

伝統芸能文楽における人形遣い間の協調の分析とモデル化

代表研究者	植 田 一 博	東京大学 大学院総合文化研究科 教授
共同研究者	櫻 哲 郎	東京大学 大学院総合文化研究科 特任研究員
共同研究者	渋谷 友 紀	独立行政法人高齢・障害・求職者雇用支援機構 障害者職業総合センター 研究員

1 背景と目的

人形浄瑠璃文楽（以下、文楽）は、大阪で成立し本拠地とする、浄瑠璃と人形によって演じられる人形浄瑠璃の系譜の一つである。重要無形文化財に指定され、ユネスコの無形文化遺産に登録されている、日本を代表する伝統芸能の一つである。文楽は、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ている。そのうち人形操作については、一体の人形を三人の人形遣いが遣う三人遣いという、世界的にも珍しい操作方法が採られており、それにより多彩かつ情緒深い動作が実現されていると言われていた（松平, 2003）。その際に、人形の頭と右手を遣い、全体の司令塔でもある主遣いから、人形の左手を遣う左遣い、および足を遣う足遣いに「ず」と呼ばれる合図が送られていると言われていた。「ず」は人形動作自体に隠されており、発話や事前の打ち合わせなしでも、様々な動作を演技中にその場で決定、実行できる特徴があると言われていた（後藤, 2002; 櫻・成田・沢井・森田・植田, 2013）。この点で、文楽人形操作以外の協調タスクと一線を画すと考えられる。その一方で、このような「ず」を介した人形の協調操作は暗黙的な技に支えられ、習得のためのマニュアルも存在せず、人形遣いたちは10年以上に亘る修行によってこの技術を体得すると言われていた（後藤, 2002）。本研究では、主遣いと左遣いの間の協調に焦点を絞り、「ず」の合図としての役割と構造を、左遣いに対する行動実験と、実際の人形動作の分析により明らかにすることが目的である。具体的には、全体の司令塔である主遣いの意図を、特に主遣いが行おうとしている動作（型と呼ばれる動作パターン）の種類を、人形動作中に埋め込まれている「ず」と呼ばれる合図から、左遣いがどのように判断しているのか、そして人形動作にそのような判断を実現できる情報（「ず」の要素）が含まれているのかどうかを分析した。そのために、左遣いの型の判別について、2-1 で述べるような仮説を立て、演技状況を再現した左遣いの型判別に関する実験（詳細は2節を参照）、および隠れマルコフモデルによる人形動作の分析（詳細は3節を参照のこと）を行い、左遣いの型判別の仕組みについて検討した。

2 左遣いの型判別に関する実験

2-1 目的と仮説

左遣いが「ず」からいかに動作の型判別を行っているのかを検討するための仮説として以下の3つを立て、これらを実験的に検討した。

仮説1：左遣いは時間的に徐々に「ず」が内包する動作の型を判別している。

仮説2：動作の型によって左遣いの判別のタイミングが異なる。

仮説3：経験によって判別の正確性が上がる。

2-2 方法

（1）実験参加者

現役の人形遣いで芸歴30年以上の熟達群5名と15年以下の準熟達群6名の合計11名が実験に参加した。

（2）実験条件

本実験では、型の冒頭部分の左遣い視点の動画を各実験参加者に一定時間提示した上で、行われようとしている型を推定してもらい、その確信度を予め提示した型の回答候補に分配するように求めた。動画提示の対象とした型の種類は8種類で、それに動画提示を行わないダミー2種類を加えた全10種類を、回答候補の型として提示した（表1）。公演で多用される型の他、芸談や動作観察から似ていると判断できる型や、特徴的な動きを含む型を回答候補とした。表1の右側に動作の簡単な説明と、似ている可能性のある型（下線部太字）を記載した。提示する時間は0.75/1.00/1.25/1.50(s)の4種類とした。

表 1 型候補の名称と動作内容

“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。 拝みと類似 。 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。 お辞儀と類似 。 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。 指すと類似 。 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。 招くと類似 。 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) 実験手続き

実験参加者は、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行った (図 1)。①2 秒間の注視点表示の後、左遣い視点の動画 (型の冒頭部分) が提示された。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% になるように割り振った。

