

MoodGroove: テキストと身体動作による対話型音楽生成・編集システム

土屋 実結* 清水紘輔† 鈴木健嗣‡

概要. ダンスは、人間が感情を表現するための典型的な手段である。これまでの研究では、振付情報に基づいて音楽を作曲するなど、人間の体の動きを音楽に変換する方法が探求されてきた。しかし、従来の研究では、動きと感情と音楽のマッチングに焦点が当てられておらず、機械的に動きを音楽に変換する傾向があった。本研究では、ユーザーの感情を表現し、それとインタラクションするための音楽を、ユーザーの動きを用いて創作するシステム「MoodGroove」を開発した。このシステムは、まずユーザーが声で入力した感情情報に基づいて15秒のループ音楽を生成し、その後ユーザーの動きによって音楽のパラメータを編集することで、音楽の印象を変化させる。この研究は、ユーザーの感情を楽曲の形で表現することが、曖昧な感情を非言語的に表現し、その感情への没入とコミュニケーションを促進する可能性があることを示唆している。

1 はじめに

人々は非言語的手段として表情や身体表現を用いて感情を表出し [1]、ダンスは一つの代表的な例と言える [2]。Wallbott ら [3] の研究は、特定の感情が身体の特定位の独特な動きと関連することを示した。ダンスにおける身体動作の定量化研究では、Sawada ら [4] が DLT 法を用いて上肢動作と感情の関係を解明し、速度、加速度、移動距離の総和が感情識別の手がかりとなることを実証した。Camurri ら [5] は、Laban [6] のダンス動作における時間、空間、重さの概念を基に、感情の身体動作を重さ、時間、空間、流れの4要素で分類した。

一方、身体動作を音楽に変換する方法については、主に作曲支援、物理的音響メディア、Human-Robot Interaction の文脈で研究が進められている。Morales ら [7] は、ダンサーの動きに合わせて音楽を作り出すソフトウェア「SICIB」を開発した。また、Cammuri ら [8] はダンス、演劇、アートインスタレーションなど多様な場面で、人間の動きや感情を解釈し、それに応じて音楽や視覚効果を生成する高度なインタラクティブシステムの開発を行った。

既存研究は感情、身体動作、音楽の関係を個別に探究してきたが、これらの要素を統合した包括的システムについては十分な検討がなされていない。本研究はこの課題に対し、自然言語処理、AI による音楽生成、リアルタイムモーションキャプチャを組み合わせた人間の情動を音楽に変更するインターフェースを提案する。さらに、高橋ら [9] は「専門知

識を持たないユーザーが自動生成されたコンテンツを『出発点』として直感的に『修正』できるインターフェースの開発」の重要性を指摘している。音楽の完成度とユーザーが編集することのできる自由度を両立させるうえで、本研究では高橋らの指摘に基づき、ユーザーが生成された音楽を容易に編集できるシステムの構築を目指す。我々は二段階の方法を提案する。システムはまず、感情体験に関するテキスト記述からベースの音楽を生成し、次にリアルタイムの身体動作を通じてこの音楽を動的に変調する。

2 提案手法

本研究は、感情体験、音楽生成、身体動作を統合した包括的システムの構築を目指している。このシステムは、自然言語処理による感情分析、AI を用いた音楽生成、リアルタイムモーションキャプチャによる音楽変調を組み合わせ、従来個別に研究されてきた要素を統合する。テキストによる感情の入力から音楽を生成し、身体動作によってその音楽を動的に変調させる二段階のプロセスを採用している。第一段階では、システムはユーザーの音声入力から感情情報を抽出し、15秒のループ音楽を生成する。これにより、ユーザーは専門知識がなくても自分の感情を反映した音楽を作成できる。第二段階では、生成された音楽をユーザーの身体動作に応じてリアルタイムで「編集」し、より個人的な感覚に合わせることを可能にする。効果的な音楽「編集」のため、音楽の構造的特徴と感情表現の関係に着目した [10]。Gabrielsson [11] らの研究によれば、loudness, timbre, pitch, tempo, melody などの要素が音楽の感情価に重要な影響を与えることが示されている。また、検出する身体動作には Camurri ら [5] の研究を参考にし、感情、音楽、動きが連携されるようにマッピングした。本アプローチにより、ユーザーの動

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 江戸川女子高等学校

† 筑波大学

‡ 筑波大学

きは感情表現の形態であると同時に、音楽パラメータの調整手段としても機能する。

2.1 背景情報入力によるループ音楽の生成

第一段階では、ユーザーの感情状態を反映した基本ループ音楽の作成に焦点を当てた。ユーザーの入力をもとに GPT は、format, genre, sub-genre, mood, style, tempo, BPM, duration の情報を JSON 形式で出力する。生成された JSON ファイルは MusicGen[12] へ入力され、MusicGen は、この情報を用いてユーザーの入力に合致する 15 秒のループ音楽を生成する。その後、音楽編集システムで効果的に音楽を編集するために AI 技術を用いた音源分離ツールである Spleeter[13] を使用し、生成された音楽を Melody, Backing, Bass, Drum の 4 つの異なるトラックに分割する。

