Polyhymnia: 多重音演奏の統計モデルと演奏記号解釈による自動ピアノ演奏表情付けシステム*

○金 泰憲, 深山 覚, 西本 卓也, 嵯峨山 茂樹 (東大院・情報理工)

1 はじめに

本稿では我々が開発した自動ピアノ演奏表情付けシステム Polyhymnia を紹介する．本システムは演奏記号の自動解釈と多重音演奏の統計モデルに基づいた演奏表情の推定によって，与えられた未知のピアノ曲に対して，表情豊かでかつ多様なスタイルの演奏の MIDI データを自動的に生成することを目的とする．本システムを用いれば，ピアノが弾けない人でも著作権フリーの演奏が手軽に得られる．また，得られたモデルとそれを用いて生成された演奏の分析による人間の音楽演奏に関する新しい知見が期待される．Polyhymnia は 2010 年の表情付けシステムの為のコンテスト (Rencon2010) で自律部門の一位を獲得した．

2 関連研究

ピアノ演奏の自動演奏表情付けを行う様々なシステムが提案されている [1]．Director Musices は専門家によって抽出された様々な演奏規則を実装することによって演奏表情を生成する．Kagurame と COPER は人間の演奏表情の DB を検索することによって演奏表情を生成する．ESP, YQX そして usapi はピアノ演奏の統計モデルを用いて人間の演奏から演奏表情を学習し，未知の曲に対する演奏表情を生成する．今まで提案されたシステムは，主に単旋律の演奏に関して良い結果を報告しているが，演奏記号の解釈，多重音演奏に関する表情付け，そして多様な演奏表情の生成に関してはあまり議論されていなかった．Polyhymnia はその三つの問題を取り扱うことによって，より人間の演奏に近い演奏表情の生成することを目的とする．

3 ピアノ演奏における演奏表情

ピアノ演奏は瞬時テンポ，音量，演奏された音長 (以下，演奏音長) と言った 3 つのパラメータで表現できる．人間のピアノ演奏を見ると，演奏記号による大間かな演奏表情が観測され，さらに音符パターンによる細かい演奏表情が観測される．旋律の演奏では瞬時テンポ，音量，演奏音長が常に変動しているが，多重音演奏ではさらに各旋律の表情がお互いに異なる．また，和音の演奏では各音符の発音時間のずれ，音量，演奏音長が異なる [2]．この知見から，より表情豊かなピアノ演奏を生成するためには，演奏記号の解釈と多重音演奏の特徴を再現する必要があると考えられる．

4 演奏記号の自動解釈

楽譜には *cresc.*, *dim.* などの強弱記号や *rit.* などのテンポ記号がある．これらの記号は MIDI 形式では表現し難いため，Polyhymnia は MusicXML 形式の楽譜を入力とする．演奏記号の人間の解釈を分析した結果，テンポと音量の制御が決して線形ではなく，時間に対して指数的に変化していることがわかった．

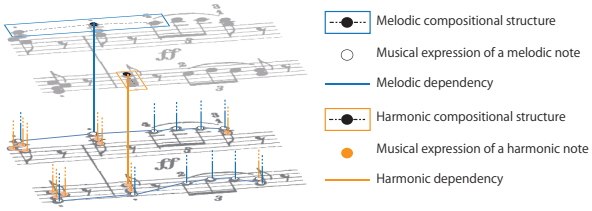

Fig. 1 外声部と和音の演奏の組み合わせを用いた近似により単純化された演奏表情の依存関係．

さらに，*rit.* によるテンポの低下では，多くの場合音量も共に低下することがわかった．

このような演奏を再現するために，我々は次のような数理モデルを提案する．時刻 t でのテンポや音量などの演奏表情を d_t とした時，その時間変化の制御は

$$d_t = d_0 \cdot (\beta \cdot t^{\alpha} + 1) \tag{1}$$

によりモデル化した．ここで t は演奏記号の始まりを 0, 終わりを 1 とした時の時刻である．α は制御の形を決めるパラメータであり，$0 < \alpha < 1$ の場合は凸形，$\alpha > 1$ の場合は凹形のように制御される．β は制御の方向とその深さを表すパラメータであり，$\beta > 0$ の場合は増加，$\beta < 0$ の場合は低下する．

