

ダンスの脳内表現に生成 AI で迫る ——音楽・身体・感情にかかわる脳応答を包括的モデル化——

発表のポイント

- ◆ダンス映像視聴中に fMRI で計測した脳応答を、音楽と身体運動を統合した「クロスモーダル特徴」（深層生成モデル由来）で説明できることを示しました。低次の運動特徴や音響特徴に比べ、クロスモーダル特徴のほうが人間の脳活動をよく説明することを発見しました。
- ◆美しさ、躍動感などダンス鑑賞中に生じる数十項目にわたる感情表現をモデル化することで、ダンス視聴中に生じる感情の違いに応じて異なる神経パターンが得られることを発見しました。
- ◆熟達ダンサーはダンス未経験者に比べて脳の広範にわたってダンスの情報が表現されている一方、熟達ダンサーの方が脳内表現の個人差が大きいことを定量的に示しました。
- ◆本研究は神経美学・芸術科学やダンス教育・訓練支援、人間と AI の創造的インタラクション設計に資する基盤知見を提供します。



ヒトと生成 AI のダンスに対する情報処理を、世界で初めて定量的に比較

概要

東京大学大学院人文社会系研究科の今水 寛教授、若林 実奈大学院生（当時）、名古屋工業大学大学院工学研究科の高木 優准教授、神戸大学大学院人間発達環境学研究科の清水 大地助教、産業技術総合研究所人間情報インタラクション研究部門の大畑 龍研究員らは、自然なダンス映像を視聴している参加者の fMRI（注 1）脳活動データと、音楽からダンスの振り付けを生成するクロスモーダル深層生成モデル（EDGE）（注 2）から得た Transformer（注 3）の特徴量を結び付け、ダンスが脳内でどのように表現されるかを定量的に明らかにしました。

本研究では、ダンス動画データベース（AIST Dance DB）（注 4）に含まれるストリート／ジャズを含む多様なダンス動画を視聴する 14 名（熟達ダンサーを含む）の脳活動を計測し、ダンスに含まれる特徴から、脳活動を予測する「エンコーディングモデル」（注 5）を構築しました。その結果、音と動きを統合したクロスモーダル特徴が、運動・音響特徴単体よりもダンスが誘発する脳活動を良く説明できることを示しました。さらにエンコーディングモデルをコンピュータ上の「脳活動シミュレータ」として用いることで、異なる感情を喚起するダンスが異なる神経パターンをもたらすことを定量化しました。熟達ダンサーではダンス特徴による説明範囲が広い一方、個人差が大きいことも見いだされました。これらの成果は、動き・音楽・熟達度が美的・感情体験を形作る仕組みの理解に資するものです。

発表内容

本研究は、自然視聴下の fMRI 脳活動と、音楽からダンスの振り付けを生成するクロスモーダル深層生成モデル（EDGE）を統合し、ダンスが脳内でどのように表現されるかを定量化しました。参加者は 14 名（熟達ダンサー 7 名／未経験者 7 名）で、各参加者は約 5 時間にわたり AIST Dance DB のダンス動画を視聴しました。EDGE は過去のモーションと音楽から次の動きを予測する Transformer に基づくモデルで、ここから得た音×動きのクロスモーダル特徴を用い、全脳ボクセル単位（注 1）のエンコーディングモデルを構築しました（図 1）。

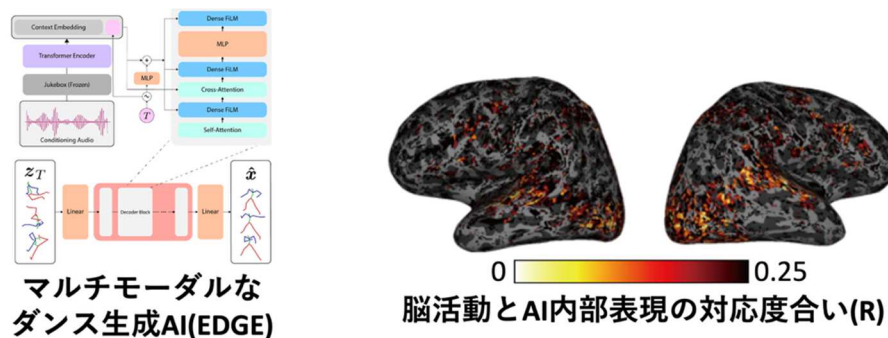


図 1：ダンス生成 AI (左) と、AI の内部表現と対応する脳部位 (右)

