7ZE-02

ダンス映像からの盛り上がり度推定に基づくループシーケンサーを用いた背景楽曲生成

杉山　紘次郎 †1　　白松　俊 †1　　北原　鉄朗 †2

†1 名古屋工業大学　　†2 日本大学

1　はじめに

近年，ダンス情報処理 (Dance Information Processing)[1] という研究分野が新たに提唱されている．一口にダンスといっても様々であるが，本研究では近年流行しているストリートダンスに着目する．ストリートダンスでは通常，先に音楽があり，音楽に合わせて踊るのが一般的である．しかし我々は，逆にダンスから楽曲を生成する技術の開発を目指す．これにより，ダンサーは音楽に縛られないダンスをすることができ，また聴衆視点ではダンサーの動きから即興的に生成される音楽という，新たな即興芸術になる可能性がある．これまでにも，ダンスの類似性から楽曲を検索するダンス楽曲インターフェース Query by Dancing[2] や，身体動作認識に基づくインタラクティブサウンド生成システム [3] といった研究は存在したが，ダンスからの楽曲生成を目指すものではなかった．本稿では，ダンスの雰囲気や盛り上がりからの楽曲生成を目標に，まずはダンス動画の骨格情報からの盛り上がり度計算手法を検討する．

ダンス動画からの音楽生成手法を検討する上で，安坂ら [5] が開発した，映像の盛り上がり度に基づくループシーケンサーのアプローチを参考にする．安坂らは，動画の変化の速さから盛り上がり度を段階的に算出し，その段階ごとに音楽を生成していた．それに対し本研究では，ダンス動画の骨格情報から盛り上がり度を計算し，安坂らのシステムを用いて楽曲生成を行う．

本研究では，ダンス動画のデータセットとして AIST Dance Video Database[1] と，骨格情報がアノテーションされた AIST++ Dataset[4] を使用する．将来的には骨格情報の推定や盛り上がり度計算をリアルタイム化する予定だが，本稿では AIST++ Dataset の骨格情報を使用した楽曲生成を試み，まずは手法の妥当性を検証する．骨格情報からの盛り上がり度計算には，舞踏学や身体動作の解析に広く用いられているラバン特徴量 [6] を用いる．ラバン特徴量は，大別すると動きを扱う Effort と形状を扱う Shape に分かれるが，本稿では特に Effort に分類される Weight（速度；動作の物理特性），Time（加速度；動作の切迫感），Flow（躍度；動作の連続性），Space（動作の迂回度；動作の非直接性）の 4 特徴量を用いる．これら 4 特徴量からダンスの盛り上がり度を計算し，ループシーケンサーを用いて背景楽曲生成を試みる．

2　提案手法

2.1　盛り上がり度

本研究での盛り上がり度とは，ダンスの雰囲気や盛り上がりを定量化したものとする．具体例として次のものがあげられる．

(1) 盛り上がり度が高い例
  (a) 音楽でいうサビに当たる部分
  (b) 明るい、激しいといったポジティブな印象を受ける部分
  (c) 笑顔、表情が柔らかい
(2) 盛り上がり度が低い例
  (a)　ゆったりとした動作
  (b) 暗い、静かといったネガティブな印象受ける部分
  (c) 真顔、緊張

2.2　盛り上がり度の計算

本研究の目的はダンサーがする即興的なダンスに対してリアルタイムに追随する背景楽曲生成であるが，本稿ではその第一段階としてダンス動画からのオフラインでの楽曲生成を試みた．ダンスがどれくらい盛り上がっているかの指標としてラバン特徴量を採用し，盛り上がり度の定量化を行った．

まずラバン身体動作表現理論という心理状態と身体動作の相関関係を規定する理論があり，このラバン理論で用いられているラバン特徴量の Effort の Weight，Space，Time，Flow の 4 つを計算する．計算方法は Caroline Larboulette ら [6] の計算方法を一部変更したものを採用した．まず，ダンス動画の骨格情報を 2 小節ごとのブロック s に分割して，s 毎に盛り上がり度を決定する．各ブロック s は，時刻 t_i のフレーム T 個からなる集合 $\{t_1, \cdots, t_T\}$ と見なせる．時刻 t における関節 k の 3 次元座標を $\boldsymbol{n}_t^k$ とする．また時刻 t における全関節の速度の重み付き線形和を $v(s)$，加速度の重み付き線形和を $a(s)$，躍度の重み付き線形和を $j(s)$，動作の迂回度/方向の偏りの重み付き線形和を $d(s)$ とする．全関節の特徴量の重み付き線形和をとるにあたり，関節 k の重み α_k は，表 1 のように設定した．これは，聴衆側はダンサーの顔に最も注目し，その次に両手足であることが多いと考えたためである．ただし，$\sum_k \alpha_k = 1$ となるように設定した．1 秒間のフレーム数は 30 フレームで，フレームの間隔を δ で記述する．

