Vol.2010-MUS-85 No.2
2010/5/27

単旋律と和音の確率モデルの組み合わせによるピアノ曲演奏の自動表情付け

金　泰憲[†1]　深山　覚[†1]
西本卓也[†1]　嵯峨山茂樹[†1]

楽譜情報を基に人間らしい演奏表情を自動的に生成する問題に対して，確率モデルに基づいた機械学習手法が応用されて来た．しかし演奏楽譜に多重音が含まれる場合はモデルが複雑になるため，膨大な学習データが必要，ないし計算が困難になるといった問題があった．本稿では単旋律と和音の確率モデルの組み合わせによりデータスパースネス問題を避けながら，多重音を含むピアノ曲演奏の表情を自動的に付ける手法を提案する．評価実験の結果，多重音を含むピアノ曲に対して，人間らしい表情を持った演奏が生成されることが分かった．また心理実験による主観評価では，提案手法を用いて生成した演奏表情が人間らしく，さらには音楽的に自然に聴こえることが確認された．

Performance Rendering for Piano Music with a Combination of Probabilistic Models for Melody and Chords

Tae Hun Kim,[†1] Satoru Fukayama,[†1]
Takuya Nishimoto[†1] and Shigeki Sagayama[†1]

In this paper, we present a method to generate human-like performance expression for polyphonic piano music with a combination of probabilistic models for melody and chords to avoid data sparseness problems. Probabilistic models and machine learning have been applied to solve the problem of generating human-like expressive performance given a music score. In case of polyphonic music, however, it was difficult to make a tractable model and a huge amount of training data was necessary. The results of the experiments show that the proposed method is able to generate fluctuations of performance parameters for polyphonic piano music such like human performers do. The results of subjective evaluations are also reported which indicate that the generated performance expression sounded human-like and have certain degree of musicality.

1. はじめに

人間の演奏は表情を持つと言われているが，人間の演奏表情付けに関するメカニズムはまだ不明な点が多く，それを計算機で再現するのはとても難しい問題である．しかし，楽譜を入力として自動的に演奏表情を生成するシステムがあれば，動画やホームページなどの背景音楽として名曲の著作権フリー演奏が自動的に得られ，著作権の問題をさけて楽曲の使用が可能になる．また楽器演奏が出来ない一般人への作曲支援や音楽教育支援も考えられる．

本研究ではピアノ演奏の表情付けを扱う．ピアノは，演奏表情の表現力に優れた広く普及している楽器であり，その演奏表情は管楽器や弦楽器に比べて比較的単純なパラメータで表現できる特徴を持つ．

人間のピアノ演奏を観察すると，旋律においては音符ごとの瞬時テンポ，音量，そして演奏音長が常に変動していることが分かり（**図 1**），和音は発音時間のずれ，各音符の音量，そして演奏音長が和音構成音ごとに異なることが分かる（**図 2**）．ピアノ曲に対する自動演奏表情付けは，これらの変動を楽譜を入力として生成できればよい．しかし，楽譜と演奏表情の関係の多くは明示的ではないため，ピアノ曲の自動演奏表情付けは非常に難しい問題となっている．

この問題を解くために，従来確率モデルに基づいた実演奏データからの機械学習を用いる手法が応用されてきた．しかし多重音が演奏楽譜に含まれる場合，そのモデルが非常に複雑になり，膨大な学習データが必要となる．そのため，場合によっては自動表情付けが困難な場合がある．本稿では，多重音を含むピアノ曲を対象として，データスパースネス問題を避けながら人間らしい演奏表情を自動的に生成する手法を提案する．

2. 関連研究

1983 年の J. Sundberg らの研究[1] をはじめ，自動演奏表情付けに関する様々な計算機モデルが提案されている．ここでは，その中でも確率モデルに基づく実演奏データからの機械学習手法を用いた関連研究を簡単にまとめる．

S. Flossmann らは演奏文脈 (performance context) を導入し，ピアノ曲を対象とした単

†1 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, the University of Tokyo

