

人形浄瑠璃のマルチモーダル AI 学習データセットおよび大規模モーションモデルの構築

代表研究者

董 然

中京大学 工学部 准教授

1 研究概要

本研究は、日本の伝統芸能である人形浄瑠璃の所作と、音楽・語りに含まれる「序破急」のテンポ変化を、AI が学習可能なマルチモーダルデータセットとして整備し、それを基盤として大規模モーションモデル（Large Motion Model: LMM）を構築することを目的とする。ChatGPT に代表される大規模言語モデルは、言語的対話能力を急速に向上させている。一方、ロボットや CG キャラクターが人間に自然に受け入れられるためには、言語情報に加え、視線、姿勢、動作速度、間合いといった非言語的表現を備えることが不可欠である。

人形浄瑠璃では、精密な顔表情や複雑な機械的関節に依存せず、三味線、義太夫、人形遣いの協働により、観客に強い感情移入を生じさせる。この表現は、物語の展開に応じて緩急を変化させる「序破急」と、主遣いから左遣い・足遣いへ伝達される「ホド」「ズ」といった非言語的合図に支えられている。本研究では、この伝統的なインタラクション原理を、生成 AI、モーション生成、ヒューマンインタフェース、ロボットインタラクションなどの情報通信技術へ接続することを試みた。

助成期間中に、文楽、八王子車人形、阿波人形浄瑠璃を対象として、動作と音楽を対応付けたデータ整理を実施し、3.67 時間、40 種類の所作、120Hz で約 158 万フレームからなる JoruriPuppet データセットを整備した。データには熟練者と初心者者の演技を含め、骨格データおよび SMPL 形式への変換を行った。また、音楽テンポと動作速度の同期を定量化する Jo-Ha-Kyu スコア、滑らかな S 字軌跡を評価する S-curve スコア、頭部と手部の対照的運動を評価する Contrast スコアを提案した。さらに、既存の音楽から動作を生成する 7 種類の SOTA 手法を本データセットで学習・評価し、従来指標だけでは捉えられないテンポ変化に基づく表現の不足を明らかにした。

さらに、文楽における非言語的協調コミュニケーションを対象に、主成分分析（PCA）とヒルベルト-ファン変換（Hilbert-Huang Transform: HHT）を用いて「ホド」と「ズ」の時間周波数特性を解析した。その結果、主遣いの背中・首の動きと左遣いの左前腕の追従関係を、瞬時位相および瞬時周波数として定量的に観測し、熟練した人形遣いが約 0.2 秒程度の遅れを伴いながら同期的な協調動作を実現していることを示した。これらの成果は、文化的身体知をデータ化し、生成 AI に組み込むための基盤を提供するものである。図 1 は、収集した人形浄瑠璃の所作データ、音声、骨格表現、SMPL 表現、およびテンポ変化特徴量を組み込んだ音楽-動作生成の全体像を示す。

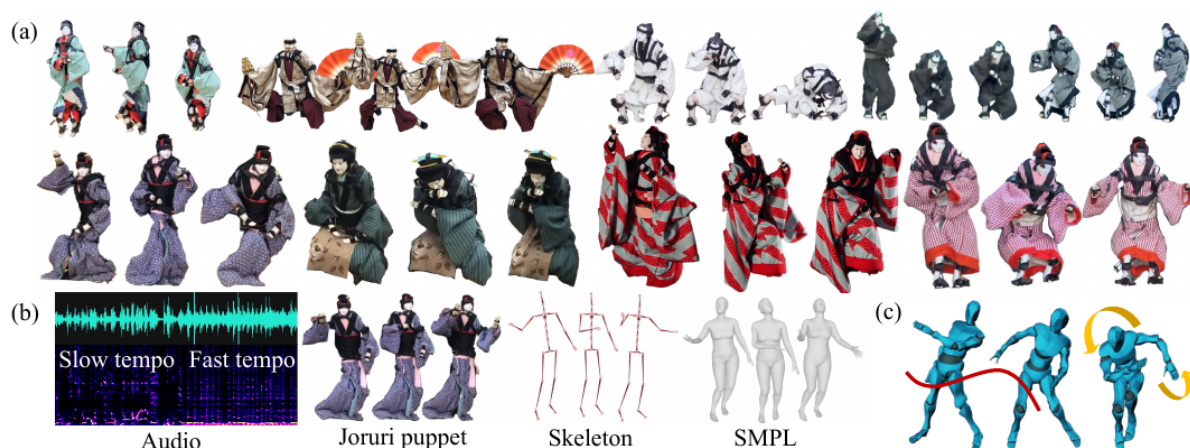


図1 JoruriPuppet データセットと音楽-動作生成の概要。(a) 人形浄瑠璃の演技例と所作データ、(b) 音声・人形画像・骨格・SMPL 表現の対応、(c) テンポ変化特徴量を用いた音楽-動作生成の概念 (Dong et al. [1]より)