R 値は脳活動と AI 内部表現の時間変化の一致度を示す相関係数

主な結果として、

- エンコーディングモデルは大脳皮質のさまざまな場所の脳活動を予測することができました。運動特徴は主に視覚～背側経路、音響特徴は腹側視覚野と聴覚皮質、そしてクロスモーダル特徴は頭頂間溝・楔前部といった高次連合領域を特によく予測しました。側頭-頭頂連合領域は幅広い情報を表現していました。また、脳領域ごとによく反応するダンスが異なることも示されました。
- 大規模オンライン評価を行い各ダンス動画に 42 の感情や審美に関する印象を収集しました（図 2 左）。エンコーディングモデルを用いて、それら印象とモデル推定脳活動を対応付けることで、躍動感や審美といった、ダンスから惹起されたそれぞれの印象がどのような脳部位を賦活させるかがわかりました（図 2 右）。これによって、ダンスから生じる印象は広い範囲の脳部位と複合的に結び付くことがわかり、単純な次元では捉えきれないことが示唆されました。

ダンス鑑賞体験で生じる多様な印象

崇拝 美しさ 共感 嫌悪感 退屈さ 侮辱
 ぎこちなさ 罪悪感 忘我感 穏やかさ 興奮
 混乱 誇り 感嘆 恐怖 楽しさ
 渴望 怒り 嫉妬 畏怖 不安
 失望 興味 様かしさ

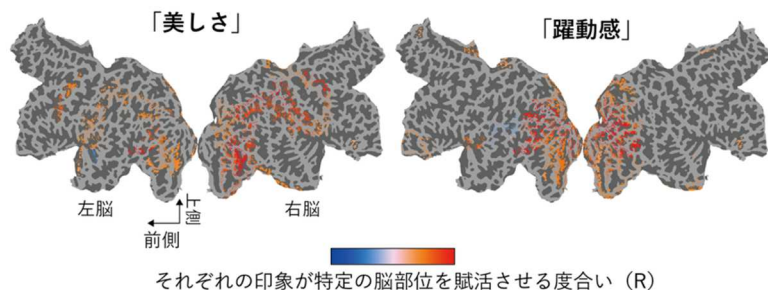


図 2：ダンス鑑賞体験で生じる印象を収集し（左）、
それぞれの印象がどのような脳部位を賦活させるかを明らかにした（右）。

- 熟達ダンサーと未経験者の比較では、熟達ダンサーは特に運動特徴を幅広い脳領域で表現する傾向にあった一方、被験者間の活動パターンの個人差が大きかったことが分かりました。
- 実在のモーションに他ジャンル音楽を人工的に組み合わせた仮想のダンス動画を作り出して、脳シミュレータによって脳活動を推定しました。すると、実際のダンス動画は視覚皮質を強く賦活するのに対し、人工的に作られたダンス動画では相対的に前頭領域の活動が高まる傾向が見られ、不一致／予測誤差などが引き起こす脳内のプロセスを示唆しました。

以上より、音と動きが織りなす高次の時系列構造をもつクロスモーダル表現が、自然なダンス観察時の脳活動をよりの確に説明すること、審美・感情経験や熟達度がその表現と結び付くことを示しました。生成 AI×自然視聴 fMRI という枠組みは、神経美学・ダンス科学・創作支援 AI などへの応用に向けた有力な手がかりを提供します。

なお、本研究は東京大学大学院人文社会系研究科心理学研究室事前審査制度の承認のもとに実施されました。

発表者・研究者等情報

東京大学

大学院人文社会系研究科

今水 寛 教授

若林 実奈 研究当時：博士課程学生

名古屋工業大学

大学院工学研究科

高木 優 准教授

神戸大学

大学院人間発達環境学研究科

清水 大地 助教

産業技術総合研究所

人間情報インタラクション研究部門

大畑 龍 研究員

論文情報

雑誌名：Nature Communications

題名：Cross-modal deep generative models reveal the cortical representation of dancing