2.2 動きを通じた音楽編集システム

第二段階では、ユーザーの動きに連動させる形で音楽パラメータを動的に変更し、自分の感覚と合致した表現をユーザーが探索できることを可能としている。動きの検出には、モーションキャプチャシステム「mocopi」を使用し、ユーザーの身体の各部位の加速度と角速度をリアルタイムで取得する。取得した動作データを基に、音楽パラメータを調整することで、ユーザーの動きに応じた音楽生成を実現する。動きと音楽のパラメータのマッピングには Wallbotら [3] や Camurriら [5] の研究を参考にした。動作パラメータとして、腕の曲率および振率、足の躍度、腰および全身の運動量を計算する。腕の曲率 κ と振率 τ は、フレネ・セレの公式に基づいて導出される。具体的には、腕の位置ベクトル $\vec{p}(s)$ に対して、曲率 κ は以下の式で表される。

$$\kappa = \frac{|\vec{v} \times \vec{a}|}{|\vec{v}|^3} \quad (1)$$

ここで、 \vec{v} は速度ベクトル、 \vec{a} は加速度ベクトルである。振率 τ も同様に、フレネ・セレの公式により計算される。足の躍度 J は、足部分の加速度 \vec{a}_f を時間で微分することで求められる。

$$J = \frac{d\vec{a}_f}{dt} \quad (2)$$

速度に関しては、腰および全身の速度 v_{waist} と v_{fullbody} を以下の式で計算する。速度は時間積分により求められる。

$$v_{\text{waist}}(t) = v_{\text{waist}}(t_0) + \int_{t_0}^t \vec{a}_{\text{waist}}(\tau) d\tau \quad (3)$$

$$v_{\text{fullbody}}(t) = v_{\text{fullbody}}(t_0) + \int_{t_0}^t \vec{a}_{\text{fullbody}}(\tau) d\tau \quad (4)$$

ここで、 $v_{\text{waist}}(t_0)$ および $v_{\text{fullbody}}(t_0)$ は初期時刻 t_0 におけるそれぞれの速度である。

音楽パラメータの調整は、各動作パラメータに対応する音量の変化として実装する。各音量は以下のように線形変換される。システムの音声再生においては、各種の音声を 30% の音量で常時再生しつつ、動作に応じて音量パラメータを動的に調整する。この設計により、基盤となる音楽のリズムを維持しながら、ユーザーの動きに応じた音の変化を実現する。

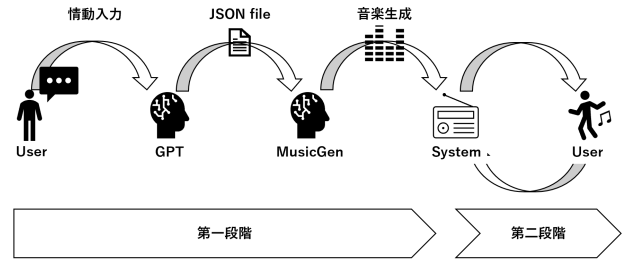


図 1. 開発したシステムの図

3 おわりに

本研究では、ユーザーの感情状態を音楽形式で表現するシステムを開発した。このシステムは、ユーザーの感情に基づいて音楽を生成し、さらにユーザーのダンス表現によって生成された音楽のパラメータを変更することを可能にした。しかしながら、本システムには検討すべき課題が残されている。人間のダンス表現の多様性に対し、音楽の変数が固定された一連の身体動作と常に連動してしまう可能性があるためである。そのため、音楽、感情、動きが適切に一致しているか、また本システムがユーザーの自由な身体表現を制限していないかを慎重に検証する必要がある。

謝辞

本研究では、Stanford 大学の田中翔大氏に人間の体の動きと音楽変換の助言を頂いた。感謝する。

参考文献

- [1] Zoltán Kövecses. *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*. Cambridge University Press, Cambridge, UK, 2000.
- [2] M.N. H'Doubler. *Dance: A Creative Art Experience*. University of Wisconsin Press, 1998.
- [3] Harald G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896, 1998.

- [4] M. Sawada, K. Suda, and M. Ishii. Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and Motor Skills*, 97:697–708, 2003.
- [5] A. Camurri, I. Lagerlöf, and G. Volpe. Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human Computer Studies*, 59:213–225, 2003.
- [6] Rudolf Laban. *Modern Educational Dance*. Macdonald & Evans, Ltd., London, 1963.
- [7] Roberto Morales-Manzanares, Eduardo Morales, and Roger Dannenberg. Sicib: An interactive music composition system using body movements. *Computer Music Journal Article) Computer Music Journal*, 25:25–36, 06 2001.
- [8] Antonio Camurri, Shuji Hashimoto, Matteo Ricchetti, Andrea Ricci, Kenji Suzuki, Riccardo Trocca, and Gualtiero Volpe. Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal*, 24(1):57–69, 2000.
- [9] 高橋 拓椰, Christoph Wilk, 嵯峨山 茂樹, and 宮下 芳明. 旋律と「和声進行の緊張感」を入力する対話型和声付け支援システム. *インタラクシオン 2020 予稿集*, pages 3A–08–, 3 2020.
- [10] Klaus R Scherer and Marcel R Zentner. Emotional Effects Of Music: Production Rules. In *Music And Emotion: Theory and research*. Oxford University Press, 08 2001.
- [11] Alf Gabrielsson and Lindström. *The Influence Of Musical Structure On Emotional Expression*, pages 223–248. 08 2001.
- [12] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. Simple and controllable music generation, 2024.
- [13] Romain Hennequin, Anis Khlif, Felix Voituret, and Manuel Moussallam. Spleeter: a fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, 5(50):2154, 2020.