ただし，特徴量毎に値の範囲が異なるため，値の範囲を揃えるため、正規化を行う関数 can を適用してから重み付き線形和をとった．

Table 1: 関節 k の重み α_k

関節 k	α_k
頭	0.2
左手, 右手, 左足, 右足	0.1
その他の 8 関節	0.05

- Weight（速度；動作の物理的特性） $v_(s)$:

$$v(s) = \sum_k \frac{\alpha_k}{T} \sum_{t \in s} \mathrm{can}\left(\left\| \frac{\boldsymbol{n}_{t+1}^k - \boldsymbol{n}_t^k}{\delta} \right\| \right)$$

- Time（加速度；動作の切迫感）$a(s)$:

$$a(s) = \sum_k \frac{\alpha_k}{T} \sum_{t \in s} \mathrm{can}\left(\left\| \frac{\boldsymbol{n}_{t+2}^k - 2\boldsymbol{n}_{t+1}^k + \boldsymbol{n}_t^k}{\delta^2} \right\| \right)$$

- Flow（躍度；動作の連続性）$j(s)$:

$$j(s) = \sum_k \frac{\alpha_k}{T} \sum_{t \in s} \mathrm{can}\left(\left\| \frac{\boldsymbol{n}_{t+4}^k - 2\boldsymbol{n}_{t+3}^k + 2\boldsymbol{n}_{t+2}^k - \boldsymbol{n}_t^k}{2\delta^3} \right\| \right)$$

- Space（動作の迂回度；動作の非直接性） $d(s)$:

$$d(s) = \sum_k \frac{\alpha_k}{T} \sum_{t \in s} \mathrm{can}\left(\frac{\sum_{t=2}^{T} ||\boldsymbol{n}_{t+1}^k - \boldsymbol{n}_t^k||}{||\boldsymbol{n}_t^k - \boldsymbol{n}_1^k||} \right)$$

正規化関数

$$\mathrm{can}(x) = \mathrm{cdf}(\mathrm{std}(log(x)))$$

cdf は標準正規分布の累積密度関数であり，std は標準偏差が 1 になるよう標準化する関数である．

最後に上記の式で得られた 4 つのラバン特徴量の重み付き線形和を盛り上がり度とした．4 つラバン特徴量の重み付き線形和をとるにあたり，Weight と Space は視覚的にもわかりやすい情報であるため，盛り上がり度に大きな影響があると考え高く設定した．具体的には，$\beta_w, \beta_s = 0.4$，$\beta_t, \beta_f = 0.1$ と設定した．ただし，$\sum_k \beta_k = 1$ となるように設定した．

ただし，単純に重み付き線形和をとっただけでは盛り上がり度の振れ幅が 0.4 から 0.6 と小さくなってしまい，生成される楽曲が変化の乏しいものになってしまうため，振り幅を調整するためのロジスティック関数 g を適用し，ブロック s における盛り上がり度 E_s を以下のように算出する．ロジスティック関数における c はゲインを示し，この値を変更して振れ幅を調節した．x_0 はシグモイド中点を示しており，値を左右に調節することができる．

$$E(s) = g\left(\beta_w \cdot v(s) + \beta_s \cdot d(s) + \beta_t \cdot a(s) + \beta_f \cdot j(s)\right)$$

$$g(x) = \frac{1}{1 + e^{-c(x - x_0)}}$$

3　生成結果例

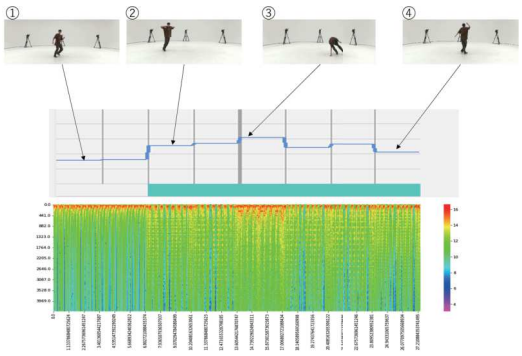

図 1: 盛り上がり度とダンスの比較

2 節で計算した盛り上がり度からループシーケンサーを用いて実際に楽曲生成を行った．図 1 の①ではその場で基本的なステップを踏むようなダンスで，②は横にスライドしたり回転をするダンス，③は地面に手をついて足を大きく回しながら回転するようなダンス，④は体を左右に揺らしたり小さく回転するようなダンス，をしていた．図 1 の中央のグラフは提案手法によって計算した盛り上がり度を表しており，下は生成した楽曲のスペクトログラム示す．

