

文楽人形操作における“型”動作伝達の分析：人形動作に含まれる非言語的合図の検討

An Analysis of “Kata” Motion Transmission in Bunraku Puppet Manipulation: Examination of Nonverbal Cues Embedded in Puppet Movements

櫻 哲郎[†], 渋谷 友紀[‡], 植田 一博[†]

Tetsuro Sakura, Tomonori Shibuya, Kazuhiro Ueda

[†] 東京大学, [‡] 障害者職業総合センター

The University of Tokyo, National Institute of Vocational Rehabilitation

sakura-999@g.ecc.u-tokyo.ac.jp

概要

本研究では、文楽における三人遣いの協調操作に着目し、主遣いの意図（次に行う“型”動作）を左遣いがどのように人形動作から読み取るのかを分析した。具体的には、左遣いが取りうる型の候補を絞り込む仮説を立て、現役の人形遣いによる型判別課題と人形動作のHMM解析によって検討を行った。その結果、左遣いの判断は固定的なルールに依存せず、人形動作に内在する情報をもとに柔軟に行われている可能性が示唆された。これは、文楽の持つ即興性に適応した仕組みと考えられる。

キーワード：協調操作, 非言語情報, 伝統芸能, 文楽

1. はじめに

文楽は、三味線による演奏、太夫による語り、人形による芝居の三要素から成る日本の伝統芸能である。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 (Goto, 2016)。

この協調操作は、発話や事前打ち合わせを用いることなく、人形動作自体を合図の媒介とし、即時的な意図共有と動作の実行が可能となっている点に特徴がある。このような形態は、joint action (Sebanz et al., 2006) や interpersonal coordination (Keller et al., 2014) といった既存の協調行為研究の枠組みに部分的に含まれる一方で、意図共有に使われるチャンネルの特異性、動作のバリエーションの多様さ、そして即興性の高さから、それらとは異なる構造的特性を有する可能性がある。さらに、当事者間のみで意味を持つ合図がやりとりされる点で covert signaling (Pinker et al., 2008) の概念とも親和性を持つが、従来の covert signaling 理論が主に社会的地位や属性の選別的開示に関わるのに対し、本研究では即時的かつ動作ベースの非言語的合図による協調という、より動的かつ手続き的な側面に焦

点を当てている点で独自の知見を提供しうる。

しかしこの協調技術は、暗黙的に伝承され、明確なマニュアルも存在せず、人形遣いたちは10年以上に及ぶ修行を通じて体得するとされる。なかでも、本研究では左手を操作する“左遣い”に注目する。左遣いは、両足を操作する足遣いと異なり全体の意思決定を行う主遣いと身体の接触を持たず、やや離れた位置から人形全体の動作を観察し、主遣いの操る頭部や右腕の動作中に埋め込まれた“ず”と呼ばれる合図から次に行う動作（型）を判断していると考えられ、先行研究では左遣いの視線計測により演技中は頭部を中心とした人形動作に注視点が集中していることが確認されている (櫻 et al., 2009)。

本研究では、この“ず”が左遣いの判断を可能とする情報をどの程度含むのかを明らかにすることを目的とし、主遣いの意図（型）を左遣いがどのように読み取るかに着目した。まず、左遣いの型の判別に関する仮説を立て、演技状況を再現した行動実験を実施した。その上で、一般化線形混合モデルとクラスタ分析を用いて判別の仕組みを検討し、さらに実測された人形動作に対して隠れマルコフモデルによる状態表現を行い、合図としての構造的特徴を分析した。

2. 方法

左遣いによる型の判別の仕組みについて検討するため、まず演技中の人形動作の特徴に基づいて型の候補を絞り込む仮説モデルを構築した (図1)。

このモデルは、演技上の制約（多くの型で頭や右手がわずかに先行するが、違和感のない演技のためには左手も早期に動き出す必要がある）、主遣いの操る人形頭部や右手の動作観察（型の冒頭が一部類似するものがある）、および合図となりうる情報を統制した先行研究 (櫻 et al., 2013) の結果に基づいている。これらを踏まえ、左遣いは主遣いの動作特徴から、次の型