2 研究背景と目的

2-1 AI エージェントにおける身体表現の課題

近年、音声対話型 AI や LLM を搭載した AI エージェントは、検索、翻訳、文章生成、教育支援など多様な用途で利用されている。しかし、対話の質を人間同士の相互作用に近づけるためには、発話内容だけでなく、相手の感情や会話の文脈に応じた身体動作、反応速度、間合い、リズムが重要となる。特に、ロボットやデジタルヒューマンを医療・介護・教育・観光・文化産業などの現場で活用するためには、ユーザが自然に感情移入できる非言語的表現を設計する必要がある。

一方で、人間の顔表情や骨格構造を忠実に再現する CG キャラクターや人型ロボットの制作には高いコストが伴う。精密な眼球、表情筋、皮膚変形、全身関節を再現しても、不自然さが残る場合があり、いわゆる *uncanny valley* の問題も生じる。この課題に対して、人形浄瑠璃の所作は重要な示唆を与える。人形は無機物であり、顔表情も限定的であるにもかかわらず、観客は物語の登場人物として受け入れ、強い情動反応を示す。これは、外形の写実性ではなく、運動の緩急、視線に相当する首の向き、手の軌跡、呼吸のような間合いが感情表現を支えているためである。

したがって、本研究の中心的な問いは、人形浄瑠璃に内在する非言語的表現原理を、どのようにデータセットとして記述し、AI モデルに学習させるかである。従来の音楽から動作を生成する研究では、西洋ポップ音楽のような比較的一定の拍子に基づくデータセットが多く用いられてきた。これに対して、人形浄瑠璃の音楽と所作は、物語の展開に合わせてテンポが柔軟に変化し、拍と拍の間の伸縮が感情表現に深く関与する。このようなテンポ変化を扱うデータセットと評価指標が不足していたことが、本研究の動機である。

2-2 序破急と人形浄瑠璃の非言語的協調

「序破急」は、古典芸能において演技や音楽の展開を構成する基本原理であり、ゆっくりと始まり、次第に速度と密度を高め、最後に鋭く収束する構造を持つ。ただし、人形浄瑠璃における序破急は、単純な三分割の形式ではなく、義太夫の語り、三味線の打音、物語展開、演者間の合図が相互に影響し合い、その場で即興的に形成されるリズム変化である。

文楽では、一体の人形を主遣い、左遣い、足遣いの三人が操作する。主遣いは首と右手を担当し、人形全体の方向性や感情表現を統括する。左遣いと足遣いは、主遣いが発する微細な合図を読み取り、人形の左手や足を同期させる。このとき用いられる合図が「ホド」と「ズ」である。「ホド」は協調動作の開始を促す誘導的合図であり、「ズ」は首の回転や傾きによって左手・足の動きを同期させる合図である。これらの合図は観客に明示されるものではなく、演技の流れの中に埋め込まれている。

本研究では、この「観客には見えにくい、演者間では共有される」非言語的情報を、モーションデータと音声データの時間周波数解析により明らかにすることを試みた。これは、AI エージェントがユーザに対して過度に説明的な動作を行うのではなく、相手の状態に応じて自然に反応し、微細なリズム変化を通じて協調するための設計原理を与えるものである。図2は、序破急に基づくテンポ変化の概念と、本研究で用いた IMU センサ配置および骨格表現を対応付けて示す。

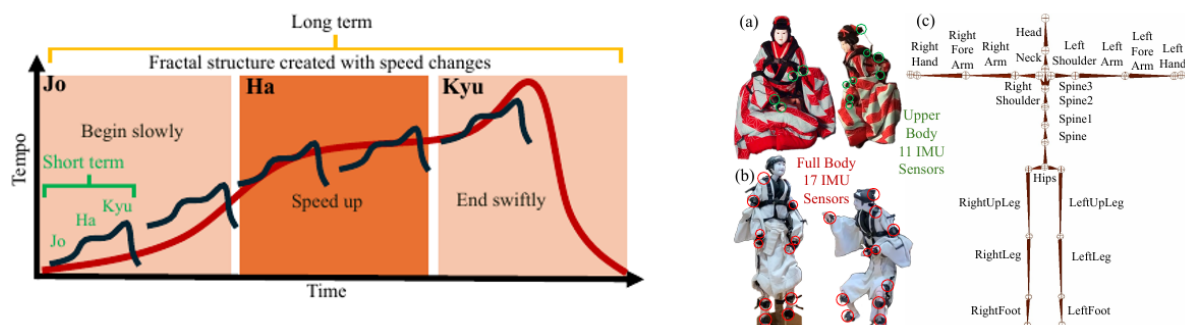


図2 序破急に基づくテンポ変化と IMU モーションキャプチャの概要。(a) 序破急におけるテンポ進行の概念、(b) 上半身型・全身型の IMU センサ配置、(c) JoruriPuppet データセットの骨格構造 (Dong et al. [1]より)

