

事例に基づく演奏表情生成システムにおける 複旋律を考慮した表情生成式

原田 杏奈[†] 清水 厚志[†] 鈴木 泰山^{††} 野池 賢二^{†††} 金子 雄介^{††††} 徳永 幸生[†]

芝浦工業大学 情報工学科[†] 株式会社ピコラボ^{††}

株式会社トランス・ニュー・テクノロジー^{†††} 株式会社日本総合研究所^{††††}

1.はじめに

本稿では、事例に基づく演奏表情生成システムにおける旋律類似性評価式の改良と評価について述べる。事例に基づく演奏表情の自動生成では、事例に用意した演奏データ集の中から対象曲に類似した事例を検索し、それらを用いて演奏データを生成する。現在までに我々が用いてきた、類似事例の検索や演奏表情の生成のための旋律類似性評価式は、複旋律の楽曲であっても単旋律とみなして扱っており、生成演奏表情の質を下げる一因となっていた。そこで本研究では、複旋律を考慮した旋律類似性評価式の改良を行い、生成された演奏の評価実験を行った。

2.事例に基づく演奏表情生成システム

本研究で用いた事例に基づく演奏表情生成システム”Kagurame Phase-II”[1][2][3]の構成を、図1に示す。

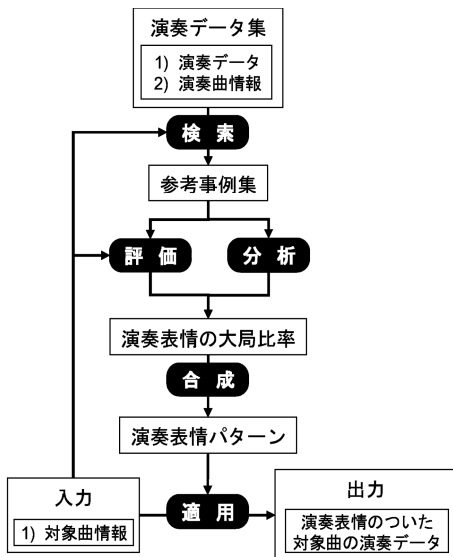


図1 Kagurame Phase-II の構成

本システムでは、あらかじめ演奏データ集と対象曲ともに楽曲を階層的な旋律断片に分割する。そして演奏データ集から対象曲に類似する旋律を検索し、その旋律の類似度を評価する際に、以下の4つの旋律類似性評価式を用いている。

- 1) LengthEvaluator : 旋律の長さの類似性 D_l
- 2) KeyRateEvaluator : 音高の類似性 D_k
- 3) RhythmEvaluator : リズムの類似性 D_r
- 4) HarmonyEvaluator : 音階の類似性 D_h

これら4つの旋律類似性評価式をさらに式(1)により正規化して合計し、旋律の類似性 D を算出する。

$$D = A_l D_l + A_k D_k + A_r (1 - D_r) + A_h (1 - D_h) \quad (1)$$

D は0以上の値であり、0に近いほど類似している。 $A = (A_l, A_k, A_r, A_h)$ は重み付け変数であり、それぞれ0以上の値をとる。この変数を変化させることで、それぞれ異なる演奏表情をもつ演奏が生成される。一方、 A の値が大きくなると、最高重要度の参考事例のみをもとに演奏表情が生成される。

3.複旋律を考慮した旋律類似性評価式

音高の類似性 D_k は、比較する2つの旋律断片 s の音高比の特徴量 $C_k(s)$ の、差の絶対値で表される式(2), (3)。

$$C_k(s) = \frac{\sum_{n \in N_s} p(n)}{|N_s|} - \frac{\sum_{n \in N_p} p(n)}{|N_p|} \quad (2)$$

$$D_k(s_1, s_2) = |C_k(s_1) - C_k(s_2)| \quad (3)$$

ここで、 N_s は対象旋律断片 s の音符の集合、 N_p は旋律断片 s の1つ上の親に当たる旋律断片の音符の集合、 $p(n)$ は音符の音高を示す。

式(2)では、対象となる旋律断片に含まれる音符全ての音高の平均を算出しているため、楽曲を単旋律とみなしていることになる。このため楽曲の構成によっては、主旋律の音高と異なる音域が強調された演奏が生成されてしまう。

そこで、ピアノ譜の上段、下段をもとに楽曲を主旋律と副旋律に分け、旋律ごとに音高の特徴の類似性 D_k を算出するように旋律類似性評価

Generation Algorithm considered homophony in a Case-Based Performance Rendering System

[†] Department of Information Science and Engineering, Shibaura Institute of Technology

^{††} PicoLab Co., Ltd.

^{†††} Trans New Technology, Inc.