を段階的に絞り込む判別を行っていると考えられる。また、人形操作には高度な熟練を要することから、①左遣いの熟達度、②観察時間、③型の種類が判別精度に影響するという仮説が導かれる。

この仮説の検証のため、まず行動実験を実施した。左遣い視点の動画（型の冒頭部分、0.75、1.00、1.25、1.50 秒）を提示し、現役の人形遣い（熟達群 5 名、準熟達群 6 名）に対して型の推定課題を行った（図 2）。判別者は 10 種の候補（2 種のダミーを含む、表 1）から、確信度を割り当てる形式で回答した。型候補は類似動作を含むものをベテランの助言のもと選定した。実験では、熟達度（2 水準）、提示時間（4 水準）、型の種類（8 水準）を操作し、1 名あたり 96 試行分の正解確信度に対して一般化線形混合モデル（GLMM）を用いて分析を行った。また、確信度の分布にクラスタ分析を適用し、型の混同傾向や、識別の手がかりとなる動作特徴の探索を行った。

さらに、解析的アプローチとして、文楽人形の動作波形を対象として、各型に対応する動作要素の時系列構造を抽出するために、ガウス混合モデル（GMM）を用いた隠れマルコフモデル（GMM-HMM）を用いた。本手法は、音声認識をはじめとする時系列データ解析で広く用いられている (Rabiner, 1989)。具体的には、人形の頭部・首部・肩部・右手に内蔵されたモーションキャプチャによって得られた 6 軸（位置・回転）情報を、時間に応じた離散状態（隠れ状態）系列に写像し、型ごとの状態遷移パターンの構造を分析した。今回は事前に型の数が明確であることから、複数の状態数で検証し、モデルの妥当性を ELBO (evidence lower bound) と型ごとの再現率で評価して状態数を決定した。パラメータ推定には EM アルゴリズムを用いた。再現率の評価には、訓練・検証データそれぞれで推定された状態系列の遷移パターン（状態の出現順）を比較し、レーベンシュタイン距離に基づいて類似性を定量化した。状態系列の比較に際し、動作速度の揺らぎに対応するため、状態系列を連長圧縮したうえで距離計算を行った。なお、データ数が限られていたため、10 種類の型動作を 1 セットとした 9 セットのデータを、学習用 7 セットと検証用 2 セットに分け、9 分割の交差検証（K-fold cross-validation）を実施した。学習時には 70 動作をまとめて状態推定・学習し、検証時には学習済みモデルを用いて 20 動作の状態を推定した。この手順を全 36 通りについて繰り返し、型ごとの再現性も評価した。以上の手順により、人形動作の類似性に基づく判別が可能であること、すなわち動作中に音素のような構成要素が存在し、その順序の組