また，楽譜には *trill*, *turn* などの装飾記号がある．我々はこれらの記号により展開される音符の音量は正規分布に従うと仮定し，展開される音符 i の音量 d_i を

$$d_i = d_0 + \mathcal{N}(0, \sigma^2) \tag{2}$$

によりモデル化した．ここで，d_0 は展開される音符の親の音符の音量を表す．

5 多重音演奏の統計モデル

一つの音符に対する演奏表情は周りの音符のパターンに依存しており，前後の演奏表情にも依存していると考えられる．多重音演奏のモデリングでは，音符パターンが複雑であり，各音符の演奏表情間の依存関係も複雑になるため，計算不可能なモデルになるか，膨大な学習データが必要であると言った問題がある．この問題を解決するために，我々は外声部の演奏と和音の演奏の組み合わせによる近似に基づいた新しい手法を提案した [2]．Fig. 1 は提案した近似を表す．人間の聴覚は

- 複旋律の演奏から各旋律の表情が認知できる，
- 和音の異なる演奏表情が認知できる，
- 内声部より外声部の演奏表情がより認知し易い

といった特性を持つと言われている [3]．そのため，前に述べたような近似を用いると人間の聴覚に一番相応しい演奏表情の生成が可能であると考えられる．また，多重音演奏のモデルを旋律演奏のモデルと和音演奏のモデルといった単純なモデルの組み合わせとして表現するため，少ないデータを用いても多重音演奏の学習および推定が可能である．

我々は単旋律と和音の演奏を Linear Chain Conditional Random Fiels [5] としてモデル化した．音符 i の演奏表情を d_i とする．音符パターンを表す楽譜素

---

*Polyhymnia: Automatic Piano Performance System with Statistical Modeling of Polyphonic Expression and Interpretation of Musical Symbols. by Taehun KIM, Satoru FUKAYAMA, Takuya NISHIMOTO and Shigeki SAGAYAMA (The University of Tokyo)

Table 1　生成実験で用いたテスト曲

ID	作曲家	曲名	テンポ
CF	F. Chopin	Mazurka no. 5	fast
CS	F. Chopin	Trauer Marsch	slow
MF	W. A. Mozart	Sonatina no. 5-3	fast
MS	W. A. Mozart	Marche Funebre	slow
RT	S. Joplin	The Entertainer	middle
GR	E. Grieg	7 Ly. Pieces op. 71	slow

性ベクトルは S_i とし，その要素を s_{ik} とする．演奏表情列 D に対する楽譜素性ベクトルの列を $\mathbf{S}$ とし，d_i は d_{i-1} のみに依存すると仮定すると，$\{d_{i-1}, d_i, s_{ik}\}$ の j 番目の組み合わせの有無を表す局所素性関数 f_j と，その組み合わせが D と $\mathbf{S}$ の中に出現する回数を表す大域素性関数 F_j は

$$f_j(d_{i-1}, d_n, \mathbf{S}, i) = \delta(\{d_{i-1}, d_i, s_{ik}\}_j) \tag{3}$$

$$F_j(D, \mathbf{S}) = \sum_{i=1}^{I} f_j(d_{i-1}, d_i, \mathbf{S}, i) \tag{4}$$

のように定義てきる．各 F_j の重み θ_j を導入すると，$\mathbf{S}$ が与えられた時の D の条件付き確率分布 $P(D|\mathbf{S})$ は情報量最大原理により

$$P(D|\mathbf{S}) = \frac{1}{Z(\mathbf{S})} \exp \sum_j \theta_j F_j(D, \mathbf{S}) \tag{5}$$

のように定義できる．ここで $Z(\mathcal{S})$ は正規化因子であり，

$$Z(\mathcal{S}) = \sum_{D'} \exp \sum_j \theta_j F_j(D', \mathbf{S}) \tag{6}$$

のようにで定義できる．ただし D' は学習データに含まれている演奏表情列である．$Z(\cdot)$ は Forward-backward アルゴリズムで効率的に計算でき，Θ は Stochastic Gradient Descent アルゴリズム [6] による最尤推定法を用いて学習データから求められる．未知の曲に対する演奏表情の推定は，$P(\cdot)$ を最大にするような D を求める最適化問題に帰着し，Viterbi アルゴリズムで計算できる．

6　評価実験

6.1　多様な演奏表情の生成

Polyhymnia を用いて様々な曲の演奏表情を生成した．演奏表情の生成では CrestMuse PEDB [4] に含まれている F. Chopin の 15 曲に対する V. Ashkenazy の演奏[1]を学習させたモデル (*Ashkenazy-model*) と W. A. Mozart の 7 曲に対する G. Gould の演奏[2]を学習させたモデル (*Gould-model*) の二つを用意した．テスト曲としては Table 1 のような 6 曲を用いた．