Vol.2010-MUS-85 No.2
2010/5/27

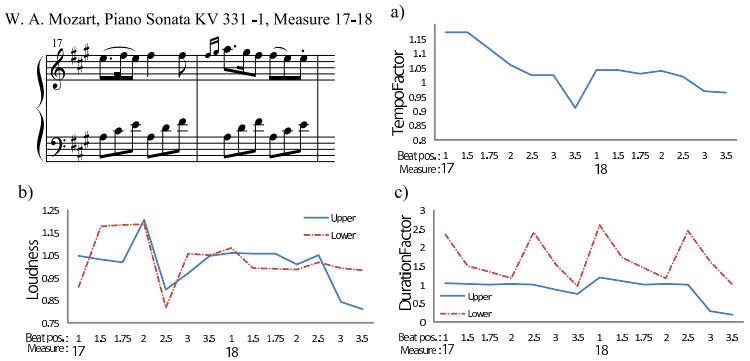

図 1 旋律に対する人間の演奏表情（Ingrid Haebler, CrestMuse PEDB より）．a) は式 (1) を用いて求めた瞬時テンポの変動，b) は式 (2) を用いて求めた音量の変動，c) は式 (3) を用いて求めた演奏音長の変動を表す．

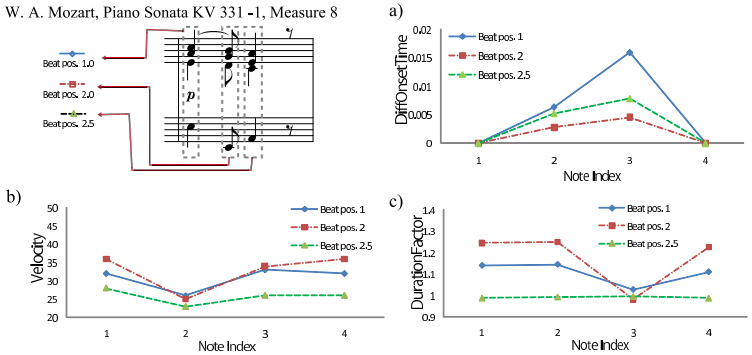

図 2 和音に対する人間の演奏表情（Ingrid Haebler, CrestMuse PEDB より）．a) は式 (4) を用いて求めた各音符の発音時間のずれ，b) は式 (5) を用いて求めた各音符の音量，c) は式 (6) を用いて求めた各音符の演奏音長を表す．Note Index は和音の構成音を表す（1 が最も高い音符）．

旋律メロディーの確率モデルに基づく機械学習手法を用いた自動演奏表情付け手法を提案している[2]．モデルの学習は 2 人の有名ピアニストの演奏を用いて行われ[⋆1]，演奏表情は音符ごとの瞬時テンポ，音量，そして演奏音長の三つのパラメータの推定によって生成されている．K. Teramura らはガウシアンプロセスを用いて演奏表情の学習と推定を行い，演奏表情パラメータには打鍵開始時間，離鍵時間，音量，瞬時テンポの四つを用いている[3]．特に瞬時テンポについては，テンポの周期性と連続性を考慮した手法を提案している[4]．これらの研究はともに単旋律メロディーモデルに基づいた手法であり，多重音を含む曲に関しては論じていない．

G. Grindlay らは HMM(Hidden Markov Model) に基づいたピアノ曲の自動演奏表情付け手法を提案し，伴奏パートの演奏表情付けに関しても論じているが，多重音の場合は音符ごとの発音時間のずれ，音量，そして演奏音長を生成することは出来ない[5]．

⋆1 N. Magaloff, R. Batik

3. 提 案 手 法

本研究では多重音を含むピアノ曲演奏の自動表情付けのために，以下の 2 点を考慮した手法を提案する．

音楽心理学によると，人間が多重音を含む楽曲を聴く時には特に外部声部が聴き取り易いと言われている[6]．従って，多重音を含むピアノ曲の演奏表情は上下の外部声部と和音の演奏表情の組み合わせとして近似することができる．また図 1 に見て取れるように，人間の演奏を観察すると，最も高い声部と最も低い声部の演奏表情は同じではない．

ピアノ曲の演奏表情には発想記号や構造分析に基づいた大局的な演奏表情と，図 1 のように前後の文脈や拍に基づいた局所的な演奏表情がある．自動演奏表情付けにはこれらの演奏表情を両方生成する必要があるが，本研究では局所的な演奏表情の自動生成手法を論じる．局所的な演奏表情付けでは，演奏文脈との依存関係の多くが明示的な規則で書けない．しかし確率モデルを用いればそれらの依存関係の傾向が効果的に扱える可能性がある．このとき，局所的な演奏表情の自動生成問題は，豊かな楽譜素性で表せる演奏文脈（performance context）の列が与えられた時，最も確率が高い演奏表情の列を推定する問題であると捉えることができる．

