楽曲演奏学習のための時間的揺らぎの特徴抽出

水野美沙 † 鈴木優 † 川越恭二 †
† 立命館大学 情報理工学部

1 はじめに

近年，様々な楽曲演奏における学習が行われている．この手法の一つとして，演奏学習の手本を真似ることによる学習が行われており，楽譜には含まれない演奏者独自の演奏表現を学習者に伝えることが重要となる[1]．そのため，演奏者独自の演奏表現に含まれる特徴を抽出し，数値化することを本研究の目的とした．本研究では，演奏者独自の演奏表現と発音時刻には相関関係があることに着目した．そこで，楽譜に記載されている発音時刻と演奏者独自の発音時刻との差に対して時系列データ特徴抽出手法を適用することによって，演奏者独自の演奏表現における特徴を抽出できると考えた．評価実験において，楽譜と異なる発音時刻で演奏されている楽曲と原曲との直感的な類似度を測定することにより，提案手法によって得られた楽曲間の類似度が利用者の直感に合致していることを示した．

2 提案手法

本研究では，演奏者独自の演奏表現により生じる，人間が直感的に感じる差異を揺らぎと定義する．我々は，揺らぎの要因として発音時刻が重要であると考えた．そこで，発音時刻の差を時系列データと捉え，時系列データ特徴抽出手法を適用する．

まず，楽譜と異なる発音時刻で演奏されている楽曲において，ある音とその直後の音の発音時刻の差をとる．さらに，楽譜通りに演奏された楽曲の発音時刻の差との比較により正規化を行う．この正規化されたデータを揺らぎデータとする．次に，算出した揺らぎデータに対して時系列データ特徴抽出手法を適用し，揺らぎ特徴データを抽出する．最後に，各手法によって抽出された揺らぎ特徴データ間のユークリッド距離を算出し，各手法における原曲と演奏者独自の演奏表現により演奏された楽曲との類似度を求める．

2.1 揺らぎデータの算出方法

ある演奏表現 k によって演奏された楽曲 G^k を考える．G^k の持つ発音開始時刻の集合を $T^k = \{t_1^k, t_2^k, \dots t_N^k\}$ とする．N は楽曲が持つ音の要素数である．なお，$k=0$ は楽譜どおりに演奏された原曲と定義する．このとき G^k の揺らぎデータを M^k とし (1) 式に示す．

$$M^k = \left\langle \frac{t_{n+1}^k - t_n^k}{t_{n+1}^0 - t_n^0} \;\middle|\; n = 1, 2, \dots, N-1 \right\rangle \quad (1)$$

2.2 揺らぎ特徴抽出手法

M^k に対して，時系列データ特徴抽出手法を適用することによって，M^k が持つ揺らぎの特徴を抽出する．以下では，利用可能な 3 種類の時系列データ抽出手法を述べる．

2.2.1 離散フーリエ変換

M^k に離散フーリエ変換を適用し M^k の周期を抽出する．M^k における離散フーリエ変換の式を (2) 式に示す．n を要素数番号，Δ_n^k を揺らぎデータの n 番目の要素とする．

$$X_\alpha = \sum_{n=1}^{N-1} \Delta_n^k \cdot \cos\left(\frac{2\pi n\alpha}{N-1}\right) - j \cdot \sum_{n=1}^{N-1} \Delta_n^k \cdot \sin\left(\frac{2\pi n\alpha}{N-1}\right) \quad (2)$$

本研究では，$P_\alpha = \sum_{n=1}^{N-1} \Delta_n^k \cdot \cos\left(\frac{2\pi n\alpha}{N-1}\right)$, $\alpha = 0, 1, 2, 3$ とし，M^k における揺らぎ特徴データを $F^k = \{P_0, P_1, P_2, P_3\}$ とする．

2.2.2 ウェーブレット変換

M^k において，近似関数と差分関数を用いて近似した値を抽出する．M^k における次数を i とした近似関数を A_i $(i = 1, 2 \dots, a)$，A_i の m 番目の要素を $A_i(m)$ とする．ここで $a = \log_2(N-1)$ とする．このとき A_i を (3) 式に示す．なお A_i の要素数は 2^{a-i}，$A_0 = M^k$ である．

$$A_i = \left\langle \frac{A_{i-1}(2m-1) + A_{i-1}(2m)}{2} \;\middle|\; m = 1, 2, \dots, 2^{a-i} \right\rangle \quad (3)$$

次に，次数を i とした差分関数を D_i $(i = 1, 2 \dots, a)$，D_i の m 番目の要素を $D_i(m)$ とする．このとき D_i を (4) 式に示す．

$$D_i = \left\langle \frac{A_{i-1}(2m-1) - A_{i-1}(2m)}{2} \;\middle|\; m = 1, 2, \dots, 2^{a-i} \right\rangle \quad (4)$$

本研究では，M^k における揺らぎ特徴データを $W^k = \{A_a, D_a, D_{a-1}\}$ とする．

2.2.3 区分定数近似法

Chakrabarti[2] によって提案された区分定数近似法を利用した．M^k において離散幅が小さい値を近似していく．このとき近似値と近似値が連続する区間を抽出する．近似値が連続する区間 div_x は，高さ h_x，幅 w_x，開始地点 w_{x-1} で表される．なお x を区間番号，$w_0 = 0$

