音楽解釈研究のための演奏 deviation データベースの作成

豊田健一[†]　片寄晴弘[†,††]　野池賢二[††]

演奏分析・音楽解釈を根本的に発展させるものとして，情緒あふれる演奏がどのような逸脱 (deviation）を持っているかを記述した演奏データベースに対する期待は大きい．本論文では，ガイドとなる音列情報を与え，小節情報と演奏表現傾向を利用し，DP と HMM を用いて，効率的に，MIDI 入力された演奏データから deviation データを作成する手法について述べる．

Utility System for Constructing Database of Performance Deviation

Kenichi TOYODA, Haruhiro KATAYOSE and Kenzi NOIKE

Database which contains deviations from the normalized notes are indispensable for the study of performance analysis and rendering. This paper describes a procedure to produce a deviation data efficiently, based on Dynamic Programming and HMM, utilizing measure information and expression tendency model.

1. はじめに

近年の音楽情報処理研究の主要研究領域に，Performance Rendering（情緒あふれる演奏の自動生成）がある[1)2)3)]．現在の Performance Rendering の研究の興味の中心は，学習型や自律型のシステムに移りつつあるが，学習システムに与えるデータにしても，あるいは，人間が分析を行っていくにしても，情緒を含んだ演奏の deviation 記述（正規化された演奏をリファレンスとして演奏特徴を記述）したデータの用意は，研究を進めていく上での前提条件となる．

いうまでもなく，情緒あふれる演奏のテンポは揺らぎが大きく，単純な閾値処理で拍の正規化・量子化を行うことはできない．この問題は，音楽に関連するパターン認識の主要テーマであり[4)5)]，最近でも活発に研究が行われている[6)7)]．

上記の拍の量子化に関する研究は，さまざまな工夫により認識率の向上に大きく貢献したが，揺らぎの大きな演奏データに対して，100 %の認識率を得ることは容易なことではない．我々は，人間の介在を前提に，エラーフリーのデータを作成することを目標にかかげ，その上で，できる限り作業効率を上げる支援システムの作り上げることを主題としている．

本稿では，deviation 記述に関する基本的な概念を示した上で，SMF 形式の MIDI 情報から，ガイドを利用して，効率的に deviation データを得る手法について述べていく．

2. 演奏表現とデータ記述

2.1 演奏表現の記述

自然楽器の演奏制御対象とそのレベルにはさまざまなものがあり，信号レベルでそのすべてを制御することはきわめて困難である．その中で，ピアノを代表とする打鍵楽器における制御対象は，各音の発音時刻，消音時刻，音の強さ (MIDI での velocity 値）に，ほぼ簡約される☆．一方，音楽的な演奏表現については，スラーやテヌート等の表意記号に関する表現，演奏者によって意識下あるいは無意識下で理解される個々の音符レベル，拍節レベル，フレーズレベルの表現が畳積され，その結果が，それぞれの制御対象に投影されている．ここでは，運動制御の時定数の視点から，拍打（およびそれ以上）レベルでの表現とそれ以下のレベルでの表現（拍内表情）に分離して（図 1），データ記述方式（note 形式）を設計することにした

データの記述例を図 2 に示す．この図において，基本的な演奏データ（機械的演奏に相当する）は太字で示されるものであり，「各音符の発音時刻，音高（ノートネーム），持続時間」の組として記述される．それ以

† 関西学院大学, Kwansei Gakuin Univ.
†† さきがけ研究 21, PRESTO, JST.

☆ 実際には velocity による音量制御には限界がある．また，ペダル操作についても考慮する必要がある．厳密にいうと，近似的な簡約である

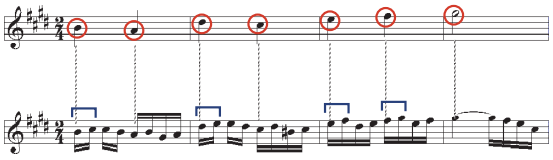

図 1　拍打による演奏制御

```
.....
.....
2.00 BPM 126.2 4
2.00 (0.00 E3 78 3.00 -0.11)
=2
1.00 TACTUS 2 4
1.00 BPM 128.1 4
1.00 (0.00 C#4 76 0.75 -0.09) (0.04 E1 60 1.00 -0.13)
1.75 (0.10 D4 77 0.25 -0.14)
2.00 BPM 130.0 4
2.00 (0.00 B3 75 1.00 -0.03) (0.00 G#3 56 1.00 0.03)
3.00 BPM 127.7 4
3.00 (0.00 B3 72 1.00 0.00) (0.09 G#3 56 1.00 -0.12) (0.14 D3 57 1.00 -0.21)
=3
1.00 TACTUS 1 4
1.00 BPM 127.6 4
1.00 (0.00 B3 77 2.00 -0.05) (0.00 G#3 47 2.00 -0.05) (-0.06 D4 57 2.00 -0.32)
3.00 BPM 129.7 4
3.00 (0.00 F#4 75 1.00 -0.15) (0.00 D4 54 1.00 0.03)
=4
1.00 BPM 127.7 4
1.00 (0.00 D#4 73 0.75 -0.38) (0.02 C4 65 0.75 -0.08)
.....
.....
```

図 2　演奏表情の記述

外が，演奏表情に関わる deviation 項である．テンポに関わる情報は,「時刻，BPM，テンポ値，単位となる音価」の組として記述される．各音の拍内表情に関するデータとしては，括弧中に，発音時刻の deviation，当該音符の velocity 値，持続時間の deviation として記述される．

