣いは黒装束に黒頭巾を纏い、声を発することもできないため、発話やアイコンタクトによる合図のやり取りは存在しない。
- (e) 3人の人形遣いは、熟練していればどのような組み合わせであっても即座に操作できる。
- (f) 三味線と太夫（床）は、音楽、ナレーション、台詞の役割を持ち、主遣いの動きに大きな制約を与え、左遣い、足遣いに対しても、直接的、間接的に影響をもたらすと言われる。

したがって、限られた情報伝達手段である“ず”が鍵になる協調システムだと考えられる。

次に、人形遣いの3人が連携をとるための合図である“ず”の概略を説明する。文楽人形の操作に関しては主遣いが意思決定を行い、あらゆる動作を好きなタイミングで実行できると言われており、この際に用いられる合図が“ず”と呼ばれるものである。ただし、“ず”の体系や仕組みについても明文化されたマニュアルのようなものではなく、後藤ら[3]によって左遣いの役割の核心は「主遣いの意図を読み取ってその望む形を実現すること」などの指摘がされるように、各人形遣いは主遣いの意図を察することが重要だとも言われている。本論文では主遣いから他の人形遣いへの合図を指す言葉とする。

定義の曖昧な“ず”であるが、上述した特徴から、“ず”には以下の情報が含まれていると考えられる。

- ・次に行う動作の内容

- ・動作の速度や大きさ

動作の内容については、典型的な動作パターンは先述した通り“型”と呼ばれ、多くの動作は型の組み合わせで表現される。

そして、“ず”の伝達媒介としては、以下のものが挙げられる。

- ・主遣いが制御する人形自体の動き（頭や体の向き、右手の動きなど）
- ・主遣い自体の動きや接触反力（腰の動きなど）

主遣い自体の動きについては、足遣いが自分の腕を主遣いの腰に当てることによって、人形の歩行のタイミングなどを、人形を介さずに直接知ることができる仕組み（腰当たり）になっていることが知られている。しかし、人形自体の動きに含まれる“ず”については明示的なマニュアルもなく、目視しただけではどれが“ず”であるかわからない。そのため一步離れた位置から人形の左手を担当する左遣いは、人形動作の観測から“ず”を見極めるという高度な判断を行う必要があると考えられる。この左遣いが受け取る情報と左遣いの役割については、2章でより詳細に述べる。

この非明示的な要素を含む非言語情報“ず”と、人形自体の構造や動作の工夫（また床の特徴）とあいまって、文楽人形は分割操作とは思えない自然な動作や感情豊かな表現を実現するものと考えられる。特に、マスターである主遣いが出力となる人形の一部を操作しながら合図を送り、他がリアルタイムにその合図を読み取り操作する仕組みは、マスター—スレーブとしても特徴的といえる。ただし、繰り返し説明しているように、これらの技術の習得のためのマニュアルは存在せず、「足遣い10年、左遣い10年」[10]と言われるように、足遣いを10年経験した後、左遣いを10年経験することで、主遣いに必要なノウハウを獲得できるという、長期に渡る段階的な修行による技の習得が必要だと考えられている。

1.3 文楽の持つ特殊性とその成立過程

1.2 節で舞台や人形構造など、文楽の基本的な構成要素について説明した。本節では、中でも固有のシステムである三業と三人遣いについて、その成立過程を含めて詳細に説明する。

三業とは、先述した通り、三味線の演奏・太夫の語り・人形芝居の三要素を指す。この三要素は、ただ同期して演奏・演技するだけという関係にあるだけではなく、三味線と太夫は互いにタイミングを測り、時に崩したりしながら演奏を行い、人形はそれによって作りだされた間（浄瑠璃の間）を読んで演技を行う[6]。そして三人遣いとは、高さが約 1.5[m]、重さが 6~12[kg]、時に 20[kg]にもわたる大きな人形を、繊細にかつ大きく動かすために 3 人で操作を分担する人形操作方式のことを指す。文楽では、三人遣いによって、三業（三味線と太夫）がもたらす毎回変化する音楽要素（床）の中（特殊な文脈状況下）で、多彩な演技を人形遣いが公演中に自由に決定し、実行することができる。では、なぜ文楽では三業と三人遣いを含む形態となったのか。その一因は文楽の成立過程にあると言われている[10]。

伝統芸能文楽の成立は、江戸時代前期にまでさかのぼる。はじめは浄瑠璃に合わせて簡単な仕組み（棒遣い）の人形が芝居を行う比較的シンプルなものであった。やがて、1.2.1 節でも述べたように、1684 年ごろの義太夫節と呼ばれる浄瑠璃の誕生によって、大きく流行することになる[5]。文楽が流行すると共に、より多くの人間に、より繊細な表現を見せるため、人形は大型化し、さらに細かい表現を突き詰めていくこととなった。当時の人形遣いはからくり人形の細工師を兼業するものも多く、多くの技術が試行錯誤的に取り入れられていくこととなった。例えば、子ザルと呼ばれるレバーによるワイヤ駆動や、くじらのひげを素材としたばねを用いた反力調整（図 1-5）などによって表現の自由度を増し、より繊細に操作できるような工夫がなされていった[11]。

3 人遣いは、1734 年の「蘆屋道満大内鑑」と呼ばれる演目において、「差し駕籠」で、高く手をあげ、腹を膨れたようにみせる演技を行うために実装したのがきっかけとなって、それ以降定着したと言われている[12]。これより 40 年ほど前にも 3 人で人形を操作した事例はあるものの、操作分割方法が異なり、定着はしなかった。分割方法の関係上、移動を伴う演技に大きな制約が生じたことが定着しなかった理由だとも言われている[13]。この時代には人形の仕組みも更新が行われているので、現在に伝わる文楽人形の仕組みがこの瞬間に完成したとは断定できない。つまり、文楽人形の構造と操作システムは、義太夫節に乗せた高度な演技を実現し、多くの観客に見せるという目的のた



図 1-5 文楽人形のからくり（“政岡”のかしら）

持ち手となる左手で、チョイや子ザルを引くことで、ワイヤ駆動により首の上下（うなずき）や瞼の開閉（ねむり）を操作する．文楽人形の構造ならびに操作方法の詳細は付録 A.1 と A.2 を参照．

めに、試行錯誤的な“進化”が行われた結果として成立したと言えよう．

この時期は人形浄瑠璃が大流行し、全国各地で多くの“座”（公演団体）が誕生し、切磋琢磨する関係にあった．この時に各地に広まった人形芝居は、文楽以外の、例えば阿波人形などとして、民俗芸能の形で現在でも残っている．

文楽は職人集団としてこの時代を生き残った．観客を魅了する演奏技術、人形構造、そしてその操作技術を確立した．これらは極めて習得が困難な芸能であるが、それ故に多くの人間を惹きつける情緒深さや表現自由度を有すると考えられる．文楽人形が三味線と太夫が作るシナリオに合わせながらも、さまざまなアドリブ動作を行えることも、進化の末に獲得した人形と操作技術によるものと考えられる．

次節では、この特殊性の内容について、先行研究と関連研究を踏まえて説明する．

1.4 文楽を対象とした先行研究と関連する先行研究

前節では、文楽には三業と三人遣いという歴史的な背景から成立した固有のシステムがあり、三人遣いによって、高い演技要求を満たすために、高度な協調操作システムが確立したことを説明した。本節では、その操作システムの特殊性について整理するため、文楽に関する既存の研究と、文楽人形操作と類似する協調タスクに関する研究について整理する。

1.4.1 文楽の操作システムと先行研究

文楽を対象とした研究例は少なく、床を対象としたものとして、渋谷ら[6]による浄瑠璃と人形遣いの呼吸の同期の分析などがある。文楽人形の動作を対象とした先行研究の例としては、文楽がもつ情緒深い動きに着目し、その動作と感性情報との関係性を探るもの[14]などが挙げられる。

一方で、文楽の協調操作システム、つまり3人での分割操作の技法に関しては、1.2節で述べたように芸談に基づく定性的な記述がほとんどで、実験的検証などを含む先行研究は少ない。

既存の先行研究として、櫻らは1.2節で述べられた協調操作の特徴から、文楽人形遣い間の情報伝達を情報伝達モデルによって捉え、人形操作が図1-6のように行われているという仮説を提出した[7]。このモデルでは、まず“床”を参考にして、主遣いが人形の役や表現したい心情を考慮し、実現したい動作をイメージしながら人形の頭、右手を操作する（図1-6左上）。そして残りの人形遣いが、それぞれ人形の動きなどに含まれる合図である“ず”や、“床”を参考に、主遣いが表現したい動作のイメージを推察し、残りの人形部位を操作する。この3者の操作が一致することで、“イキのあった”動作が出力される（図1-6右上）。つまり、“ず”と“床”を入力としたフィードフォワード系で記述可能という仮説である。この仮説に基づけば、本システムはマスターが明確なマスター—スレーブシステムともいえるが、マスターである主遣い自体が出力動作の一部を担当することや、スレーブへの合図の伝達方法、またタイムラグのない出力が行える点が特徴的である。しかし、このモデル内の入力にあたる“ず”や“床”の連携への影響について検証した先行研究は存在せず、このモデルの妥当性についても議論できていない。

そこで、櫻ら[15]は、床や事前の打ち合わせを統制した実験により、これらの要素が

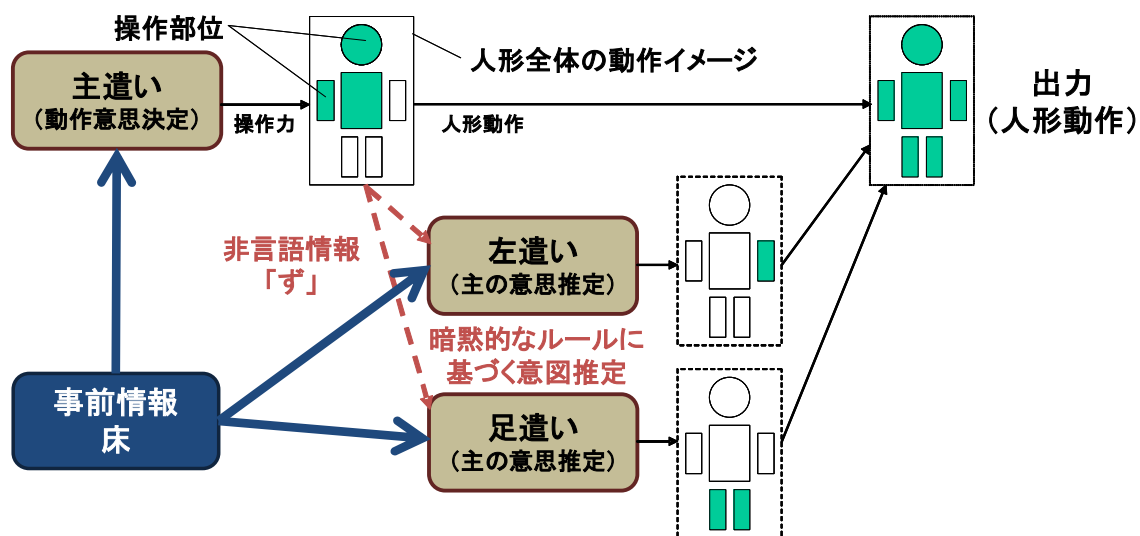


図 1-6 文楽の情報伝達モデル（論文より改編）[8]

なくても，“ず”のみで動作決定を含めた協調操作が行えることを実験的に示した。もし人形操作システムの大枠が図 1-6 に示す通りであれば，“ず”とその判別が協調操作の鍵になると考えられる。

したがって、文楽人形の操作システムには、楽曲や打ち合わせによる補助なしに、任意のタイミングで身体の一部が同期した演技を主遣いが人形の一部を操作しながら自由に決定、実行できる特別な仕組みがあると考えられる。この仕組みが一般的な協調タスクに対し、どのような特殊性を持つのかについては次節で述べる。

また櫻らはこの仮説に基づき、主遣いの動作と左遣いの動作の同期性を解析することで、その関係性を分析する研究も行っている[8]。この先行研究では、3人の人形遣いのうち、主遣いの操る右手と左遣いの操る左手の動作が特定の型で同期することに着目し、この同期性を協調性（“イキ”が合う状態）の指標とした。そしてセンサを内蔵した人形（詳細は、付録 A.3 節および A.4 節を参照のこと）で、型動作を計測し、人形の右手と左手の動作をウェーブレット変換によって位相解析することで、主遣いと左遣いの操作の同期性を主要周波数成分の位相の遅れの形で評価した。その結果、1つの型動作（“お辞儀”）を実行する主遣いと左遣いの間には、例えば図 1-7 のように両手を最後に同期させて動かすような動作を行う場合、①主遣いが先行して“ず”を伝え、②左遣いが動作を揃え、③一致した動作を実現する3つの段階が存在することを示した。また、

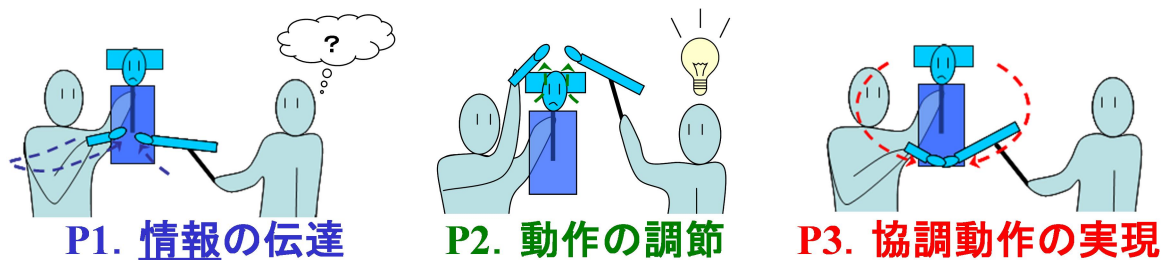


図 1-7 お辞儀における協調段階の変化[8]

人形遣い同士の事前の打ち合わせや、楽曲要素（床）などの文脈要素を統制した先行研究[15]が行われている。この先行研究では、3 人の人形遣いが打ち合わせをする、しないを統制した上での比較と、スピーカーを使って 3 人が楽曲（床）を聞ける状態、ヘッドフォンにより主遣いだけが聞ける状態を統制した上での比較を行った。その結果、打ち合わせがあってもなくても、また 3 人全員が楽曲（床）を聞ける場合でも主遣いしか聞けない場合でも、動作の高い再現性が確認されたため、主遣いは“ず”のみで任意の型動作を実行でき、一般的な芸能において用いられている楽曲要素（床）や事前の打ち合わせは、動作実行には必須ではなく、細かい動作のタイミングや大きさなどを揃えるための補助要素として使われる可能性が明らかになっている。

しかし、この“ず”の具体的な内容や判別の仕組みは未だ工学的には解明されていない。

このように、文楽は、移動を含めて多彩な演技動作を実行可能で、事前の打ち合わせなしに、合図の発信側と受信側により分担操作される人形動作自体を合図の媒介にすることで、動作種類の決定・共有とタイミング調整を同時に、かつ瞬時に行える特性をもっているが、そのための合図とその判断方法がいまだに明らかになっていない。1.4.2 節では、文楽人形遣い以外の協調タスクに関する研究を参考に、文楽人形遣いの仕組みに近いと考えられる協調操作を含むタスクを取り上げ、それらが文楽とどう異なるのか、および文楽ではどのような仕組みがあり得るのかについて比較考察を行う。

1.4.2 文楽とそれ以外の協調タスクの比較

先述したように、文楽には、事前の打ち合わせなしに、合図の発信側と受信側により分担操作される人形自体の動作を媒介に、動作種類の決定と共有、ならびにそのタイミング調整を、同時にかつ瞬時に行えるという特性がある。他の芸能や協調を要する作業（以降、協調タスクと呼称する）ではこのような現象が起こっているのか、また起こっているとすればどのような手段で行われているのかについて考察し整理する。

一般に協調タスクは複雑な要素から構成されており、協調のための作業が行われるまでには様々な段階が存在する。T. W. Malone は人間の協調活動において **Coordination**（調整）を行うために意思決定が、意思決定を行うためにはコミュニケーションが、コミュニケーションを行うには共通オブジェクト（共通の目標や課題、文脈情報など）の認知が必要であるという階層モデルを提案している[16]。このモデルに当てはめて文楽が持つ特性を簡単にまとめると、「主遣い（意思決定者）からの“ず”による指示（限られたコミュニケート）で、打ち合わせなどの文脈情報なし（限定された共通オブジェクト）で様々な人形動作を行うこと」となる。よって、コミュニケート、および共通オブジェクト・作業対象（出力）に着目すれば、文楽と他の協調タスクとの違いを整理できると考えられる。

本来、共通オブジェクトには事前の打ち合わせのほか、人形遣い達が修行によって体得した膨大な情報や明文化されていないルールなども含まれる。しかし、ここでは修行や訓練を終えた人間同士が協調タスクを実行するために必要な情報を、共通情報として定義し、先行研究における文楽以外の協調タスクとの比較、整理に用いる。よって、ここでは意思決定、コミュニケーション、共通情報および操作対象が重要なポイントであるとし、タスク遂行のための意思決定を行う人物（リーダーと呼称する）の役割と、コミュニケーションのために用いる合図、共通の情報とそして出力される動作（演技）に着目して、以下のポイントから先行研究を整理した。

①リーダー（命令者・先行者）の存在の有無

- ・ 一般に協調と呼ばれる作業では、作業遂行のために、**明確なリーダーシップ**をとる（意思決定を行う）者が、①存在する／②しない／③するが順次入れ替わる、という3つのタイプが存在する。
- ・ これらは主にタスクの目的達成に意思決定が必要かどうかで決定される。
- ・ 文楽では“主遣い”と呼ばれる人形遣いが、演技中も意思決定を行い指示を与える。

②シナリオ・打ち合わせ（文脈情報）の有無

- ・ 通常，多くの動作が必要なタスクでは，あらかじめその順番などが定義，共有される．
- ・ この順番（シナリオ）の影響力は**タスクが持つ即興性**に大きく左右される．
- ・ 文楽では“床”（演奏と語り）は重要な要素だが，先行研究[15]から，これらの要素が共有されていない状態や，打ち合わせが行われていない状態での動作も行える．

③合図に利用可能なチャンネル

- ・ タスクの種類や命令の種類によって**合図を行う際に利用可能なチャンネル**も変化する．
- ・ これらは**タスクの出力形態**に大きく左右される．芸能の場面では限られることが多い．
- ・ 文楽人形操作の現場では，発話やアイコンタクトによる合図が困難であることから，人形遣い同士の身体接触や人形の動作自体に巧妙に合図が埋め込まれている．

④出力の特徴

- ・ タスクとしてが行う内容そのものである．この自由度や分割方式，動作変更（即興）の頻度が，タスクに必要な合図や文脈情報の量と深い関連がある．
- ・ 特に，行う作業内容があらかじめ決まっている協調タスクでは，合図による動作開始タイミングの調整が重要となる．
- ・ 文楽では多自由度の人形動作を，3 人の人形遣いによって実現する．後述するが，人形の移動動作を伴った動作を行える点と，アドリブで好きな動作を決定できる点が特徴的である．
- ・ また，文楽の芸談によれば，“ず”の中には“ホド”と呼ばれる動作開始直前を示す動き（合図）があるといわれている[17]．これは必ず主遣いから出されるため，一連の動作を行うたびに，少なくとも一度は主遣いによる意思決定と合図が行われることを意味する．
- ・ また，“ず”は細かい動作要素に分かれているとも言われており[3]，人形遣いによっては“ホド”と他の人形遣いの裁量に頼る遣い方もよく行われる[18]が，打ち合わせなしでアドリブを行う際には，高頻度での合図が行われると考えられる．

これらに基づいて整理すると表 1-1 のようになる。基本的には出力の多様性（意思決定の頻度や必要な行動の種類）に応じて、文脈情報や合図のチャンネルが増える傾向がある。しかし、文楽における協調操作では、使える合図や文脈情報の少なさに対して、動作決定や同期などの要求が多いという一見相反する特性がある。

以降では、それぞれ他の協調タスクでの先行研究について触れ、文楽の協調操作システムでは、どのような要素によってこれを可能にしていると考えられるかについて考察する。

表 1-1 協調タスクの分類

名称・分類		リーダー		文脈情報	
タスク名	有無と専属性	シナリオ	打ち合わせ		
協調歩行	なし※	なし	なし		
協調運搬	なし※	なし	あり		
セッション	なし※	あり	あり		
クラシック	あり(指揮に専念)	あり	あり		
文楽	あり(操作にも参加)	あり(なくても動作可)	ほとんどなし※		
※状況により変化		※但し、床本は存在する			

名称・分類		作業出力		出力以外の合図		出力の特徴			
タスク名	媒体	出力(方法)		動作の方向や速度の変更		作業開始時以降の同期指示		動作の決定と指示	
協調歩行	身体運動	なし		身体運動で実行		なし		なし	
協調運搬	物体運動	発話など		物体反力で実行		発話で実行		発話で実行	
セッション	音情報	アイコンタクトなど		音情報で実行		音情報やアイコンタクトで実行		文脈情報や音情報で実行	
クラシック	音情報※	指揮棒・指揮者全身		指揮者の全身を用いて実行※		指揮者の全身を用いて実行		なし	
文楽	人形動作	ほぼなし※		人形動作で実行		人形動作で実行		人形動作で実行	
※指揮者は指揮動作のみ		※左遣いの場合		※指揮者は音を出さない					

1.4.2.1 協調歩行(Synchronization)

身近な協調現象として、無意識のうちに友人と足音がシンクロする、観客の手拍子のリズムが自然に揃う現象が挙げられる。

このような現象 (synchronization) は振動子の相互作用として表現される。この現象は、社会的な自己組織化プロセスと関連付けても考えられており[19]、その発生メカニズムや、この現象が対人場面でどのような役割を果たすのかについて研究が行われてきた。

例えば、Yun ら[20] は、タッピング課題[21]を対象に、同期変化の評価と共に同期と関連する脳領域の機能的接続性を評価することで、無意識的な身体同期性と対応する脳間 (inter-brain) の同期的つながりを見出した。本人は意識せずとも無意識的な過程により同期が発生するこの現象は、ヒトの社会性と脳科学、組織科学の関係性を示す鍵になると考えられる。また、Zivotofsky ら[22]は、2 人の人間の歩行を対象に、同調をフィードバックループとして考え、その同調を発生させる要因について検討を行った。その結果、触覚フィードバック（手を握ること）が要因となることを示し、歩行時に上肢と下肢が連動するためと考察した。さらにそれが視覚や聴覚によるコミュニケーションがなくても同期することを示し、歩行障害者へのパターン改善の応用可能性を示した。

このように、synchronization に相当するタスクでは、無意識的に同調が発生（し、かつ社会的な関係などによりその特性が変化）するケースが多く、それらのケースには、①リーダー／フォロワーの関係は相互作用的で、明確な主従関係が存在する場合としない場合（社会的関係で変化し得る）があるが、②無意識的過程で、事前の打ち合わせなどはほぼ不要となる、③意識的に複雑な命令や合図を行う必要がなく、④同期のためのチャンネルは本人の身体動作そのものである、という特徴がある。

合図と出力については、以下のようにまとめられる。

- ・観測可能な出力（合図）で 2 者のパターンが受動的に一致する過程である。
- ・合図（同期用入力）は出力そのものであり、同期状態を導く。
- ・動作は単一の単純な動作である。

文楽の人形操作と比較してみると、単純なモデルではあるが、複数の出力が同期するという意味で、文楽と共通性をもつ（例えば、文楽でも左右の腕動作などは同期する（櫻ら[8]））。ただし、文楽では、タスクを行うたびに意識的に合図を送り、人形の身体での同期を実現させている点で、文楽と単純な同調と

は大きく異なるといえる。

1.4.2.2 協調運搬

文楽では、主遣いの意思決定により、1つの人形を自然にコントロールできている。

文楽に類似した身近な例として、複数の人間で1つの重量物を運搬するシーンが挙げられる。このようなシーンでは、上述の身体動作だけでなく、タスクの対象物の状態（位置や移動する方向など）が意図伝達の重要な対象となる。また、重要な点として、明確な意思決定を行っているが、その合図は意識的には出していない場合もあることが挙げられる。

このような協調現象については、人間がコミュニケーションに利用しているチャンネルの検討が行われている。例えば、協調運搬[23]では、人間同士が1つの重量物を運搬する際、互いに合図を行わなくても、重量物へ与えているトルクから、移動したい方向を伝達できる点に着目し、運搬物を介したシステムに人間のインピーダンス特性を考慮することで、リアルタイムな協調運搬が可能になることを、ロボットへの実装で示した。

ただし、現実的な場面での人間同士の協調運搬タスクを考えると、（同時刻で近くの人間が同じ対象物を操作するため）①リーダーが明確に存在することが多く、②大まかな移動ルートなどはあらかじめ打ち合わせて決めておき、③同時に発話なども行いながら、対象物の運動を媒介にして（つまり意図的な指示を用いずに）細かな動作方向などを伝達することで、④複数人での重量物の運搬が成立している。

発話による合図も多く行われるが、このうち重量物を媒介とする意図的でない合図は“インフォーマティブモーション”[24]と呼ばれ、ユーザーが意図的に操作しなくても先回りして行動するインタフェースへの応用が期待されている。簡単にまとめると、以下のような特性がある。

- ・観測（計測）可能な信号で方向や速度を常時制御するシステムである。
- ・命令者は速度や方向を無意識に伝達し、大きなルート変更などの意思決定は、発話などの別の合図を組み合わせで行われる。
- ・単純な運搬動作などに適用される。

このように主従関係が明確かつ、操作対象を利用した合図で実時間の調整を可能とする点は文楽と類似しているが、協調運搬では動作方向の大がかりな変更などは打ち合わせや発話を併用することでスムーズな連携を行う。一方で、文楽では徹底した修行とプロトコル構築によって、これらを撤廃していると考えられる。

また文楽では、合図はいずれもリーダーから出される意図的なものとされているが、

これらの合図は熟達した人形遣いが感知でき、素人には自然な動作にしか見えない。また、文楽では様々な動作命令を打ち合わせなしに次々に出すことができる。よって、人形の移動中の姿勢維持などについては文楽でも協調運搬と似たようなプロセスが行われている可能性があるが、様々な動作（自由な演技動作）を意思決定し、その情報を共有する機能は、協調運搬システムでは説明できない。

1.4.2.3 ジャズセッション

協調運搬では、即興性（自由に演技を決定し実行できる）の部分が文楽との相違点として挙げられたが、身体運動の同期の他に、即興性をもち、内容をその場で決定でき、道具を用いた表現芸能として、ジャズセッションが挙げられる。