このような議論に基づいて，多重音を含むピアノ曲の自動演奏表情付けのために単旋律と和音の確率モデルを導入し，次のような戦略で演奏表情の学習と推定が行える．

学習フェーズ

(1) 学習対象の楽譜を右手と左手に分ける．
(2) 右手の場合は最上の音符列を外部声部として抽出し，左手の場合は最下の音符列を外部声部として抽出する．各手ごとに和音を抽出する．
(3) 抽出した外部声部とその演奏データを用いて，右手と左手の単旋律モデルを学習する．
(4) 抽出した和音とその演奏データを用いて，右手と左手の和音モデルを学習する．

推定フェーズ

(1) 入力された楽譜を右手と左手に分ける．

(2) 右手の最上の音符列の演奏表情を学習された右手の単旋律モデルを用いて推定し，左手の最下の音符列の演奏表情を学習された左手の単旋律モデルを用いて推定する．

(3) 右手の和音の演奏表情を学習された右手の和音モデルを用いて推定し，左手の和音の演奏表情を学習された左手の和音モデルを用いて推定する．

(4) 推定された四つの演奏表情を統合する．

この戦略では，楽譜を外部声部と和音に分離したことによって，単純な楽譜素性を用いた線形マルコフモデルで問題を解くことが出来る．

以下では単旋律モデルと和音モデルを用いた演奏表情の学習および推定に関する具体的な定式化を論じる．

3.1 単旋律モデル

3.1.1 演奏表情パラメータ

単旋律の演奏表情のパラメータは瞬時テンポ，音量，そして演奏音長の三つが考えられる．本研究で用いた具体的な演奏表情パラメータは次の通りである．

瞬時テンポ

$$\mathrm{TempoFactor}_t = \log(\frac{\mathrm{Tempo}_t}{\mathrm{Tempo}_{\mathrm{avg}}^{\mathrm{scope}}}) \tag{1}$$

ここで $\mathrm{Tempo}_{\mathrm{avg}}^{\mathrm{scope}}$ は n_t を現在の音符とした時の $n_{t-3}, n_{t-2}, n_{t-1}, n_t, n_{t+1}$ の平均テンポである．

音　　量

$$\mathrm{Loudness}_t^{\mathrm{melody}} = \log(\frac{\mathrm{Velocity}_t}{\mathrm{Velocity}_{\mathrm{avg}}^{\mathrm{scope}}}) \tag{2}$$

ここで $\mathrm{Velocity}_{\mathrm{avg}}^{\mathrm{scope}}$ は n_t を現在の音符とした時の $n_{t-3}, n_{t-2}, n_{t-1}, n_t, n_{t+1}$ の平均ベロシティーである．

演 奏 音 長

$$\mathrm{DurationFactor}_t^{\mathrm{melody}} = \log(\frac{\mathrm{Duration}_t^{\mathrm{real}}}{\mathrm{Duration}_t^{\mathrm{score}}}) \tag{3}$$

ここで $\mathrm{Duration}_t^{\mathrm{score}}$ と $\mathrm{Duration}_t^{\mathrm{real}}$ はそれぞれ時間 t の楽譜に明示されている音長と実際に演奏された音長である．

演奏表情パラメータ列は現在のパラメータ値が直前のパラメータ値のみに依存すると仮定する事で，1 次マルコフ連鎖でモデル化できる．

表 1 単旋律モデルで用いた楽譜素性

瞬時テンポ	音量	演奏音長
-	Pitch	-
-	$\mathrm{Duration}^{\mathrm{score}}$	$\mathrm{Duration}^{\mathrm{score}}$
NoteInterval I,II,III,IV	NoteInterval I,II,III,IV	NoteInterval I,II,III,IV
DurationRatio I,II,III,IV	DurationRatio I,II,III,IV	DurationRatio I,II,III,IV
Metric I,II	Metric I,II	Metric I,II
ArticuationMarks	ArticulationMarks	ArticulationMarks

