

ダンスモーションに同期した 表情自動合成のための楽曲印象解析手法の提案

朝比奈わかな[†] 岡田成美[†] 岩本尚也[†] 増田太郎[†] 福里司[†] 森島繁生^{††}

早稲田大学/JST[†] 早稲田大学理工学術院総合研究所/JST^{††}

1. はじめに

近年、3DCG 制作ツール(MikuMikuDance 等)の普及により、楽曲に合わせて CG キャラクタを踊らせる作品が増加傾向にある。このようなダンス動画においてキャラクタの表情はダンスモーション全体の印象に大きく影響する。例えば、楽曲やダンスモーションの印象と全く異なる印象の表情を付与した場合、その動画は違和感のある作品になってしまう。また、楽曲の印象が一定でも、例えば、ダンスモーションの激しさの度合いやキャラクタの姿勢によって、全体の印象が大きく左右される場合がある。そのため、作品全体の印象を決定する要素として、楽曲の印象だけでなく、ダンスモーションの情報も考慮する必要がある。しかし、一つの作品を作るために、それらの要素に関する専門的知識を取得することは容易ではない。

そこで我々は、キャラクタの表情を自動合成するための、ダンスモーションに同期した楽曲印象推定手法を提案する。本手法を用いることで、楽曲とダンスモーションにマッチした自然な表情の合成を自動で行うことが可能となる。

2. 楽曲とモーションからの印象推定

本章では、楽曲とモーションから印象を推定する具体的な手法について説明する。

2.1. 使用するモデル

既存手法として DiPaola らは Thayer モデルを用いた楽曲印象推定及び表情自動合成手法を提案した。Thayer モデルによる音楽解析は、音響特徴量を計算し、GMM を用いて階層的にクラスタリングを行うことで楽曲の印象を推定する。Thayer モデルは心理学で用いられる Russell モデル^[2]と対応しているため、各クラスタに Russell の印象語を割り当てることで、心理学に基づいた楽曲の印象推定が可能である。一方で、同一クラスタ内においてそのクラスタに属する確率の大きいものと小さいものが混同してしまうという問題点がある。そこで本手法では、重回帰分析に基づく VA 平面座標値(印象)を推定する手法^[3]を用いた。ここで、VA 平面とは、興奮 -

弛緩を表す Arousal (活発度)の軸および、快 - 不快を表す Valance (感情価)の軸からなる感情平面である。重回帰分析は、GMM などのクラスタリング手法とは異なり、直接 VA 座標値を求める手法であるため、Thayer モデルのようなクラスタリングによる問題が生じることがなく、各印象の微妙な違いも推定することができる。

さらに、従来研究^[1,3]で考慮されていないモーションの特徴量を用いることで、ダンス動画のためのより精細な印象推定を行う。具体的には、音響特徴量とモーション特徴量の両方を用いて重回帰分析を行うことで、1 秒毎の VA 座標値を求める。最終的に、この VA 座標値に対応した表情をダンスキャラクタに合成する。

2.2. 使用する特徴量

本章では、印象推定の際に使用するモーション特徴量および楽曲特徴量について説明する。モーションの印象を表す要素として、Laban の身体理論^[4]を基に「空間性」と「力動性」の2種類に大別する。「空間性」は身体の全体的な形状についての幾何学的特徴を意味し、「力動性」は動作の力学的特徴を意味する。そこで、モーション特徴量として、沼口ら^[5]が用いた Laban の身体理論に関する特徴量を参考に、表 1 の 13 次元の特徴量を選んだ。また、音響特徴量に関してはリズム、音色に関する特徴量 80 次元を用いた(表 2)。

表 1. 使用するモーション特徴量

モーション特徴量	概要
空間性	ルートから両手足間の距離、両手足で囲まれた領域の面積(平均・分散)
力動性	全身の角速度、身体移動速度(平均・分散)

表 2. 使用する音響特徴量

楽曲特徴量	概要
リズム	オンセット振幅、テンポ
音色	Mfcc(20 次元)、スペクトル重心、スペクトルフラックス、スペクトルロールオフ、スペクトルフラットネス、スペクトルコントラスト(8 音域)

2.3. 重回帰分析による印象推定

具体的な印象推定方法として、音響特徴量とモーション特徴量を説明変数、1 秒毎の VA 座標値を目的変数とする、以下の式で表される式で重回帰分析を行う。

“Automatic Generation of the facial animation of dancing characters considering music emotion and motion features”

[†]Wakana ASAHINA, Narumi OKADA, Naoya IWAMOTO, Taro MASUDA, Tsukasa HUKUSATO, Waseda University/JST

^{††}Shigeo MORISHIMA, Waseda Research Institute for Science and Engineering/JST

$$V = a_0 + \sum_{i=1}^N a_i x_i, \quad A = b_0 + \sum_{i=1}^N b_i x_i \quad (1)$$

ここで、 x_i は音響特徴量およびモーション特徴量、 a_i, b_i は重回帰分析の偏回帰係数である。後述の主観評価実験を行うことで取得したVA座標値を正解値データとして用いることで、最適な重回帰式の偏回帰係数を求める。