3 研究方法

3-1 データ収集と前処理

本研究では、助成申請時に計画した映像ベースの所作抽出を検討し、SAM2 などを用いた人形領域の抽出および 3 次元骨格推定を試みた。しかし、人形浄瑠璃では人形遣いの身体、衣装、道具が頻繁に重なり、さらに人形そのものが人体の標準骨格とは異なるため、一般的な映像ベースの姿勢推定のみでは安定した骨格再構成が困難であった。そこで、既に確立していた IMU センサを用いたモーションキャプチャ手法を中心に据え、映像データは補助的な教師信号および検証資料として利用する方針に整理した。

IMU センサは、人形の操作点および主要な回転中心に対応する位置へ装着した。女性型の文楽・阿波人形浄瑠璃では足部が衣装に隠れるため、上半身を中心とした 11 センサ構成を用いた。一方、八王子車人形などでは人形遣いの足さばきが人形の下半身運動と密接に関係するため、必要に応じて 17 センサ構成を用いた。これにより、頭部、胸部、肩、腕、手、腰、脚の回転を高い時間分解能で取得した。

取得したデータは、BVH 形式で整理した後、各関節のオイラー角を軸角表現へ変換し、SMPL 骨格へ対応付けた。SMPL 形式へ変換することで、既存のモーション生成モデルや評価指標を適用しやすくなり、他の公開データセットとの比較も可能となる。さらに、同時に記録した三味線・義太夫の音声データから拍位置、BPM、BPM 変化量を推定し、動作データと時間的に対応付けた。

3-2 テンポ変化特徴量と評価指標

従来の音楽から動作を生成する研究では、beat alignment score のように、動作のピークと音楽拍の時間的一致を評価する指標が用いられることが多い。この指標は、一定拍子のダンスやポップ音楽には有効であるが、人形浄瑠璃のように拍間隔そのものが伸縮する音楽では十分ではない。なぜなら、重要なのは「拍に合っているか」だけでなく、「テンポが変化したときに動作速度も対応して変化しているか」であるためである。

そこで本研究では、音楽拍に基づく各区間について瞬時 BPM を算出し、その区間における全身または主要関節の平均動作速度との相関を Jo-Ha-Kyu スコアとして定義した。テンポが速くなると動作速度も高まり、テンポが緩むと動作も緩む場合、この相関は高くなる。さらに、動作の美しさを評価するために、頭部および手部の軌跡が S 字曲線にどの程度近いかを測る S-curve スコア、演劇的な見え方として頭部と手部の対照的な動きを測る Contrast スコアを定義した。

これらの指標は、単なる機械的な同期ではなく、伝統芸能の身体知に根ざした表現を評価することを目的としている。例えば、文楽では頭部が先に動き、手部がわずかに遅れて追従することで、人形が意思を持って動いているように見える。この遅れや軌跡の対比は、動作生成モデルが単に音楽拍に合わせ

て手足を動かすだけでは再現しにくい。したがって、提案指標は、文化的表現の質を AI モデルの評価へ導入するための基礎的試みである。図 3 は、従来の beat alignment が拍との時間的一致を評価するのに対し、提案する Jo-Ha-Kyu スコアが拍間隔内の動作速度変化を評価することを示す。

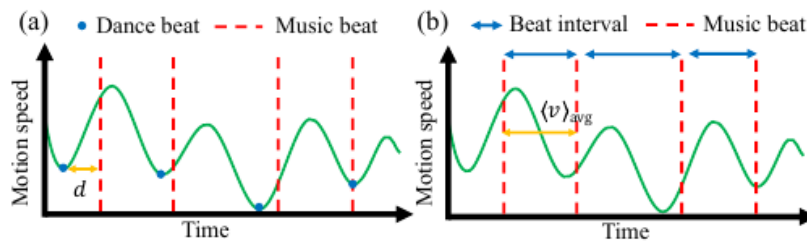


図 3 従来指標と提案指標の比較。(a) 従来の beat alignment 指標、(b) 拍間隔内の動作速度変化を用いる Jo-Ha-Kyu スコアの考え方 (Dong et al. [1]より)

3-3 「ホド」「ズ」の時間周波数解析

人形遣いの協調動作を解析するため、文楽人形の各関節の 3 軸回転データに主成分分析を適用し、所作の主要な回転方向を抽出した。そのうえで、ヒルベルト-フアン変換を用いて、非定常かつ非線形な動作信号を固有モード関数 (IMF) へ分解し、各 IMF の瞬時周波数と瞬時位相を求めた。HHT は、固定基底を用いるフーリエ変換とは異なり、局所的なテンポ変化や短時間の合図を解析するのに適している。