3.1.2 楽 譜 素 性

単旋律の場合，学習と推定でパラメータごとに異なる楽譜素性を用いることが有効であると考えられる（**表 1**）．ここで Pitch と $\mathrm{Duration}^{\mathrm{score}}$ はそれぞれ音高と音長である．NoteInterval I,II,III,IV は現在の音符を n_t とした時，それぞれ $(n_{t-3}, n_t t-2), (n_{t-2}, n_{t-1}), (n_{t-1}, n_t), (n_t, n_{t+1})$ の相対音程を，DurationRatio I,II,III,IV はそれぞれの音長比である．Metric は現在の音符が小節内で強迫か弱拍かを表す変数であり、$\{very\_strong, strong, weak\}$ のどれかを値とする．Metric I,II はそれぞれ n_{t-1} と n_t の Metric である．ArticulationMarks は *staccato, accent, fermata* などのアーティキュレーション記号を表す．

3.2 和音モデル

3.2.1 演奏表情パラメータ

和音の演奏表情パラメータとしては発音時間のずれ，音量，そして演奏音長の三つを用れば良い．音高の順番に並べた演奏表情パラメータ列は単旋律モデルと同じく 1 次マルコフ連鎖でモデル化できる．

発音時間のずれ

$$\mathrm{DiffOnsetTime}_i = \mathrm{OnsetTime}_o - \mathrm{OnsetTime}_i \tag{4}$$

ここで $\mathrm{OnsetTime}_o$ は和音の構成音中，外部声部に属する音符の発音時間，$\mathrm{OnsetTime}_i$ は現在演奏する音符の発音時間である．

音　　量

$$\mathrm{Loudness}_i^{\mathrm{chord}} = \log(\frac{\mathrm{Velocity}_i}{\mathrm{Velocity}_o}) \tag{5}$$

ここで $\mathrm{Velocity}_o$ は和音の構成音中，外部声部に属する音符のベロシティー，$\mathrm{Velocity}_i$ は

表 2 和音モデルで用いた楽譜素性

瞬時テンポ	音量	演奏音長
Pitch	Pitch	Pitch
$\text{Duration}^{\text{score}}$	$\text{Duration}^{\text{score}}$	$\text{Duration}^{\text{score}}$
NoteDistance	NoteDistance	NoteDistance
OuterNote	OuterNote	OuterNote

現在演奏する音符のベロシティーである．

演 奏 音 長

$$\text{DurationFactor}_i^{\text{chord}} = \log(\frac{\text{Duration}_i^{\text{real}}}{\text{Duration}_o^{\text{real}}}) \quad (6)$$

ここで $\text{Duration}_o^{\text{real}}$ と $\text{Duration}_i^{\text{real}}$ はそれぞれ和音の構成音中，外部声部に属する音符の演奏音長と現在演奏する音符の演奏音長である．

3.2.2 楽 譜 素 性

和音の場合は三つの演奏表情パラメータ共に同じ楽譜素性を用いて学習と推定を行うことが有効であると考えられる（**表 2**）．ここで Pitch は音高，$\text{Duration}^{\text{score}}$ は音長である．NoteDistance は和音の構成音中，外部声部に属する音符と現在演奏する音符との距離を相対音程を用いて表す．OuterNote は現在演奏する音符が和音の構成音のなかで一番外側の音符であれば true，そうではなければ false が値となる．

3.3 学習と推定

本研究では，演奏表情推定の問題を豊かな楽譜素性で表せる演奏文脈列が与えられた時の最も確率が高い演奏表情列を推定する最適化問題として扱うため，Conditional Random Fields(CRFs)[7] を用いて演奏表情の学習と推定を行うことができる．単旋律モデルと和音モデルともに，Stochastic Gradient Descent アルゴリズム[8] による最尤推定法を用いてモデルパラメータの学習を行い，Viterbi アルゴリズムを用いて演奏表情を推定する．提案手法の実装では León Bottou らの “crfsgd” パッケージ[⋆1]を用いた．

3.4 演奏表情パラメータの量子化

CRFs の特性上，演奏表情の学習および推定を行うためには演奏表情パラメータを量子化する必要がある．実験では各演奏表情パラメータを k-means アルゴリズムを用いて 32 段階に量子化した．各クラスタの初期値は学習データから求めた演奏表情パラメータの事前確率分布に基づいてランダムに与えるため，パラメータの事前生起確率を保ちながら，生起確率が高い演奏表情をより細かく分割するような非線形量子化が可能である．

⋆1 http://leon.bottou.org/projects/sgd

表 3 実験 1 で用いた学習データ．1 曲に対する 4 演奏を用いた．

曲名	演奏者
Piano Sonata KV331, 1st Mov.	Hiroko Nakamura
Piano Sonata KV331, 1st Mov.	Norio Shimizu
Piano Sonata KV331, 1st Mov.	Ingrid Haebler
Piano Sonata KV331, 1st Mov.	Lily Kraus