型 8 種類と提示時間 4 種をランダム化して 32 回を 1 セットとし、それぞれ同じ条件で録画した別の型の動画を使用した 3 セット分 (計 96 試行) について回答してもらった。なお、確信度の割り振りは 0%~100% の 5% 刻みの 21 段階とした。

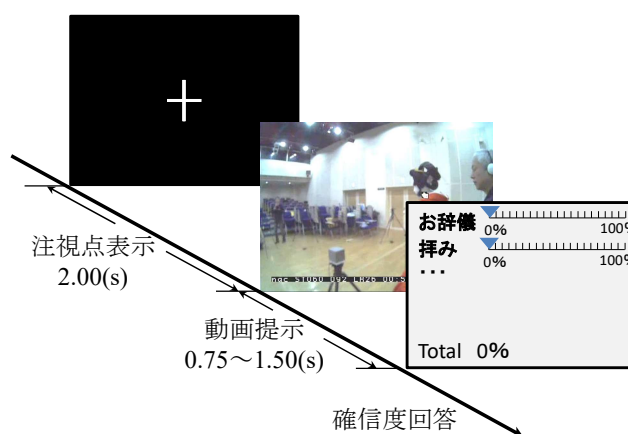


図 1 実験手続き

2-3 分析方法

(1) 熟達度、提示時間、型の種類の影響の分析の方法

2-1 で述べた仮説 1~3 を検討するために、型の推定の正確性を示す指標として、各試行において正解の候補に割り振った確信度の平均値を正解確信度と呼び、従属変数とした。分析では、まず実験参加者の熟達度 (2 水準)、型の提示時間 (4 水準)、型の種類 (8 水準) の 3 要因の影響を一般化線形混合モデルにより検討した。これにより、熟達差や提示時間の長さ、また型の種類が判別に影響するのかどうかを確認した。

(2) 提示時間ごとの型判別の正解確信度のクラスタ分析の方法

舞台上で多種多様な型を素早く判別しなければならないことや、似た動きをする型が存在し左手の動き出

しのタイミングもまちまちであることを考えると、型の動作内容と提示時間には密接な関係があると推測される。そこで、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（正解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、正解確信度の全体の分布に対してクラスタ分析を行った。これにより、左遣いの型の判別が、似た動作特徴を利用して、時間的に進行しているかどうかを確認した。

2-4 分析結果

（１）熟達度、提示時間、型の種類の影響に関する分析結果

判別者の熟達度、型の提示時間、型の種類の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り検討した結果、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。したがって、左遣いは時間的に徐々に「ず」が内包する動作の型を判別しているという仮説１と、動作の型によって左遣いの判別のタイミングが異なるという仮説２は支持されたが、経験によって判別の正確性が上がるという仮説３は支持されなかった。また、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。

図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。熟達度の主効果が有意ではなく仮説３が支持されなかった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、熟達効果も一定程度存在すると考えられる。

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。これらの結果と図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であり、提示時間が長いほど正解確信度が高かったので、型の判別は瞬間的なものではないと考えられる。

以上より、型の提示時間、型の種類、判別者の熟達度の３要因は、特に前二者は、いずれも型判別に重要な要素であり、左遣いはある程度の時間窓の中で判別をしていると推測される。

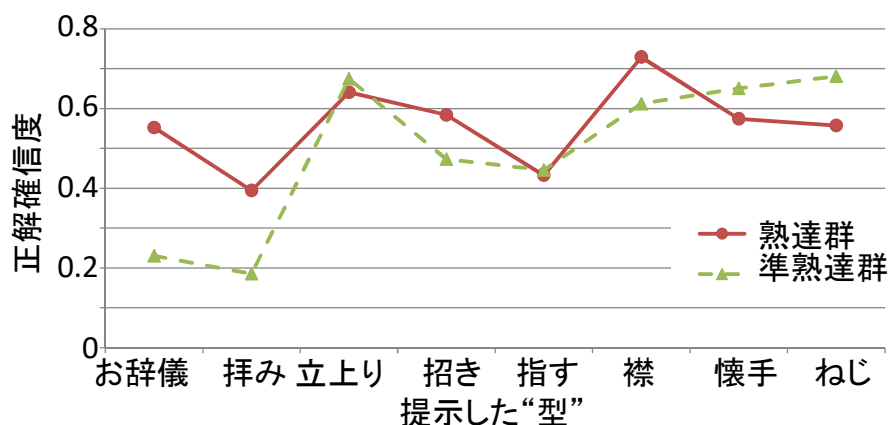


図２ 熟達度ごと、型の種類ごとの正解確信度の平均値

（２）提示時間ごとの型判別の正解確信度のクラスタ分析の結果

提示時間ごとに８個の型に対する正解確信度分布（１０次元）のデンドログラムをWard法で作成した（図３）。似た正解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けば、型の判別が時間とともに似たような動作特徴を利用して進んでいく可能性が高いと考えられる。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だが、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。