解析対象として、文楽の代表的な所作である「手を合わせる」場面を取り上げた。この場面では、三味線の打音と義太夫の裏拍子を契機として、主遣いが人形の肩・背中をわずかに動かす「ホド」を発し、その後、首の回転によって左遣いへ動きの方向を示す「ズ」を発する。左遣いはこれに追従し、人形の左手を右手へ合わせる。この一連の動作は数秒の短い時間で行われるが、三味線、義太夫、主遣い、左遣いの間には高度な時間的關係が存在する。図 4 は、「手を合わせる」所作において、三味線・義太夫のリズム、主遣いの「ホド」「ズ」、左遣いの追従動作が時間的に連鎖する過程を示す。

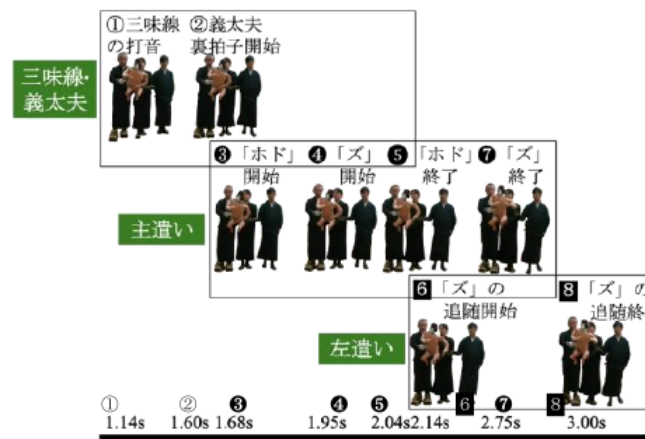


図 4 文楽における「ホド」「ズ」を含む非言語的協調コミュニケーションの時系列 (劉・董・蔡 [2]より)

4 研究成果

4-1 JoruriPuppet データセットの構築

第一の成果は、人形浄瑠璃の音楽と所作を対応付けた JoruriPuppet データセットの構築である。本データセットは、文楽、八王子車人形、阿波人形浄瑠璃という 3 種類の人形浄瑠璃を対象とし、合計 3.67

時間、40種類の所作、120Hzで約158万フレームを含む。演者には、50年以上の経験を持つ熟練者と、学習初期段階の初心者が含まれており、匠の技と学習過程の差異を比較可能な構成とした。

本データセットの特徴は、非ポップ音楽である伝統的な語りと三味線を含み、テンポ変化が所作の中核にある点である。既存の音楽-動作データセットの多くは、一定テンポの西洋音楽やダンス動作に基づいている。それに対して、本データセットは、物語の展開に応じてリズムが伸縮し、動作速度もそれに応じて変化する。したがって、単なるダンス生成ではなく、感情表現や物語性を伴う演劇的モーション生成の基盤として利用できる。

また、データは骨格表現と SMPL 表現の両方で整備した。これにより、従来のモーション解析、深層学習による生成、ロボットへのリターゲットング、CG キャラクターへの適用など、多様な応用が可能となった。データセットおよびソースコードは、研究発表後に公開する方針で整理しており、AI 研究者が伝統芸能に基づく非言語的インタラクションを再利用できる基盤を提供する。図5は、画像系列、骨格表現、SMPL 表現、音声波形およびスペクトログラムを対応付けた学習用データの具体例を示す。