4. 評 価 実 験

提案手法の多重音を含むピアノ曲に対する演奏表情生成の性能を検証するためにテスト曲が既知である場合と未知である場合の二つの実験を行った．学習データは CrestMusePEDB ver. 2.3 を用いた[⋆2]．

4.1 実 験 環 境

実験 1–テスト曲が既知である場合

実験 1 ではテスト曲がシステムに対して既知である場合の演奏表情生成の性能を検証する．この実験では，多重音を多数含んでいる W. A. Mozart, Piano Sonata, KV331, 1st Mov. の 4 人の演奏を用いて学習を行った (表 3)．テスト曲は同じ曲を用いた．

実験 2–テスト曲が未知である場合

実験 2 ではテスト曲がシステムに対して未知である場合の演奏表情生成の性能を検証する．この実験では Vladimir Ashkenazy が演奏した F. Chopin の 14 曲を用いて学習を行った (表 4)．テスト曲は学習データには含まれていない F. Chopin, Nocturne No. 10, Op. 32, 2nd Mov. を用いた．

4.2 生 成 結 果

システムに対してテスト曲が既知である場合，4 人の演奏から共通の演奏表情が学習および生成された．人間の演奏を見ると外部声部ごとに音量や演奏音長の変動が異なることが分かるが，同様に提案手法を用いて生成した演奏表情も**図 3** のように外部声部の音量と演奏音長の変動が異なることが分かった．

⋆2 http://www.crestmuse.jp/pedb

Vol.2010-MUS-85 No.2
2010/5/27

表 4 実験 2 で用いた学習データ．14 曲に対する 14 演奏を用いた．

曲名	演奏者
Prelude Op. 28 No. 1, 4, 7, 15, 20	V. Ashkenazy
Etude Op.10-3, 10-4, 25-11	V. Ashkenazy
Waltz Op. 18, 34-2, 64-2, 69-1, 69-2	V. Ashkenazy
Nocturne No. 2 Op. 9-2	V. Ashkenazy

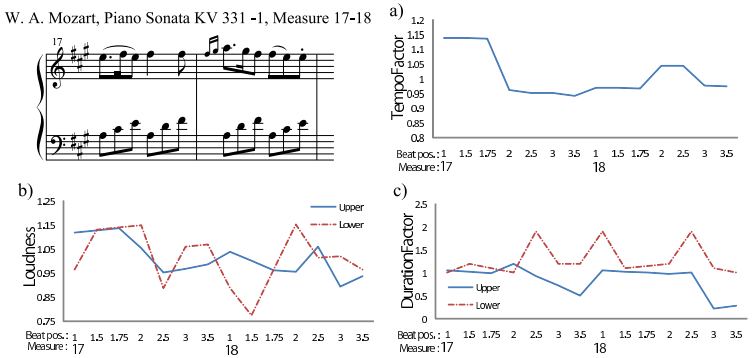

図 3 実験 1 の結果（旋律）．a) は式 (1) を用いて求めた瞬時テンポの変動，b) は式 (2) を用いて求めた音量の変動，c) は式 (3) を用いて求めた演奏音長の変動を表す．

人間の和音の演奏表情を見ると和音を構成する音符ごとに発音時間のずれ，音量，演奏音長が異なることが分かるが，同様に提案手法を用いても**図 4** のように異なる発音時間のずれ，音量，演奏音長が和音構成音ごとに生成された．この結果から，曲が既知である場合は，提案手法により多重音を含むピアノ曲の演奏表情が自動的に生成されることが確認された．

システムに対してテスト曲が未知である場合も，提案手法を用いて**図 5** のような瞬時テンポの変動や上と下の外部声部の異なる音量や演奏音長の変動が生成され，和音の場合も発音時間のずれ，音量，演奏音長が和音構成音ごとに異なって生成された．このように，曲が未知である場合も提案手法により多重音を含むピアノ曲の演奏表情が自動的に生成されることを確認した．

4.3 主観評価

提案手法を用いた生成結果が実際にどのように聴こえるか調べるために実験を行った．心理実験では実験 1 と実験 2 で用いた W. A. Mozart Piano Sonata KV 331-1(以下，SNT331-1) と F. Chopin Nocturne No. 10 Op. 32-2(以下，NCT010) と，新しく W. A. Mozart,