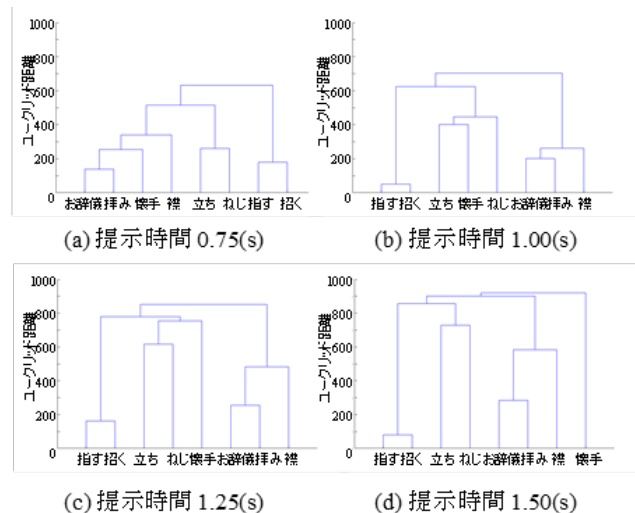


図3 提示時間ごとのデンドログラム

以上から、似た動作を行う型のクラスは、動作判別困難な型のそれとほぼ一致し、提示時間とともに似た動作同士の判別が可能になることが明らかとなった。

2-5 まとめ

以上の結果をまとめると、左遣いによる型の種類の判断には以下の特徴があることがわかる。

- ・ 正解確信度には一部の型のみ熟達差が存在する。
- ・ 提示時間が増えると、正解確信度が大きく伸びる。
- ・ 提示時間が短い間は、似た動作特徴をもつグループに、正解確信度が割り振られる。
- ・ 提示時間が長くなると、似た動作特徴をもつグループの中で、正解確信度が偏っていく。
- ・ 似た動作特徴をもつグループ内では、1.50秒経過しても、型が一意に定まらないことがある。

したがって、左遣いの型の判別は、時間的に進行するが、一定期間の動作から一意に決まるパターンマッチングのようなものではなく、時間とともに似た動作特徴をもつグループから正解を絞るような構造をもっていると考えられる。

3 隠れマルコフモデルによる人形動作の分析

3-1 目的

左遣いによる実際の型判別が時間的に行われていることは、2節の実験により支持されたとと言える。後藤(2002)の主張を併せて考えると、それが、事前の打ち合わせや「床」(太夫による語りや三味線による音楽)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、「ず」が人形動作のどこに存在しているのかを判別することは困難であり、「ず」が何であるかを示す明文化されたマニュアルも存在しない。そこで本節では、このような「絞り込みを実現するような情報」が、人形の型動作の中に含まれているかどうかを機械学習によって検討した。

3-2 分析方法と入力データ

入力データとして、人形各部位の動作データを取得した。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)情報から構成されていた。櫻・成田・沢井・森田・植田(2013)における分析を考慮して、人形の動作データのうち、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、首、肩、右手にセンサを設け、それらの時系列動作情報を入力データとした。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。2節の実験で検討した10の型に関する動作データをデータ・サンプルとした。これ

ら人形動作データを入力として隠れマルコフモデル（Hidden Markov Model；以下 HMM）に与え、動作要素の組み合わせ（隠れ状態の遷移）の抽出を行った。入力が多次元の連続値のため、混合ガウス分布モデル（GMM）を用いた（河原，2016）。

データ数が十分ではないため、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（K-fold cross-validation）を行った。学習時は 70 の動作をまとめて状態推定ならびに学習を行い、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。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）のみを利用した。再現性の評価方法として、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。そして、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。状態数を決定する際に、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を評価した。HMM のパラメータ推定には EM アルゴリズム（Dempster, Laird, & Rubin, 1977）を用いた。