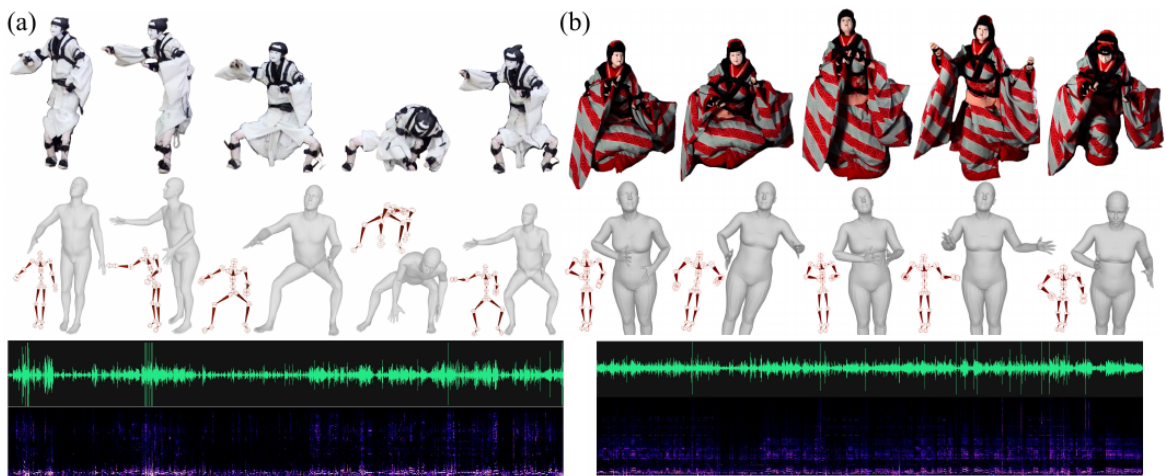


図5 学習用データ例。(a) 人形浄瑠璃所作の画像系列・骨格表現・SMPL 表現・音声スペクトログラムの対応、(b) 別所作における同様のマルチモーダル対応関係 (Dong et al. [1]より)

4-2 序破急評価指標によるデータセット比較

第二の成果は、序破急に基づく新規評価指標の提案と、それによるデータセット比較である。AIST++、FineDance、Motorica などの既存データセットと比較した結果、JoruriPuppet はテンポ変化に対する動作速度の追従が顕著であり、提案した Jo-Ha-Kyu スコアにおいて高い値を示した。特に Madmom を用いた拍抽出では、JoruriPuppet の平均 BPM 変化量は 50.8 ± 21.3 と大きく、平均 Jo-Ha-Kyu スコアは 0.1486 (95%信頼区間 0.0486-0.2456) となった。これは、音楽テンポと動作速度の間に正の相関が存在し、人形浄瑠璃の所作がテンポ変化を強く反映していることを示している。

従来の beat alignment score では、拍に対する動作ピークの近さが評価される。しかし本研究の結果から、テンポ変化を伴う伝統音楽では、拍への一致だけでなく、拍間隔の変化に伴う速度変化を評価する必要があることが明らかになった。例えば、一定テンポのダンスでは beat alignment が高くても、Jo-Ha-Kyu スコアは必ずしも高くない。一方、人形浄瑠璃では、拍の伸縮に合わせて頭部や手部の速度が変化するため、提案指標が表現の本質をよりよく捉える。

この結果は、音楽-動作生成タスクの評価において、文化的・芸術的文脈に基づく指標を導入する必要性を示すものである。特に、伝統芸能、演劇、ジャズ、即興舞踊など、テンポが固定されない表現では、提案した評価枠組みが有効であると考えられる。表1は、既存の音楽-動作データセットと JoruriPuppet を比較し、JoruriPuppet が平均 BPM 変化量および Jo-Ha-Kyu スコアにおいて特徴的な値を示すことを整理したものである。

表 1 JoruriPuppet と既存データセットの比較で用いた主な指標の例 (Madmom による拍抽出の結果を中心に要約)

データセット	平均 BPM 変化量	平均速度変化率	Jo-Ha-Kyu スコア	特徴
AIST++	0.4 ± 0.4	1.3 ± 0.7	-0.3739	一定テンポ傾向が強い
FineDance	13.9 ± 25.6	10.6 ± 9.4	0.0483	一部テンポ変化を含む
Motorica	21.0 ± 26.2	15.2 ± 18.9	0.0326	動作速度変化は大きい が音楽との相関は限定的
JoruriPuppet	50.8 ± 21.3	11.0 ± 6.2	0.1486	音楽テンポと所作速度の同期が強い

4-3 音楽からの所作生成モデルの実装と評価

第三の成果は、構築したデータセットを用いた音楽からの所作生成モデルの実装と評価である。本研究では、FACT、DeepPhase、Bailando、LDA、EDGE、TM2D、Bailando++など、既存の音楽-動作生成手法を JoruriPuppet データセットで学習し、FID、Div、beat alignment といった従来指標に加え、提案した Jo-Ha-Kyu スコア、S-curve スコア、Contrast スコアを用いて評価した。表 2 は、7 種類の SOTA 音楽-動作生成手法を JoruriPuppet で評価した結果を示し、従来指標と提案指標の両方から各手法の特性を比較している。

結果として、従来指標では一定の性能を示す手法であっても、序破急に基づく細かなテンポ変化や演劇的な頭部・手部の関係を再現することは難しいことが分かった。例えば、TM2D は Jo-Ha-Kyu スコアで比較的高い値を示したが、ground truth の表現には及ばない。また、LDA や EDGE など一部の拡散モデル系手法は、全体的なモーション品質と多様性の点で一定のバランスを示したものの、急激なテンポ変化に対する速度調整はなお不十分であった。

この結果は、音楽から動作を生成する研究において、単に音楽特徴量を入力するだけでなく、テンポ変化そのものをモデルが明示的に扱う必要があることを示している。本研究では、瞬時 BPM および BPM 変化量を音楽特徴として追加し、拡散モデル系手法への統合を進めた。これにより、生成された所作が音楽の緩急により敏感に反応する可能性が示された。今後は、LLM による会話文脈理解と、LMM による所作生成を統合し、対話内容に応じて序破急を変化させるリアルタイムインタラクションへ発展させる。