ジャズセッションでは、演奏者がアイコンタクトなどの他、各パートが奏でる音のメロディー、ハーモニー、リズムなどが合図となり意図が伝達される。これらの協調状態を可視化するため、森ら[25]は、音情報から、音楽要素（コード進行など）と感性情報（緊張感や主張度など）をそれぞれのエージェントが個性（嗜好）に応じて抽出する機能を取り入れたマルチエージェントモデルによる表現を行うことで、演奏の進行に従いながらも、多様性のあるセッションシステム構築を行っている。また、Nicholas P. Dempsey ら[26] は、演者の背景を含む様々な文脈情報に着目し、これらを受け取る能力が演奏者ごとに異なることや、それらの情報が演奏にどのような影響を与えるかについて議論を行っている。ジャズの現場では、①リーダー（主となる演奏者）があり、②楽譜などの文脈情報に沿いつつも、③アイコンタクトや演奏の音情報そのものを媒介として、④タイミングの調整やアドリブ、演奏者（リーダー）の交代などを伝達している。アドリブには論理的制約が存在しており[27]，これらを活用することで、同期のタイミングやアドリブのスムーズな導入を可能としている。まとめると以下のような特徴を持つことがわかる。

- ・複数の演奏者が、楽曲と論理的制約を頼りに、協調する現象。
- ・合図は楽曲自体＋同期トリガとなるアイコンタクトなど
- ・後述するクラシックより、いくつかの制約の上で即興的なアドリブ要素が存在する

[28]

文楽と比較すると、同じ芸能分野で、リアルタイムに協調演奏ができ、即興性も認められる。異なる点としては楽譜などのシナリオにそって進行することがあげられる。文楽にも床本は存在しており、三味線と太夫はこれらの制約の中で演奏と語りをを行うが、文楽人形は演奏と語りによるシナリオに従いつつも、動作要素の決定を含む多くの即興

動作をシナリオなしに自由に決定、実行することができる。また、文楽では、高い即興性を保つために、人形の動作ごとに合図を出す必要がある。また、JAZZ を含めタスク内容やタイミングの多くを決定する、床（楽曲）などの文脈情報は補助的に利用されるにとどまる[15]。

1.4.2.4 クラシック音楽

芸能分野で、より多くの表現自由度をリーダーが決定できる例として、クラシックの指揮者があげられる。クラシックの演奏現場では、指揮者は事前打ち合わせに加え、各パートの同期タイミングの他、感情表現などを専用の指揮棒と身体動作を駆使して伝え、全体の意思を統一する。

一般にプロ集団であれば、指揮者なしでもオーケストラや合唱は成立するが、指揮者の有無によって様々な面でパフォーマンスが変化する。この効果に関する研究は多く行われており、指揮者には以下のような役割があるとされている[29]。

- ・指揮者はオーケストラと同期状態を作り、指揮者が表現の方向を決める。
- ・指揮者の合図は、①開始／終了指示、②テンポ／拍子制御、③強弱・音色の制御等
- ・結果として、演奏表現や全体の方向性をコントロールできる。
- ・動作要素の追加はできないが、他の制御項目は文楽と類似する。
- ・出力（楽曲）は定まっていて、そこに見える合図を大量に組みこむ。
- ・打ち合わせなどの文脈情報も最大限活用し、十分なリハーサルも行う。

このようにクラシックの現場では、①演奏自体には参加しない明確なリーダーが存在し、②はっきりとした楽譜で多くの演奏内容が固定されている。③しかし、各パートの同期タイミングだけでなく、表現に関する様々な指示を、④全身と指揮棒を使ってリアルタイムに他の演奏者に伝達することができる。

文楽と比べてみると、クラシックの指揮者は多くの非言語情報で、全体の方向性をパートごとへの指示で決定できる点が文楽と似ているが、文楽と異なり、明示的な専用の合図動作を使って行われる。一方、文楽でも多くの意思決定を行うが、これらは限られたチャンネルを用いて自分が担当する演技動作の中に隠す形で行い、これらは演技補助ではなく演技内容を決定するものである。さらに、楽譜などがない状態でも、ひとつひとつの動作を決定することができる点が異なる。

1.4.3 まとめ

以上、表 1-1 にまとめられた通り、多くの協調タスクでは、スピードや移動方向などの情報は、音や操作対象物などの媒体を利用して伝達することが可能であるが、(最初の動作開始を除く)同期タイミングを合わせようとする、発話や専用の合図を利用するものが多い。また、多くの協調タスクでは、途中での動作変更などの複雑なプロトコルを実現するために、予め打ち合わせや楽譜などを活用する。このように協調タスクに必要な(意識的/無意識的)命令の種類や頻度と、その伝達チャンネルと同時に使える文脈情報の量と質は、多くの協調タスクである程度対応しており、人間はこれらを駆使することで目的を実現している。

しかし、伝統芸能文楽では、①意思決定者が一部を操作する人形自体を合図の媒介とし、アイコンタクトなどは用いない、②動作内容の決定もリアルタイムで行える、というように、命令のために使える合図の媒介の自由度が狭く、事前情報も限られているにもかかわらず、動作をリアルタイムに自由に決定できるという、一見相反する特性を実現している。これは文楽人形が、舞台上で多くのアドリブ動作を行うという特徴に起因すると考えられる。通常楽譜や台本なしに様々な動作を行うには、発話などの合図や打ち合わせが必要となるが、文楽ではこれを“ず”という特殊な合図を動作に含めることと、長年の修行によってこの“ず”を見分けることによって可能としていると考えられる。

人形の協調操作のうち、合図を用いない姿勢の維持や身体動作の同調などについては他の協調現象に関する先行研究でも取り扱われており、文楽でも同様の特徴を備えている可能性がある。しかし、事前の打ち合わせや楽曲などの情報なしに即興で動作を決定し実行できる点については、他に類を見ない特徴といえる。

そのため、文楽の協調操作を解明するには、多くの動作を実現する要因である“ず”と、そこからどのように情報を判断するのかを検証していく必要がある。

1.5 本論文の目的と構成

ここまで述べたように、文楽人形の芝居は類を見ない特殊な協調操作によって成立している。この協調操作システムにより、人形遣いは公演内で、分割された身体をリアルタイムに同期させつつ、様々な動作を、感情を込めて自由に行うことができる。そして、そのシステムを実現可能としているのが、人形動作中に埋め込まれる“ず”と呼ばれる

信号であり、人形遣いはこれを瞬時にもしくは時間的に判断することで、他の協調タスクでは考えにくい文脈情報なしでの同期と動作情報の伝達を、人形自体の動きを媒介として観客に見せるための人形動作そのものと同時に実現し、リアルタイムに動作決定を行える自由演技（アドリブ性の高い演技）を可能としている。これは舞台上という発話などができない状態で、床に合わせてアドリブを含んだ多彩な動作をするためと考えられる。しかし、“ず”の構成内容やその伝わり方、“ず”を受け取る左遣いと足遣いの判断方法などについては未だ工学的に明らかにされていない。

そこで本研究では、人形遣い同士の情報伝達のうち、“ず”を受け取って左手を操作する“左遣い”に着目して、この情報伝達の仕組みの一部を解明する。左遣いはアドリブの対応などのために、多彩な動作を逐一判別しながら左手を操作しなければならない。そこで特に左遣いに対する“ず”の中でも、動作の種類である型の判断に絞り、その判別方法について議論する。具体的には、2章で芸談や先行研究をもとに左遣いの“ず”の判断について仮説を立て、さらに3章で左遣いが“ず”を判断する状況を再現した実験を行い、仮説の妥当性を検討する。また4章では隠れマルコフ過程をベースにした確率モデルにより、仮説で挙げたリアルタイムな判別を行えるだけの情報（“ず”）に当たるものが人形動作に含まれているかどうかを検討する。5章でここまでの結果から考察と議論を行い、左遣いの型判別の仕組みについて議論する。

第2章 左遣いの“ず”判別の仮説

2.1 はじめに

前章では、文楽全体の構造とその特殊性について解説し、文楽人形操作における高度な協調では、主遣いのリアルタイムな意思決定を他の人形遣いに伝える特有の合図である“ず”とその判別方式が重要な役割を占めている可能性を述べた。

本章では、左遣いに着目し、“ず”の伝達方式とそれを判別する人間の判断プロセスに関する仮説を立てる。この仮説に基づいて、次章での実験条件を設定し、“ず”とその判別方式の解明へ向けて検討を行う。

2.2 文楽における左遣いの役割

本研究では、左遣いの動作判別を対象としているが、左遣いは演技上どのような情報を受け取り、どのような役割を果たすのかを整理する。これは、協調操作にあたり“ず”以外に利用可能な情報がないかどうかを確認するためでもある。

文楽での“左遣い”とはそれ専門の役職ではなく、人形遣いの担当配置の一つである。文楽の人形遣いは、その経験などに応じて、公演ごとに登場する各人形の“主遣い”，“左遣い”，“足遣い”のいずれかに任命される[9]。

経験の浅い人形遣いは、1人で操作できる端役のツメ人形¹や、足遣いを担当することで、経験を積み、次第に左遣い、さらには主遣いも担当するようになる。また左遣い

¹ 文楽人形のうち、百姓、腰元、兵卒、捕手、人足などの端役の人形をいう。不恰好な顔つきをしており、衣裳も簡単である。一人の人形遣いが左手を人形の背後から差込んで支え、右手で人形の右手を遣う。

を主に操作する人形遣いも、主役でない人形の主遣いを担当するなど、徐々にステップアップする仕組みが採用されている。

その中で左遣いは、操作対象である人形から比較的離れた位置から、人形の姿勢や舞台全体を見渡せる立場にあり、人形の左手操作や重い人形の把持の補助の他、扇子などの小道具の取り出しなども行う。多くの視覚情報を利用可能な左遣いだが、人形や主遣いから少し離れた位置にいるため、足遣いが行う“腰当たり”（主遣いの腰に腕で触れ、歩行のタイミングを測る）のように主遣いからの反力による情報伝達は期待できない。人形の身体についても同様で、左遣いは差し金（図 1-4 を参照のこと）と呼ばれる棒で左手を離れた位置から操作するが、左手と胴体はヒモでつながっているものの、これを引っ張ることは主遣いの演技の妨げおよび人形のプロポーションを崩すことになるので禁じられている。着物の重さなども身体表現や操作時の抵抗として利用し得るが、“腰当たり”のように主遣いからの直接的な合図として利用するのは現実的ではない。

以上、文楽における人形遣いの入出力をまとめると、表 2-1 のようになる。このように、左遣いは人形の周辺を含め、多くの情報を見ることができ一方で、反力を利用するような情報には頼れない。したがって、“ず”も人形の動作中もしくは主遣いの身体動作に含まれると考えられる。次節では、先行研究から左遣いへの“ず”の伝達媒介について考え、その伝達形式に基づいて 2.4 節で“ず”に関する仮説を提示する。

2.3 左遣いへの“ず”の伝達媒介

前節では、左遣いが“ず”の判別に視覚情報を利用している可能性を説明した。本節では先行研究から、実際に左遣いがどのような情報を参照可能なのかについて説明する。

櫻らは、人形動作解析時のパラメータを検討するために、視線計測装置を利用した予備実験を行った[8]。左遣いに視線計測装置を取り付け、演技実験を行うことで、左遣いの取得している情報について推定を行った（図 2-1）。取得した左遣いの視界データを図 2-2 に示す。ピンク色のマーカが左遣いの視線が向いている（中心窩で捉えている）対象となる。図 2-2 に円として示すような、頭部、頭部前方、肩、頭部-右手付近、その他の 5 つの領域区分に対する、左遣いの視線を示すマーカの時間変化を観察することで、左遣いの注視対象を推定した。

結果の一部（基本的な型動作「お辞儀」と「拝み」を行う部分）を表 2-2 と図 2-3 に示す。様々な“型”において、左遣いの視線は主遣いが操作する頭部と右手に集中し、その他の領域については滞留時間の比率は低く、連続滞留時間はほぼ 100ms を下回る

という結果を得た．この滞留時間は，既存の研究[30], [31] [32]における成果によれば，注視点に当たらないと判断でき，また実験後の人形遣いのコメントからも同様のことが推察されたことから，右手や頭部の絶対座標と速度変化を，左遣いが主として取得する物理量だと仮定して解析を行っている．

表 2-1 人形遣いの入出力

	共通情報	入力(合図)	出力
主遣い	床・台本	無し	かしら・胴・右手の操作
左遣い		人形動作・姿勢	左手の操作 小道具の出し入れ
足遣い		腰当たり 人形の向き	両足の操作 膝の表現(女方)

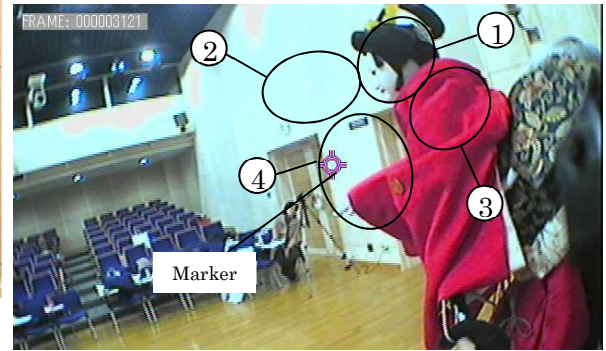
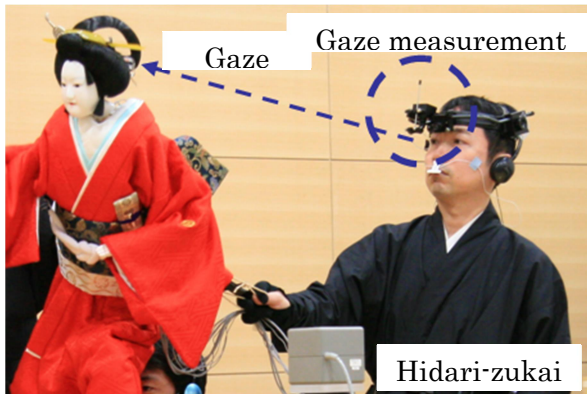


図 2-1 左遣いの視線計測実験

図 2-2 左遣いの視界

表 2-2 演技中の視線計測結果（注視点）[8]

Area number		Time(s)	Rate(%)
①	Head	4.47	32.8
②	Area in front of the Head	3.47	25.06
③	Shoulder	0.60	4.34
④	Area Between Head and Right hand	1.93	14.94
⑤	Other	0.13	0.96
⑥	Marker lost (blink)	3.1	22.41

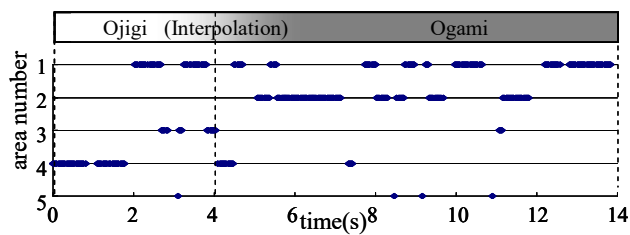


図 2-3 視線の時間分布

公演中の左遣いは、次に行うべき動作の“型”の種類を、“ず”から読み取る以外に、演目のシナリオから推測する、主遣いとのお打ち合わせからあらかじめ決めておく、などの方法も考えられる。しかし、1章で述べたように、櫻らの先行研究によって、この2つの方法が利用できる状態とできない状態のいずれの場合でも型動作を自由に行うことができること、これらの情報は動作の大きさやタイミングの調整のため、補助的に利用されている可能性が高いことがわかっている[15]。

以上、左遣いは人形の操作に関する情報の多くを、主遣いが操作する人形の頭部や右手の動きから読み取っていると考えられる。芸談においても、人形の合図である“ず”は人形の頭部である頭（かしら）から「ズ」と呼ぶようになったとされており、その多くは人形の頭部動作にあると推測される。

では、人形の頭部動作をみるだけで、次に行うべき動作の“型”を多くの候補から瞬時に判断し、観客からみて遅延のない全身動作を実現することが可能なのだろうか。次節では、これらの時間的制約や操作の制約について考慮し、左遣いの“ず”による動作判別の仮説を提示する。

2.4 左遣いによる動作判別の仮説

先行研究から、左遣いは主遣いの操作する人形の頭や右手の動きを観察することで、事前のお打ち合わせがなくても、“ず”によって行われようとしている“型”動作の種類を判別できることがわかっている。また、前節で左遣いへの“ず”のうち、少なくとも動作を指示する部分は、主遣いが操る人形の動作に含まれることを説明した。本節では、1.4.1 節でも触れた文楽における情報伝達に関する先行研究や芸談をもとに、左遣いがどのような判別を行っているのかについて、仮説を提示する。

2.4.1 表現上の制約

文楽人形の型は日常動作や象徴的な動作（座る動作やお辞儀など）を中心に 50 種類程度存在し、中には途中までの動作が互いに類似している型も存在する。にもかかわらず、多くの型では動作を開始してから、遅くとも 2 秒以内には左手を動かし始めなければならない。このことから、型の判別に与える制約は以下の 2 つが主と考えられる。

- ・型の種類の制約
- ・型の観察時間の制約

まず、前者に関して言えば、型は 50 種類程度あり、かつ部分的に似た動作もあることから、単純に 50 種類の動作が用意され、そのはじめの部分で分類しているとは考えにくい。もし“ず”が、「型の種類」ではなく「部分的に似た動作」を指し示す合図だとすると、型の判別は複数の“ず”を利用した時間的な絞り込み作業になると考えられる。この場合、型はそれぞれ動作の特徴や動作長の点で異なることから、型の種類によって、判別できるまでの時間が異なる仕組みになっていると考えられる。逆に、“ず”がこのような動作を利用せず、極めて微小な動作などで型の情報を伝えているのであれば、このような制約は存在しなくなるが、左遣いの観察は極めて困難になり、型の判別に要する時間も一定に近いものになると考えられる。

次に後者について考えると、左遣いは型の判別を人形動作の観察によって素早く行わなければならない。中には動作開始から 0.5(s)後には左手が動きだす動作もある。このことから、型の判別は最初の一瞬 (0.5(s)以内)で行うか、左手を動かしながら判別を続けられる仕組みのどちらかをとっていると考えられる。この場合も、型によって左遣いが判別を行うまでの時間に大きな違いがある。もし“ず”が最初の一瞬 (0.5(s)以内)で型の種類を判別させるような明示的なサインであれば、それよりも長い時間、観察を行っても正確性への影響は少ないと考えられる。逆にそうでないならば、時間的な猶予が増えれば増えるほど判別の正確性も上昇すると予想される。

このように、左遣いの型判別方法の検証には、その観察時間に着目することが有効だと考えられる。3 章では、これらの要素を検証可能な実験を行う。

2.4.2 左遣いの型判別仮説

前節から、文楽においては、以下の動作判別方法が考えられる。

- ①：最初の一瞬（～1 秒前後）で、すべての型を判別する
- ②：決定的なツリー構造によって時間的に判別する
- ③：ツリー構造で時間的に推測するが、各動作のアレンジも考慮する

これらは排他的で、どれか一つを仮説とする。このうち①については、型の種類が動作開始までの時間に対して多すぎることや、動作の大きさなどのアレンジを主遣いが頻繁に行うことから、現実的ではない。②については、動作観察を続ける中で、動作の明確な分岐点が時間的に発生することを意味するが、操作手法が明文化されない状況や、『ズを受けた左遣いは、人形全体の動作や視線、あるいは右手の動きといった状況を判

断して左手を遣います[9]』など、より複雑な判断が行われていることを報告する芸談と一致しない。そのため、③に示した、動作の速度や細部が主遣いのアレンジで自由に変えられても、左遣いがこれを踏まえて操作する形式が、もっともよく状況を説明できる判別方法だと考えられるが、実際に左遣いはそのような時間的に曖昧な判断を行っているのか、またそもそも人形動作にそのような分類が可能な情報が含まれるのか。現時点ではどちらも不明である。本研究では③を仮説とし、3章以降ではこの2点を検証していく。

③の立場を採用した場合、左遣いの型の判別は図2-4のように進行すると考えられる。主遣いが行う型を決めて動き始めた時点（図の左側）では、左遣いの中では、すべての型の確からしさ（以降、確信度）は等しく低いが、動作がある程度進むと動作中の“ず”を左遣いが受け取り、映像の中で実際に主遣いが意図し実行した型（以降、正解の型）の確信度（以降、正解確信度）が増していく。図2-4では、正解の型は“お辞儀”であるため、確信度の中で“お辞儀”に対する確信度（下線）が正解確信度となる。ただし、型には似た動作特徴を持つ（動作が途中まで似ている）ものが複数存在するので、正解の型と動作の点で似た型に対する確信度も同時に増していく（図の中央）。そして、決定的な情報が得られ、似た動作から正解の型を判別可能になった時点で、正解確信度が100%に上昇する（図の右側）。よって時間ごとの確信度の分布からは、似た動作特徴を持つグループが観測できると考えられる。また、各時点の正解確信度は、判別する左遣い（以降、判別者）が熟達するにつれて高くなると考えられる。

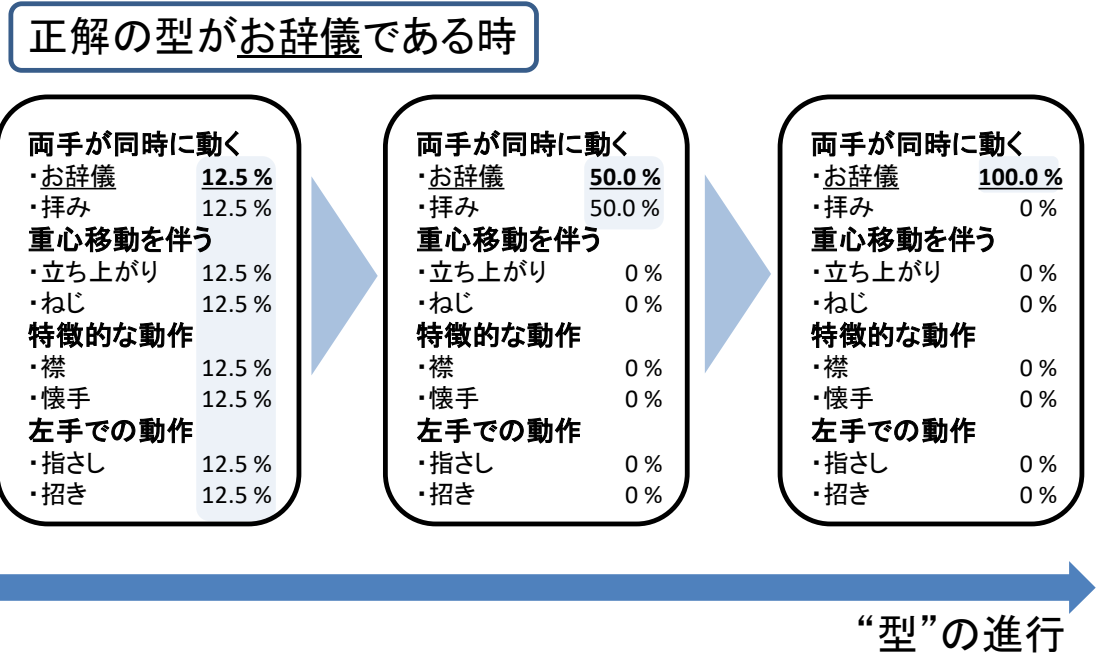


図 2-4 期待される左遣いの確信度の時間変化

2. 4. 3 左遣いの型判別に影響を与える要因

前節の仮説を実験的に検証するにあたって、統制実験を行い、確信度の分布をいくつかの実験条件により変化させて観測する方法が有効と考えられる。そこで、有効と考えられる実験条件について考える。

2. 4. 3. 1 仮説 1：型判別の正確性は、熟達の影響を受ける。

芸談で「足遣い 10 年、左遣い 10 年」[33]と言われている通り、左遣いに至るまで 10 年程度の修行を、“ず”を出す主遣いになるまでにさらに 10 年程度の修行を要するとされる。また、芸談においても左遣いと足遣い（あわせて“手伝い（てつたい）”と称する）への“ず”の出し方は十人十色と言われており[9]、状況に応じて“ず”の出し方を変えることがあるとも言われることから、左遣いとしての経験の長さに応じて“ず”の読み取りの精度が変化すると考えられる。すなわち、左遣いとしての経験が浅い人形

遣いと，“ず”を出す主遣いにまで習熟した人形遣いとは、型判別の正確性に違いがある可能性がある。

2.4.3.2 仮説2：型判別の正確性は動作の観察時間と共に上昇する。

先述したように型の判別は人形動作の観察によって素早く行わなければならない。中には動作開始から 0.5(s)後には左手が動き出す場合もある。もし“ず”が最初の一瞬(0.5(s)以内)で型の種類を判別させるような明示的なサインであれば、それよりも長い時間、観察を行っても正確性への影響は少ないと考えられる。逆にそうでないならば、時間的な猶予が増えれば増えるほど判別の正確性も上昇すると予想される。実験条件として、型の動画冒頭部分の提示時間を変化させれば、この左遣いの観察時間をコントロールできる。

2.4.3.3 仮説3：型の種類によって判別に必要な時間が異なる。

型にはさまざまな種類が存在し、それぞれ動作の特徴や動作長も異なる。もし“ず”が型の種類や動作の特徴を利用したものであれば、正確に判別できるまでの時間は、型の種類によって異なると予想される。

2.5 まとめ

文楽における左遣いの型判別は、主遣いの動作を逐一観察し、その時その時にどんな動作があり得るか候補を絞りながら、左手を操作している可能性が高い。さらに、主遣いが動作の速度やタイミングを変更しても対応できると考えられる。これは左遣いへの負荷が大きく、一見非効率的な判別方式だが、もしこれが可能であれば、主遣いの表現の幅を確保しつつ多彩な演技が可能になると考えられる。

次章では、実際に左遣いに、演技中の左遣い目線の映像を視覚情報として提示し、型動作の絞りこみが行われるか、またどのように行われるかを直接検証する。さらに4章では、そもそも主遣いが操る人形の動作内に、型判別を可能にするような情報が含まれているかどうかについて、機械学習を用いた検討を行う。この2つによって、左遣いが仮説通りの時間的な判別を行っているのかどうかについて考察し、さらにそのような分類を可能にする情報が人形動作だけに含まれているのかどうかを確認する。

第 3 章 左遣いの型判別実験

3.1 目的と概要

第 1 章で述べたとおり，左遣いは視覚情報を頼りに“ず”を読み取り，左手の操作を行う．しかし，この方法に明文化されたマニュアルは存在せず，また観客側から見ても“ず”を認識することは難しい．第 2 章では，芸談や先行研究から，左遣いがどのように判別を行っているのかに関する仮説を提示した．本章では，この仮説に基づいて左遣いがどのように“ず”を認識して，行うべき動作を判断しているのかを実験的に検討する．

具体的には，型の冒頭部分の左遣い視点の動画を実験に参加した各判別者（現役の人形遣い）に一定時間提示し，動画内で行われようとしている型が何であることを，自分が左遣いになったと仮定して推定してもらう実験を行った．この実験を実施するにあたり，第 2 章で述べた通り，左遣いの型認識が，候補となる型の確信度が時間とともに変化する絞り込みのプロセスであると仮定し，型の判別（正確性）に影響を与えられ①左遣いの熟達度合い，②型動作の観察時間，③型の種類の 3 つを要因として，実験条件として設定した．これにより，実験結果から，左遣いが，主遣いの動作開始からいつどの程度型を判別できるのか，また熟達度などの要素が左遣いの型判断にどのように影響するのか，すなわち 2.4.2 節のように，型の確信度が時間とともに変化していくプロセスに則っているかについて分析することが可能となる．