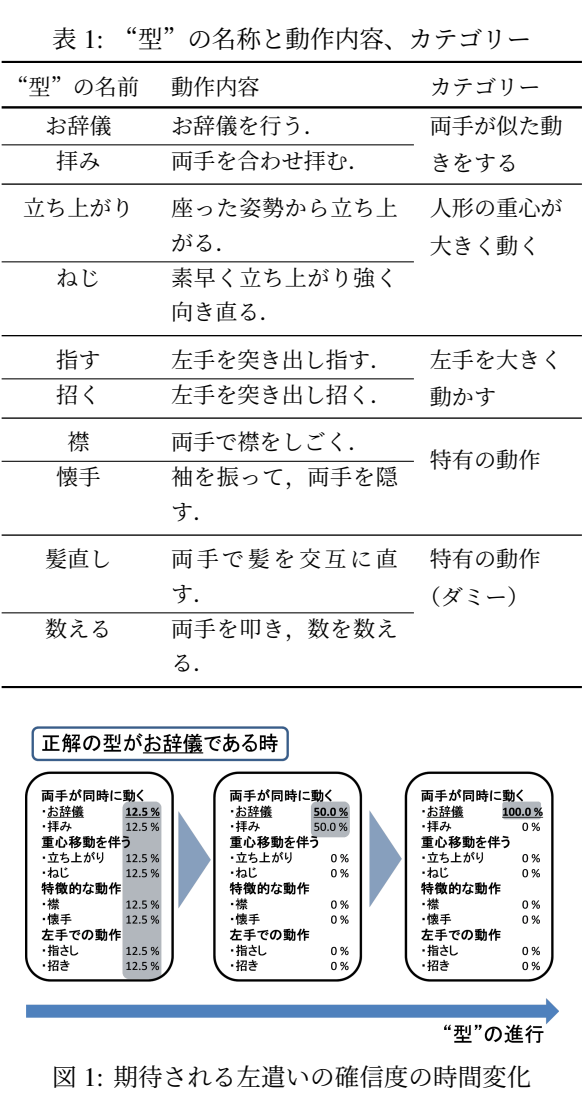


図 1: 期待される左遣いの確信度の時間変化

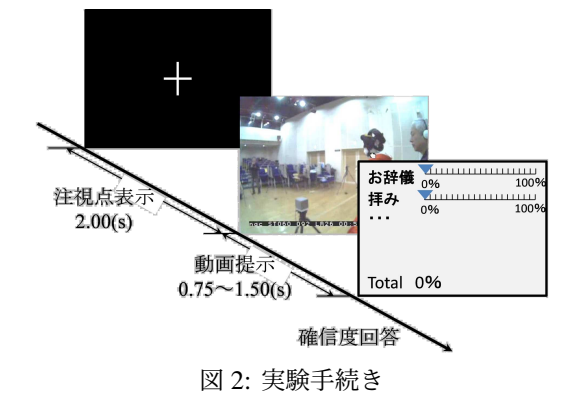


図 2: 実験手続き

み合わせが型の識別に寄与しているかを検討した。

3. 結果

3.1 行動実験の分析

実験結果として、各提示時間ごとの正解の型への確信度（正解確信度）の平均を図 3 に示す。

一般化線形混合モデルの分析の結果、型の提示時間および型の種類の主効果が有意であった（提示時間： $F(3,27) = 37.52, p < .001, \eta^2_G = .09$ ；型の種

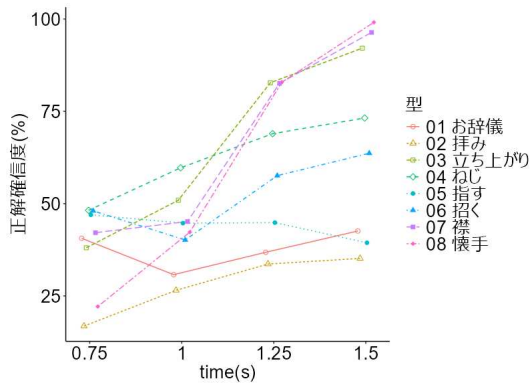


図 3: 提示時間ごと、型ごとの正解確信度の平均値

類: $F(7, 63) = 9.28, p < .001, \eta_G^2 = .10$). 一方、熟達度の主効果 ($F(1, 9) = 2.46, p > .1, \eta_G^2 < .01$) および二次の交互作用 ($F(21, 189) = 0.675, p > .1, \eta_G^2 = .01$) は有意ではなかった. このことから、提示時間が長いほど正解確信度が上昇し、またその効果は型によって異なると言える. 実際に、熟達度×型の種類 ($F(7, 63) = 2.78, p < .05, \eta_G^2 = .03$) および提示時間×型の種類 ($F(21, 189) = 4.38, p < .001, \eta_G^2 = .08$) の交互作用は有意であった. 一方で、熟達度×提示時間の交互作用は有意ではなかった ($F(3, 27) = 0.32, p > .1$).

提示時間×型の種類の単純主効果の検定の結果、“立ち上がり”、“襟”、“懐手”において提示時間の影響が有意だった. Shaffer の方法による多重比較の結果、以下のような提示時間間の有意差が確認された: 襟: 0.75–1.50 s ($p < .05$), 1.00–1.50 s ($p < .05$), 0.75–1.25 s ($p < .1$) 懐手: 0.75–1.50 s ($p < .0001$), 1.00–1.50 s ($p < .001$), 0.75–1.25 s ($p < .05$)

これらの結果から、特定の型では 1.00 s 以降に判別が進行し、確信度が大きく上昇する傾向が示された. また、“お辞儀”の型においては熟達者の正解確信度が高く、熟達によって困難な型に対する判別能力が向上することが示唆された.