表 2 7 種類の SOTA 音楽-動作生成手法を JoruriPuppet で評価した結果の概要 (Dong et al. [1]より)

Method	FID_g ↓	Div_g ↑	FID_k ↓	Div_k ↑	Beat Alignme nt ↑	Jo-Ha- Kyu Score (r_avg) ↑	Speed change ↑	S-curve Head ↑	S- curve Hand ↑	Contrast Score ↑
Ground truth	64.30	3.0820	51.49	4.5642	0.1897	0.1511	12.02	0.3993	0.4442	0.5172
FACT [Li et al. 2021]	90.21	5.4281	94.73	4.7862	0.1703	-0.1334	4.72	0.4423	0.3949	0.5244
DeepPhase [Starke et al. 2022]	104.05	0.2949	117.14	3.9937	0.2415	-0.0244	2.30	0.4134	0.4186	0.4297
Bailando [Siyao et al. 2022]	83.54	2.1526	66.42	5.1693	0.2078	-0.3388	8.60	0.3335	0.2862	0.3123
LDA [Alexanderson et al. 2023]	64.14	3.4984	50.14	5.3771	0.1010	0.1016	7.72	0.4507	0.4708	0.6124
EDGE [Tseng et al. 2023]	111.61	4.3906	53.83	4.0469	0.1761	0.0382	6.76	0.4567	0.4091	0.3906
TM2D [Gong et al. 2023]	89.98	2.9699	113.68	4.4167	0.1962	0.1451	4.86	0.4296	0.3704	0.5444
Bailando++ [Siyao et al. 2023]	78.57	2.5335	75.41	4.6154	0.1805	-0.0899	2.43	0.3151	0.3149	0.1662

注：太字は生成手法中の最良値を示す。

4-4 「ホド」「ズ」の非言語的協調メカニズムの解析

第四の成果は、PCA と HHT を用いた「ホド」「ズ」の非言語的協調メカニズムの解析である。文楽人形の「手を合わせる」所作を対象とし、三味線の打音、義太夫の裏拍子、主遣いの背中・首の動作、左遣いの左前腕の動作を時系列として対応付けた。解析では、背中の IMF5 が「ホド」に対応し、首および左前腕の IMF6 が「ズ」に対応することを確認した。この解析は、文楽における「ホド」「ズ」の検出と時間周波数解析に関する先行研究[3]に基づくものである。

具体的には、義太夫の裏拍子が始まった後、主遣いは約 1.68 秒から 2.04 秒の間に「ホド」を発し、その後、約 1.95 秒から 2.75 秒の間に首の「ズ」を発する。左遣いは約 2.14 秒から 3.00 秒の間に左前腕を動かし、主遣いの「ズ」に約 0.2 秒程度遅れて追従する。これは、左遣いが単に同時に動いているのではなく、主遣いの微細な動作信号を読み取り、自然な時間差で協調していることを示している。図 6 は、首と左前腕の IMF6 に現れる瞬時周波数および瞬時位相の対応関係を示し、「ズ」に対する左遣いの追従を可視化している。

さらに、Procrustes Analysis、Mutual Information、Canonical Correlation Analysis を用いて、首と左前腕、右肩と左前腕、左肩と左前腕の関係を比較した。その結果、首と左前腕の関係が最も強く、左遣いの動作が主遣いの首の「ズ」に追従していることが統計的にも支持された。この成果により、従来は職人技として暗黙的に理解されていた「ホド」「ズ」の一部を、時間周波数領域における定量的なコミュニケーション信号として記述できた。

この発見は、ロボットや CG エージェントの協調動作設計にも応用できる。例えば、複数のロボットや人間とロボットが共同作業を行う場面では、明示的なコマンドだけでなく、動作の向き、速度、開始前の微細な予備動作を合図として利用できる。文楽の「ホド」「ズ」は、そのようなナッジ的で柔軟な非言語的インタラクション設計のモデルとなる。また、本研究で用いた HHT に基づく動作解析の枠組みは、先行する舞踊動作解析において所作の固有モード抽出に有効であることが示されており[4]、序破急の時間構造に関する理論的整理[5]とも接続する。