もし図 4 に示すような判別が動作の類似性に基づいて行われていれば、抽出される隠れ状態は、動作の類似性を決定する動作要素、すなわち音声認識でいう音声波形から得られる音素にあたるものになると考えられる。そして、それらの特定の順序の組み合わせが、単語のような意味、すなわち人形動作の「型」としての意味を持つはずである。その場合、左遣いは、型ごとに異なる動作要素（隠れ状態）の遷移によって、主遣いによってどの型が行われようとしているのかを絞り込みことが可能になるはずである。そのような隠れ状態とその遷移が抽出できるかどうかを検証した。

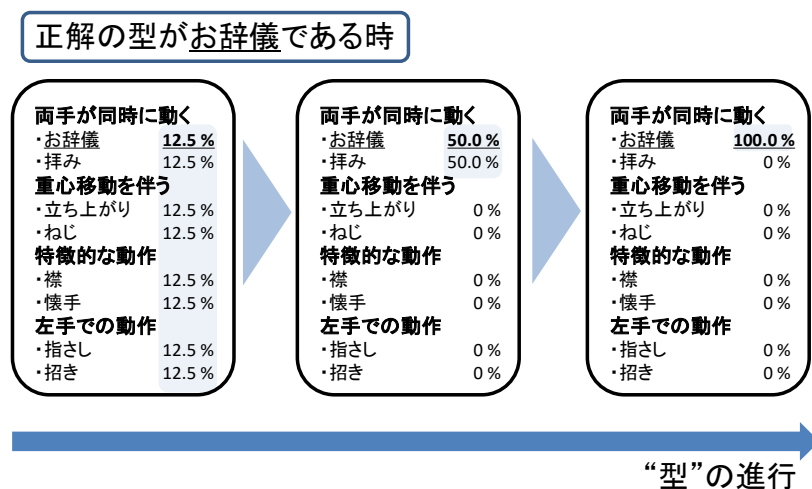


図 4 期待される左遣いの判断の時間変化

3-3 分析結果

HMM の学習にとって最適な状態数を、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）、および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から検討した結果、状態数 9 が最も適当だと言えることがわかった。そのため、以下では状態数 9 での計算結果を示す。

図 5 に実際の動作波形と、HMM によって抽出された状態遷移の一例を示す。この図では、頭と右手の Y, Z 座標の速度のみを示しているが、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 5 において、実際の動作波形と抽出された状態遷移とを比較すると、人形の各部位の動き方と状態と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」）では、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。詳細は省略するが、すべての型動作について上記が再現性を有する形で確認できた。

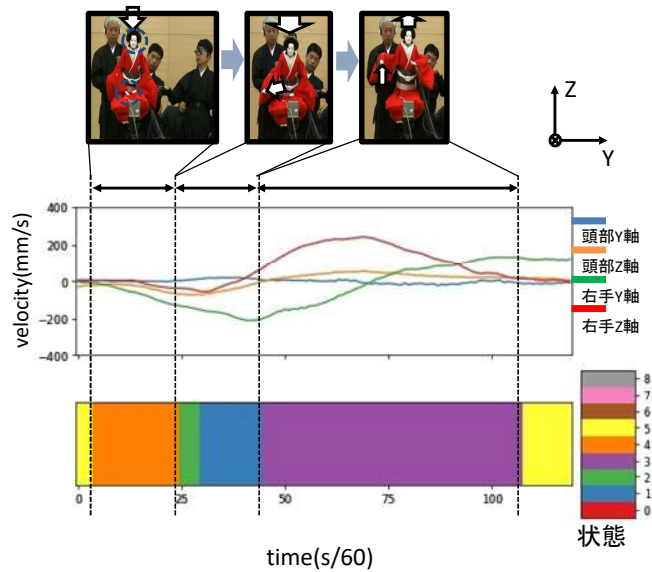


図5 人形動作と状態遷移の関係
上から連続写真、人形の部位速度、HMMによる状態遷移

すべての型について、型ごとに抽出された状態の推移（順序）を調べると、状態推移そのものは複数の動作試行間で類似していたが、試行ごとの動作のタイミングなどの違いにより、それぞれの状態の長さや開始のタイミングは異なっていた。型の種類の判別には順序が重要だと考えられるため、この状態推移に対し連長圧縮（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抽出した。その上で、型ごとに各状態の出現頻度を計算し、型同士のレーベンシュタイン距離（36通りの学習の平均）を求めた。その結果を図6に示す。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がわかる。また、「指す」と「招く」のような類似した型動作の距離がある程度近くなっていること、「指す」および「招く」と、これら以外の型との距離が遠いこともわかる。この結果は、状態遷移には型判別を可能にするだけの情報が含まれることを支持するものと言える。