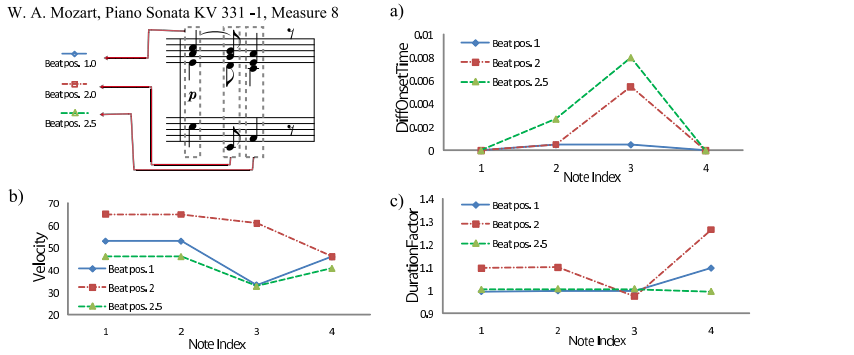

図 4 実験 1 の結果（和音）．a) は式 (4) を用いて求めた各音符の発音時間のずれ，b) は式 (5) を用いて求めた各音符の音量，c) は式 (6) を用いて求めた各音符の演奏音長を表す．Note Index は和音の構成音を表す（1 が一番高い音符）．

Piano Sonata KV545, 3rd Mov.（以下，SNT545-3）の 3 曲を用意した．各曲ごとに表情なし (deadpan)，人間の演奏表情，比較演奏表情，そして提案手法による生成結果の四つのサンプルを被実験者 25 人[⋆1]に聴かせた．SNT331-1 と NCT010 の提案手法による生成結果はそれぞれ実験 1 と実験 2 の生成結果を流用し，SNT545-3 は W. A. Mozart, Piano Sonata の 6 曲[⋆2]の学習によって得られた生成結果を用いた．なお比較演奏表情は，提案手法の有効性を確認するため，単旋律モデルを用いて上の外部声部のみを生成し，その演奏表情を他の音符にコピーした演奏表情である．評価では演奏表情の人間らしさと音楽的自然さを 6 段階で評価させた．

主観評価の結果を**図 6** に示す．SNT331-1 と NCT010 では提案手法による生成結果が人間の演奏表情に近いという評価が得られた．SNT545-3 では提案手法による生成結果が比較的低く評価されたが，その理由として，モーツァルトのテンポの速いピアノソナタでは局所的な演奏表情より，発想記号や構造の解釈による大局的な演奏表情がより重要であるためと考えられる．しかし，3 曲の平均を見ると，提案手法を用いて生成された演奏表情が人間の演奏表情に最も近い評価を得ており，比較演奏より良い評価が得られることが分かった．この結果から，提案手法を用いて生成された演奏表情は人間らしく，音楽的に自然に聴こえる

⋆1 実験には音楽に関する非専門家が 6 人，準専門家が 17 人，専門家が 2 人参加した．
⋆2 M.J. Pires が演奏した W. A. Mozart Piano Sonata, KV279-1, 279-2, 279-3, 331-1, 545-1, 545-2 の 6 曲．KV545-3 は学習データには含まれていない．