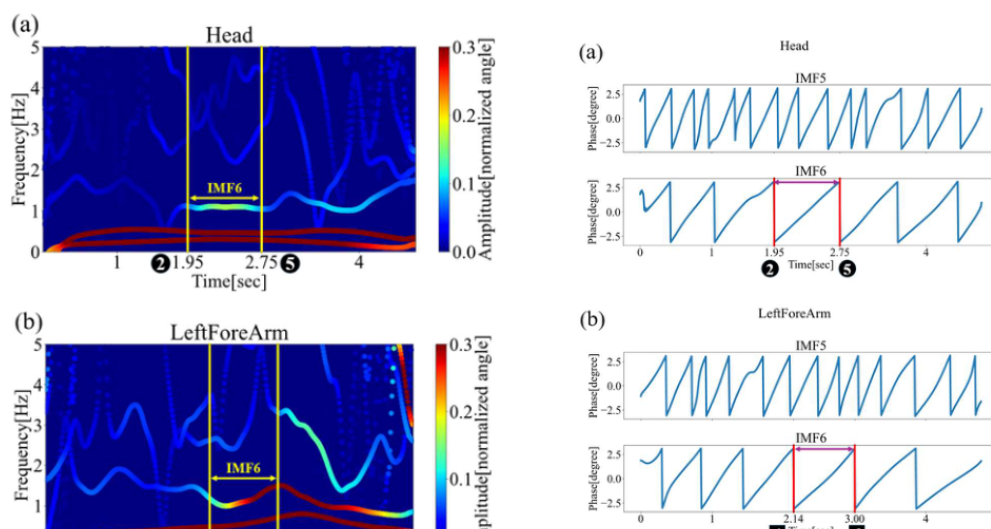


図 6 HHT による「ズ」の瞬時周波数・瞬時位相解析。(a) 首の瞬時周波数・瞬時位相、(b) 左前腕の瞬時周波数・瞬時位相。首・左前腕の IMF6 が「ズ」に対応する (Liu, Dong, Cai [3]より)

5 情報通信分野への意義

本研究の意義は、伝統芸能の解析に留まらず、情報通信における新しい身体的インタラクション基盤を提示した点にある。従来の情報通信は、文字、音声、画像を中心として発展してきた。しかし、AI エージェント、メタバース、遠隔コミュニケーション、介護支援ロボットなどでは、情報の意味だけでな

く、相手がどのように反応し、どのようなテンポで動作するかが重要となる。

本研究で構築したデータセットと評価指標は、AI が人間の感情や文脈に応じた動作速度を学習するための基盤となる。特に、序破急に基づくテンポ変化は、対話の盛り上がり、沈黙、躊躇、共感、驚きなどを表現するための時間設計として利用できる。これにより、発話内容に合わせて身体動作を生成するだけでなく、会話全体のリズムを調整する AI エージェントの実現が期待される。

また、文化遺産のデジタルアーカイブとしても意義がある。人形浄瑠璃の所作は、師弟関係の中で長年にわたり身体的に伝承されてきた。これをセンサデータ、音声データ、骨格表現、SMPL 表現として整理することで、伝統芸能の保存、教育、国際発信に貢献できる。さらに、公開データセットとして AI 研究者へ提供することで、文化的多様性を反映した生成 AI 研究を促進できる。

6 実行過程で生じた課題と対応

本研究の実行過程では、主に三つの課題が生じた。第一に、GPU の型番変更と調達遅延である。採択後、当初予定していた GPU より高性能なモデルが発売されたため、研究効率を高める目的で最新 GPU への変更を検討した。その結果、調達手続きに時間を要したものの、導入後は大規模な拡散モデル学習と複数モデルの比較実験が可能となった。

第二に、映像ベースの骨格再構成における遮蔽の問題である。人形浄瑠璃では、人形遣いの身体や衣装が人形の一部を頻繁に隠すため、一般的な 3D 姿勢推定手法では関節位置の推定が不安定になった。これに対して、IMU データを教師信号とし、映像から得られる部分的な情報を補助的に用いるハイブリッドな枠組みを構築した。これにより、完全な映像再構成には至らない場合でも、既存のセンサデータと映像を対応付け、将来的な人形特化型姿勢推定モデルの学習へつなげる道筋を得た。

この課題に関連して、SAM を用いた人形領域抽出と、SOTA の SMPL 再構成手法を組み合わせた予備実験も実施した。図 7 は、その一例を示す。図 7(a) では、プロンプトに基づく SAM により人形領域をおおむね正しくセグメンテーションできている。また、図 7(b) では、抽出結果をもとに単一フレームの姿勢と形状を SMPL 表現として概ね再構成できている。しかし、現状の結果は静止画レベルの確認に留まり、連続フレームにおける時間的一貫性、操り手や衣装による遮蔽、人形固有の関節構造への適合性にはなお大きな課題が残る。したがって、汎用的な SAM や既存の SMPL 再構成手法のみでは、人形浄瑠璃映像から高精度にモーションを抽出するには不十分である。今後は、本研究で整備した IMU データセットと対応映像を教師信号として活用し、人形に特化した映像ベース骨格推定モデルを構築・微調整することで、多様な人形映像から実用的な精度で所作データを抽出できる方法へ発展させる必要がある。