3.2 左遣い型判別実験

3.2.1 本実験の目的と概要

本実験の目的は，左遣いの型判別仮説について検証するため，2.4.3 節で検討した正

解確信度に影響をもたらす 3 要因を実験的に検討することである。そのために、左遣い視点の動画の提示による実験を行い、型判別が 2.4 節で述べた通りに行われているかどうかを検討した。具体的には、型の冒頭部分の左遣い視点の動画を実験に参加した現役の人形遣い（以下、判別者）に一定時間提示し、動画内で行われようとしている型が何であるかを、自分が左遣いになったと仮定して推定してもらう実験を行い、型判別の正解確信度への、判別者の熟達度（2 水準）、型の提示時間（4 水準）、型の種類（8 水準）の 3 要因の影響を検討した。

左遣いにとって型判別が可能と考えられる範囲で動画の提示時間を変化させることで観察時間を操作し、ダミー（動画では提示しないが、選択肢には存在する型のこと）を加えた 10 種類の候補に確信度を合計 100%になるよう割り振る形式で回答してもらった。そして、複数候補からの型の判別の正解確信度を、上述した 3 つの要因間で比較した。動画提示に用いる型は、似た動作特徴を持つものを含んだ 8 種類の型とした（残り 2 種類が上述のダミーになる）。3 つの要因とその交互作用が判別の正解確信度へ与える影響を、3 要因を独立変数、正解確信度を従属変数、判別者を変量効果とする一般化線形混合モデル（Generalized linear mixed model, GLMM）により検討した。

3.2.2 実験計画

第 2 章で提示した仮説通りの仕組みが実際にあるかどうかを検討するため、判別の確信度（どの型を行おうとしているか、その確からしさ）の変化に着目した。そして、左遣いの型判別の正確性に関わる以下の 3 つの要因について評定水準を設定した。設定にあたって、第 2 章で提示した判別の仕組みの仮説はもちろんのこと、芸談や演技上の制約、実験に参加しないベテランの人形遣いによるアドバイス、予備実験も参考にした。具体的には以下の通りである。

判別者の熟達度要因として、芸歴 30 年以上の熟達群（5 名、平均 34.4 年）と 15 年以下の準熟達群（6 名、平均 8.5 年）の 2 水準を設定した。

（動画での）型の提示時間要因として、0.75, 1.00, 1.25, 1.50(s)の 4 水準を設定した。予備実験から、提示時間が短すぎる（0.50(s)以下）と直観的な回答が行えない場合があることが判明した。一方で、現役の人形遣いに長い時間（1.50(s)以上）提示すれば比較的容易に判別可能なことも判明した。また、実際の公演中の型も多くは 2 秒以内に左手の動作や追従を促すものが多い。そのため、0.50(s)と 1.50(s)の間で等間隔になるように提示時間を設定した。

型の種類要因として 8 水準を設定し、動画で提示した。直観的に回答しやすいように、

それにダミー2種類を加えた10種類を回答候補の型として提示し（表3-1）、判別者には動画を見た後で10種類に対して確信度を合計が100%になるように割り振ってもらった。目的が図2-4に示したような確信度変化の観測であるため、“型”は確信度が分かやすい似た動きとそうでないものを含むことが望ましいと考えられる。“型”の選出にあたっては、公演で多用される型のほか、動作特徴の点で似ている型や、特徴的な動作を含む型を実験に参加しないベテラン人形遣いの助言のもと選定し、動画で提示した。表3-1の右側に型の動作内容と、似ていると考えられる型（太字下線）を示した。使用した動画は、事前に現役の人形遣いに様々な型をランダムな順序で実行してもらい、その際に左遣いに装着した視線計測装置（㈱ナックイメージテクノロジー社製NAC-EMR9）のカメラで収録されたものを実験に使用した。また、その際の主遣い、左遣い、足遣いの配役はそれぞれ一人で固定した。つまり、左遣いの注意が向いていた対象を含む一定範囲の視野を動画として収録し、実験で提示した。ベテランの主遣いと左遣い各1名による演技の動画を提示した。

なお、本実験で採用する、100%になるように確信度を割り振る評価方式は、評価尺度法（rating scale method）の1つである数値配分法にあたる[34]。評価法形式が数値間の心理的な等間隔を保証することに疑問がもたれることもあるが、この点に関して、田中は『カテゴリ幅は経験的にはほぼ等しいことが多く、もともと序数尺度ではあるが、距離尺度として用いても大過はないことが多い』[35]としており、慣例的に評価法形式によって得られた評価データは間隔尺度として分析されている。今回扱う評価データは、もとの事象（型）が発生する確率自体は0%～100%の比例尺度であるが、取得している評価データは「どの型がどの程度起こりうるか」の確信度である²。これは主観的にどの程度起こると考えているかの確信度を問う課題であるため、人を対象として得た評価データの慣例に倣い間隔尺度として扱う。

² 妻藤によれば、確信度の評価の取り扱いについては、特に再認記憶の分野において、信号検出理論の枠組みで理論化されている[36]。この理論では、記憶系の出力が何らかの一次元の量（熟知度、既視感の程度など）に変換され、それがある基準値よりも大きいかどうかで、「見たことがある」「ない」が判定される仮定であり、その量の大きさの基準値からの距離に、尺度値をあてはめたものが確信度だと仮定されてきた。現代では、実験データ解釈の前提として特に言及することなく間隔尺度として使われている。

表 3-1 型候補の名称と動作内容
(連続写真は研究室ホームページに掲載[37])

“型”の名前	動作内容と似ている可能性のある“型”
01お辞儀	お辞儀を行う。 <u>拝みと類似</u> 。 右手を振り、両手で円を描いて、頭を下げる。
02拝み	両手を合わせ拝む。 <u>お辞儀と類似</u> 。 右手を振り、両手を正面にそろえ、目を閉じる。
03立ち上がり	座った姿勢から立ち上がる。重心移動を伴う。 軽く首と体をひねり、ゆっくりと立ち上がる。
04招き	左手を突き出し招く。 <u>指すと類似</u> 。 頭を動かし、その視線方向へ手を伸ばして招く。
05指す	左手を突き出し指す。 <u>招くと類似</u> 。 頭を動かし、その視線方向へ手を伸ばす。
06数える	両手を叩き、数を数える。(ダミー) 右手を振り、視線の先に添えた左手をゆっくり数度叩く。
07襟	両手で襟をしごく。 右手で襟をしごいた後、左手、両手と順にしごく。
08懐手	袖を振って、両手を隠す。 右手で袖を隠した後、両手を隠す。
09髪直し	両手で髪を交互に直す。(ダミー) 頭の後ろにゆっくり手を添え、両手で髪を直す。
10ねじ	素早く立ち上がり、強く向き直る。重心移動を伴う。 右手と体をねじりながら、立ち上がり、ポーズをとる。

3.2.3 実験手続き

判別者は、用意されたノート PC を通じて、1 名ずつ、以下の手順を繰り返し実施した (図 3-3)。

①2 秒間の注視点提示の後、左遣い視点の動画 (型の冒頭部分) が提示された (図 3-3 中央)。

②型の候補 10 種類が画面に提示され、①で提示された型が何であるかについての確信度を、合計が 100%になるようにそれらに割り振った (図 3-3 右下)。

型 8 種類と提示時間 4 種類からなる 32 回を 1 セットとし、同じ条件で録画した異なる動画を用いた 3 セット (計 96 試行) を提示した。各セット内で、型と提示時間の順番はランダム化した。なお、確信度の割り振りは 0%~100%の 5%刻みの 21 段階とした。

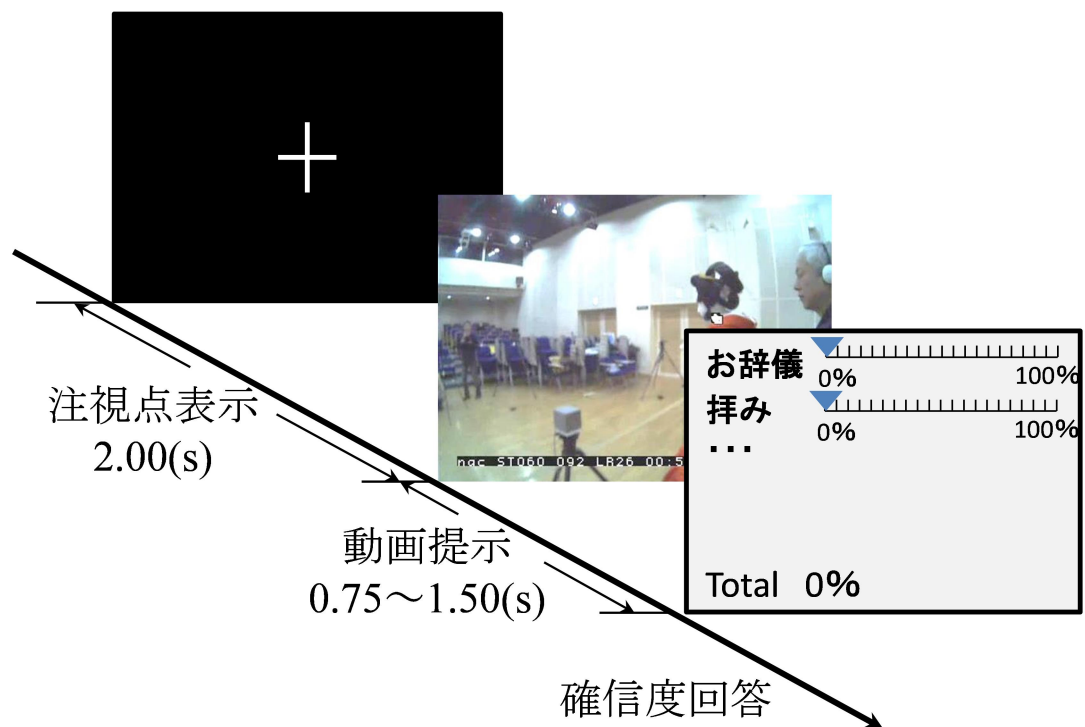


図 3-1 実験手続き

3.3 実験結果

3.3.1 熟達度，提示時間，型の種類の正解確信度への影響

判別者の熟達度，型の提示時間，型の種類の3要因が型判別の正解確信度に与える影響を分析するために，3要因を独立変数，正解確信度を従属変数，判別者を変量効果とする一般化線形混合モデルによる分析を行った。

その結果，型の提示時間，型の種類の主効果は有意であったが，熟達度の主効果と2次の交互作用に有意差は認められなかった（熟達度： $F(1, 9) = 2.46, p > .1$ ，型の提示時間： $F(3, 27) = 37.52, p < .001$ ，型の種類： $F(7, 63) = 9.28, p < .001$ ，2次の交互作用： $F(21, 189) = 0.675, p > .1$ ）。熟達度×型の種類，ならびに型の提示時間×型の種類の1次の交互作用は有意であった（熟達度×型の種類： $F(7, 63) = 2.78, p < .05$ ，型の提示時間×型の種類： $F(21, 189) = 4.38, p < .001$ ）が，熟達度×型の提示時間の1次の交互作用は有意ではなかった（ $F(3, 27) = 0.32, p > .1$ ）。

図3-2に型の種類の正解確信度の平均値を熟達度ごとに示す。熟達度×型の種類の交互作用における単純主効果の検定を行った結果，型の種類が“お辞儀”における熟達度の効果が有意であった（ $F(1, 9) = 5.47, p < .01$ ）。熟達度の主効果が有意ではないので強い主張はできないが，型の種類によっては熟達度による違いが見られたため，仮説1は部分的に支持されたと言えよう。なお，熟達群および準熟達群における型の種類の単純主効果は準熟達群のみ有意であった（熟達群： $F(7, 28) = 1.82, p > .1$ ，準熟達群： $F(7, 35) = 14.79, p < .001$ ）。これらの結果と図3-2から，準熟達群にとって判別の難しい型（“お辞儀”）において，熟達により正解確信度が上昇するものと考えられる。また，型の提示時間の主効果が有意だったので，仮説2も支持されたと言える。

図3-3に型の種類の正解確信度の平均値を提示時間ごとに示す。型の提示時間×型の種類の交互作用における単純主効果の検定を行った結果，型の種類が“立ち上がり”，“襟”，“懷手”における提示時間の効果が有意であった（“立ち上がり”： $F(3, 27) = 11.24, p < .001$ ，“襟”： $F(3, 27) = 26.92, p < .001$ ，“懷手”： $F(3, 27) = 38.92, p < .001$ ）。さらに，提示時間0.75(s), 1.00(s)において型の種類の効果が有意傾向，1.25(s), 1.50(s)において有意だった（0.75(s)： $F(7, 63) = 2.16, p < .1$ ，1.00(s)： $F(7, 63) = 2.00, p < .1$ ，1.25(s)： $F(7, 63) = 8.74, p < .001$ ，1.50(s)： $F(7, 63) = 16.73, p < .001$ ）。したがって，仮説3も支持されたと言える。図3-3を見ると，“立ち上がり”，“襟”，“懷手”などの複数の型において1.00～1.25(s)の間に正解確信度が大きく伸びていると考えられる。そこで，1.00(s)

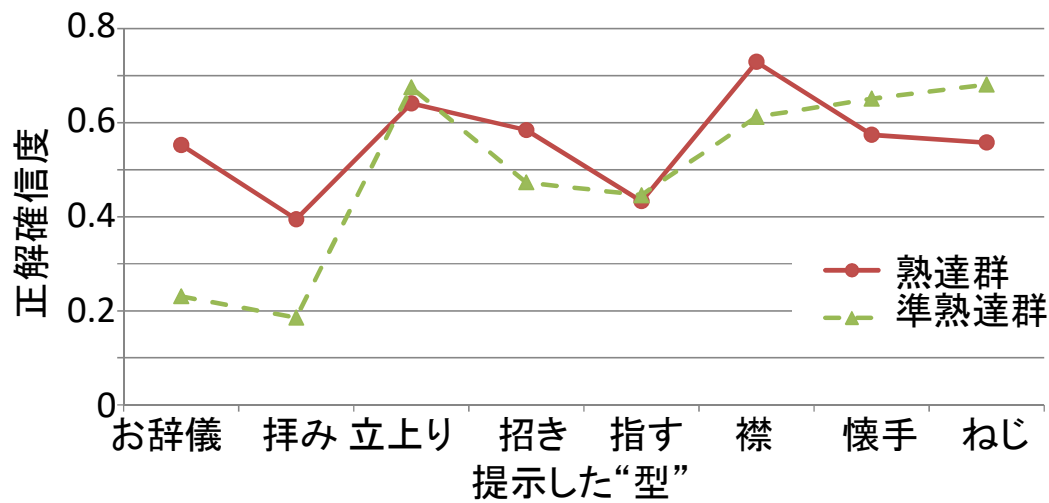


図 3-2 熟達度ごと、型の種類ごとの正解確信度の平均値

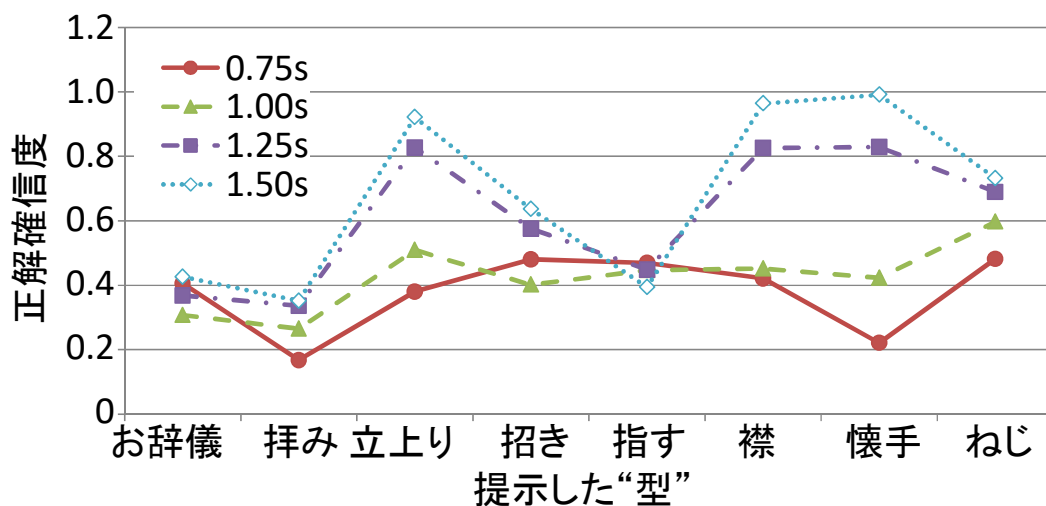


図 3-3 提示時間ごと、型の種類ごとの正解確信度の平均値

を対照群とした Dunnet 法による比較を行った。その結果，“立ち上がり”において 1.00～1.50(s)間に有意傾向($p < .1$)，“襟”において 1.00～1.25(s)間に有意傾向($p < .1$)，1.00～1.50(s)間に有意差($p < .01$)，“懐手”においては 0.75～1.00(s)間に有意傾向($p < .1$)，1.00～1.25(s)間および 1.00～1.50(s)間に有意差がみられた(1.00～1.25(s) : $p < 0.01$, 1.00～1.50(s) : $p < 0.001$)。したがって，提示時間が進むにつれ，複数の型において時間の影響が発生し，特に“襟”と“懐手”では 1.00～1.25(s)の影響が大きく，“懐手”では，0.75～1.00(s)の影響も大きいことが判明した。“立ち上がり”では 1.00～1.25(s)の間では有意差はなかったが，1.00～1.50(s)の間に正解確信度が大きく伸びることがわかった。

以上により，3 要因はいずれも型判別に重要な要素であり，仮説 1～3 はおおむね支

持されたと言える。

3.3.2 提示時間ごとの確信度のクラスタ分析

前節では、8つの型のうち“お辞儀”において熟達度による差がみられ、他の型動作に比べて重心移動や特徴的な動作をする複数の型（“立ち上がり”や“懐手”など）において提示時間の影響が大きいなど、型の種類やその動作特徴が重要と考えられる結果が得られた。しかし、その判別プロセスが図 2-4 に示したように型の動作特徴にもとづいて進行しているかどうかはそれだけでは判断できない。そこで、図 2-4 に示すような似た動作特徴をもつ（似た動作をする）型のグループが、途中まで判別に迷う（確信度を同時に割り振る）型のグループとして存在するかどうかを検討するために、確信度の全体の分布に対してクラスタ分析を行った。

分析のため、表 3-2 に各提示時間におけるそれぞれの型に対する確信度（全参加者平均）を行列表示した。

例えば、“お辞儀”の映像を 0.75(s) 提示した時に“拝み”だと回答した割合は、1 行 2 列目に表示されている（14.8%）。太字で示された対角成分が、映像と回答が一致した正解確信度となる。この行列を利用することにより、正解確信度のみの分析ではわからない、各提示時間でどの型を候補としていたのかが明らかにできる。例えば、0.75(s) 時点の 4-5 行 4-6 列（中央の四角で囲った部分）を見ると、“招き”と“指す”の動画を提示した際には、“招き”，“指す”，“数える”のいずれかであるという回答が多く、早い段階でこの 3 つのいずれかに絞り込まれているが、1.50(s) 時点でも“招き”と“指す”のどちらであるかを判別しきれていないことがわかる。“招き”と“指す”は似た動作をする型であり、左遣いが受け取る“ず”も似たものである可能性が考えられる。

そこで、実際に判別に迷った型のクラスタがどのようなものであるか、類似する複数の型の組み合わせと関連しているかどうかを分析するため、提示時間ごとのデンドログラムを分類感度に優れる Ward 法で作成した（図 3-4）。上述した“招き”と“指す”をはじめとして、判別に迷った結果、似た確信度分布が回答された型候補同士がクラスタとして視覚化されている。もし、このクラスタが表 1 に示す似た動作の型の組み合わせと一致し、時間とともにクラスタ内での距離が離れていくなれば、図 2-4 で示したように型が判別されている可能性が高いと考えられる。

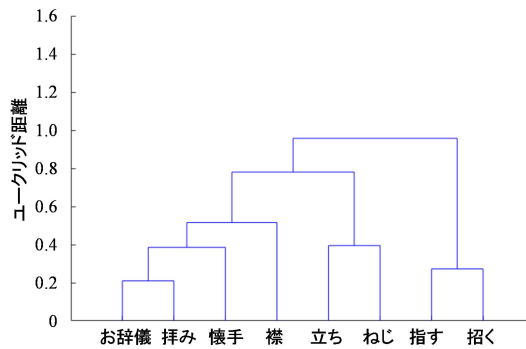
表 3-2 各提示時間での各型の確信度の平均値(%)

提示した型	提示時間 0.75(s)での 回答確信度の平均値(%)									
	01お辞	02拝み	03立ち	04招き	05指す	06数え	07襟	08懐手	09髪直	10ねじ
01お辞儀	40.6	14.8	7.9	0.3	1.8	5.3	14.4	3.0	11.5	0.3
02拝み	24.1	16.8	8.6	2.4	8.8	24.8	4.1	2.7	5.5	2.1
03立ち上がり	0.0	3.0	38.0	1.4	23.0	13.2	0.6	2.3	1.5	17.0
04招き	0.0	0.0	0.0	48.0	39.8	11.1	0.0	0.0	0.0	1.1
05指す	0.0	0.0	0.0	21.8	47.0	26.4	1.8	0.0	3.0	0.0
06数える	—	—	—	—	—	—	—	—	—	—
07襟	1.1	17.7	6.7	0.6	5.2	11.4	42.1	2.1	12.6	0.6
08懐手	7.6	12.6	8.2	10.0	9.8	7.3	5.6	22.1	12.6	4.2
09髪直し	—	—	—	—	—	—	—	—	—	—
10ねじ	4.8	3.3	26.4	3.5	2.6	9.5	0.3	0.3	1.1	48.2

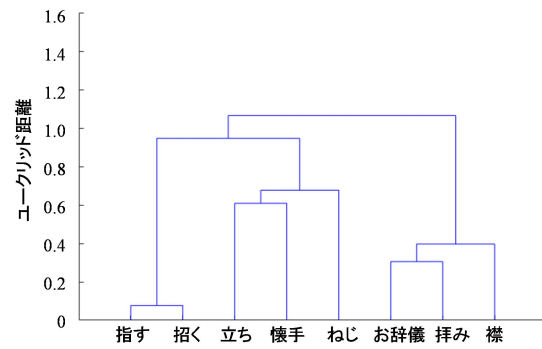
提示した型	提示時間 1.00(s)での 回答確信度の平均値(%)									
	01お辞	02拝み	03立ち	04招き	05指す	06数え	07襟	08懐手	09髪直	10ねじ
01お辞儀	30.8	17.9	2.9	0.6	1.4	3.2	38.8	1.7	2.1	0.8
02拝み	27.0	20.5	1.8	0.9	0.9	21.5	10.0	5.3	4.5	1.5
03立ち上がり	1.5	0.0	50.9	1.5	17.6	15.9	2.3	1.5	1.4	7.4
04招き	0.0	0.0	0.0	40.2	45.8	14.1	0.0	0.0	0.0	0.0
05指す	0.3	0.3	0.3	32.6	44.7	20.6	0.3	0.3	0.3	0.3
06数える	—	—	—	—	—	—	—	—	—	—
07襟	3.3	31.5	0.3	0.3	0.3	7.0	45.2	1.2	10.6	0.3
08懐手	5.5	5.2	8.5	2.4	6.4	4.8	10.6	42.3	8.0	6.4
09髪直し	—	—	—	—	—	—	—	—	—	—
10ねじ	0.3	0.3	17.6	2.1	13.9	4.5	0.3	0.9	0.3	59.7

提示した型	提示時間 1.25(s)での 回答確信度の平均値(%)									
	01お辞	02拝み	03立ち	04招き	05指す	06数え	07襟	08懐手	09髪直	10ねじ
01お辞儀	36.8	16.5	0.0	0.0	0.0	0.0	43.8	0.5	2.4	0.0
02拝み	23.2	33.6	3.5	0.3	0.3	14.7	12.1	0.8	11.2	0.3
03立ち上がり	0.6	0.0	82.7	0.0	0.0	2.1	0.6	2.1	6.1	5.8
04招き	0.0	0.0	3.0	57.6	34.8	4.5	0.0	0.0	0.0	0.0
05指す	3.3	0.3	3.3	35.3	44.8	8.9	0.3	0.3	3.0	0.3
06数える	—	—	—	—	—	—	—	—	—	—
07襟	1.5	10.3	0.0	0.0	0.0	0.0	82.6	0.9	4.7	0.0
08懐手	0.0	0.0	6.1	0.0	2.3	3.2	4.1	82.9	0.0	1.5
09髪直し	—	—	—	—	—	—	—	—	—	—
10ねじ	0.0	0.0	14.4	4.2	6.1	4.1	0.0	0.0	2.3	68.9

提示した型	提示時間 1.50(s)での 回答確信度の平均値(%)									
	01お辞	02拝み	03立ち	04招き	05指す	06数え	07襟	08懐手	09髪直	10ねじ
01お辞儀	42.6	10.6	0.0	0.0	0.0	0.0	43.8	0.0	3.0	0.0
02拝み	20.0	35.2	0.6	0.6	1.8	11.2	16.5	0.6	12.9	0.6
03立ち上がり	0.0	4.5	92.1	0.0	0.0	1.2	0.0	0.0	2.1	0.0
04招き	0.0	0.0	0.0	63.8	33.3	3.0	0.0	0.0	0.0	0.0
05指す	0.0	0.0	0.0	53.0	39.4	7.6	0.0	0.0	0.0	0.0
06数える	—	—	—	—	—	—	—	—	—	—
07襟	0.0	0.6	0.0	0.0	0.0	0.0	96.4	3.0	0.0	0.0
08懐手	0.0	0.0	0.0	0.0	0.0	0.0	0.3	99.1	0.6	0.0
09髪直し	—	—	—	—	—	—	—	—	—	—
10ねじ	3.0	0.0	9.8	1.2	5.2	4.5	3.0	0.0	0.0	73.2



(a) 提示時間 0.75(s)



(b) 提示時間 1.00(s)

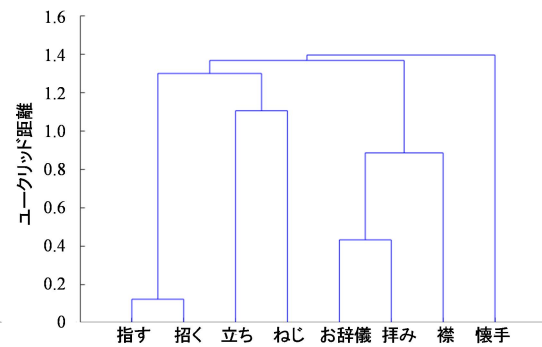
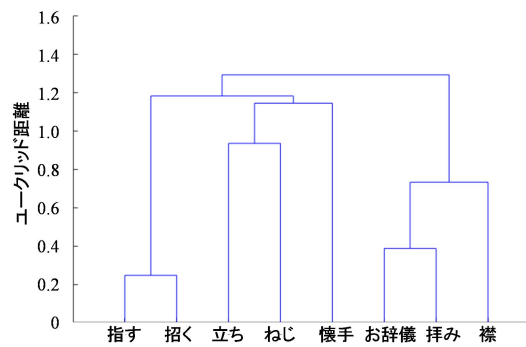