Time Feature Extraction of Temporal Music Tune
Misa MIZUNO†, Yu SUZUKI†, Kyoji KAWAGOE†
†College of Information Science and Engineering,Ritsumeikan University.
†{misa,suzuki,kawagoe}@coms.ics.ritsumei.ac.jp

とする．
本研究では，M^k における揺らぎ特徴データを $K^k = \{h_1, w_1, h_2, w_2, h_3\}$ とする．

3 評価実験

3.1 実験条件

ある一つの楽曲において，異なる発音時刻で演奏された楽曲を人間が聴き比べた際に，原曲に近いと感じた順と，提案手法を適用後の原曲と揺らぎ特徴データとの類似度により算出した順が，一致するかどうかを確かめることにより本研究の有効性を示す．本実験では，人間が揺らぎを感じる基本的な条件を，揺らぎを持つ音の音長の差，位置の差，個数の差であると仮定し，パターンIとパターンIIを設定した．パターンIは音長に着目した特徴を抽出するために，パターンIIは揺らぎを持つ音の位置と個数に着目した特徴を抽出するために設定している．なお，原曲 M^0 の音の要素数は33音，データ全体の長さは21秒である．

1. パターンI：特定の8音をそれぞれ0.25秒延長した楽曲を d_1，0.50秒延長した楽曲を d_2，1.0秒延長した楽曲を d_3 とする．
2. パターンII：0.25秒延長した音が8音の楽曲を d_1，延長する音を d_1 から3音減らした楽曲を d_4，d_1 と異なる8音を0.25秒延長した楽曲を d_5 とする．

なお，著者が実際に楽曲を聴き，パターンIにおける原曲との類似度が高い順を d_1，d_2，d_3，パターンIIにおける原曲との類似度が高い順を d_4，d_1，d_5 と決定した．

3.2 実験結果および考察

パターンIにおける原曲との類似度を表1，パターンIIにおける原曲との類似度を表2，原曲に近い楽曲順を表3に示す．なお表3において，F はフーリエ変換，W はウェーブレット変換，K は区分定数近似法を表す．

表3より，パターンI，IIの各手法とも人間が感じた揺らぎを特徴として抽出できた．また表1，2から，特に揺らぎを持つ音要素の個数が異なる場合に，類似度の差が大きくなることがわかった．そのため，揺らぎの特徴が個数に依存する場合に，これらの手法は有効であると考えられる．特に区分定数近似法を利用した場合に，最も個数の差による揺らぎの特徴が考慮されていた．しかし，区分定数近似法は音長を特徴とした場合，個数を特徴とした場合よりも揺らぎの特徴を考慮できていなかった．そのため，揺らぎの特徴抽出において有効なパターンとそうでないパターンの差が他の手法よりも大きいことがわかった．これに対しウェーブレット変換は，類似度の変化の割合は区分定数近似法よりも小さいが，音長，位置，個数の差の全てを特徴として考慮することができていた．そのため，個数の差を特徴とする場合には区分定数近似法が有効であり，音長，個数，位置の全ての差を特徴とする場合にはウェーブレット変換が有効である．

表1: パターンIにおける実験結果

フーリエ変換		ウェーブレット変換		区分定数近似法	
曲番号	類似度	曲番号	類似度	曲番号	類似度
d_1	0.060	d_1	0.57	d_1	0.060
d_2	0.049	d_2	0.53	d_2	0.060
d_3	0.042	d_3	0.47	d_3	0.059
平均	0.050	平均	0.52	平均	0.060

表2: パターンIIにおける実験結果

フーリエ変換		ウェーブレット変換		区分定数近似法	
曲番号	類似度	曲番号	類似度	曲番号	類似度
d_1	0.060	d_1	0.57	d_1	0.060
d_4	0.42	d_4	0.95	d_4	0.50
d_5	0.039	d_5	0.52	d_5	0.055
平均	0.17	平均	0.68	平均	0.21

表3: 原曲に近い実験データ順

	パターンI				パターンII			
手法 / 順位	人	F	W	K	人	F	W	K
1	d_1	d_1	d_1	d_1	d_4	d_4	d_4	d_4
2	d_2	d_2	d_2	d_2	d_1	d_1	d_1	d_1
3	d_3	d_3	d_3	d_3	d_5	d_5	d_5	d_5

4 おわりに

本研究では，楽曲の時間的揺らぎにおける特徴抽出を行った．その結果，各手法によって揺らぎの特徴を抽出可能であることがわかった．今後は多様な揺らぎパターンを用いることや，異なる曲同士との比較評価を行う必要がある．さらに，ウェーブレット変換と区分定数近似法を組み合わせることによって，より精度の高い結果を得られるようにな手法を提案する予定である．

謝辞 本研究の一部は文部科学省科学研究費補助金（課題番号：20500104）によって行われた．ここに記して感謝の意を表す．

参考文献

[1] 大島千佳, 宮川洋平, 西本一志：“Coloring-in Piano: 表情付けに専念できるピアノの提案”, 情報処理学会音楽情報科学研究会研究報告, **vol.2001**, No.103, pp. 69–74 (2001).

[2] K. Chakrabarti, E. Keogh, S. Mehrotra and M. Pazzani: “Locally adaptive dimensionality reduction for indexing large time series databases”, ACM Transactions on Database Systems (TODS), **27**, 2, pp. 188–228 (2002).