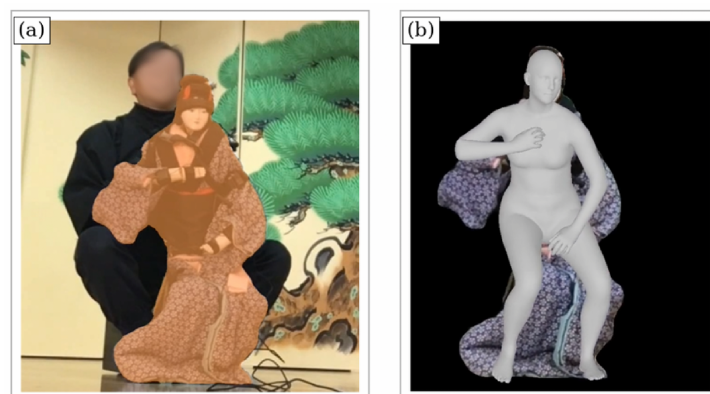


図 7 SAM を用いた人形領域抽出と SMPL 再構成の予備結果。(a) 人物顔部分をマスクした SAM セグメンテーション結果、(b) SMPL 再構成による単一フレームの 3 次元形状・姿勢推定結果

第三に、IMU センサのドリフトや装着位置の制約である。人形の構造は人体と異なり、衣装の下にセンサを固定する必要があるため、長時間の収録ではノイズや姿勢のずれが生じる。これに対して、MEMD や HHT を用いた信号分解によりノイズを低減し、テンポ変化に対応する主要な IMF を抽出することで、解析の安定性を高めた。今後は、IMU、映像、物理モデルを組み合わせたより堅牢な骨格推定手法を開発する予定である。

7 今後の展望

今後は、JoruriPuppet データセットをさらに拡張し、より多様な演目、演者、舞台環境を含めることで、大規模モーションモデルの学習基盤を強化する。特に、熟練者と初心者との差異、文楽・八王子車人形・阿波人形浄瑠璃の所作差、女性型と男性型の人形構造の違いを体系的に整理し、生成モデルが演技スタイルを制御できるようにする。

モデル面では、拡散モデルを基盤として、音楽テンポ、台詞内容、感情ラベル、所作カテゴリを統合するマルチモーダル LMM の構築を進める。LLM が会話テキストや状況を解釈し、LMM が序破急に基づく身体動作を生成することで、発話と所作が一体となったロボット・CG キャラクターのインタラクションが実現できる。さらに、人間参加者による評価実験を行い、生成された動作がどの程度自然で感情豊かに受け取られるかを検証する。

社会実装の面では、医療・介護現場における感情豊かなコミュニケーションロボット、観光・文化産業におけるデジタルキャラクター、伝統芸能教育のための動作教材、メタバース空間における文化体験コンテンツなどへの応用が考えられる。本研究は、人形浄瑠璃という日本独自の身体知を、情報通信技術と生成 AI の基盤へ接続する試みであり、文化と AI の双方に新しい展開をもたらすと期待される。

【参考文献】

- [1] Ran Dong, Shaowen Ni, and Xi Yang. “JoruriPuppet: Learning Tempo-Changing Mechanisms Beyond the Beat for Music-to-Motion Generation with Expressive Metrics.” SIGGRAPH Asia 2025 Conference Papers, 2025.
- [2] 劉 旭楠, 董 然, 蔡 東生. 「序破急を用いた文楽における非言語的協調コミュニケーション」. ヒューマンインタフェース学会論文誌, Vol.28, No.1, 2026.
- [3] Xunan Liu, Ran Dong, and Dongsheng Cai. “Detecting Hodo and Zu in Bunraku Nonverbal Communication using Hilbert-Huang Transform.” The Journal of the Society for Art and Science, Vol.24, No.2, pp.7:1-7:10, 2025.
- [4] 董 然, 蔡 東生. 「ヒルベルト変換を用いたダンスモーション解析」. 電子情報通信学会和文論文誌(D), Vol.J102-D, No.12, pp.843-853, 2019.
- [5] Tamba, A. Japanese Musical Structure and Jo-Ha-Kyu related studies, 1981.

〈発 表 資 料〉

題名	掲載誌・学会名等	発表年月
JoruriPuppet: Learning Tempo-Changing Mechanisms Beyond the Beat for Music-to-Motion Generation with Expressive Metrics	SIGGRAPH Asia 2025 Conference Papers	2025 年
Detecting Hodo and Zu in Bunraku Nonverbal Communication using Hilbert-Huang Transform	The Journal of the Society for Art and Science, Vol.24, No.2	2025 年
序破急を用いた文楽における非言語的協調コミュニケーション	ヒューマンインタフェース学会論文誌, Vol.28, No.1	2026 年