図 3-4 提示時間ごとのデンドログラム

まず 0.75(s)の結果を見ると、組み合わせによって距離はまちまちではあるが、この時点で既に表 1 で示した似た動作の型の組み合わせと一致するクラスタ（両手を動かす“お辞儀”と“拝み”，左手を突き出す“指す”と“招く”，重心を大きく移動する“立ち上がり”と“ねじ”）が形成されていることがわかる。また，提示時間が増えるにつれて，各クラスタの距離と各型の距離が増え，クラスタ内の型の判別が進行していることがわかる。ただし，両手を使って特徴的な動作を行う“襟”や“懷手”については，特定のクラスタに属さず，時間ごとに他のクラスタに近づいたり離れたりしながら，最終的に明確に区別されることが明らかとなった。

以上から，似た動作を行う型のクラスタは，動作判別に迷う型のそれとほぼ一致し，提示時間とともに似た動作どうしの判別が可能になることがわかった。しかし，特徴的な動作をする型である“襟”や“懷手”の存在が，他の型の判別を惑わしていることも示唆された。例えば，表 2 の“お辞儀”と“襟”の関係をみると，“お辞儀”が正解の時（1 行目）は“襟”と区別がつかず正解確信度が低くなるが，逆に“襟”が正解の時（7 行目）は，“お辞儀”と間違えることがなく高い正解確信度を得ている。他にも，提示時間が 1.50(s)になっても正解確信度が 100%にならない点が図 2-2 と異なる。これらの原因については 5 章で考察する。

3.3.3 各提示時間における確信度の主成分分析

クラスタ分析によって図 2-4 のように似た動作候補から回答を絞っていく様子は明らかになったが，各時間でどのような動作特徴（似ている部分）を重視しているのかは明らかになっていない。そこで，確信度に対して主成分分析により次元圧縮を行うことで，時間ごとに注目している動作特徴の抽出を試みた。もし左遣いが動作の特徴を利用して型を絞り込んでいるのであれば，縮約された主成分は，似ている部分，すなわち共通する動作特徴を示す可能性が高い。時間ごとにこの主成分係数を比較することで，各時間でどのような動作特徴が注目されているのかを明らかにできる。

表 3-3 に提示時間 0.75(s)での主成分分析の結果を示す。上から順に確信度の主成分係数と固有ベクトル，寄与率が主成分ごとに記されている。すべての提示時間において，第 4 主成分までで 90%程度の寄与率が得られていることから，今回提示した 8 つの型の回答は 4 つの特徴（説明変数）で説明できると言える。

表 3-3 提示時間 0.75 秒での確信度の主成分分析

回答項目	提示時間0.75sでの主成分係数					
	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分
01お辞儀	-0.294	-0.296	0.732	0.003	-0.264	0.230
02拝み	-0.207	-0.218	-0.026	-0.008	0.066	-0.152
03立ち上がり	-0.210	0.474	-0.056	-0.337	0.239	0.587
04招き	0.561	-0.048	0.078	0.584	-0.073	0.188
05指す	0.625	0.031	-0.026	-0.388	-0.028	0.140
06数える	0.118	-0.020	0.038	-0.491	-0.239	-0.562
07襟	-0.232	-0.370	-0.669	0.073	-0.319	0.201
08懐手	-0.055	-0.063	0.020	0.161	0.748	-0.277
09髪直し	-0.114	-0.172	-0.068	0.057	0.195	-0.059
10ねじ	-0.193	0.683	-0.023	0.345	-0.325	-0.297
固有ベクトル	0.073	0.048	0.020	0.012	0.008	0.005
寄与率	0.436	0.289	0.120	0.070	0.046	0.030

表 3-4 各提示時間における主成分係数

主成分係数 回答項目	提示時間0.75s		提示時間1.00s		提示時間1.25s		提示時間1.50s	
	第1主成分	第2主成分	第1主成分	第2主成分	第1主成分	第2主成分	第1主成分	第2主成分
01お辞儀	-0.294	-0.296	-0.274	-0.101	0.206	0.033	0.176	-0.029
02拝み	-0.207	-0.218	-0.365	-0.159	0.203	0.016	0.078	-0.022
03立ち上がり	-0.210	0.474	0.107	0.475	-0.345	-0.502	-0.200	-0.104
04招き	0.561	-0.048	0.414	-0.385	-0.285	0.561	-0.383	-0.357
05指す	0.625	0.031	0.562	-0.263	-0.247	0.449	-0.240	-0.225
06数える	0.118	-0.020	0.104	-0.110	-0.034	0.044	-0.027	-0.040
07襟	-0.232	-0.370	-0.500	-0.195	0.775	0.021	0.825	-0.058
08懐手	-0.055	-0.063	-0.085	0.073	-0.093	-0.454	-0.181	0.895
09髪直し	-0.114	-0.172	-0.092	-0.018	0.028	-0.019	0.022	-0.003
10ねじ	-0.193	0.683	0.130	0.683	-0.208	-0.149	-0.070	-0.058
寄与率	0.436	0.289	0.424	0.273	0.308	0.240	0.284	0.252
重視する 動作特徴	左手動作 重心移動		左手動作 重心移動 両手動作		襟動作 左手動作 重心移動		襟動作 懐手動作	

また、それぞれの主成分を構成する主成分係数を見ると、前節で述べた似た動作の型の組み合わせが高い値を示していた。例えば、提示時間 0.75(s)での第一主成分では、左手を突き出す（人形の右手を動かさず左手のみを遣う）“招く”と“指す”の係数が高い絶対値を示している。これは、人形遣いが 0.75(s)時点で、左手を突き出す動作であるかどうかを回答に強く反映させていることを意味する。各提示時間において高い絶対値を示す主成分係数をまとめると表 3-4 のようになる。表の下部に高い絶対値を示す型の動作特徴をまとめた。この結果から左遣いはまず左手を突き出すかどうか、重心を移動させるかどうかをごく短時間の間に判断し、次に、両手を動かす動作や、特徴的な動作についての判断を順次行っている可能性が高い。

この順序は、それぞれの型において左手をどのくらい早く動かさなくてはならないか

に関連していると考えられるが、詳細は5章で考察する。

3.4 考察

3.4.1 各要因の影響について（一般化線形混合モデルによる分析）

一般化線形混合モデルによる分析の結果、明らかとなった左遣いの型判別の特徴は、以下のとおりである。

- ・準熟達群にとって判別が困難な特定の型では、熟達により判別の正解確信度は上昇する。
- ・特定の型では1.00～1.25(s)の間に正解確信度が大きく上昇する。

前者については、2.5.1 節での分析結果では、1つの特定の型（“お辞儀”）でしか有意差が検出されなかった。この原因については以下の通りだと推察される。準熟達群も、熟達群に比べて短いものの修行を行った現役の舞台で活躍する人形遣いであり、基本動作は把握している。そのため、今回の実験における型の判別だけであれば、多くの型を熟達群と同程度に判別可能であったと考えられる。そのため、熟達群と準熟達群の差がさほど明確には明らかにならなかったと考えられる。しかし、2.3.1 節でも述べた「足遣い10年、左遣い10年」[33]という芸談や、「ズは個人によってかなり変化や程度の差があり、従って各人のズの出し方をすべて憶え、それに的確に対応するには相当の期間を必要とする。」[3]と言われていることを鑑みると、“お辞儀”の熟達差は偶然発生したものではなく、経験による主遣いの癖の把握の熟達差などを反映したものだと考えられる。このように左遣いの型判別の精度は修行によって増すと考えられる。さらに、型判別は時間的に進行し、型の種類によって難易度や必要な時間が異なると言える。また、合図とされる“ず”は、型の種類を100%同定できるような明示的な信号ではないと考えられる。

3.4.2 動作特徴の利用について（クラスタ分析）

3.3.2 節では、判別に迷う型（確信度を同時に割り振る型）の組み合わせが同じ動作特徴を持つ型（動作が似ている型）の組み合わせと一致するかどうかを調べるために、各提示時間における確信度に対してクラスタ分析を行った。その結果、以下が明らかとなった。

- ・瞬間的な提示（0.75(s)）の時点で、同じ動作特徴を持つ型の組み合わせと判別に迷う型のクラスタがある程度一致する。

- ・時間の経過とともに、クラスタ内での判別が進む。
- ・他とあまり類似しない固有な型動作は、1.00～1.25(s)時点までは、他のクラスタの型動作と混同される傾向にある。
- ・1.50(s)時点では、上記の固有な型も判別され、比較的明確な分類が成立する。

以上の結果から、左遣いは人形の動きを見て、どのような動作特徴を持つ型かを素早く判別し、そこから似た動作のうちのいずれであるかを絞り込むようなプロセスで判断していると推定できる。ただし、“襟”などの固有な動きを持つ型は、ある程度時間がたないと判別が困難だと推測される。

3.4.3 重要視される特徴について（主成分分析）

3.3.3 節では、左遣いが各提示時間においてどのような動作特徴に着目しているのかを検討するために、各提示時間における確信度の主成分分析を行った。その結果、左遣いの確信度は、いくつかの主成分に縮約されたが、その主成分係数は提示時間によって大きく異なった。その推移から、左遣いは、①動作のごく初期（0.75(s)）では、人形の左手を突き出す“指す”と“招く”の可能性を最も高く見積り、②動作の初期（1.00(s)）では、重心移動を伴う“立ち上がり”や“ねじ”の可能性も考慮し、③動き出し（1.25(s)）てからは、特徴的な動作である“襟”などに注意を払いつつ、人形の左手を操作していると推測される。

実際に人形の動作を観察すると、人形の左手を突き出す型や重心移動を伴う型は、他の型と比較して左手の動き出しが早い傾向にあり、上記手順と合致すると言える。その一方で、正解確信度はすぐには100%に収束せず、左遣いは長い時間、動作判別に注意し続けなければならない。この点が、“ず”の理解に長い修行を要する一因だと考えられるが、この理由については5章で考察する。

3.5 まとめ

本章では、左遣いの“ず”による型判別の仕組みについて明らかにするため、重要と考えられる3要因を統制し、型判別を再現する実験を行った。

まず、実験結果の一般化線形混合モデルによる分析により、以下が明らかになった。

- ・特定の型では、熟達により判別の正解確信度が上昇する。
- ・特定の型では、1.00～1.25(s)の間に正解確信度が大きく上昇する。

次に、クラスタ分析により、以下が明らかになった。

- ・瞬間的な提示 (0.75(s)) の時点で、同じ動作特徴を持つ型の組み合わせと判別に迷う型のクラスタがある程度一致する.
- ・時間の経過とともに、クラスタ内での判別が進む.
- ・他とあまり類似しない固有な型動作は、1.00～1.25(s)時点までは、他のクラスタの型動作と混同される傾向にある.
- ・1.50(s)時点では、上記の固有な型も判別され、比較的明確な分類が成立する.

そして、主成分分析の結果から、左遣いは、①動作のごく初期 (0.75(s)) では、人形の左手を突き出す“指す”と“招く”の可能性を最も高く見積り、②動作の初期 (1.00(s)) では、重心移動を伴う“立ち上がり”や“ねじ”の可能性も考慮し、③動き出し (1.25(s)) からは、特徴的な動作である“襟”などに注意を払いつつ、人形の左手を操作している可能性が示唆された.

この結果は、第2章で提示した判別方法の仮説と合致するものである. ただし、第2章で例示したように全ての型を効率的に分割し、短時間で型を100%特定するものではなかった. いくつかの型は素早く分類が行えるようになっている一方で、実験時の最大提示時間をもってしても不定性が残っていた. これは、今回の実験環境が実際の演技よりも厳しい状況にあったためと考えられる. 例えば、実際の文楽の演技では、PC画面よりかなり広い周辺視野も利用でき、画面上の注視点ではなく、待機状態の人形の頭を注視している状態から動作が始まる. 今回の実験上の制約については5章にて詳しく議論する.

以上、左遣いの型判別についての特性が一部明らかになった. 一方で、本実験では映像データから型判別を再現してもらったが、2章で述べたように人形の動作自体に型を分類する情報が含まれているかを証明するものではない. 次章では、そもそも主遣いが操る人形動作に、2章で述べたような分類が可能になるような情報が含まれているかどうかについて、実際の人形の動作データから隠れマルコフモデルで隠れ状態を推定することにより検証する.

第 4 章 人形動作からの型判別情報の抽出

第 2 章では、左遣いの判別は、型ごとに固有のサインがあるのではなく、人形動作の時系列から得られる動作特徴を示す情報の連続から、動作候補を絞り込むような仕組みであると想定し、仮説を提示した。

第 3 章での結果から、実際の判別が時間的に行われていることは支持された。先行研究と併せて考えると、それが、事前の打ち合わせや“床”ではなく、主遣いが出す“ず”によって行われていると考えられる。しかし、実際に人形の型動作の中にこのような絞り込みを実現するための情報が存在するかどうかはこれまで示されていない。また、少なくとも素人目には、このような“ず”がどこにあるのかを判別することは困難であり、“ず”が何であることを示す明文化されたマニュアルも存在しない。

そこで本章ではこのような「絞り込みを実現するような情報」が型動作に含まれるかどうかを機械学習によって検討する。

まず「絞り込みを実現する情報」について定義するため、第 2 章で述べた人形動作による情報伝達の仮説を簡単にまとめる。「絞り込みを実現する情報」は、行う型を指示するための“ず”とする。「“ず”によって型動作を指示する」というと、型単一を示す“ず”（を含んだ動作）があると考えるのが普通だが、後藤[3]によれば、『ズが人形遣いたちに共通している以上、これも一種の（型）であろうが、より細かな基本的で自然な動きに分解されているために、一般的な（型）のような固定化はおこなっていない。』と指摘しており、“ず”は単一の型を示すための単一の合図（動作）というより、動作要素（小さな動作）の時間的な組み合わせ（順序）から成り立っていると考えた方が妥当である。また、このような仕組みをとることで、主遣いが動作中に細かな速度変化や動作のアレンジをアドリブで行っても、左遣いが型の判別を行うことができるアドリブ

追従性の高いシステムになると考えられる．なお，“ず”は本来もっと広い意味での指示を示しているが，本論文では型の指定のための情報を“ず”と表記する．

すなわち，“ず”の中で型の種類を指示する情報とは，「連続した小さな動作要素の組み合わせ（順序）」だと考えられる．

このような時間的に連続した入力から，要素の組み合わせを推測する類似例として，音声認識が挙げられる．音声認識は，マイクで拾った音情報（時系列の波形）から，状態（50音やその元となる音素）を取り出し，その遷移（順序）でどんな単語を話そうとしているのかを推測することで実現しており，シンプルな隠れマルコフモデル

（Hidden Markov Model，以下 HMM と略記）の学習によってこの状態遷移を可視化できる[38]．もし，型が同様の情報を持っているのであれば，HMMによって状態の可視化（組み合わせによって型を決定できるような動作要素の可視化）を行えると考えられる．そこで，本章では第3章で用いた映像の収録時に磁気式モーションキャプチャによって計測した人形動作のデータから，HMMによって動作要素の組み合わせを可視化できるかどうかを検討することで，第2章で提示した左遣いの判別を可能とする情報が，主遣いが操る人形動作に含まれているか検討する．

実際には左遣いは動作分類だけでなく，人形の姿勢維持やイメージを共有した動作出力など多くのタスクを行う必要があるが，ここでは時間的な型動作の判別のみに焦点を当てる．

4.1 使用する隠れマルコフモデルとパラメータ設定

先述した通り、第2章で述べた左遣いの時間的な判別の仕組みは、音声認識の仕組みに類似している。時系列の音声データから、音素の組み合わせを抽出するように、人形の多次元の時系列動作データから、動作要素の組み合わせを取り出せると期待できる。そこで、HMMモデルによって、多次元動作の入力から1次元の状態遷移を隠れ状態として得ることを本章の目的とする。

音声の分析との違いとして、入力が多次元であることと、型のマニュアルがないことから、得られた隠れ状態が何であるかの判断が難しい点が挙げられるが、ここでは「型の分類ができる程度の動作要素の組み合わせが抽出できるかどうか」を主眼とし、型ごとに固有の動作要素の組み合わせが抽出されるかどうかをまずは観察する。この動作要素の組み合わせの内容と型との相関については、その後考察を行う。

注意すべきは、この解析で得られる動作要素の組み合わせは、入力データである実験用に演技してもらった10種類の型に対応した情報であり、左遣いが暗黙的に使用している“ず”とは必ずしも一致しない可能性がある点である。実際には、同定された動作要素のほんの一部分だけが“ず”であるという可能性もあるし、今回抽出されなかった部分が10種類以外の型を分類するために重要な要素かもしれない。また、人形遣いごとに“ず”の出し方は異なるとも言われ[3]、マニュアルが存在しないことから、“ず”の判別はより複雑であると推察される。そこで本章では、あくまで人形動作中に時間的な型分類を実現する情報（動作要素）が含まれているかどうかを検討することを目的とし、単純に型の種類と動作要素の組み合わせが一致するかどうかを評価基準とした。

この目的に沿って、本章で使用する隠れマルコフモデルと各パラメータを以下の手順で決定した。

人形動作から、動作要素の組み合わせの抽出までをHMMモデルに当てはめた。入力が多次元の連続値であることから、混合ガウス分布モデル（GMM）を用いた。GMM-HMMは音声認識などの分野においてよく用いられ、今回の文脈的な意味を伴った動作波形に対しても有効に機能すると考えられる[39]。有限の状態推移から、ガウス分布の重ね合わせとして、動作が出力されていると仮定して、学習を行った。この際のモデル構成は、①状態の有限集合、②出力の集合、③状態遷移確率分布、④出力確率分布、⑤初期状態確率分布である。本章の目的は、多次元の時系列人形動作から、“ず”を含んだ動作要素の状態遷移を得ることなので、人形の動作データのみを入力することが望ましい。ただし、状態数（動作要素の組み合わせを構成する動作要素の種類）につ

いては注意して設定する必要がある。ノンパラメトリックベイズ法などを採用した階層ディリクレ過程隠れマルコフモデル（HDP-HMM）などを用いれば、状態数も自動で決定（無限の状態から削減）されるが、今回は動作要素の組み合わせから最終的に見分けた10種類の型が明確に決まっているので、GMM-HMMによる推定の他に、複数の状態数を設定した分析を個別に行い、検証データ分類時の ELBO (evidence lower bound) と、型ごとに得られた状態の再現率（レーベンシュタイン距離）によって適切な状態数を評価することとした。ELBO が十分収束し、型ごとの再現率が良好であれば、得られた状態遷移が型ごとの特徴を良く表現していると期待できる。検証データやレーベンシュタイン距離などの詳細は後述する。実際の左遣いは、公演中の型の動き出しの傾向や、型の選択傾向などから、初期状態確率分布を持っているとも考えられ、芸談からある程度 HMM パラメータの特徴についても予想可能と考えられるが、ここではあくまで得られた計測動作データのみから学習させ、実際の人形遣いとの違いは5章で議論の対象とする。

次に入力データの詳細を説明する。入力データとして、第3章で使用した映像データと同時に収録した人形各部位の動作データを与えた。人形の動作データは、正規の文楽人形と同等のスペックをもつ計測用文楽人形の各部位に内蔵された磁気式モーションキャプチャによって計測された120Hzの6軸（X, Y, Z 軸方向の変位と回転）情報から構成される。人形の動作データのうち、先行研究から[15]、左遣いの位置から目視で観測可能と考えられる剛体部位4か所、すなわち人形の頭部、首部、肩部、右手の時系列動作情報を入力データとした（図4-1左）。X, Y, Z 各軸の方向（世界座標系）は図4-1右の通りとなっている。この半球内のデータを取得可能である。センサの埋め込み位置の詳細は付録A.3節およびA.4節に記載する。ただし、ロール、ピッチ、ヨー³の各回転軸については、埋め込まれたセンサを中心とした部材座標系の角度であるため、これらとX, Y, Z 軸とは必ずしも一致していない。しかしながら、本論文の目的は、先述した“ず”の構成要素にあたる状態を音声認識でいう音素のように観測することであるため、入力データの部位座標系の軸方向の定義が世界座標系と一致していなくとも大きな問題にはならない。またセンサから得られる情報では、初期姿勢とセンサ原点との位置が試行ごとに異なるため、速度と角速度を入力情報として使用した。3章の実験で準備された10の型の動作データをデータ・サンプルとした。データセットについては、デ

³ ロール、ピッチ、ヨーはそれぞれ部材座標系 x, y, z 軸周りの回転角を示す。本計測システムでは、世界座標系 X, Y, Z 軸方向の変位と、部材座標系 x, y, z 軸周りの回転角を取得する。

ータ数が少ないため、10 の型動作を 1 セットとして計測した 9 セットを、学習用の 7 セットと検証用の 2 セットに分け、 k -分割交差検証⁴ (*K-fold cross-validation*) を行うことにした。学習時は 70 の動作をまとめて状態推定、学習し、検証時は学習した HMM で 20 の動作をまとめて推定した。そして、すべての 36 通りの組み合わせについて分析を行い、型ごとの再現性を評価した。

評価方法については、試行ごとに動作の大きさや速度にばらつきがあり、人形動作の特性上正規化が難しいことから、隠れ状態の長さが可変のものとし、得られた状態遷移を連長圧縮し、状態の遷移（出現順序）のみを利用することとした。再現性の評価方法として、音声認識で一般的に用いられる単語の類似性の評価を参考に、型ごとに、訓練データから計算される隠れ状態の遷移と検証データから計算される隠れ状態の遷移との類似性を、レーベンシュタイン距離として計算した。これは、2 つの単語の片方をもう一方に修正するために必要なコスト（単語の修正や追加などの必要回数）を示し、同一の単語であれば 0、「ぶんらく」と「ぶんらくざ」であれば 1、「にんぎょうし」と「にんきょう」であれば追加と修正で $1+1=2$ となる。利用例としては、方言の分類[40]などがあげられる。この距離を両者の長い方で正規化した値を類似度（0 に近いほど類似）として利用した。先述した 36 通りの学習で得られた HMM での検証結果に対して、このレーベンシュタイン距離を求めて、型の再現性として評価した。もし、過学習なく“ず”の抽出ができていれば、高い再現率が期待できる。また状態数の決定においては、この 36 通りの学習のそれぞれについて、訓練データによる学習結果と検証データによる検証結果とのすべての組み合わせの間でもレーベンシュタイン距離を求め、状態数ごとに過学習が起きていないかどうか評価した。

HMM の学習については、先述した入力から動作要素の組み合わせとしての状態遷移を得るために、確率分布としてガウス分布を用いた HMM により EM アルゴリズム[41]でパラメータ推定を行った。EM アルゴリズムは反復法的一种であり、隠れ状態の確率分布と各パラメータ（ガウス分布の平均と分散など）の最尤推定を行う。どちらも一般的な手法であるが、両者を利用した例としては、視覚音声認識のための HMM のパラメータ推定法[42]などが挙げられる。そして、Viterbi アルゴリズム[43]により学習した状態に基づき、時系列の状態への記号化を行った。こちらも尤度に基づいた探索アルゴリズムで、音声認識などで一般的に用いられる手法である。

⁴ k -分割交差検証では、標本群を k 個に分割する。そのうちの 1 つをテスト事例とし、残る $k-1$ 個を訓練事例として、 k 回検証を行う。こうして得られた k 回の結果を平均して 1 つの推定を得る。

左遣いの型判断を完全に再現するには、「人形の時系列動作」から推測される、動作要素の組み合わせが、「主遣いが行いたい型」に依存するとし、多階層のマルコフモデルとした上で、型の種類に相当する状態の設定、および多くの演技サンプルが必要となる。しかし今回は型ごとに固有の動作要素の組み合わせが取り出せるかどうかを確認することが目的であり、左遣いの判断内容の再現までは行わない。今回 HMM 解析のため、Mike Hughes ら[44]によるライブラリ `bnpy` を使用した。

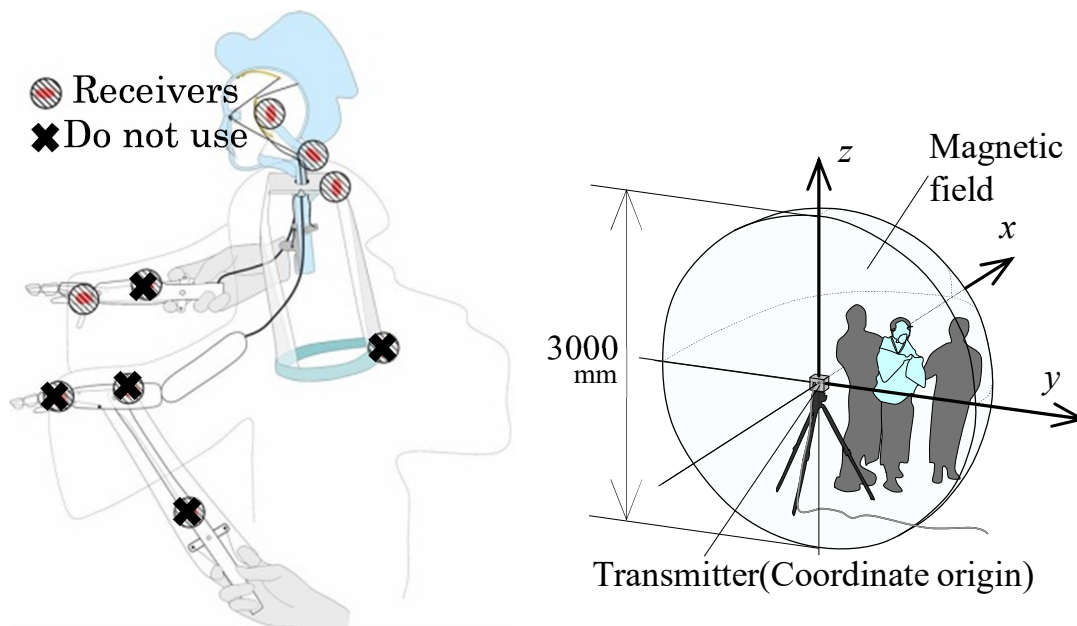


図 4-1 計測用文楽人形の座標系と採用する計測箇所

(左図が計測用文楽人形の計測可能箇所と本節で使用する箇所, 右図が計測システムの原点および座標系と計測範囲)