3.2 クラスタ分析

提示時間ごとのクラスタリング結果 (図 4) から、似た動作要素を持つ型が早期に同一クラスタに分類される傾向が見られた. 例えば: “お辞儀”と“拝み”(両手を使う)、“指す”と“招く”(左手を突き出す)、“立ち上がり”と“ねじ”(重心移動が大きい). これらは、判断の際に迷いの対象となる型でもあり、同時に確信度が割り振られやすい (= 混同されやすい) 傾向が確認された. さらに、提示時間が長くなるにつれて、クラスタ間・クラスタ内の距離が拡大し、型どうしの判別が進行している様子が見られた. これは、仮説モデ

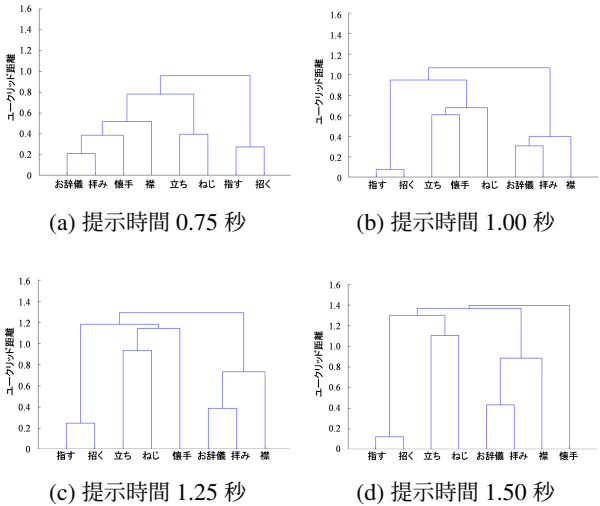


図 4: 提示時間ごとのクラスタ分析結果 (a~d)

ル (図 1) における「動作的な類似性に基づく候補の絞り込み」というプロセスと整合的である.

一方、“襟”や“懐手”のように動作が特徴的な型は、クラスタ構造に明確に組み込まれず、提示時間とともに複数のクラスタを渡り歩く傾向が見られた. これらは最終的には独立した判断に至る点からも、他の型より遅れて識別される型と考えられる. また、一部の型では確信度が 100 % に収束せず、最終的な判断に曖昧さが残るものもあり、判別が「段階的で確率的」な仕組みである可能性が示唆された.

3.3 HMM 解析

モーションキャプチャによる人形動作データを用いた HMM 解析では、各型の動作が型の種類の教示無しで 9 つの状態からなる状態遷移パターンとして表現できることが明らかとなった (図 5).

状態数 K を変化させて学習を行った結果 K=10 で ELBO loss がもっとも低い値をとった. また、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って型の再現性を評価した結果、K=9~10 で十分な収束がみられた. 本研究で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 のモデルを採用した.

図 5 では、人形の頭部と右手の Y, Z 座標の速度変化を時系列で示し、これに対応する HMM の状態遷移を可視化している. 動作波形と抽出された状態との対応をみると、動きの局面ごとに異なる状態が割り当てられていることがわかる. また、“お辞儀”と“拝み”、“指す”と“招く”のような特に類似する動作を行う型では、共通する状態やそこからの分岐などが見られた. これは、左遣いが段階的な判別を行う上で必要な構造的な手がかり (= 音素的な要素) が人形動作