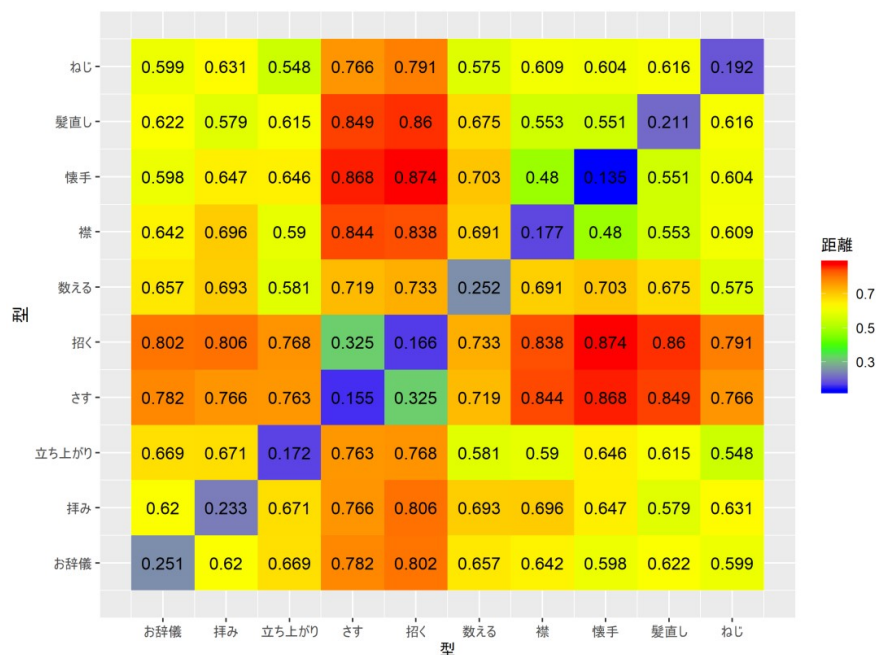


図6 型同士のレーベンシュタイン距離

3-4 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、型の分類を可能にしている人形動作の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、人形動作には、図 4 に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と言える。

4 結論

本研究では、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。その上で、人形動作の隠れマルコフモデルによる分析を行い、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（2 節）の結果、型の判断は、その種類ごとに時間的に進行すること、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（2 節）の結果、型ごとの動作特徴が型判別において重要なことが明らかとなった。
- ・ 隠れマルコフモデルを用いた状態（動作要素）の抽出とその遷移の分析（3 節）の結果、型判別が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、文楽の人形操作では、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、主遣いから手掛かりとなる細かい動作要素の連続情報を受け取りながら、実施すべき型を推定して操作を続けている可能性が高いことが明らかになった。このように、文楽のもつ特殊性である、人形の協調操作を実現する仕組みを、主遣いと左遣い間の情報伝達の特徴と、型動作の特性の観点から明らかにできた。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行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、そこに含まれる「ず」の組み合わせによって、左遣いに難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それに基づく（アドリブを含めた）演技動作を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知る限り存在せず、このことが、協調操作の習得に長い時間を要する理由となっている可能性が考えられる。

【参考文献】

- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society (Series B)*, 39(1), 1–38.
- 後藤 静夫 (2002). 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』, 86, 281–293.
- 河原 達也 (2016). 『IT Text 音声認識システム (改訂 2 版)』. オーム社.
- 松平 盟子 (2003). 『文楽にアクセス』. 淡交社.
- 櫻 哲郎・成田 友紀・沢井 和也・森田 寿郎・植田 一博 (2013). 文楽人形遣いの協調操作を実現する非言語情報通信. 『電子情報通信学会和文論文誌 D』, J96-D(1), 195–208.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
文楽人形遣いの非言語情報を用いた“型”動作の実時間判別	電子情報通信学会 HCG シンポジウム 2019	2019 年 12 月 11 日