^{††††} The Japan Research Institute, Limited

式の改良を行った．具体的には，旋律断片内の主旋律と複旋律とでそれぞれ音高の類似性 D_{kmain} ， D_{ksub} を算出し，それらを加重平均した値を旋律断片全体の音高の類似性 D_k とした．(式(4))．

$$e^{D_k} = a \cdot e^{D_{kmain}} + (1-a) \cdot e^{D_{ksub}} \quad (4)$$

ここで a は重み付け変数であり， $0 \leq a \leq 1$ である．

4.改良したシステムの評価

改良した表情生成式の有効性を評価するために，改良前と後のそれぞれのシステムが生成した演奏を用いて，定量評価と定性評価の 2 通りの評価実験を行った．

演奏データ集には Chopin 作曲の楽曲 5 曲を設定し，それらに含まれない Chopin 作曲「Etude Op.10 No.3」の冒頭 8 小節の演奏を生成した．各旋律の重み付けは，主旋律と副旋律との音符数の比をもとに，主旋律 0.9，副旋律 0.1 とした．

システムが人間の演奏にどの程度近い演奏を生成できたかを定量評価するため，システムが生成した演奏と人間の演奏との相異度 $ratio$ を，以下の式(5)により算出する．

$$ratio = \frac{1}{|N|} \sum_{n \in N} \left| \ln \frac{E_{kp2}(n)}{E_h(n)} \right| \quad (5)$$

N は楽曲に含まれる音符の集合であり， $E_{kp2}(n)$ はシステムが生成した演奏表情の値， $E_h(n)$ は人間の演奏表情の値である．

相異度が小さいほど，人間の演奏とシステムの演奏が近いことを表す．相異度が小さいほど，実演奏とシステムの演奏が近いことを表す．

改良前の演奏と改良後の演奏の音の強さ (Velocity) に対する相異度 $ratio$ を表 1 に示す．改良後の方が， $ratio$ の値が低くなっていることがわかる．

表 1 相異度 $ratio$ の比較

	ratio
改良前	0.28
改良後	0.18

定性評価の実験は，150 人収容できる教室でスピーカーから流した演奏を試聴する全体実験と，ヘッドフォンを用い被験者が任意に演奏を試聴するヘッドフォン実験の 2 種類を行った．被験者は全員工学部の学生であり，全体実験が 102 名，ヘッドフォン実験が 39 名である．

実験方法は，「蓮根」方式[4]を参考に設定した．被験者は演奏を聴き，どの程度人間の演奏に似ているかを示す”近似度(5 が似ている)”、どの程度自然な演奏に聴こえるかを示す”自然度(5

が自然)”をそれぞれ 5 段階評価で解答用紙に記入する．演奏の試聴順序や回数は，全体実験では決められているのに対し，ヘッドフォン実験では被験者の自由とした．

表 2 試聴実験の結果

	近似度		自然度	
	全体	ヘッドフォン	全体	ヘッドフォン
改良前	2.62	2.33	2.52	2.25
改良後	2.94*	2.76	2.74	3.07**

*は 5%水準で有意．**は 1%水準で有意

試聴実験の結果を表 2 に示す．

全体実験，ヘッドホン実験の両方とも，改良後の方が近似度，自然度とも高くなっていることがわかる．この結果から，表情生成式の改良によって演奏表情の質が向上したことを，人間の主観による定性評価によっても確認できた．

これらの二つの評価実験の結果から，複旋律を考慮した表情生成式への改良が有効であることが確認でき，システムの生成演奏をより人間の演奏に近づけることに寄与したことを示せた．

5.おわりに

本稿では，事例に基づく演奏表情生成システムにおける旋律類似性評価式の改良とその評価について述べた．これまで単旋律にしか対応していなかった表情生成式を複旋律に対応させたことで，人間の演奏により近い演奏が生成できたことが確認できた．

今回は音の強さをもとに定量評価を行ったが，今後は，音符の長さや発音時刻による評価も行う．また，他の演奏データ集と対象曲の組み合わせで同様の実験を行い，楽曲の特徴により生成される演奏表情の違いも考察する．

参考文献

- [1] 鈴木泰山，徳永健伸，田中穂積：事例に基づく演奏表情の生成，情報処理学会論文誌，Vol.41，No.4，pp1134-1145，2000.
- [2] 金子雄介，鈴木泰山，徳永幸生：事例に基づく演奏表情生成システムにおける表情生成式の最適化，情報処理学会研究報告音楽情報科学，2004-MUS-58，Dec2004.
- [3] 清水厚志，原田杏奈，鈴木泰山，野池賢二，金子雄介，徳永幸生：事例に基づく演奏表情生成システムにおける MusicXML からの旋律断片自動生成，情報処理学会研究報告音楽情報科学，2006-MUS-68，Dec2006.
- [4] 平賀瑠美，平田圭二，片寄晴弘：蓮根，目指せ世界一のピアニスト，情報処理，Vol.43，No2，pp.136-141，2002.