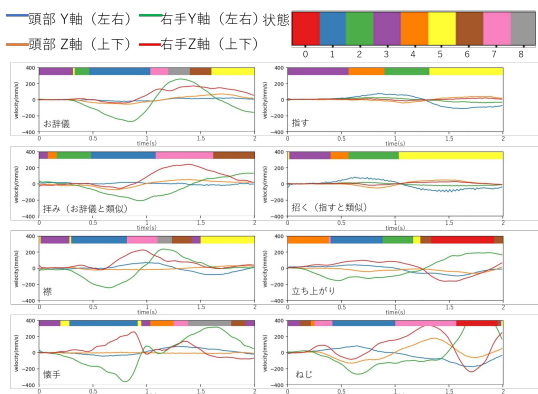


図 5: 人形動作波形と状態遷移の例

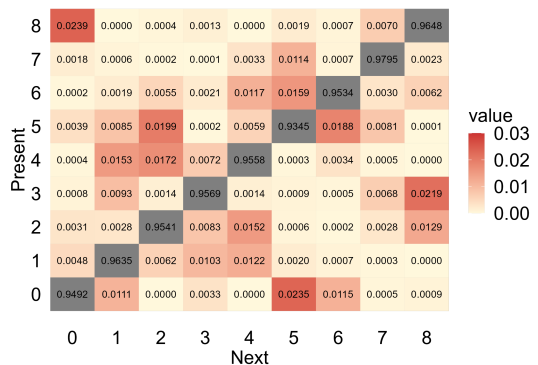


図 6: 状態遷移表

内に存在する可能性を示唆している。状態遷移確率の例（図 6）からは、いくつかの状態間に高い遷移確率が集中しており、動作に一定の進行順があることがうかがえる。

4. 考察

左遣いの型判別は、単純なルールや決まった合図に基づくものではなく、演技中に常に観察と判断を繰り返す即時かつ動的なプロセスであることが示された。確信度が時間とともに増加する一方で 100% には至らないという結果は、左遣いが常に主遣いの動作表現の意味を問いつつ、操作し続ける姿勢によるものと考えられる。これは、文楽の演技が常に即興性を伴い、太夫や主遣いの表現の揺らぎに対応する必要があるという特性に適応した仕組みと考えられる。

また、HMM を用いた動作の状態遷移解析からは、人形動作が音声言語における音素のような、離散的かつ再構成可能な動作要素の組み合わせとして記述可能であることが示唆された。これは、人形動作は単に人の身体運動を模倣するだけではなく、合図の媒介として意味を伝達する“身体言語”としての構造を備えていることを意味する。このような意味構造をもつ動作が、他の人形遣いに対する合図として機能し、それを

もとに型の判別が可能であることが示唆された。

このような高度な協調構造は、一般的な共同作業に見られるような事前の合意や言語的な打ち合わせを必要とせず、身体化された即時的な合図とその解釈によって成立している。これらは、即興的なアンサンブル演奏や熟練職人の“阿吽の呼吸”にも類似する側面が見られるが、演技においては動作要素の一つひとつが意味を担う表現行為であるため、情報の密度が高く、より明示的な意味構造を備えている点で、一線を画す可能性がある。したがって、本研究の知見は、文楽に固有の芸能構造の理解にとどまらず、非言語的協調行動や身体的知性の研究に対する新たな視点を提供するものである。

5. 結論

本研究では、文楽における三人遣いの協調操作の一端である左遣いの型判別の仕組みを、実験と HMM 解析の両面から分析した。その結果、左遣いはあらかじめ定められた固定的なルールに依存した操作ではなく、人形の動作からリアルタイムに合図を読み取り、状況に応じて柔軟に操作していることが明らかとなった。また、その判断を可能にする情報が人形動作そのものに内在していることも確認された。これらの知見は、文楽における即興かつ非言語的な協調が、高度に構造化された技法に支えられていることが示唆するものである。今後の課題としては、こうした判断モデルの定量的再現や、ベイズ推定などを用いた計算論的モデルリングが挙げられる。

文献

Goto, S., “Bunraku: puppet theatre,” in *A History of Japanese Theatre*, edited by J. Salz (Cambridge University Press, Cambridge, 2016), pp. 155–183.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70 – 76.

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130394.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105(3), 833 – 838.

櫻哲郎, 森田寿郎, 植田一博 (2009). 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, 11(3), 255–264.

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. 電子情報通信学会論文誌 D, 96(1), 195–208.

Rabiner, L. R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.