Vol.2010-MUS-85 No.2
2010/5/27

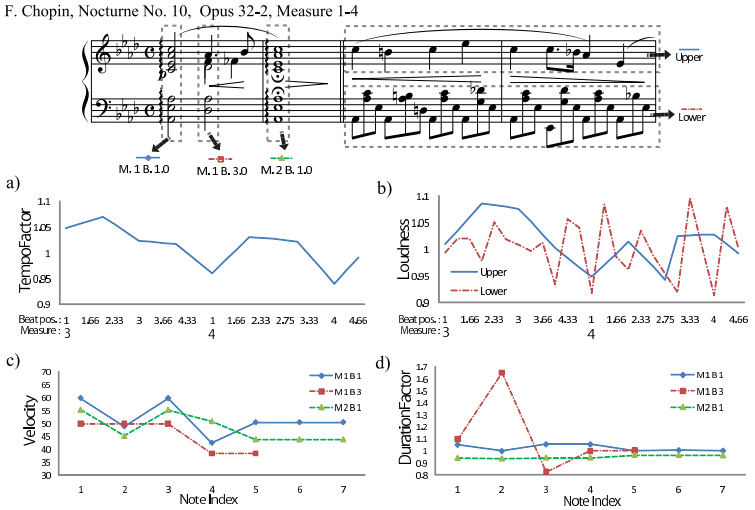

図 5 実験 2 の結果．a) は式 (1) を用いて求めた瞬時テンポの変動，b) は式 (2) を用いて求めた旋律に対する音量の変動，c) は式 (5) を用いて求めた和音に対する各音符の音量，d) は式 (6) を用いて求めた和音に対する各音符の演奏音長を表す．旋律に対する演奏音長の変動と和音に対する各音符の発音時間にずれは紙面の関係上省略した．

ことを確認した．また提案手法は多重音を含むピアノ曲演奏の自動表情付けに有効であることを確認した．

5. おわりに

本稿では，多重音を含むピアノ曲演奏の人間らしい表情を自動的に生成するために，単旋律と和音の確率モデルの組み合わせによる手法を提案し，評価実験によってその有効性を確認した．特に，システムに対して未知である曲であっても自動演奏表情付けが可能である事は，将来ピアノ曲を対象とする作曲支援システムの構築や，音楽教育への支援が可能であることを示している．

今回は前後に文脈に基づいた局所的な演奏表情の生成に関して論じた．今後，より人間らしい演奏表情の自動生成のために，発想記号や構造解釈による大局的な演奏表情の生成方法を考察する予定である．また，演奏者の個性を保った演奏表情の生成に関しても考えたい．

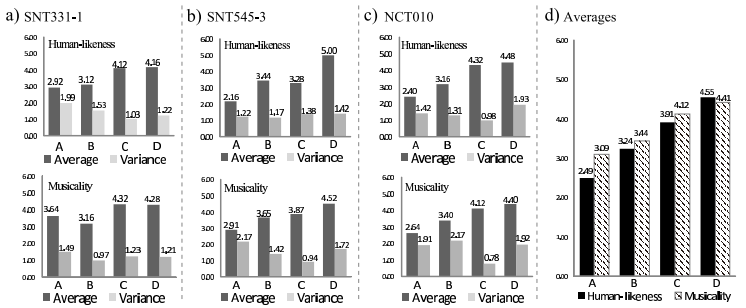

図 6 主観評価の結果．A は表情なし，B は比較演奏表情，C は提案手法を用いて生成された演奏表情，D は人間の演奏表情である．ここで人間の演奏表情には大局的な演奏表情と局所的な演奏表情の両方が含まれている．a) は W. A. Mozart, Piano Sonata, KV331, 1st Mov., b) は W. A. Mozart, Piano Sonata KV545, 3rd Mov., c) は F. Chopin, Nocturne No. 10, Op. 32, 2nd Mov. の結果である．d) は人間らしさと音楽的自然さに対する三つの曲の平均である．

参 考 文 献

1) Sundberg, J. and et al.: Musical performance: A synthesis-by-rule approach, *Computer Music Journal*, Vol.7, No.1, pp.37–43 (1983).
2) Flossmann, S. and et al.: Expressive Performance Rendering: Introducing Performance Context, *Proceedings of the 6th Sound Music and Computing Conference(SMC)*, pp.155–160 (2009).
3) Teramura, K. and et al.: Gaussian process regression for rendering music performance, *Proceedings of the 10th International Conference on Music Perception and Cognition(ICMPC)* (2008).
4) 寺村桂子，前田新一：統計的学習によるテンポの変動を考慮したピアノ演奏模写，情報処理学会音楽情報処理研究会， Vol.84, No.12 (2010).
5) Grindlay, G. and et al.: Modeling, analyzing, and synthesizing expressive piano performance with graphical models, *Machine Learning*, Vol.65, pp.361–387 (2006).
6) Parncutt, R.: Accents and expression in piano performance, *Perspektiven und Methoden einer Systemischen Musikwissenschaft* (Niemöller, K. W., ed.), Peter Lang, Frankfurt am Main, pp.163–185 (2003).
7) Lafferty, J. and et al.: Conditional random fields: probabilistic models for segmenting and labeling sequence data, *International Conference on Machine Learning*, pp. 282–289 (2001).
8) Bottou, L.: Stochastic Gradient Learning in Neural Networks, *Proceedings of Neuro-Nîmes 91*, Nimes, France, EC2 (1991).