4.2 結果

4.2.1 学習と適切な状態数

HDP-DMM での学習と状態数を 6-12 の範囲で変更したときの ELBO を求め、比較した。各状態数での学習における検証データでの ELBO の損失関数 (ELBO loss) を検証データと訓練データの組み合わせ 36 通りで求めた際の平均と標準偏差を図 4-2 に示す。ELBO loss が低い値で収束している状態数が適していると考えられる。この図から、状態数 10 で最も低い値をとっていることがわかる。このことから、状態数 9~11 の中で各型と状態数の関係と型ごとの状態の再現性をみて評価を行うこととした。

次に、図 4-3 に状態数ごとの、訓練データによる推定と検証データによる推定の間のレーベンシュタイン距離による正規化距離の平均と標準偏差を示す。正規化は長いほうの文字列数で行った。レーベンシュタイン正規化距離も、状態数 9~10 で十分な収束が行われている。今回は少ない動作要素による分岐を想定していることから、状態数 9 を採用し、以降のデータは状態数 9 での計算結果を示す。

なお、図 4-5 から図 4-9 の評価は、検証用も含めた学習データ全体で行なった。

4.2.2 抽出された状態遷移と人形動作の関係

まず、図 4-4 に実際の動作波形と、得られた状態遷移の一例を示す。図では、頭と右手の Y, Z 座標の速度のみを示すが、実際には人形の 4 部位の 6 軸の速度と角速度、計 24 次元の動作データを入力としている。

図 4-4 の動作波形との対応をみると、人形の各部位の動き方と状態がおおまかに対応していることがわかる。例えば図の例 (“お辞儀”) だと、動作開始時の動きがない状態 (状態 5 (黄)) から、①人形の頭の軽い上昇、②その後の頭の下降と右手の右振り、③右手の引き寄せて持ち上げの一連の動作が、①状態 4 (橙)、②状態 2 (緑) ~ 状態 1 (青)、③状態 3 (紫) の遷移で表現されている。

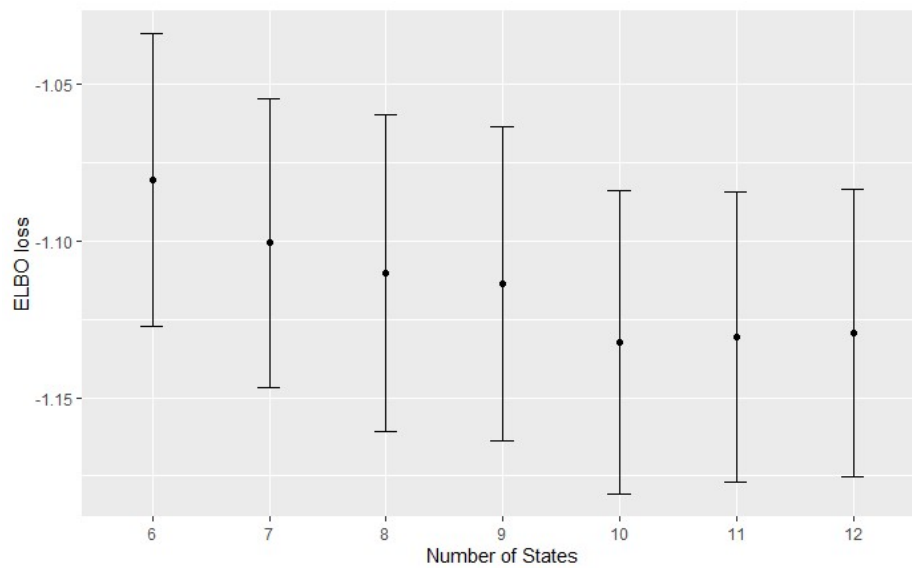


図 4-2 状態数 K に対する ELBO loss

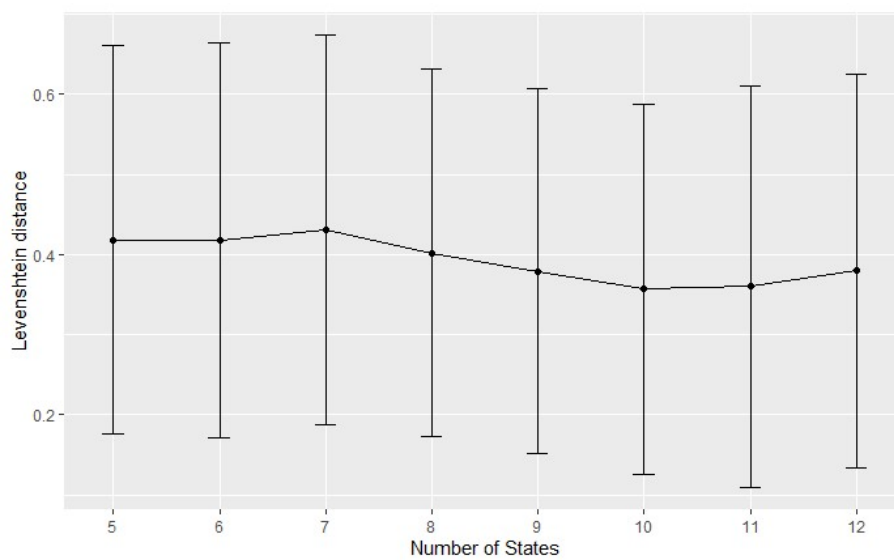


図 4-3 状態数 K に対するレーベンシュタイン正規化距離

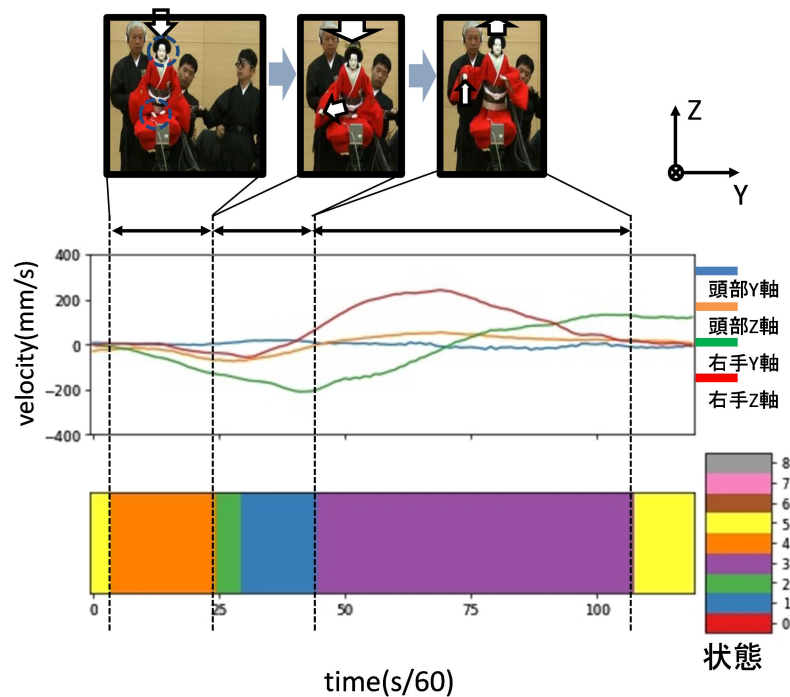


図 4-4 人形動作と状態遷移の関係

上から連続写真，人形の部位速度，HMM による状態遷移

図 4-5 にいくつかの型の動作波形と，動作要素の組み合わせを示す．他の型については付録 B.1 に示す．図では，上から順に“さす”動作の HMM 状態遷移，“さす”動作での人形の速度変化，“招く”動作での HMM 状態遷移，“招く”動作での人形速度変化を示している．各型がどのような動作をするかは，表 3-1 や研究室ホームページの連続写真[37]を参照されたい．複数の線が重なっているのは，9 回分（学習 7 回，検証 2 回）分のプロットを行っているためである．例えば，最上段のグラフでは，時間の経過（横軸）に対し，多くの線の縦軸の値が 0, 6, 8, 7 と順に推移していることがわかる．これが HMM によって得られた状態遷移である．2 段に分かれているのは，それぞれ検証データ，訓練データを示しているためである．その下の波形は図 4-4 の中段と同じく，人形頭部（かしら）と右手の Y, Z 軸方向の速度を示している．

演技中の人形の速度（下半分）について，同じ型動作であるため，9 回分の動作波形はある程度重なっているが，速度や大きさ，ならびにタイミングがデータごとにわずかに異なる（幅を持ったプロットに見える）．一方で，動作要素の組み合わせ（上半分）は，似たような遷移を行っている．

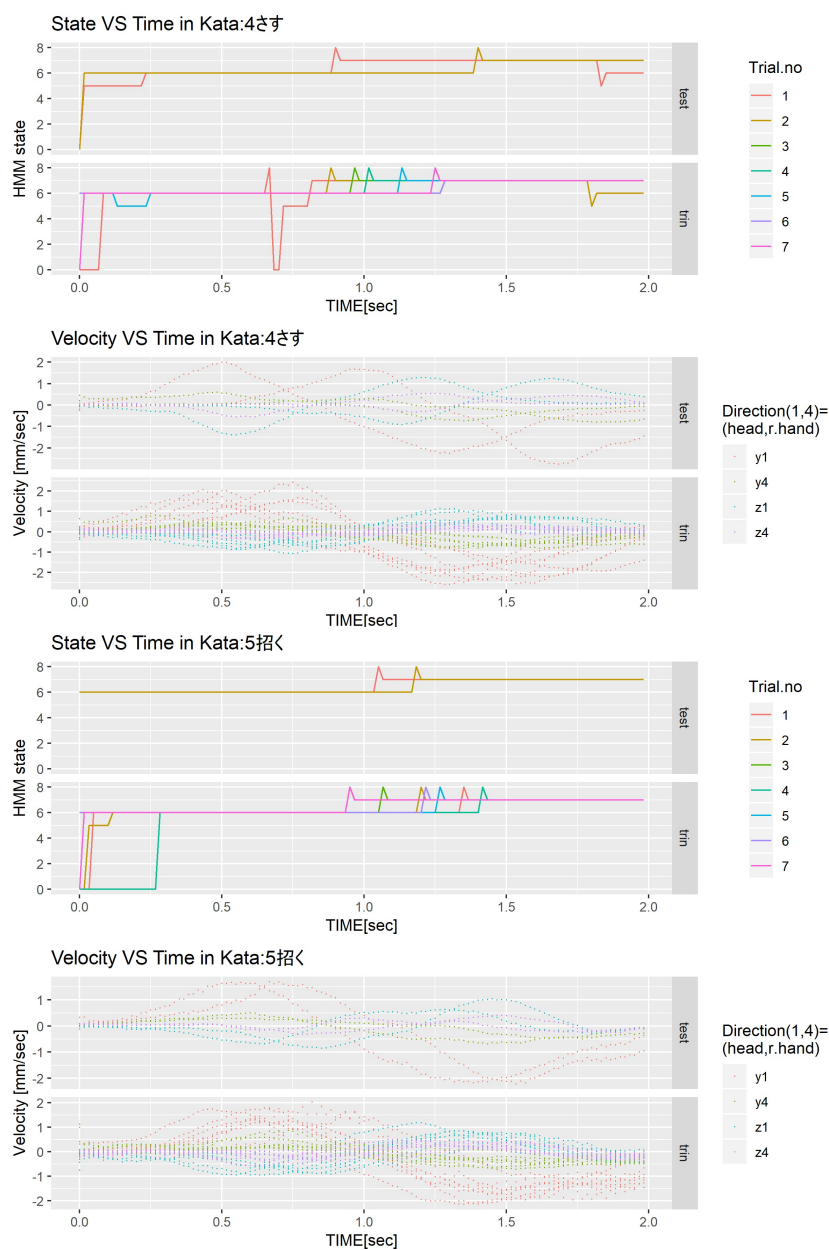


図 4-5 2 つの型での状態の遷移（上段）と動作波形の時間変化（下段）の例
 時間（横軸）に対して、上段の HMM State が HMM による状態の推移，下段の Velocity が人形の各部位の Y, Z 軸方向の移動速度を示す．人形の部位はそれぞれ頭部=1，右手=4 とし，オレンジ色（y1）は，頭部 Y 軸方向の移動速度を示す．

4.2.3 型と状態遷移の関係と連長圧縮

図 4-5 では、抽出された状態の推移（順序）は型ごとに類似していたが、試行ごとの動作のタイミングなどの違いにより、それぞれの長さやタイミングは異なっていた。一方で、型の種類の判別には、順序こそが重要であると考えられるため、この状態推移に対し、連長圧縮⁵（ランレングス圧縮）を行い、その繰り返し数を無視して出現順序のみを利用した。図 4-6 では、抽出された状態の遷移（図 4-5 での上半分）を連長圧縮し、その出現順序のみが、型ごとに示されている。例えば、「33331155」という状態の連続に対し、「315」という結果が時間長さを無視して示されている。ここでは 36 通りの学習のうちの一例のみを示す。上下段は、それぞれ検証データでの結果と訓練データでの結果を重ねて表示している。型ごとに固有の状態遷移が行われ、訓練データでの結果（各図の下半分）と、検証データでの結果（上半分）が同じ状態遷移を辿っていることがわかる。また、途中まで似た型動作である、“お辞儀”と“拝み”では途中まで似た遷移を行う。よって、これが型の判別を可能とする動作要素の組み合わせにあたる要素と考えられる。

⁵ 連長圧縮とは、連続して現れる符号を、符号とその繰り返しの回数を表す値に置き換える圧縮方式である。

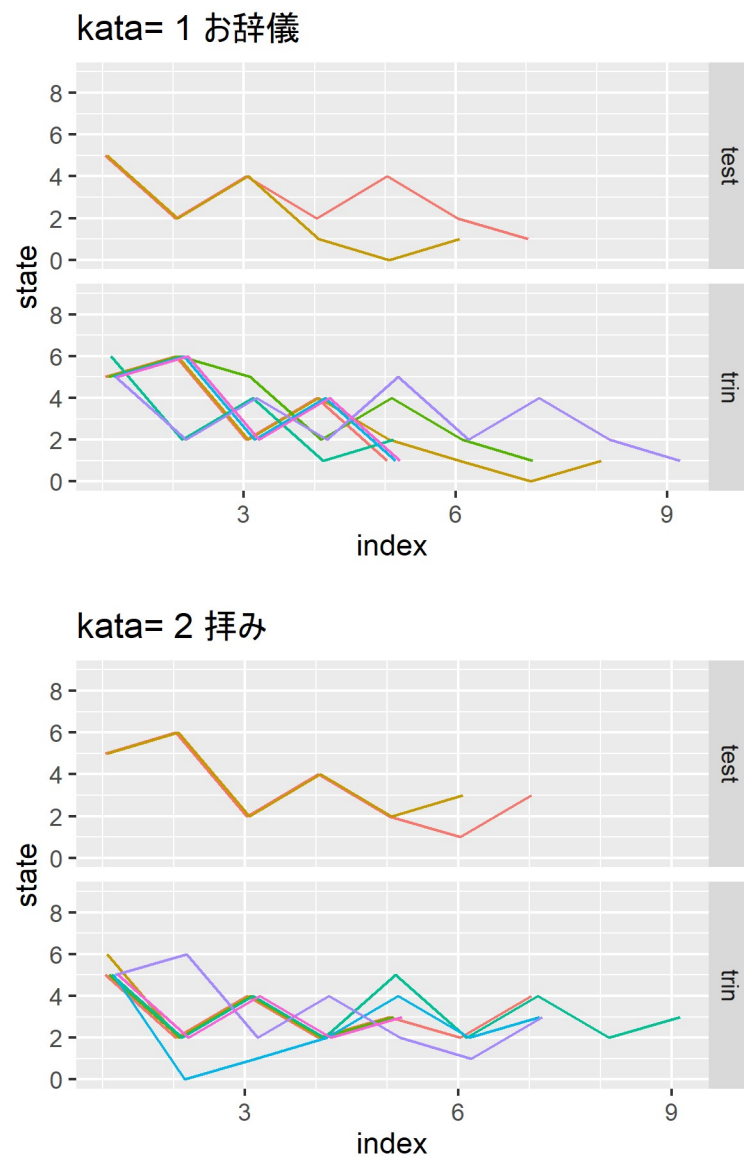


図 4-6 学習・検証データでの連長圧縮した状態の遷移
横軸は連長圧縮により時間圧縮され順序のみが保存されたインデックス。

4.2.4 抽出された状態遷移と人形動作の関係

学習状態の基礎データとして、各状態の型ごとの出現頻度をヒートマップ表示したものを図 4-7 に示す。図 4-7 の左右 2 つのグラフは連長圧縮による出現順序のみの抽出を行っていないものを行っているものを示しており、型ごとに出現頻度の合計が 1 になるように正規化されている。

状態遷移のデータを図 4-8 に示す。図の縦の列ごとにデータを見ていくと、状態 5 が他の多くの状態への分岐点となることがわかる。これは状態 5 は動き出しに多くみられる状態であり、動き初めがもっとも分岐が多いことを示唆している。一方で、これ以外に明確に 4 つ以上に数値が分配されている状態はなく、設定した状態数は多すぎないと考えられる。

さらに、図 4-9 に 10 の型全ての状態遷移の例を示す。動作が類似しているとした型同士を点線で囲ってある。型ごとにパターンがみられ、かつ類似した動作を行う“お辞儀”と“拝み”，や“指す”と“招く”，や“襟”と“懐手”において、同様の状態遷移が起こっていることがわかる。このことから、設定した状態数は少なすぎなかったといえる。教示（正解）のデータなしでの学習を行ってもこのような結果を得られたことから、HMM による“ず”を含んだ状態遷移表現はある程度成功し、型の機械的な分類も十分可能だと考えられる。

また、図 4-10 に検証データにおける、型同士のレーベンシュタイン距離（36 通りの学習の平均）を示す。同じ型同士では距離が極めて小さくなり、似たような状態遷移をたどっていることを示す。ただし、この図における同じ型同士の組み合わせには、レーベンシュタイン距離が 0 となる全く同じデータ同士の組み合わせも含まれている。同様に、“指す”と“招く”のような類似した動作も距離がある程度近くなっていることや、“指す”および“招く”と、これら以外の型との類似度が低いこともわかる。この結果も、状態遷移には型を判別するだけの情報が含まれることを支持する。また、図 4-5 に示したように、動作波形にはある程度のばらつきがあることから、動作要素の組み合わせの情報だけを追うことで、様々な速度や大きさの動作でも型の判別ができる仕組みを有していると考えられる。

なお、複数の類似する型で、状態遷移が共通の部分が見られた。例えば図 4-8 のように、すべての型は、ほぼ共通して状態 5（静止）からスタートし、指すと招くは非常に似た状態遷移を行った。これらの状態遷移が必ずしも左遣いの判断と一致するとは限らないが、時間的な判別を可能とする情報が型に含まれていることが示唆される。

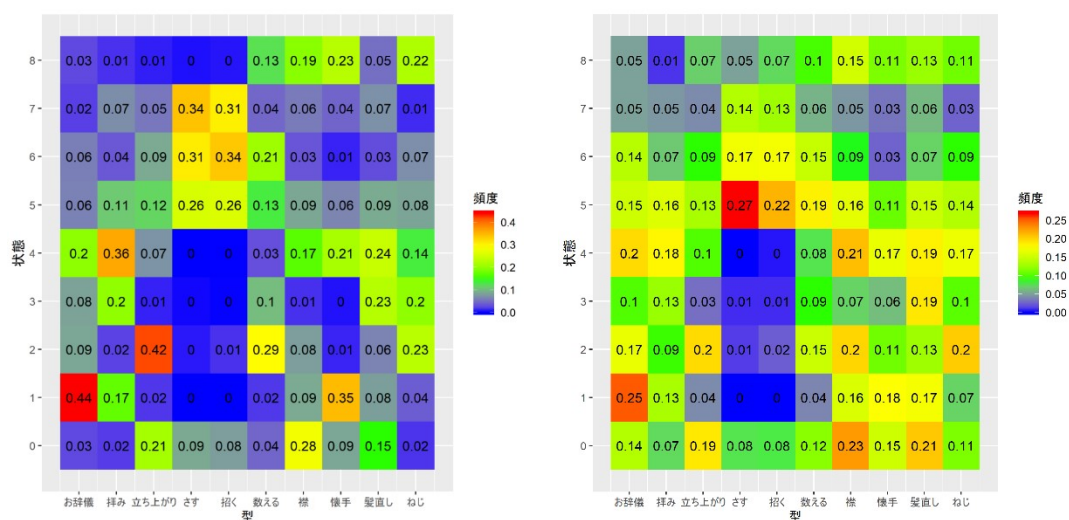


図 4-7 各型における状態の出現頻度

左が連長圧縮による順序のみの抽出をしなかった場合. 右が行った場合.

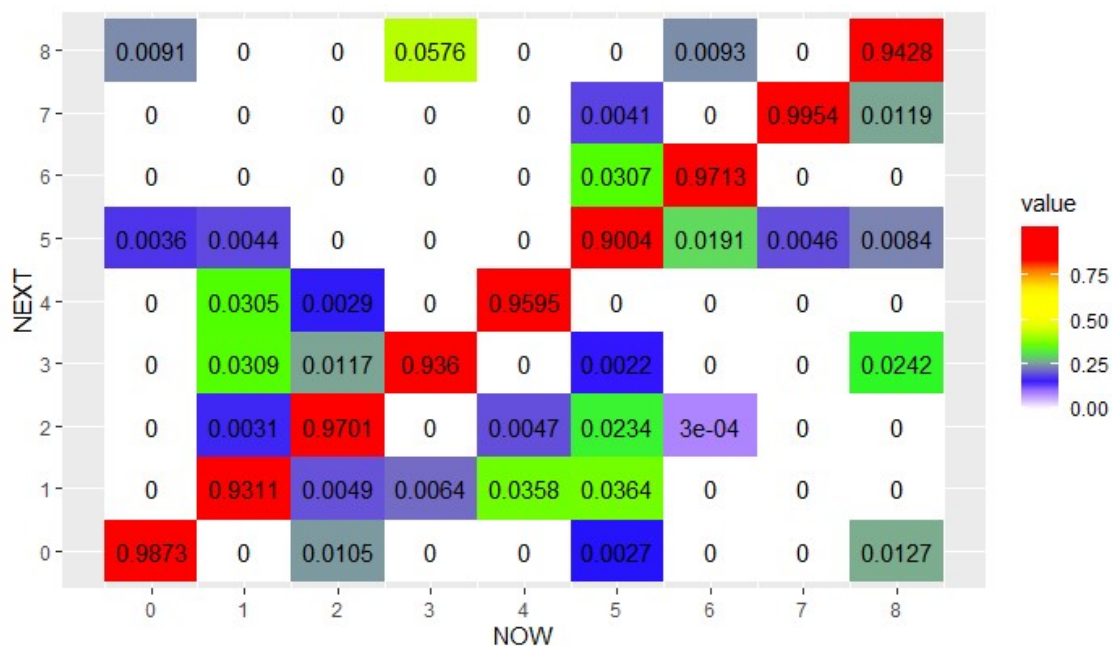


図 4-8 状態遷移表

横軸の状態から縦軸の状態へ遷移する確率を表示.

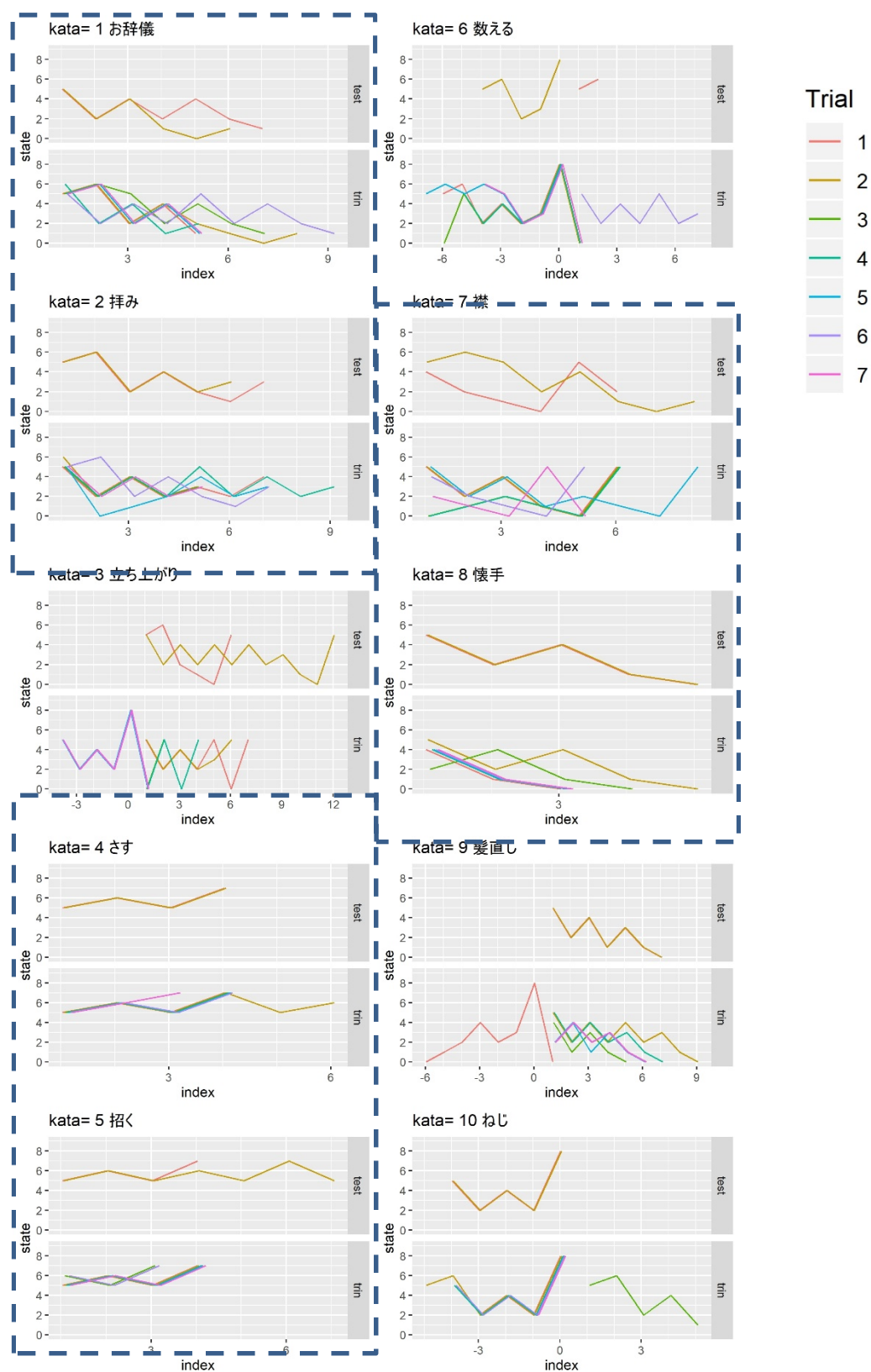


図 4-9 各型における状態遷移
類似した状態遷移を示す型を点線で囲んである。

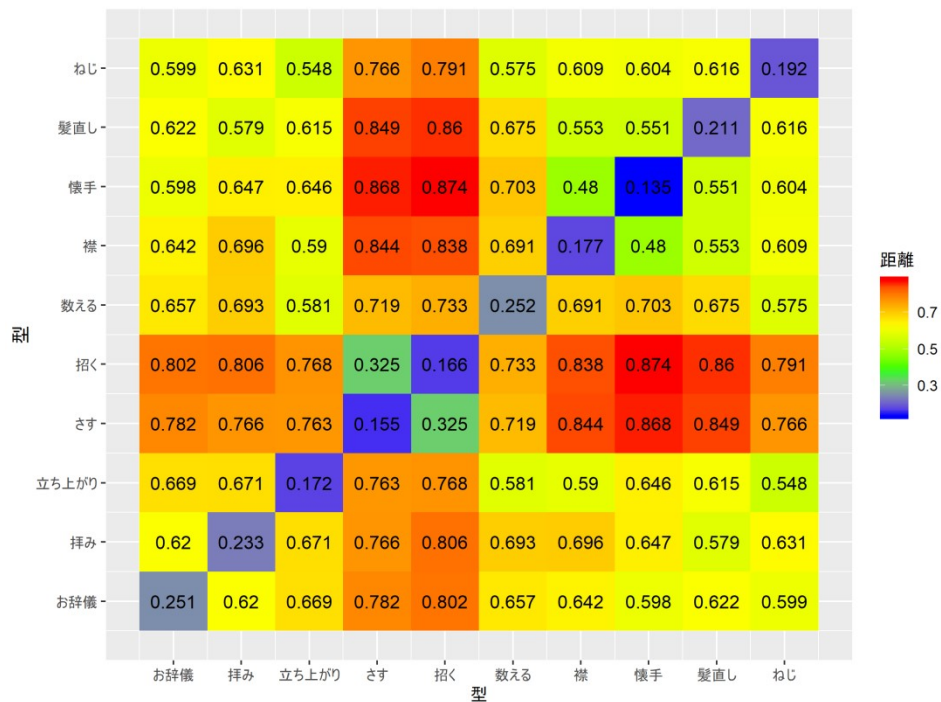


図 4-10 各型同士のレーベンシュタイン距離

4.3 考察とまとめ

以上の結果から、10 種類の型の 24 次元の人形動作データから、9 の状態からなる動作要素の組み合わせを教示なしで抽出することができた。また、動作要素の組み合わせの順序と 10 種類の型は高い相関を示したことから、型の分類に使用できる情報が得られたと考えられる。さらに、似た型同士では同様の状態遷移を行うことから、人形の動作には、第 2 章で示したような左遣いの時系列判別を実現できるだけの情報（動作要素）が含まれていることが確かめられた。今回確認された“ず”は、芸談や先行研究での議論でその存在を示唆されていたが、初めてデータから直接的に確認することができた。また、この合図の仕組みが型動作の速度や大きさなどの変動に強い（アドリブが行いやすい）ことが示唆された。