3. 正解値データ作成

本章では、上記の重回帰分析における正解値データを取得するための主観評価実験について説明する。

3.1. 評価用ダンス動画作成

重回帰分析で使用する学習データとして、各モーションセット(楽曲とモーションの組)に付与された1秒毎の印象評価値が必要である。これを得るために、主観評価実験を行う(3.2節)が、評価対象となるダンス動画は以下の手順で作成した。

- (1) いくつかのモーションセットを用意する。モーションセットの楽曲を印象推定し、VA平面の第1~4象限(4印象)に対応する楽曲を1曲ずつ用意する。なお、印象推定にはMoodSwing Turk Dataset^[6,7]を正解値データとした重回帰分析を用いた。
- (2) (1)で得られた4楽曲に対応するモーションの中で空間性・力動性の平均値が最大・最小となる領域をそれぞれ20秒間ずつ切り出す。
- (3) RWC研究用音楽データベース全100曲のポピュラー楽曲に対して(1)と同様に楽曲印象推定を行い、4つの印象のいずれかに分類する。
- (4) (2)で用意したモーション各4種類のテンポを(3)の各楽曲のテンポに同期させることで、全部で400個のダンス動画を作成した。

3.2. 実験方法

主観評価実験において、被験者に1人あたり25曲分の動画(各4種類の印象の動画、合計100動画)をランダムに再生したものを視聴してもらい、それぞれについて1フレーム毎のVA平面上の座標をマウスカーソルで指定する。これを20代の男女合計10名に対して行った。

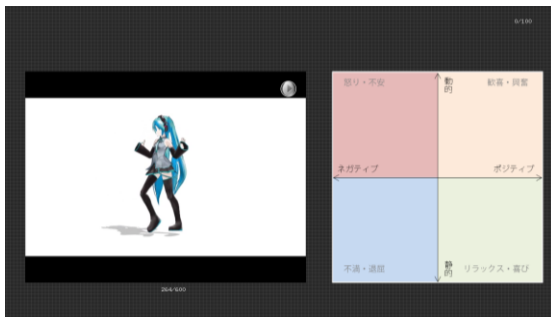


図1. 評価実験で用いたGUI

4. 結果と考察

主観評価実験から得た正解値データを用いて重回帰分

析を行った結果を表3に示す。この結果から、ダンス動画の印象推定を行う際、音響特徴量だけでなく、モーションの特徴量も考慮することで、印象推定精度が向上していることが分かった。さらに、被験者の評価結果を解析すると、モーションの力動性が大きい程VA座標の縦軸(Arousal軸)が大きく、力動性が小さい程縦軸が小さくなる傾向がみられた(図2)。なお、図2のプロット点の色はそれぞれ3名の被験者を表している。

表3. 楽曲特徴量からの印象推定と本手法の比較

印象推定法	決定係数(V座標)	決定係数(A座標)
本手法	0.2915	0.4093
音響特徴量のみ	0.2735	0.3808

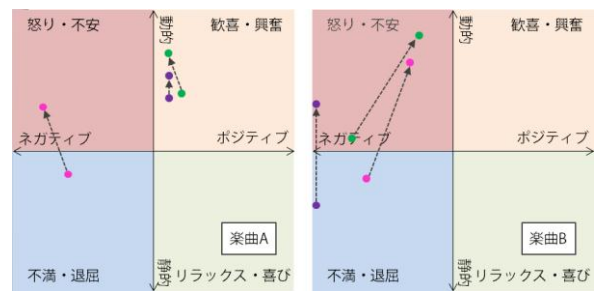


図2. 力動性大小での印象変化(力動小→力動大)

5. まとめと今後の課題

本研究では、キャラクターの表情を自動合成するための、ダンスモーションに同期した楽曲印象推定手法を提案し、ダンス動画に特化した印象推定が可能となった。今後の課題としては、印象推定の精度向上やVA平面に対応した表情の合成方法の再検討などが挙げられる。

謝辞 本研究の一部はJST CREST「OngaCRESTプロジェクト」の支援を受けた

参考文献

- [1] DiPaola, S. and Arya, A. "Emotional remapping of music to facial animation." In Proceedings of ACM SIGGRAPH symposium on Videogames, 143-149, 2006.
- [2] Russell, J.A. "A circumplex model of affect," Journal of Personality and Social Psychology, 39(6), 1161-1178, 1980.
- [3] 西川直毅, 糸山克寿, 藤原弘将, 後藤真孝, 尾形哲也, 奥乃博, "歌詞と音響特徴量を用いた楽曲印象軌跡推定法の設計と評価." 第91回音楽情報科学研究会, 91(7), 1-8. 2011.
- [4] Laban, R. and Ullmann, L. "Mastery of Movement," Princeton Book Company Publishers, 1960.
- [5] 沼口直紀, 中澤篤志, 竹村治雄, "印象語による舞踊動作データの分類法." 情報処理学会 コンピュータビジョンとイメージメディア研究会(CVIM), 167(35), 1-6, 2009.
- [6] Schmidt, E.M. and Kim, Y.E. "Modeling musical emotion dynamics with conditional random fields," In Proceedings of the 12th International Society for Music Information Retrieval Conference(ISMIR), 777-782, 2011.
- [7] Speck, J.A., Schmidt, E.M., Morton, B.G. and Kim, Y.E. "A comparative study of collaborative vs. traditional annotation methods." In Proceedings of the 12th International Society for Music Information Retrieval Conference(ISMIR), 549-554, 2011.