Fig. 2 に Polyhymnia による自動表情付け結果の一部を示す．この結果から，生成された二つの演奏表情

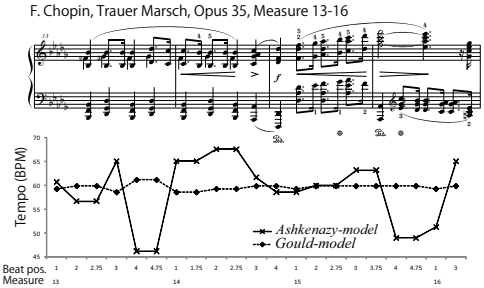

Fig. 2　F. Chopin, Trauer Marsch の自動演奏結果．紙面関係上テンポのみを示す．

[1]Prelude, no. 1 ,4, 6, 15, 20, Etude op. 10-3, 10-4, 25-11, Waltz op. 18, 34-2, 64-2, 69-1, 69-2, Nocturne no. 2, 10

[2]Piano sonata KV. 279-1, 279-2, 279-3, 331-1, 545-1, 545-2, 545-3

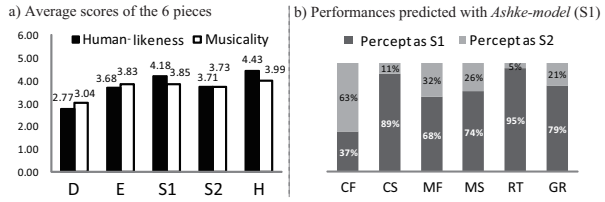

Fig. 3　6 曲の各演奏に対する主観評価の平均点数 (a) と 6 個の S1 演奏に対するスタイル識別実験結果 (b)．

はお互いに異なることがわかった．*Ashkenazy-model* による演奏表情では大幅なテンポの変動が見られたが，*Gould-model* による演奏表情ではテンポ変動の幅が小さいことがわかった．

6.2　主観評価

生成された演奏表情を評価するために主観評価を行った．各曲に対して表情なし演奏 (D)，演奏記号解釈のみの演奏 (E)，*Ashkenazy-model* による演奏 (S1)，*Gould-model* による演奏 (S2)，人間の演奏者による演奏 (H) の 5 つの演奏を用意し，19 人の被実験者に聞かせた．各演奏は human-likeness と musicality を 6 段階で評価してもらった．

Fig 3 の a) に 6 曲に対する評価結果の平均点数を示す．信頼度 95%の分散分析の結果，S1, S2, E はどちらにしても D より良い演奏のように聴こえたのがわかった．S1 と H の間は有意差がなく，S2 も H に近いように聴こえたことから，二つのモデルによる演奏は両方人間の演奏に近いように聴こえたのがわかった．

各モデルが学習した演奏スタイルを反映しているかを検証するために，もう一つの実験を行った．この実験では被実験者がテスト曲には含まれていない 3 つの曲に対する各モデルによる演奏を事前に聴く事によって各モデルの演奏スタイルを覚える．その後，12 個全ての S1, S2 をブラインドで提示し，実際に S1 と S2 のどちらの演奏の様に聴こえたかを答えてもらった．

Fig 3 の b) にその結果の一部を示す．6 つの S1 の中 5 つに対して被実験者の半分以上が正しく演奏スタイルを識別した．識別率の平均は 73.6%であった．6 つの S2 に対しても平均識別率は 73.6%であった．この結果から，Polyhymnia が生成した演奏は学習した演奏スタイルを反映していたのがわかる．

7　おわりに

本稿では演奏記号の自動解釈と多重音演奏の学習および推定可能な自動ピアノ演奏システム Polyhymnia を紹介した．評価実験の結果，本システムは学習したスタイルを保った表情豊かな演奏が生成できることを確認した．今後は楽曲の構造と演奏表情の関係をモデルに加えることによって，より表情豊かな演奏を目指したい．また，本システムはウェブサービスとして一般公開する予定である．

謝辞　本研究の一部は CrestMuse Project と Samsung Scholarship Foundation の支援を受けて行われた．

参考文献

[1] A. Kirke *et al.*, ACM Comp. Surv., 42(1), 2009.
[2] T. H. Kim *et al.*, Proc. SMC, 23–30, 2010.
[3] D. Huron *et al.*, Music Perc., 7(1), 43–48, 1989.
[4] M. Hashida *et al.*, Proc. ISMIR, 489–494, 2008.
[5] J. Lafferty *et al.*, Proc. ICML, 282–289, 2001.
[6] L. Bottou, Adv. Lectures on ML, 146–168, 2004.