この仕組みが持つ利点や左遣いの判別結果との詳しい対比等の考察は、次章にて行う。

第 5 章 総合考察

5.1 解明された左遣いの型判別の特徴

本論文ではまず、先行研究から左遣いの型判別の仕組みについて仮説を立て、映像提示を用いた実験的検証を行った。具体的には、左遣いの型判別は主遣いによって示される動作特徴の連続を利用して時間的に進行し、型判別の正確性には左遣いの熟達度、人形動作の観察時間、型の種類の 3 つの要因が影響を与えるという仮説を立てた。そして、実験参加者（判別者、現役の人形遣い）に左遣いになったと仮定して、左遣い視点の映像から型の判別を行う実験を、上記の 3 要因を統制して実施した。

実験の結果、熟達群と準熟達群のそれぞれで、様々な型について動作開始から何秒の時点でどの程度型を予想できたのかを確信度データとして取得することができた。このデータの分析により、“お辞儀”の型については熟達群の正解確信度が準熟達群のそれよりも高くなること、および特徴的な動作をする“襟”と“懐手”において 1.00～1.25(s)の間に正解確信度が大きく上昇することが明らかになった。したがって、仮説で提示した 3 要因はいずれも型の正確性に影響を与える要素であったと言える。

次にクラスタ分析によって、瞬間的な提示 (0.75(s)) の時点で、同じ動作特徴を持つ型の組み合わせと判別に迷う型のクラスタがある程度一致すること、時間の経過とともに、クラスタ内での判別が進むこと、他とあまり類似しない固有な型動作が、1.00～1.25(s)時点までは、他のクラスタの型動作と混同される傾向にあること、そして 1.50(s)時点では、上記の固有な型も判別され、比較的明確な分類が成立することが明らかとなった。これは、仮説で提示したプロセス (図 2-4) に近い判別プロセスが行われていることを示唆している。

さらに、仮説で述べた判断を可能とする“ず”が人形の動作中に含まれているかどうかを検証するため、実際の人形動作データを取得し、単純な HMM により教示なしで

の隠れ状態の抽出を行うことで、型の分類が可能な情報が出現するかどうかを検証した。その結果、多次元の時系列情報である人形動作は時系列の隠れ状態の遷移で表現され、この時間的な遷移（組み合わせ）から、判別実験で得られたクラスタの動作特徴にある程度基づいた分類が、実際の人形動作データのみからも可能なことがわかり、型判別の仮説と合致する結果を得た。

以上により、文楽人形操作における左遣いの型判別は仮説で提示したプロセス（図2-2）に近いことが示唆され、その時間特性の一部が明らかになったと言える。

今回明らかになった型判別は、時間分割された小さな動作要素（＝判別のための情報）の組み合わせから、正解の候補を絞り込むという、非効率的にもみえる特異な特徴を有することがわかった。

本章では、一見不可解なこの判断の仕組みがもつアドバンテージや、実験結果と仮説との相違点について、文楽という伝統芸能全体がもつ特徴や、類似した他の協調タスクとの比較に基づいて議論を行う。

また、今回解析の対象としなかった要素について述べ、本研究での限界と今後の展望について述べる。

5.2 型判別の仕組みが持つ利点

本研究では文楽人形の操作システムが独特なものであることを示し、その一部を明らかにした。ではこの仕組みはなぜ成立し、どのような利点を持つのか。第1章では、左遣いの型判別を含めた人形の操作方法は、文楽の表現および人形構造の進化の歴史と深く関連することを説明した。そこでは、文楽人形の操作方法は制約の厳しい舞台上で、より多くの観客に多彩な動きを見せることを目的に拡張されていった経緯を説明した。また、第1章で述べた通り、舞台上には人形遣いの他に三味線と太夫が存在し、舞台全体の流れは三味線と太夫が作り出す楽曲要素である床に従って進行する。そして、人形遣いたちはその間に合わせる形で演技を行う。しかし、床については全体の公演時間と太夫の台本（床本）こそ決まっているものの、三味線と太夫がお互いにタイミングをはかりながら進行するため、毎回一定とはならない。そして主遣いの演技も個人の裁量や床に合わせてアドリブやアレンジが行われる。そのため、左遣いはアドリブやアレンジに常に注意を向けなければならない。

今回仮説で述べた仕組みは、素早く型の種類を特定することを可能にはしないものの、左遣いが常に観察を続けることで、左手を動かしながら徐々に絞り込むことを可能にす

る仕組みである。また、動作開始の合図は頭および全体の重心を移動するものが多く、この仕組みであれば、観客から見て明らかに合図と見られるような動作を最初に抑え、かつ左手の動きの遅れを抑えつつ、細かい動作のアレンジに対応できると考えられる。そして主遣いはこれらの合図を人形の演技と同時に出すことができる。また、後藤らは「人形遣い三人の組み合わせは常に流動的である。従って三人遣いの芸のポイントは主遣いの意思の伝達と三人の連繋にある。そのために、一般的な型による処理でなく、ズと呼ばれる独特の合図（型を分解した要素）が案出された」と述べている[3]。人形遣いによって動作表現が異なる点も文楽の特徴であり、その動作が“ず”と一体化していることから、人形遣いの組み合わせを超えた連携を行うためにも、この仕組みが重要だと言える。つまり、文楽という芸態において文楽人形は、三味線や太夫の作る文脈情報に合わせつつも様々なアレンジやアドリブ動作をする必要があり、そのためにマニュアル化が困難な“ず”による制御法が誕生した。“ず”は細かく分解された動作要素などで構成され、その多様性から習得は困難であるが、“ず”と緩い文脈情報のみであらゆる演技を実現できるがゆえに、それ以上のマニュアルが不要となっていると考えられる。

このように、文楽人形の操作システムは、合図を含めた動作のアレンジに幅広く対応し、実時間での協調操作を行える他に類を見ない優れた特性を持っている。つまり、文楽が持つ演劇としての要件と合致した特性を持っていると言える。また、現在文楽人形操作についてのマニュアルが存在しないのも、この特性に起因していると考えられる。

5.3 型判別の仕組みに関する先行研究との比較

今回の結果から明らかになったのは、あくまで左遣いが型の種類を特定するまでのプロセスの特徴であり、左遣いの動作判別システムの一部であるといえる。本節では左遣いの型判別システムがどのような全体像を持つか、既存のどのようなシステムと類似しているかについて、関連する研究例を述べながら議論する。

本論文では、人形遣いの型判別を音声認識にたとえ、合図となる動作要素（を含むセグメント）が音素、品詞推定が型認識というアナログに基づいて、型判別のための情報（動作要素）の抽出を行った。その結果、人形動作データから、動作要素には型判別を実現しうる情報量が含まれていることが明らかになった。このことから、左遣いの型認識は品詞推定のような仕組みをとりうると考えられる。このような問題に対しては多様なモデルが存在し、どのモデルが左遣いの判別をよく再現するかは、今後さらなる追求が必要であるが、文脈情報も扱える代表的な方法としてはベイズ推定が挙げられる。今

回明らかになった情報をベイズ推定に当てはめると、左遣いの判別は、始めは様々な動作の可能性を確信度の事前分布として持っているが、主遣いによる動作開始から順次合図となる動作要素を受け取り、そのたびにどの動作（型）がふさわしいか、確信度分布の更新を行いながら操作を行う過程としてモデル化できる。

すなわち、「主遣いが人形の頭をあるやり方で動かした場合、100%の確率で“ず”はある型を示している」というようなルールベースのマッチングではなく、左遣いの中では様々な型の発生確率の事前分布が存在し、人形動作の観察により、その分布が徐々に変化していくプロセスだと考えられる。このようなプロセスであれば、型と1対1対応しているとは限らない“ず”の存在や、HMMの状態遷移のような時間的に細分化された“ず”も取り扱うことができる。また、いずれかの型の可能性が100%になるまで、ただ主遣いによる人形動作を観察しているだけではない。左遣いが人形の左手を突き出したり、重心移動を行う際に、それらの型の可能性がある程度高まった時点で素早く対応することで、実際の舞台で行われるアドリブやアレンジにも柔軟に対応することを実現していると推測される。

このような動作判別についての関連研究として、時系列パターン解析における「早期認識」[45]が挙げられる。多くの動作認識では、動作の系列全てを入力として取り入れてから動作の判別を行うが、例えば川島ら[46]は、人間のある動作に特有の姿勢に対し特有度というパラメータを設定することで、動作開始からの姿勢の変化より、早い段階で動作種類の特定を自己組織化写像により実現している。この研究でも、認識時間と認識精度のトレードオフや、類似した動作の誤認識について議論されており、左遣いの動作判別処理と共通点が多いと思われる。他方、大きく異なる点として、①左遣いは動作認識と並行して人形左手の操作を行わなければならないこと、②動作主自身ではなく人形の動作を操っていること、が挙げられる。動作が確定していない状態でも観客から見て違和感のない動きを実現しなければならず、①は純粋な動作判別よりも困難なタスクだと言える一方で、②については、動作判別の助けとなっている可能性が高い。文楽人形の動きは人間の動作を模したものが多いが、可動部位やプロポーションの違い、演劇性の関係で、人間の動作に比べてシンボリックな動作や、歌舞伎でいう見得のような固有の動作が含まれる。そのため人形の姿勢や初期動作は人間のそれよりも特徴が強いものになり、主遣いの動作表現の仕方によっては左遣いの動作認識の難易度を下げたり、左手の動作開始までの猶予期間を与えたりできると考えられる。このように主遣いが人間の動作を基に人形動作に施す、動作表現の加工にこそ、人形遣いが“ず”と呼んでい

るものの神髄が隠れているのではないだろうか。

“情報を含んだ身体動作”という観点では、例えばインフォマティブモーション[24]と呼ばれる人間の投擲などの準備動作から、目標とする投擲距離などを推測できるという現象についての先行研究が存在する。文楽はこの“情報を含んだ身体動作”を“舞台劇”として能動的に加工、利用することで、サイズの大きな人形の分担操作と情報伝達を実現したとも捉えられる。これは、一般的なインタラクションを利用した協調タスク、例えば協調運搬[23]のような単純な目的のために、単純に作業分担を行うタスクとは異なっている。協調運搬タスクに比べると、文楽人形は一見無謀にも思える多自由度で複雑な動作をリアルタイムに行おうとしているが、その実、人形の身体分割方法、動作方法、文脈情報、あらゆる要素がこれを実現するために働いていることになる。

まとめると、文楽における協調は“その体躯・表現を実現するのにふさわしい動き”を設計し、“会話や楽曲などの文脈情報”を補助的に利用することによって、一見困難な操作分担と複雑なリアルタイム協調を自然な見た目のまま、実現したシステムだと言える。

近年、低価格化の進むモーションキャプチャを用いた遠隔インタラクションや、VR技術の発展に伴い、より人間らしく、親しみやすい多自由度の動きがヒューマンインタフェースに求められている。本研究は、研究対象を長い歴史を持つ伝統芸能とした（三人遣いの成立は1734年と言われている[5]）が、長い年月をかけて磨かれたこの技術から得られた視点は、これらの潮流に求められる機能や協調作業を実現できる可能性を持つ大変意義あるものと考えられる。

文楽の長い修行期間には、各主遣いの癖だけでなく、これら文楽という舞台の仕組み全体を把握することによって、必要な情報を補完する技術の学習が含まれているのだろう。

5.4 本論文の限界と文楽人形の協調操作の全体像

本論文では、左遣いによる型の判別プロセスに絞って検討を行った。しかし、実際の舞台では、人形操作において左遣いは、型の種類以外にも多くの情報を読み取り、人形を操作しつつ、小道具を出し入れしたりする必要がある。また、実際の公演では床（三味線による演奏と太夫による語り）が存在し、左遣いは多くの情報の中で人形操作を行うことになる。本論文では、先行研究[15]で示された『型単体の実行においては“ず”のみで協調操作が成立する。』という前提のもとで実験やモデル化を行ったため、第2

章および第 3 章では床を構成する要素については直接検討を行わなかった。本節では型の種類以外の様々な情報が協調操作へもたらす影響について考察し、また実験条件での制約についても踏まえ、本論文の限界と、想定される人形遣いの協調操作の全体像を示す。

まず、左遣いの型判別以外の情報伝達について、主遣いは左手の動作の大きさや速度、動作に込める感情などをある程度制御できるといわれている。また、舞台上では様々な動作のアレンジや、型に分類されない細かい動作も行われる。本論文で実施した実験では、典型的な型の判別のみを対象としたが、実際の人形操作の現場では、主遣いが行おうとしている動作の速度や大きさや性根（人形の役柄や心情）などにも気をつかい、観客からみて人形動作が自然にみえるように人形を操作する必要がある。型の種類にもよるが、主遣いの動きを見てからそれに合わせては左遣いの動作は遅れがちになるため、大きさや速度に関する主遣いの意図を読み取るには多くの経験を要するものと考えられる。多くの型で主遣いのみが動作する部分が多いのも、左遣いの負担を軽減するためであろう。ただし、実際の舞台上では、主遣いからの“ず”以外に床から多くの文脈情報を得ることができる。特に音楽に合わせて舞う際の動作タイミングや、その時点でのおおまかな感情の決定には、床が非常に重要な役割を占める。櫻ら[15]の研究は、ヘッドフォンを用いて主遣いだけに床を聞かせた状態で演技をさせた場合には、人形遣い全員に床を聞かせた状態と比べ、左遣いの動作の一部に遅れが生じる現象を報告しており、実際の公演において左遣いは床からの情報も利用していると考えられる。

また、左手の操作以外にも左遣いの役割は存在する。例えば、人形が手にもって何かを操作するとき、小道具を取り出すのも左遣いの仕事であり、左遣いは舞台全体を把握する必要がある。「足遣い 10 年、左遣い 10 年」と言われるように人形遣いに習熟するには多くの年月を必要とされるが、足遣いを経験した後に左遣いを経験することで、人形遣いは舞台の仕組みを順次把握していくと考えられる。これらは左手の操作には直接かわらないが、舞台の構成において重要な点である。

次に、文楽を構成する三業について述べる。第 1 章でも述べたように文楽は三業と呼ばれる分担作業からなる。すなわち、人形遣いによる演技の他に、三味線による演奏と太夫による語りが重要な要素として挙げられる。このうち、三味線と太夫は強い関係性があり公演の前に入念な稽古を行うが、人形遣いと三味線あるいは太夫との相互のやり取りは少なく、公演直前の三者を含めた 1 度の通し稽古のみで、公演本番に臨むことになる[6]。公演中の舞台上では、三味線・太夫の作り出す床が全体の演技進行を支配し、

人形遣いはその間に合わせた演技を行う。しかし、三味線と太夫はどの公演でも同じ内容を同じタイミングで繰り返しているのではなく、重要なタイミングではイキを合わせつつも、即興的にタイミングをずらしている。また、人形遣いも重要なタイミングでは床とイキを合わせつつも、多くのアドリブを含めた演技を行っている。このタイミングについて渋谷ら[6]は、床の中の“浄瑠璃の間”と呼ばれるタイミングで人形の動作タイミングを調整している可能性があるとして述べている。

したがって、舞台上では、型を一時停止する、装飾動作を加えるなど、タイミング調整を含むアドリブが主遣いの即興的な判断で行われ、左遣いは逐一これらに対応しなければならない。この要求を満たすために、演技中の左遣いは可能な限りアドリブやアレンジの可能性を考え続ける必要がある。すなわち、舞台上での左遣いは本研究で行った実験よりも厳しい条件下で様々な注意を払い続けることになるが、型と“ず”の仕組みや床などの文脈情報を駆使することで、動作の遅れや違和感を抑えつつ、高いアドリブ性を実現していると考えられる。

これに対し、第3章で行った実験では、仮説を立てた時点で予想された結果と実際の実験結果とが一致しなかった部分がある。それが、今回の実験環境や前提に起因するものであるかどうかを考察する。2.5.2 節で述べた“お辞儀”と“襟”の関係が生じた原因、ならびに正解確信度が100%に収束しなかった原因について考える。

前者については、「“襟”動作を他の型とは見間違えないが、“お辞儀”動作を“襟”動作と見間違える」現象が発生していたことに起因する。単純に2つの動作が似ているだけであれば、このような誤警報ともいえる非対称な関係は現れないはずである。大きな要因の一つとして、実験環境によるものが考えられる。本実験では、実際の判別状況をシミュレートするため、ノートPCによる一人称視点の映像提示という手段を用いたが、実際の演技状態と異なり、ごく短時間の映像を瞬間的に提示され、確信がもてなくても答えなくてはならなかった。特に本実験では、先行研究[8]から人形を含む注視点まわりの映像情報から型が判別できると仮定して試験を行ったが、人形遣いごとに視線や視界の使い方は異なる可能性が高く、判別者が普段通りの情報獲得が行えなかった可能性は無視できない。また、動画開始前の注視点は中央に固定されているため、動画開始時に情報源である人形頭部などをみるまでに遅れた可能性がある。さらに判別者はこのようなタイプのテストに対する十分な練習も行えていないため、十分な判別をするのは困難だと考えられる。実験後のアンケートでもタスクが困難であったという意見がみられた。他に考えられる原因として、前節で述べた文楽の舞台性に基づくアドリブ性が

挙げられる。前節で述べた通り、実際の舞台では 40 を超える型に加え、人形遣いごとの個性に則った動作内容（“ず”を含む）の変更やアレンジに対応するため、本実験で扱った 10 の型と比較にならないほどの多彩なバリエーションの動作が行われる。この強力な仕組みを有するために、本実験の型を当てるだけの単純な課題に対しても、別の人形遣いや動作状況の可能性を加味し、最後まで型を特定できない事態が起こったとも考えられる。

上記の誤警報は、このアドリブへの対応により発生していると推測できる。“襟”は特徴的な動作を含む型であり、クラスタ分析からも 1.50(s)の時点で“襟”だと確信できる動作要素を含んでいると考えられるが、“お辞儀”においては、まだそのような動作が発生していない状況にある。すなわち 1.50(s)時点において、“お辞儀”は本実験の設定では存在しない「アドリブでタイミングを遅らせた“襟”」である可能性があったために、誤警報が発生したと考えられる。

第 2 章の実験では 10 の型から正しい型を選択するという単純な判別タスクとしたため、1.50(s)以内に正解が絞り込めると仮定していた。しかしながら、演技中の人形遣いにとっては、1.50(s)経過時点でもまだ他の型動作、特にアドリブを含めた他の動作の可能性を考え得る状況であったため、回答が正解の型に 100%収束しきらない結果となったと考えられる。

このように、一部仮説と異なる実験結果がもたらされたのは、本実験設定における限界によるものと解釈できる。今回、文楽の操作システムが時間的に強力なアドリブ対応（即時の動作や速度変更への対応）を有するシステムで有ることが示唆されたが、この中身までは踏み込めていない。このアドリブ対応性を明らかにするには、人形遣いの癖や、動作のアレンジ方法について仮説を立て、新たな検証実験を行う必要があるだろう。以上のように第 3 章、第 4 章で行った検討にはこれらの要素は含まれないため、このアドリブ性の内容などについては本論文ではこれ以上深い議論は行えないが、文楽の舞台を構成する三業は複雑な協調関係を築いており、左遣いは多くの情報の中で協調操作を実現していると考えられる。今回、型の判断において熟達差が大きく出なかったが、熟達した左遣いは、人形遣いの癖の対応や人形操作などに優れている可能性が高い。また、5.3 節で述べた通り、今回は型判別のための動作要素の抽出を行ったが、実際の左遣いの判別はベイズ推定に近いものと考えられる。本論文では、40 種類以上ある型のうちの 10 種類のみ、それも 9 試行分のデータしかなく、十分なデータが取れなかったため、動作要素の抽出とその判別可能性の検討にとどまったが、さらなる実験により数

多くのデータを収集し、型判別のモデル化を行うことも今後の課題だと考えられる。

5.5 今後の展望

5.5.1 アドリブと“ず”の対応関係

今回は文楽人形の型判別の仕組みが、型という単純な情報だけでなく、演技に関する様々な特徴量（動作速度や大きさ、感情など）を実時間で共有し、動作に反映させるための仕組みであると結論付けた。しかし、この特徴量が“ず”とどのように関連するのか、また動作のアレンジがどのようにして行われるのかについては未だ明らかになっていない。これらの情報が、動作要素と内部の“ず”に対してどのように影響するのかを明らかにすることは今後の課題となる。先行研究として服部ら[47]は、文楽人形の演技時に込める感情によって動作波形が変化することを示唆している。おそらく“ず”と呼ばれる要素も、同様の变化、もしくは左遣いに情報を伝えるため、よりシンボリックに動作波形が変形している可能性がある。これらの違いを観察し、“ず”となる部分と動作全体の違いについて考察するアプローチが第一歩となるだろう。

また、今回の結果は、動作要素の連続から左遣いが主遣いの意図を察するのが間に合うように人形動作が作られていることを示唆する。特に型動作では、左右の手が同時に動くものや、交互に動くものが多く、左遣いへの負担を軽減させうる要素が多い。このような型の構造からも、“ず”となる区間や大きさについてのヒントが得られるかもしれない。

5.5.2 修行の必然性と技能継承への応用可能性

実験後アンケートを行ったところ、実験の感想（自由記述）にて、

- ・動画内の主遣いの癖が強かった
- ・同じ型でもいくつかの演じ分けがあるので、特定に苦労した

などの意見があげられた。

このことから、典型的な動作パターンとする型であっても、主遣いの操作の癖や演じ方によって多くの違いが現れるものと考えられる。しかし、人形遣い達は基本的にどのような組み合わせであっても自由に演技が可能であることから、長い修行期間を使ってこれらを学習していくものだと考えられる。

伝統芸能である文楽の人形遣いの技芸員の数は現在 43 名であり、多いとはいえない。人形遣いとして入門するには、技芸員に直接弟子入りするか、研修生として入門するかの 2 つの手段が存在するが、前述した通り人形操作にはマニュアルが存在せず、人形遣いになりたくて研修生となっても、入門後の適性試験で資格を失う可能性もある。研修生の募集は 2 年に 1 度であり、その数居は低くはない。

そこで今回用いた実験動画を、初学者への人形動作の基本を理解する補助のための訓練ツールとして提供する可能性を今後検討したい。熟達のための訓練としては十分ではないと考えられるが、現役の人形遣いに確認したところ『(初学者に対して型を認識させるには) 有効だろう。』というコメントを得た。また、本実験用の映像は人形の正面を含む多視点での同時撮影を行っており、人形遣いには本来見ることのできない観客席からの見えと左遣い視点との違いなども同時に提供できると考えられる。

このような訓練ツールや映像の提供はあくまで副次的なものである上に、本研究で得られた知見も限定的なものであるが、初学者に対して“なぜ型判別が難しいのか”、“なぜ長期間の修行が必要なのか”、という疑問に答え、文楽の奥深さを理解してもらう一助になるのではないかと考えられる。

第 6 章 結論

本論文では、文楽の人形操作がもつ高度な協調操作の解明を目的に、人形遣いの一人である左遣いによる型判別がどのように行なわれているのかを実験的に検討した。また、隠れマルコフモデルによる分析を含めた検証により、以下の成果を得た。

- ・ 実験結果の一般化線形混合モデルによる分析（3.5.1 節）により、型の判断は、その種類ごとに時間的に進行することと、熟達により一部の判別精度が上昇することが明らかとなった。
- ・ 実験結果のクラスタ分析（3.5.2 節）により、型ごとの動作特徴が判別において重要なことが明らかとなった。
- ・ 実験結果の主成分分析（3.5.3 節）から時間ごとの動作特徴の有無や、型ごとの左手の操作タイミングが、左遣いの判断に重要なことが示唆された。
- ・ 隠れマルコフモデルを用いた状態遷移表現（4 章）により、型判別を時間的に進行できるだけの情報（動作要素）が、人形動作に含まれていることが示唆された。

したがって、文楽人形操作において左遣いは主遣いの操る人形部位の決まった動きから有限種類の型を一意に同定するのではなく、その動きを常に観察し、主遣いからの手掛かりとなる細かい動作要素の連続を受け取り、経過時間ごとに意識すべき動作特徴や型に注意しつつ、起こり得るすべての型の可能性を考えながら操作を続けている可能性が高いことが明らかになった。さらに、芸談や文楽の三業の性質を交えた考察から、様々なアドリブに対して左遣いの判別と操作が間に合うように人形動作自体が設計されている可能性も示唆された。