図 1 に示した盛り上がり度とダンス動画とを見比べて，その妥当性を検証する．盛り上がり度が低い①や④のダンスは比較的小さく，地味な動きや同じ位置で踊っていることが多かった．逆に，盛り上がり度が高い②や③のダンスは派手で大きな動きであったり，前後左右や上下に移動している踊りが多かった．生成した楽曲をスペクトログラムを可視化してみても盛り上がり度が高くなっている部分で赤い部分が伸びていることがわかる．よって図 1 は，算出された盛り上がり度の妥当性が示された例である．

ただし，ダンスとしては違う動きでも，盛り上がり度が同じような値になれば同じような楽曲が生成されるという問題は残る．このことから，今後は楽曲生成の幅を広げるため，盛り上がり度は一次元ではなく多次元に広げたほうが良いと考えられる．

4　まとめと今後の展望

本研究では、ダンス映像からの盛り上がり度推定に基づくループシーケンサーを用いた背景楽曲生成を行った．ダンス動画の骨格情報から盛り上がり度の計算手法を提案し，計算するモジュールをループシーケンサーに組み込んだ．本研究ではラバン特徴量を採用し，骨格情報から推定可能な速さ，加速度，躍度，動作の迂回度/方向的偏りを計算してラバン理論における Effort の Weight，Space，Time，Flow の 4 つの特徴量を推定し盛り上がり度を決定した．生成結果例で示したように，ダンスが派手（大きな動作や移動を伴う動き）の時は高い盛り上がり度の値を示し，逆にそうでない部分では低い値を示すという，妥当な結果を得た．

しかし，3 節でも述べたように盛り上がり度が一次元であることによって違う動きでも盛り上がり度が同じであれば同じような楽曲が生成されてしまう問題がある．このことから，一次元ではなく多次元に広げたほうが良いと考えられる．ほかにも今のループシーケンサーでは 2 小節の楽曲を生成する際に音素材を選択して組み合わせているので一定のリズムの楽曲が生成されてしまう．このことから，ダンサーの表現するリズムパターン解析し，楽曲にも反映させることができればより多彩な楽曲生成が可能であり，より一体感が生まれるのではないかと考えている．そのため本研究で使用しているループシーケンサ [5] を拡張する必要がある．また，OpenPose を用いた骨格情報のリアルタイム推定を行い，楽曲もリアルタイム生成することで，ダンスの動きに追随し即興性のある Adoptive Music を生成するシステムを実装する予定である．

参考文献

[1] Tsuchida, S., Fukayama, S., Hamasaki, M., Goto, M. (2019). AIST Dance Video Database: Multi-Genre, Multi-Dancer, and Multi-Camera Database for Dance Information Processing. In *Proc. of ISMIR* (pp. 501-510).

[2] Tsuchida, S., Fukayama, S., Goto, M. (2019). Query-by-dancing: a dance music retrieval system based on body-motion similarity. In *Proc. of International Conference on Multimedia Modeling* (pp. 251–263), Springer.

[3] 北洞 穂高, 前田 陽一郎, 高橋 泰岳 (2013): 身体動作認識に基づくインタラクティブサウンド生成システム, Human-Agent Interaction Symposium 2013 予稿集, (pp. 51–54)

[4] Li, R., Yang, S., Ross, D. A., Kanazawa, A. (2021). Learn to Dance with AIST++: Music Conditioned 3D Dance Generation. arXiv preprint arXiv:2101.08779.

[5] 安坂 文太, 北原 鉄朗 (2020). 動画の盛り上がり度に基づいたループシーケンサ, 情報処理学会 インタラクション 2020 (pp. 528–533).

[6] Larboulette, C., Gibet, S. (2015). A review of computable expressive descriptors of human motion. In *Proc. of the 2nd International Workshop on Movement and Computing* (pp. 21–28).