著者名：Yu Takagi, Daichi Shimizu, Mina Wakabayashi, Ryu Ohata, Hiroshi Imamizu

DOI: 10.1038/s41467-025-65039-w

URL: <https://www.nature.com/articles/s41467-025-65039-w>

研究助成

本研究は、科研費「超適応を促す身体認知・情動機構の解明（課題番号：19H05725）」、「エキスパートが認知する世界の可視化と再構成（課題番号：24H00172）」、「踊りに関するヒトの情報処理の包括的解明（課題番号：21K21309）」、JST さきがけ「大規模言語モデルとヒト脳の相互理解と新たなインタラクション創出（課題番号：JPMJPR23I6）」の支援により実施されました。

用語解説

（注 1）fMRI（機能的磁気共鳴画像法）、ボクセル単位

fMRI（機能的磁気共鳴画像法）は、酸化型および還元型ヘモグロビンの磁化率の差を利用して、脳内の血流変化を可視化する技術です。脳活動が高まると酸素代謝が変化し、血液中の酸化型／還元型ヘモグロビンの比率が変わります。この変化を画像として捉えることで、脳のどの領域がどの程度活動しているかを推定できます。fMRI における「ボクセル」は、脳の 3 次元画像を構成する最小の立方体単位のことです（2 次元画像における「ピクセル」に相当）。

（注 2）音楽からダンスの振り付けを生成するクロスモーダル深層生成モデル（EDGE）

EDGE（Editable Dance GEneration）はダンスの音（音楽）と動き（モーション）の系列から、Transformer に基づき次の動きを予測するクロスモーダル深層生成モデルです。音と動きの共同埋め込み表現を学習し、ダンスの高度な時系列構造を記述できるとされます。

（注 3）Transformer（トランスフォーマー）

Transformer は、文章や音声、動きなどの時系列データの関係性を学習するための深層学習モデルです。従来の RNN（再帰型ニューラルネット）と異なり、「自己注意（Self-Attention）」という仕組みを使って、入力の中のどの部分が互いに重要かを同時に計算できます。これにより、長い時間的関係や複雑な文脈を効率的に捉えることができます。もともとは自然言語処理のために開発されましたが、現在では音楽生成や画像生成、ダンス動作の予測など、さまざまな時系列・多モーダルデータの解析や生成に広く使われています。

（注 4）AIST Dance DB（AIST Dance Video Database）

AIST Dance Video Database（AIST DanceDB）は、産業技術総合研究所が公開している高精度かつ大規模なダンス映像データベースです（<https://aistdancedb.ongaaccel.jp>）。10 種類のストリートダンスジャンルにわたるダンサーの映像と音楽が収録されており、学術研究で自由に利用できます。ダンスの動きやリズム、音楽との対応を解析するための共通基盤として、多くの研究で活用されています。

（注 5）エンコーディングモデル（Encoding model）

刺激の特徴から脳活動（fMRI 信号）を予測する統計モデルです。本研究では、映像や音楽から得た動き・音楽・クロスモーダル特徴を入力とし、各脳領域の脳活動をどの程度説明できるかを評価します。一般的にリッジ回帰などの線形モデルを用い、学習データで重みを推定し、独立して用意された検証用データで予測精度（相関係数など）を検証します。得られたモデルを利用することで、実際に提示していないダンスでも特徴を入力すれば対応する脳活動を仮想的に再現できる“脳活動シミュレータ”として機能し、任意の感情や美的体験に関する脳活動を定量的に比較することも可能になります。

問合せ先

〈研究内容全体について〉

東京大学大学院人文社会系研究科

教授 今水 寛（いまみず ひろし）

Tel : 03-5841-3858 E-mail : imamizu@l.u-tokyo.ac.jp

〈AI を用いた脳のモデリングについて〉

名古屋工業大学大学院工学研究科

准教授 高木 優（たかぎ ゆう）

Tel : 052-735-5440 E-mail : takagi.yu@nitech.ac.jp

〈機関窓口〉

東京大学大学院文学部・人文社会系研究科 総務チーム

Tel : 03-5841-3705 E-mail : shomu@l.u-tokyo.ac.jp