以上のことから、伝統芸能文楽の持つ特殊性である、人形の協調操作を実現する仕組みのうち、主遣いと左遣い間の情報伝達の特徴と、型動作の特性の一部が明らかになった。人形遣いたちは明確な合図による、マスタースレーブ間の支配的な作業分担をおこなっているわけではなく、人間らしい動作をシンボリックにした型と、そこに含めた“ず”

の組み合わせによって、左遣いに厳しい判断を迫る反面、動作媒体を利用したリアルタイムな意図伝達とアドリブを含めた様々な演技動作を同時に実現していると考えられる。このような動作内容の変更も含めた多彩な合図を動作要素（人形動作の一部）として埋め込む例は筆者の知る限り存在せず、このことが、長い修行期間を必要とする理由となっている可能性が示唆された。

今後の展望としては、各要素に着目しやすい環境を提供することで初学者の訓練支援を行うことや、型や“ず”がもつ、情報の内容や仕組みをより深く明らかにしていくことが考えられる。

前者の訓練支援については、具体的には、客席からの人形の見え方と実際の操作視点との対応や各型の判断方法をまだ身につけていない、左遣いとしての経験が少ない初学者に対して、人形正面からの映像と左遣い視点での型動作の映像、および各時点での着目すべき注視点を提供し、各型の特徴の体系化などを支援することが挙げられる。

後者については、今回は左遣いの型判別のみに研究対象を限定したが、本研究でまだ触れていない人形操作に関わる要素として、以下が挙げられる。

- ・“ず”によって伝えられる動作の速度や大きさ
- ・“ず”によって伝えられる演じる人形の役柄や感情（性根）
- ・“ず”を構成する動作要素
- ・主遣いと足遣いの間の情報伝達

まずは、型の大きさや速度などのシームレスに伝達され则认为られる要素が、単純な伝達モデルで再現可能かどうかについて検討することが考えられる。また、計測した人形の動作波形から、動作開始時の姿勢変化や頭の動きなどの、左遣いへの“ず”を含む動作要素の内容について分析し、ベイズ推定などによる左遣いの動作判別のモデル構築を目指す必要がある。

今後これらを順に達成することで、伝統芸能“文楽”の全体構造を工学的に解析し、その技（スキル）を保存していくことは、学術的に社会的に非常に重要である。また、今回の解析ではリアルタイムな意図伝達と幅広い動作表現を両立する文楽において、マニュアルとして明文化できない程度にまで合図を曖昧にすることによるメリットが示唆された。このような技術は、長い年月かけて改良を重ねてきた文楽特有のもので、一般的な芸能では稀有なものだと考えられるが、マスターとスレーブの立ち位置がはっきりしていることから、ヒューマンインタフェースとしてみた場合、自然な見た目と即応性・応用性の面から、自然で多様な動作が可能な義手への応用などに得られた知見を

応用する可能性も考えられる．このような応用も今後視野に入れることは重要であろう．

謝辞

本研究を進めるにあたり、多大なるご指導、ご助言を頂きました東京大学大学院総合文化研究科の植田一博教授に深く感謝の意を表します。ご指導のみならず、度々こちらの我儘をきいてくださり、感謝の念に堪えません。

また本プロジェクトの共同研究者として、数多くの助言を頂きました渋谷友紀氏にも感謝いたします。文楽の三業を別の切り口から研究する共同研究者の存在は何者にも代えられないものです。

本研究は、吉田勘彌氏、吉田幸助氏、吉田清五郎氏、吉田勘市氏、吉田玉佳氏、桐竹紋臣氏、桐竹勘次郎氏、吉田玉彦氏、吉田玉路氏、桐竹勘介氏、吉田和馬氏、吉田簀之氏の12名の人形遣いの方々にご協力により実現いたしました。計測用文楽人形の開発においては人形師の甘利洋一郎先生に大変お世話になりました。これほど多くの方々が、ご多忙の中ご協力していただけたこと、誠に感謝の意に耐えません。

本研究の一部は、第48回東レ科学技術研究助成、JSPS 科研費 JP25540154、およびMEXT 科研費 JP26118002 のご支援により行われていることをここに記し、謝意を表します。

機材提供など、実験に協力してくださった Detect 社、ミユキ技研、NAC イメージテクノロジーの皆様に感謝いたします。

研究室生活においては、実験や機器のテストに積極的に協力してくれた白砂大氏、またゼミに限らず様々な意見をくれた研究室のメンバーたちに感謝いたします。メンバーは研究対象もその個性も彩に満ちていて、常に新鮮な気持ちで研究活動をすることができました。有難うございました。

最後に身勝手な生活を送っていた自分に対し、ここまでそっと応援してくださった家族に感謝の意を示します。

令和2年 11月 17日

参考文献

- [1] 松平盟子, 文楽にアクセス, 淡交社, 2003.
- [2] 豊竹咲甫大夫, 豊竹咲甫大夫と文楽へ行こう, 旬報社, 2004.
- [3] 後藤静夫, “文楽 人形浄瑠璃の現在: その多重媒介性を考える,” 人文學報, vol.86, pp.281–293, 2002.
- [4] 倉田喜弘, 文楽の歴史, 岩波書店, 2013.
- [5] 藤田洋, 文楽ハンドブック (第三版), 三省堂, 2011.
- [6] 渋谷友紀, 櫻哲郎, 佐々木正人, and 植田一博, “文楽における浄瑠璃と人形遣いの呼吸の同期,” 認知科学, vol.24, no. 4, pp.518–539, 2017.
- [7] 竹本住大夫, 人間、やっぱり情でんなあ, 文藝春秋, 2018.
- [8] 櫻哲郎, 森田寿郎, and 植田一博, “文楽人形の協調操作における動作特徴の抽出,” ヒューマンインタフェース学会論文誌, vol.11, no. 3, pp.255–264, Aug. 2009.
- [9] 細田明宏 and 後藤静夫, “文楽人形遣いの修業と配役—後藤静夫氏に聞く文楽の制作,” 帝京大学文学部紀要 日本文化学, no. 42, pp.185–220, Mar. 2011.
- [10] 桜井弘, “文楽への招待: 人形,” 文楽ハンドブック, 藤田洋 (編), pp.44–61, 三省堂, 2011.
- [11] 加納克己, 日本操り人形史: 形態変遷・操法技術史, 八木書店, 2007.
- [12] 原道生, 黄金時代の浄瑠璃とその後, 岩波書店, 1998.
- [13] 角田一郎, “浄瑠璃操り三人遣いの創始と普及について (演劇とその周辺),” 国学院雑誌, vol.85, no. 11, pp.163–175, 1984.
- [14] 元史服部, “文楽における動きの情緒表現(<特集>バイオメカニズム的 美の探究),” バイオメカニズム学会誌, vol.26, no. 3, pp.137–141, 2002.
- [15] 櫻哲郎, 成田友紀, 沢井和也, 森田寿郎, and 植田一博, “文楽人形遣いの協調操作を実現する非言語情報通信(ヒューマンコンピュータインタラクション),” 電子情報通信学会論文誌 D, 情報・システム, vol.96, no. 1, pp.195–208, Jan. 2013.
- [16] T. W. Malone and K. Crowston, “Toward an interdisciplinary theory of coordination,” 1991.

- [17] 吉田勘弥, “文楽人形の運動制御,” アート/表現する身体: アフォーダンスの現場, pp.341–259, 2006.
- [18] 吉田玉男 and 山川静夫, 文楽の男: 吉田玉男の世界, 淡交社, 2002.
- [19] Z. Nédá, E. Ravasz, Y. Brechet, T. Vicsek, and A.-L. Barabási, “The sound of many hands clapping,” *Nature*, vol.403, no. 6772, pp.849–850, 2000.
- [20] K. Yun, K. Watanabe, and S. Shimojo, “Interpersonal body and neural synchronization as a marker of implicit social interaction,” *Scientific Reports*, vol.2, p.959, Dec. 2012.
- [21] K. Dunlap, “Reaction to rhythmic stimuli with attempt to synchronize,” *Psychological Review*, vol.17, no. 6, pp.399–416, 1910.
- [22] A. Z. Zivotofsky and J. M. Hausdorff, “The sensory feedback mechanisms enabling couples to walk synchronously: an initial investigation,” *J Neuroeng Rehabil*, vol.4, p.28, Aug. 2007.
- [23] 池浦良淳, “人間とロボットによる協調作業 (人間中心型ロボティクス総合特集号),” システム/制御/情報, vol.44, no. 12, pp.682–687, Dec. 2000.
- [24] 松丸隆文, “人間機械系の情報動作学の応用展開,” バイオメカニズム学会誌 = *Journal of the Society of Biomechanisms*, vol.29, no. 3, pp.139–145, Aug. 2005.
- [25] 金森務, 片寄晴弘, 新美康永, 平井宏, and 井口征士, “ジャズセッションシステムのための音楽認識処理の一実現法,” 情報処理学会論文誌, vol.36, no. 1, pp.139–152, Jan. 1995.
- [26] N. P. Dempsey, “Hook-Ups and Train Wrecks: Contextual Parameters and the Coordination of Jazz Interactions,” *Symbolic Interaction*, vol.31, no. 1, pp.57–75, 2008.
- [27] F. Barrett, “Creativity and Improvisation in Jazz and Organization: Implications for Organizational Learning,” *Organization Science*, vol.9, pp.605–622, Jan. 1998.
- [28] B. Kernfeld and B. D. Kernfeld, *The new Grove dictionary of jazz*. Grove; London: MacMillan c2002., 2002.
- [29] J. A. Skadsem, “Effect of Conductor Verbalization, Dynamic Markings, Conductor Gesture, and Choir Dynamic Level on Singers’ Dynamic Responses,” *Journal of Research in Music Education*, vol.45, no. 4, pp.509–520, Dec. 1997.
- [30] M. Kato, “The Effects of Reading Speed on Visual Search Task,” *Cognitive Science*, pp.1102–1107., Jan. 2005.
- [31] R. Pieters, E. Rosbergen, and M. Wedel, “Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory,” *Journal of Marketing Research*, vol.36, no. 4, pp.424–438, 1999.
- [32] 福田亮子, 佐久間美能留, 中村悦夫, and 福田忠彦, “注視点の定義に関する実験的検討,” 人間工学, vol.32, no. 4, pp.197–204, 1996.
- [33] 桜井洋, “文楽への招待: 人形,” 文楽ハンドブック (第三版), 藤田洋 (編), 三省

- 堂, 2011.
- [34] J. P. Guilford, *Psychometric Methods*. Textbook Publishers, 2003.
 - [35] 田中良久, 心理学研究法 16 尺度構成, 東京大学出版会, 1973.
 - [36] 妻藤真彦, “信号検出理論と確信度: 解答の変動と正答率の理論的問題,” 美作女子大学紀要, no. 36, pp.8–15, 2003.
 - [37] 櫻 哲 郎 and 植 田 一 博 , “ 文 楽 人 形 の 型 の 連 続 写 真 ,” <http://www.cs.c.u-tokyo.ac.jp/bunraku/>, accessed Oct. 26. 2020. .
 - [38] 中川聖一, 確率モデルによる音声認識, 電子情報通信学会(コロナ社), 1988.
 - [39] 達也河原, IT Text 音声認識システム (改訂 2 版) , 株式会社 オーム社, 2016.
 - [40] W. J. Heeringa, “Measuring dialect pronunciation differences using Levenshtein distance,” PhD Thesis, University Library Groningen, 2004.
 - [41] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol.39, no. 1, pp.1–38, 1977.
 - [42] 南角吉彦, 徳田恵一, 北村正, and 小林隆夫, “隠れマルコフモデルを用いた視覚音声認識のための正規化学習,” 電子情報通信学会論文誌. D-2, 情報・システム. 2, パターン処理, vol.J86-D2, no. 2, pp.163–172, Feb. 2003.
 - [43] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE transactions on Information Theory*, vol.13, no. 2, pp.260–269, 1967.
 - [44] M. C. Hughes, W. T. Stephenson, and E. Sudderth, “Scalable Adaptation of State Complexity for Nonparametric Hidden Markov Models,” in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds. Curran Associates, Inc., 2015, pp.1198–1206.
 - [45] O. Scharenborg, L. ten Bosch, and L. Boves, “‘Early recognition’ of words in continuous speech,” pp.61–66, 2003.
 - [46] 川島学, 島田敬士, 長原一, and 谷口倫一郎, “姿勢の特有性に基づいた動作の早期認識,” 日本神経回路学会誌 = *The Brain & neural networks*, vol.19, no. 4, pp.167–174, Dec. 2012.
 - [47] 服部元史, 中坊保則, 田所諭, 高森年, and 山田和人, “文楽人形演技の動作軸の位相要素の解析 人間型アミューズメント・ロボットの情緒豊かな動作生成を目指して,” 日本機械学会論文集 C 編, vol.66, no. 644, pp.1243–1249, 2000.

付録 文楽人形の詳細

A.1 各パーツ詳細

文楽人形を構成する各部位について動きや仕組みを示す。

A.1.1 かしら

作成した文楽人形“政岡”のかしらを図 A-1 に示す。かしらは、頭・首・胴串の 3 つの部位で構成されている。胴串と首は一体となっており、首に対して頭が屈伸するようになっている。頭の屈伸は、胴串に取り付けられているチョイと呼ばれる操作部を上下させることで行うことができる。チョイは、後頭部内側と紐でつながれている。首の先端に付けられたセミクジラの髭で作られた屈伸用の板ばねの先端を通して取り回されることで、首が戻る反力を生み出している。図 A-1(b)は最大・最小屈伸角度を示しており、その差は 46° である。

胴串には小猿（小ザル）という回転レバーも取り付けられており、これを回転させることで目を開閉することができる。頭の屈伸の仕組みと同様に、板ばねを介して頭の内側と小猿が紐で連結されており、閉じた状態時には、目が開く方向に力がかかるようになっている。

口には口針と呼ばれる針が取り付けられており、手拭などの小物を引っ掛けることができるようになっている。

A.1.2 左腕

左腕の詳細を図 A-2 に示す。図 A-2(a)は各部名称、図 A-2(b)は回転範囲を示している。左腕は、手部・前腕・上腕・差し金で構成されている。差し金についている小猿を紐に手を掛けて回転させることで、手部の角度を変えることができる。手首と指の 2 関節がそれぞれ回転し、可動するが、連動しているため、小猿の操作のみで状態を決定

できる。差し金は前腕に対して回転連結しており、上腕は綿を布で巻いたものであり、紐によって胴と連結されている。

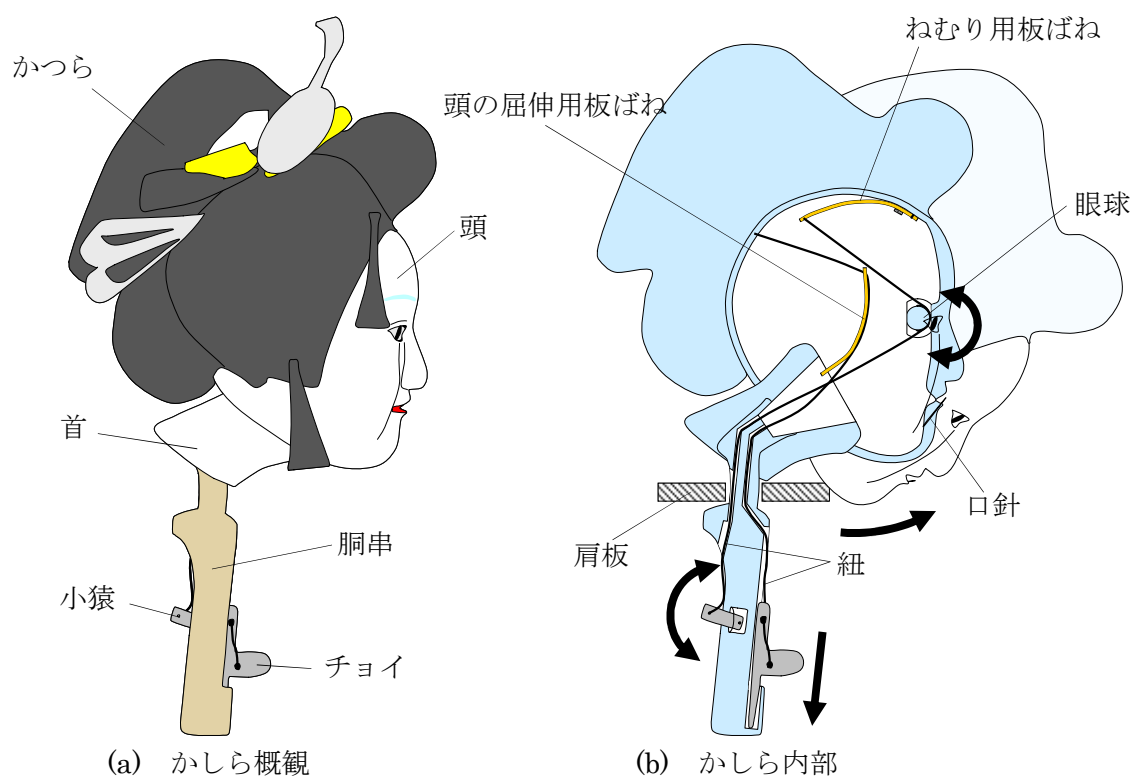
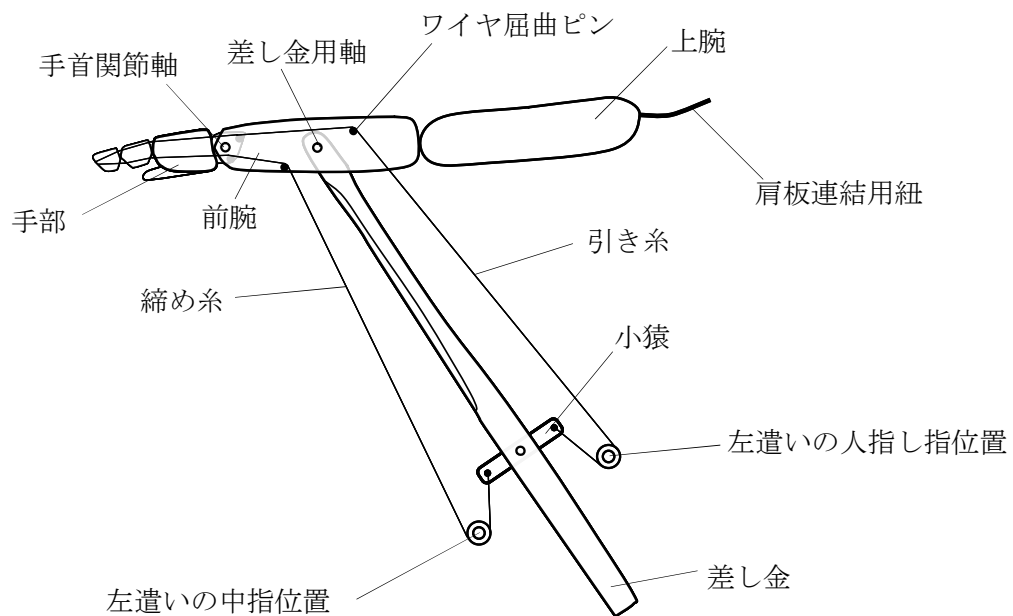
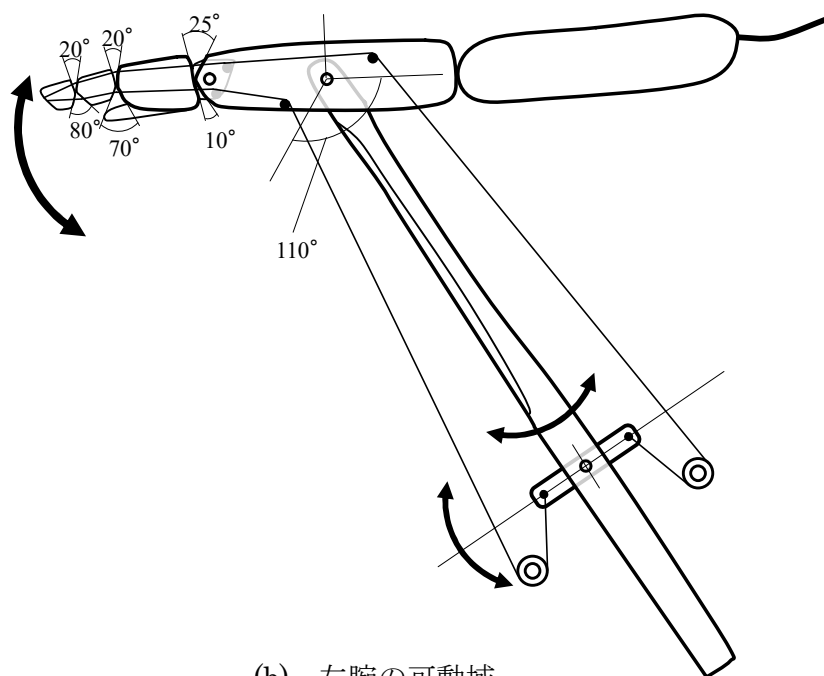


図 A-1 文楽人形のかしら



(a) 左腕の各部名称



(b) 左腕の可動域

図 A-2 文楽人形の左腕

A. 1. 3 右腕

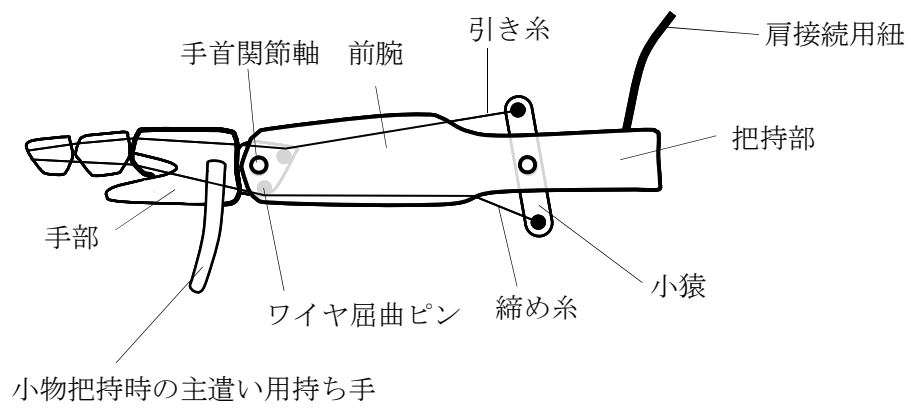
右腕の詳細を図 A-3 に示す。図 A-3(a)は各部名称，図 A-3(b)は回転範囲を示している。小猿を回転させることで，左腕と同様に手首関節と 2 つの指関節が連動する。手部には通常の操作では用いない持ち手がついており，小物を把持するときにはここを持って操作する。

A. 1. 4 胴

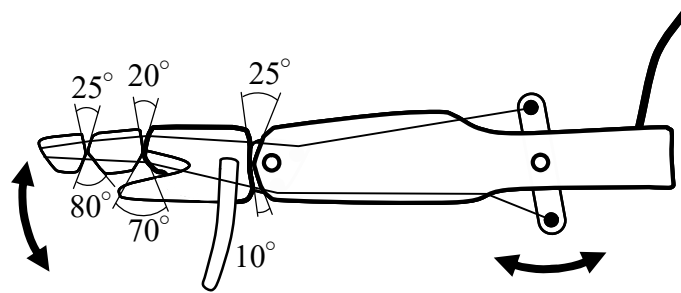
胴の詳細を図 A-4 に示す。胴には可動部分は無く，肩板と呼ばれるかしらとの連結部と，腰輪と呼ばれる部位からなっている。この 2 つの部位は布によって連結されており，肩板の左右上面には肩の形を作るためにヘチマが付けられている。また，腕や足との連結をする紐を結ぶ紐輪も付いている。

A. 1. 5 足

文楽人形の足を図 A-5 に示す。大腿と下腿の 2 つから構成されており，足遣いは操作部を把持して操作する。



(a) 右腕の各部名称



(b) 右腕の可動域

図 A-3 文楽人形の右腕

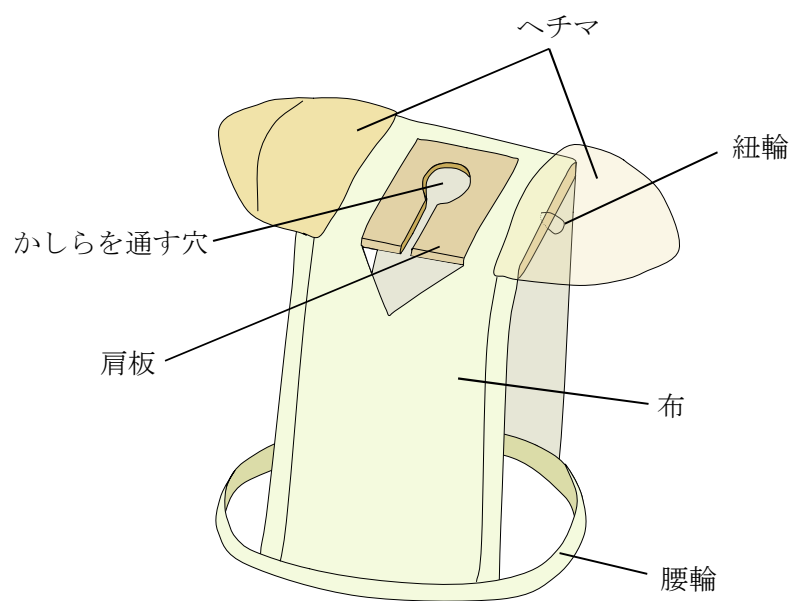


図 A-4 文楽人形の胴

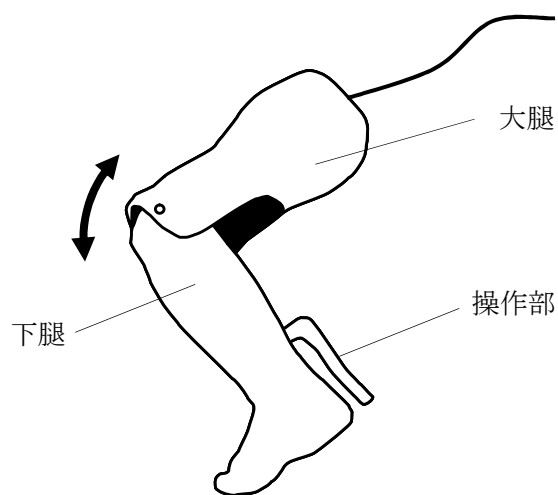


図 A-5 文楽人形の足

A. 2 各パーツの操作法

文楽人形の各部位の操作法について述べる。

A. 2.1 かしらと胴の操作法

主遣いが左手で操作する、かしらと胴の操作図を図 A-6 に示す。かしらは人形の背中から内部に手を入れ、胴串を掴むことで操作する。図 A-6(a)は胴串だけを把持して操作している場合である。このとき、胴串と肩板の間に摩擦力が働くため、かしらを回転させようとする、肩板も一緒に回転してしまう。肩板が回転してしまうと、布で繋がっている腰輪も回転し、胴全体がかしらの動きに釣られて回転してしまう。そこで、胴とかしらを別々に操作するために、図 A-6(b)のように前腕部を腰輪に当て、摩擦力を出すことで、かしらと胴串の間に働く摩擦力による胴の連動を防いでいる。

A. 2.2 左腕の操作法

左腕の操作法を図 A-7 に示す。左腕は左遣いが右手で操作する。差し金と左腕前腕は回転連結されているため、左遣いは差し金からの入力だけでは手部の動きと前腕の位置決めを行うことができないはずである。そのことから、左遣いは胴との連結されている紐の張力や、上腕部の衣装との摩擦を用いて、左腕を自由に操っていると考えられる。

このような操作しにくい仕組みになっているのは、文楽が三人遣いであることから、それぞれの人形遣いの操作位置の制約が大きいことが考えられる。

A. 2.3 右腕の操作法

右腕の操作法を図 A-8 に示す。右腕は、主遣いが右手を用いて操作する。右腕は、前腕部を握り、小猿を操作することで手部の操作ができる。左腕とは異なり、操作する右腕のみで人形の右腕の状態を完全に決定することができる。

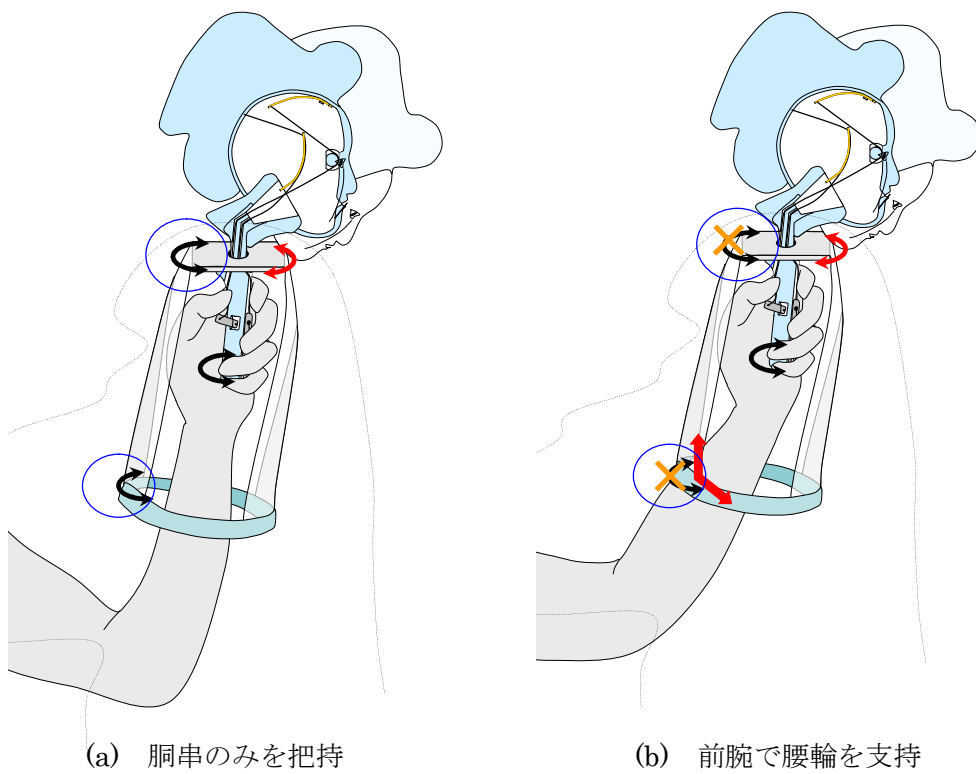


図 A-6 かしらと胴の操作法

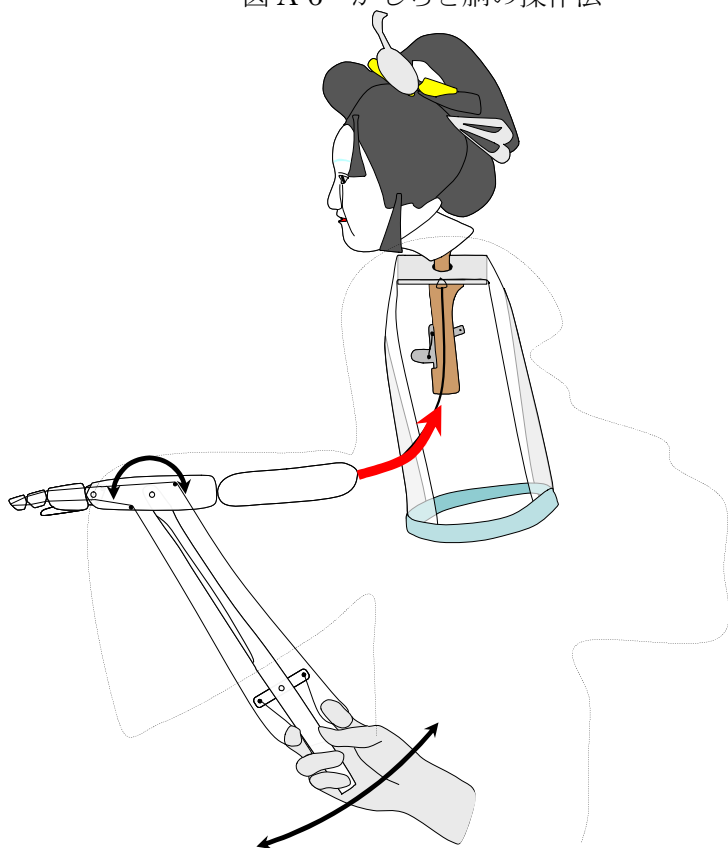


図 A-7 左腕の操作



図 A-8 右腕の操作

A. 2. 4 足の操作法（立役）

文楽人形において足は、基本的には立役（男性の人形）にしかない。立役における足の操作法を図 A-9 に示す。足は下腿についている操作部を握ることで足遣いが操作する。足は、大腿と下腿が回転連結しているため、下腿の操作だけでは大腿の位置決めが不可能である。これに対し足遣いは、胴から連結している紐の張力や、足の衣装との摩擦を利用することで大腿の位置決めを行っていると考えられる。

A. 2. 5 足の操作法（女方）

文楽人形では原則的に、女方（女性の人形）には足が無い。そのため、足遣いは、着物の裾を掴むことで、足があるように表現している。足の操作法を図 A-10 に示す。人形の着物の下部には、足遣いがつかめるように、裾の部分が段になっている。足遣いは人形の後方から手を入れ、このつまみをつかむことで操作している。このとき、右手は人差し指と中指でつまみ、左手は親指と人差し指でつまむ。また、足遣いの手が見えないように、着物の内側には腰巻と呼ばれる布が取り付けられている。

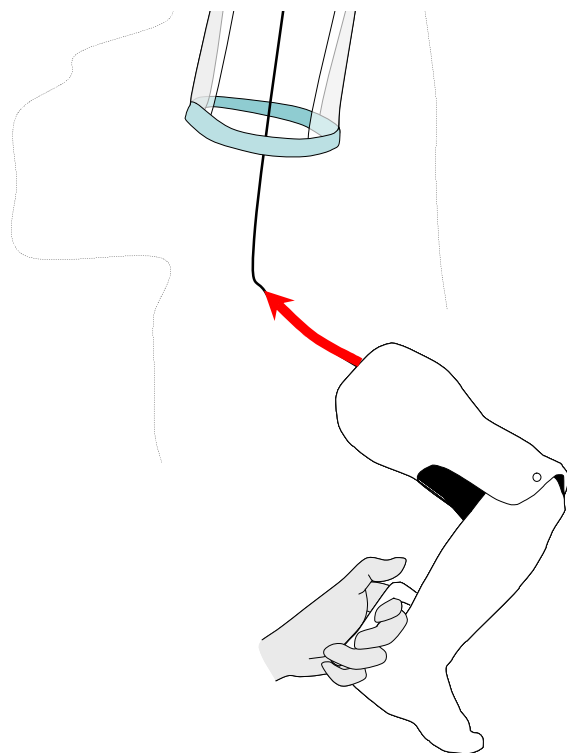


図 A-9 足の操作法（立役）

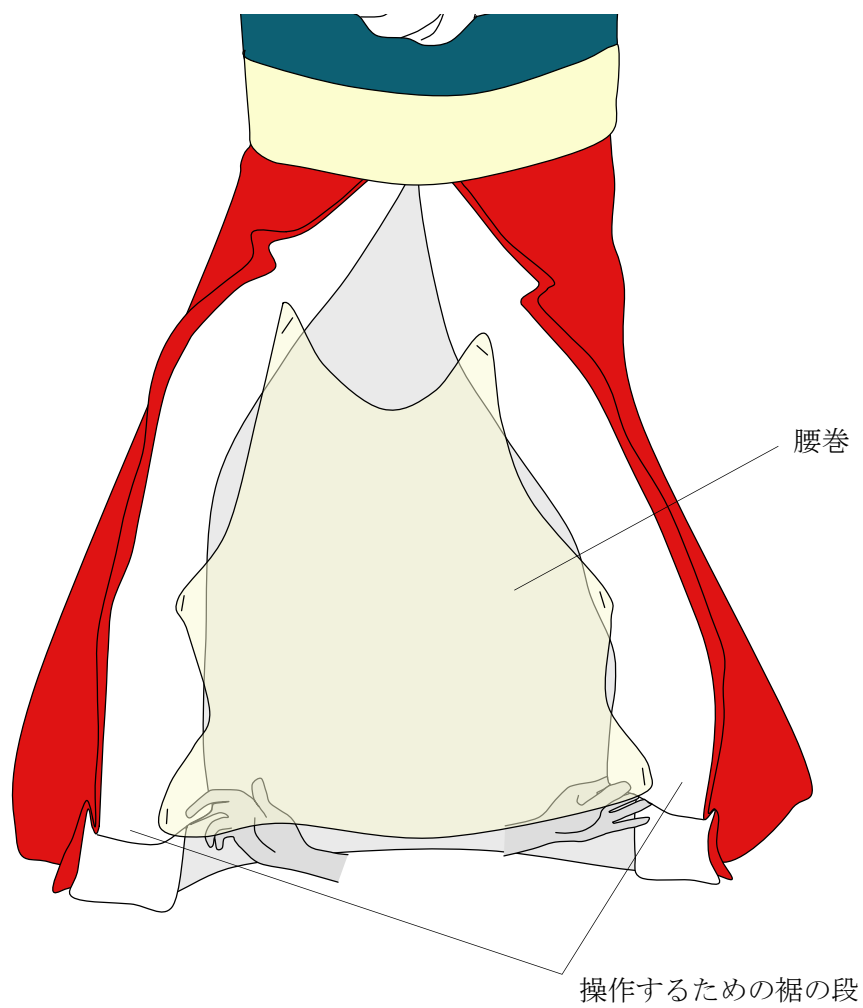


図 A-10 足の操作法（女方）

A.3 センサ配置詳細

本研究で開発した文楽人形“政岡”の組み立ての全体図と埋め込まれたセンサを図 A-11 に示す。

センサ配置の決定にあたっては、現役の人形遣いと人形師との相談の上、着物のような柔軟物でない剛体部位であること、主遣いと左遣いが操作する各部位の重要な動きが取得できること、センサ本体やケーブルによって人形遣いの操作感に影響を与えないことを条件に決定した。その結果、A.1 および A.2 から、主遣いと左遣いが把持して操る部位とワイヤ駆動によって動く部位、それに胴体を加えた全 9 点を計測箇所とした。

A.4 文楽人形のモデル化

文楽人形をリンクモデル化したものを図 A-12 に示す。リンクは太い黒線で示し、操作部は赤い線で示している。また、青い線は布・綿を表しており、黒い点線は紐を表している。ここでは、連動する指の動きなどは一つのリンクとしてモデル化している。

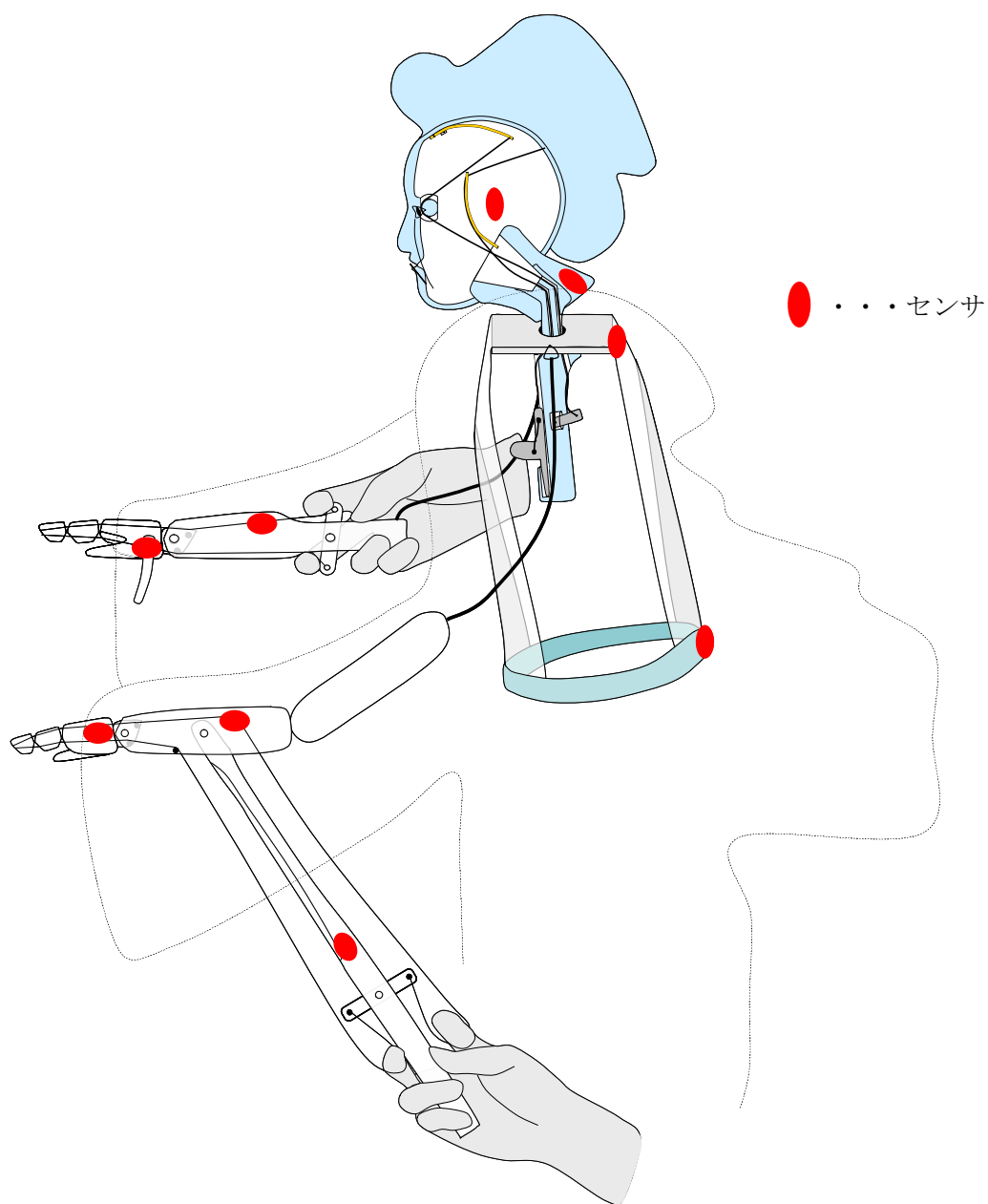
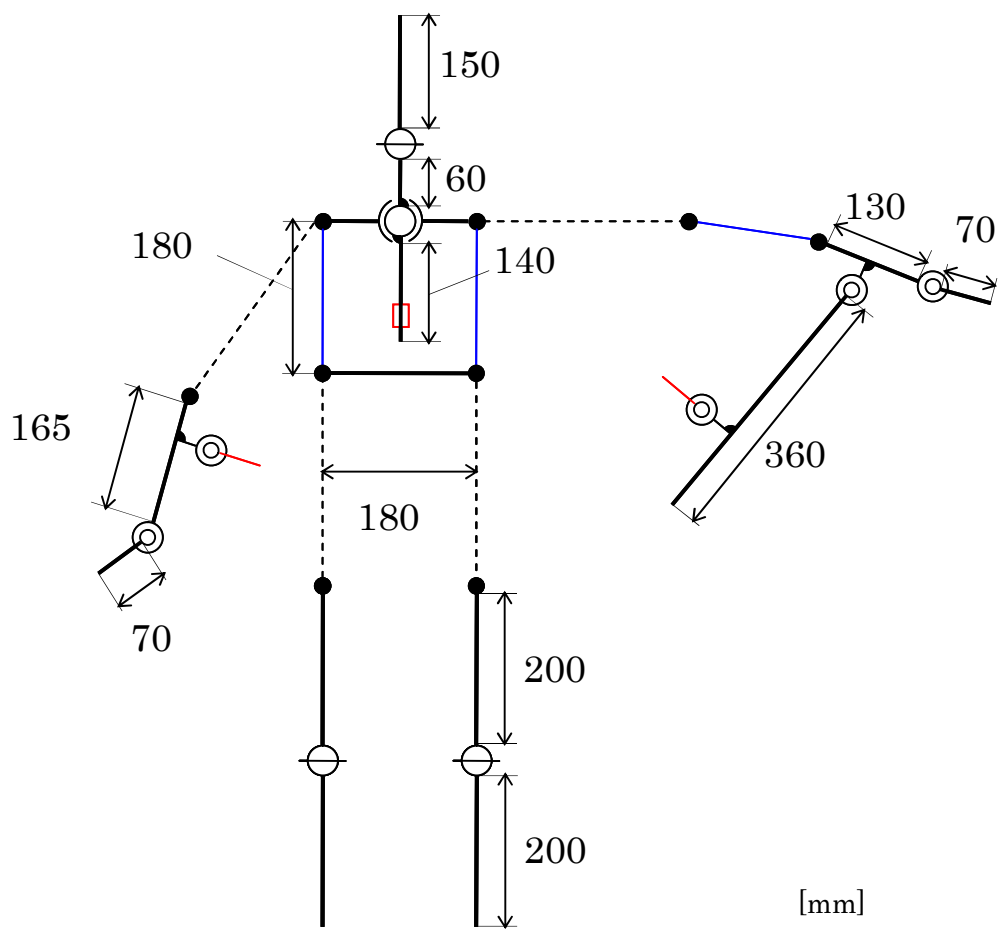


図 A-11 “政岡” の組み図とセンサ配置



[mm]

図 A-12 文楽人形のリンクモデル

付録 実験データの補足

B.1 HMM による状態の遷移と動作波形の時間変化の例

実験で用いた 10 の型の動作波形と，HMM によって符号化された状態の推移を図 B-1-2 に示す．型の内容については，本文の表 3-1 を参照のこと．

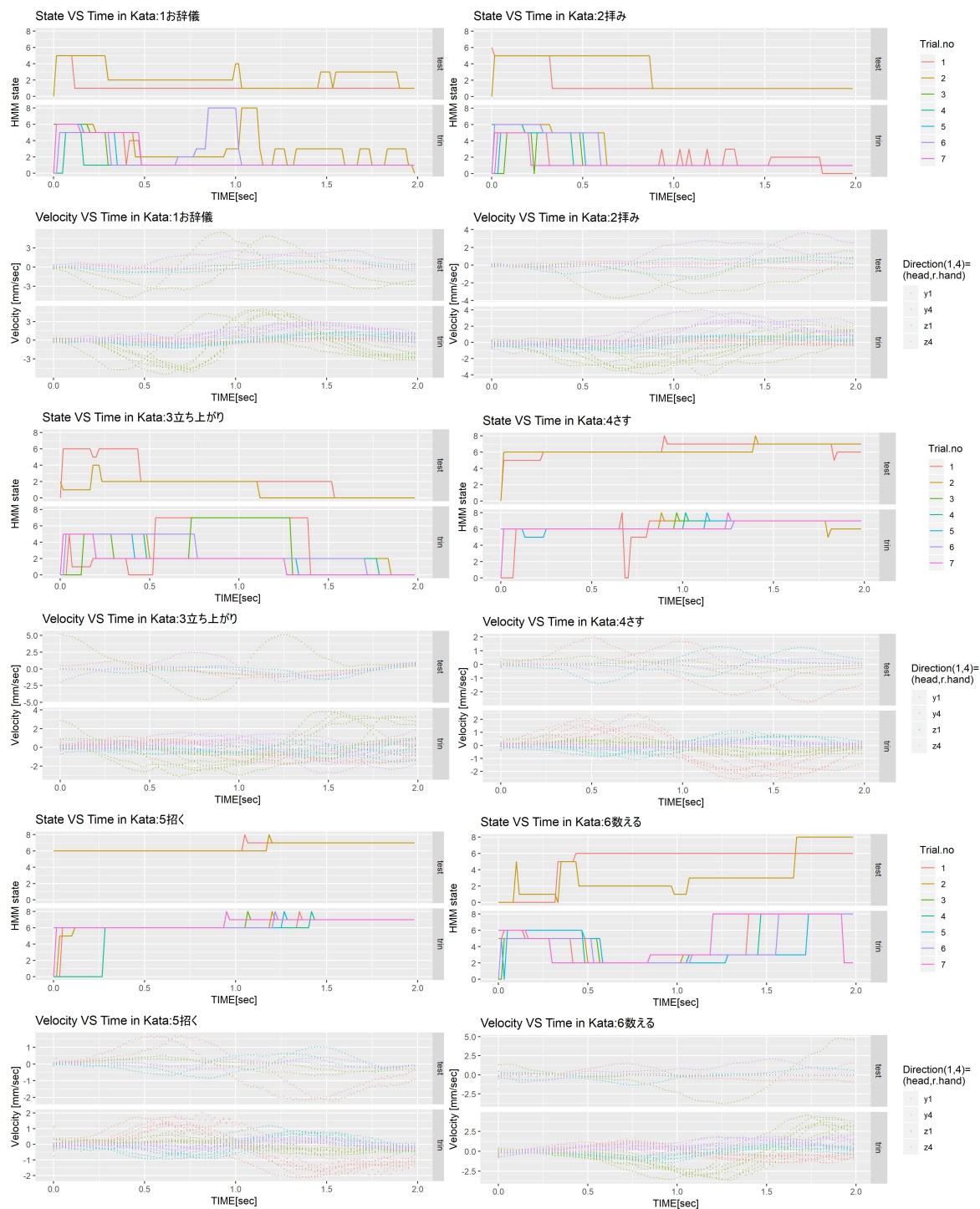


図 B-1 各型での状態の遷移（上段）と動作波形の時間変化（下段）の例 1

時間（横軸）に対して，上段の HMM State が HMM による状態の推移，下段の Velocity が人形の各部位の Y, Z 軸方向の移動速度を示す．人形の部位はそれぞれ頭部=1，右手=4 とし，オレンジ色（y1）は，頭部 Y 軸方向の移動速度を示す．

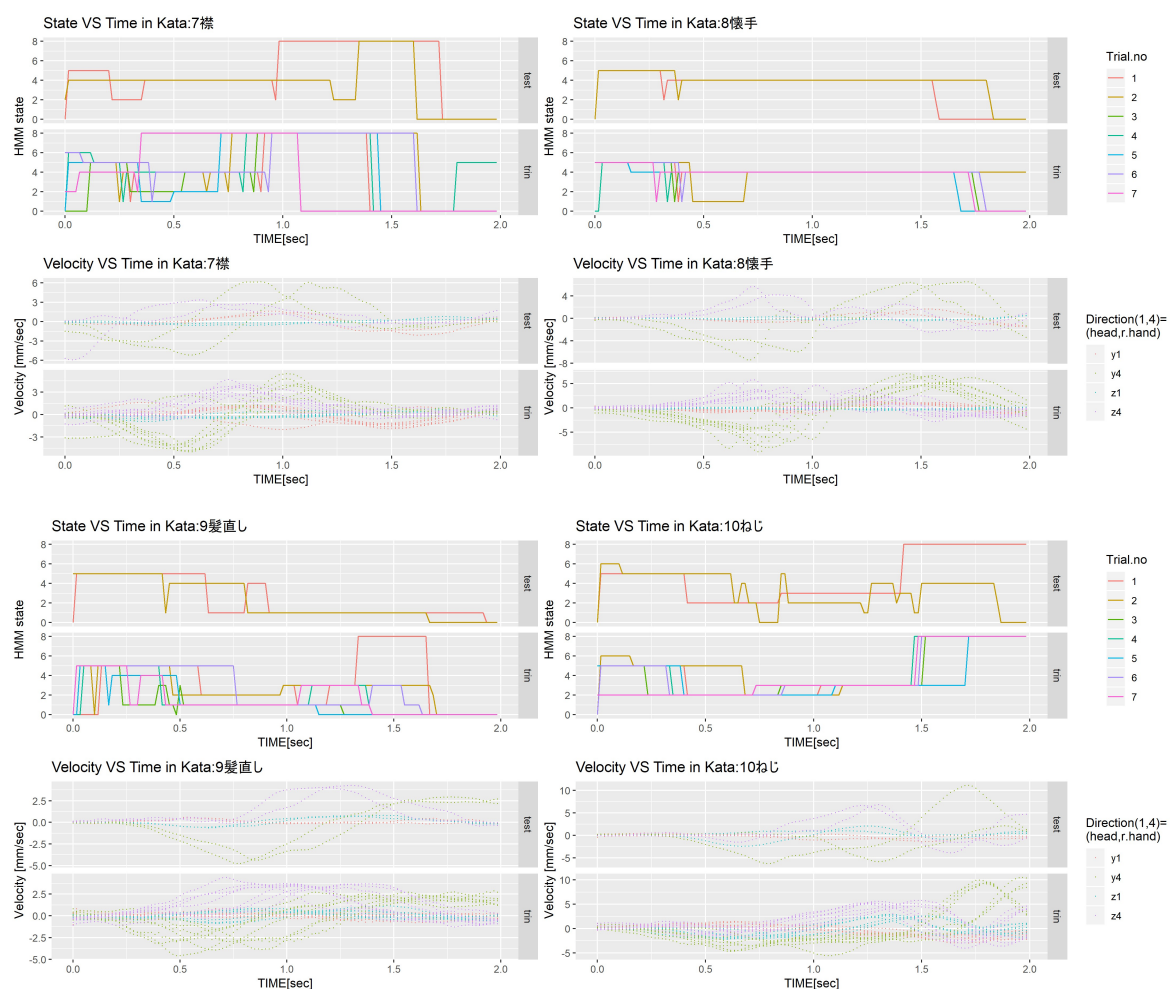


図 B-2 各型での状態の遷移（上段）と動作波形の時間変化（下段）の例 2

時間（横軸）に対して，上段の HMM State が HMM による状態の推移，下段の Velocity が人形の各部位の Y, Z 軸方向の移動速度を示す．人形の部位はそれぞれ頭部=1，右手=4 とし，オレンジ色（y1）は，頭部 Y 軸方向の移動速度を示す．