2.2　テンポとタクトス

通常，テンポとは楽曲を演奏する際の速度を意味し，単位としては，1 分間あたりの単位音符の演奏する数 (bpm) が用いられる．単位音符には，四分音符が割り当てられることが多いが，リズムの表現にあわせ，他の音符が用いられることもある．

テンポが一定の時でも，例えば，ある音楽的な区間においては，半拍レベルで拍打ちをしたり，あるいは，小節レベルで拍打ちをしたいといった要請がある．ここでは，規準拍（例えば，四分音符）に対して，何回の拍打をするかを明示的に記述するタクトス☆という概念を導入することにした．データ中では，TACTUS という記述子を用い，例えば，

1.00　TACTUS　2　4

のように記述する．この例では，時刻 1.00 以降，四分音符に対し，2 回の拍打を対応させる，すなわち，八分音符レベルでの拍打を対応させるということを表している．

☆ 楽典では，腕の動きによって計量される時間の単位として定義される．その定義からすれば，ここでの定義は，逆数的なものである．

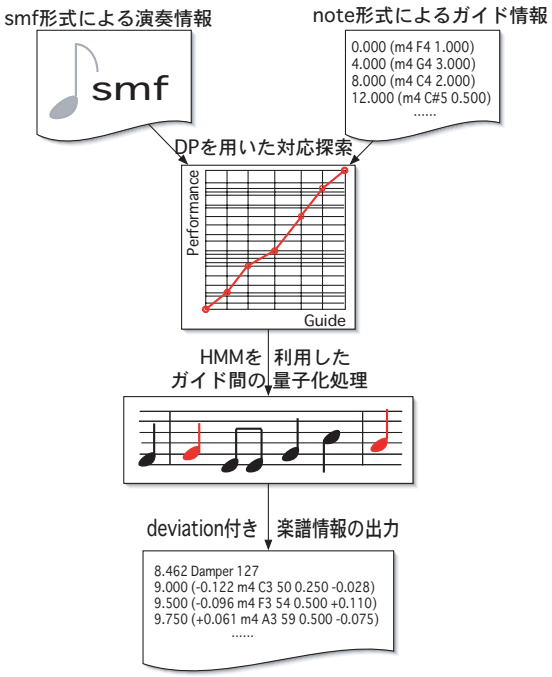

図 3　処理の流れ

3.　deviation データ抽出の概要

情緒あふれるデータを，上記に示すような deviation 項を持つデータ記述表現におきかえるためには，拍の量子化を行う必要がある．この作業を手作業で実施するのは非常に煩雑である．この作業を支援するものとして，正規化された楽譜情報をガイドとして利用する DP マッチングを用いた自動化の研究等が報告されてきた．例えば，高見らの研究[9] においては，楽譜となる情報をガイドとして与え，音響信号から，2 段の DP マッチングにより演奏表情を抽出する方法が報告されている．

DP マッチングを用いる方法は，エラーフリーのデータを得るものとして有効であるが，楽譜情報を用意すること自体が非常に煩雑である．我々は，ガイドとなる必要最小限の楽譜情報を与え，量子化処理と組み合わせることにより，効率的に deviation 項を持つデータ列を作成する手法について検討を行うことにした．処理の流れを図 3 に示す．この図に示すように，ユーザは，例えば，メロディあるいはメロディの一部を ガイドとして与える．このデータは，deviation 項を含まない note 形式である，text エディタで容易に作成できるほか，一般のシーケンサや記譜ソフト上で作成したデータを用意して，変換ツールを使って得ることもできる．このデータと表情付きの演奏データ (SMF)

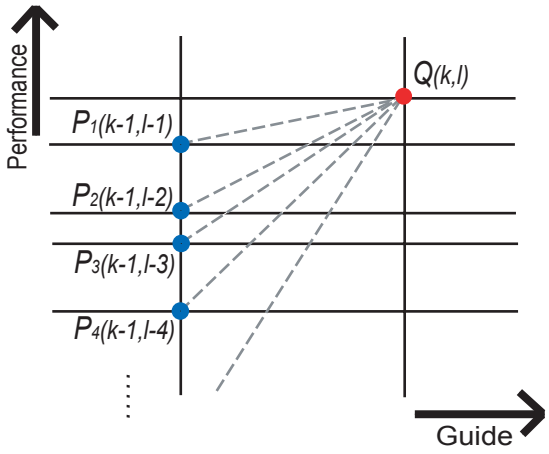

図 4 DP マッチング

とのマッチングを行い，ガイドにし たがって，表情付きの演奏データを deviation 項を含んだ note 形式として配置していくのが基本的なアイデアである．

処理の第一段階としては，DP を用いた対応探索により，ガイドに相当する音符の演奏上での時刻を割り当てる．次に，ガイド以外の音符に関して，それぞれのガイド間で HMM を利用した量子化処理により，正規化された発音時刻を与える．この際，小節（強拍）に関する制約（音量情報の利用），フレーズの演奏傾向，隣接する音符の長さが有理数比になりやすいというようなヒューリスティクスを利用する．処理の最終段階として，指定した音価毎にテンポを計算し，算出したテンポに基づいて，発音時刻の deviation，持続時間の deviation を計算し，さらに，velocity を割り当てて，演奏表情を含んだ note 形式のデータを得る．

以下の章では，DP を用いた対応探索，HMM を利用した量子化処理について具体的に説明する．

4. DP を用いた対応探索

図 3 の中央にある図は，DP を用いた演奏情報とガイド情報の対応探索を表している．縦軸には演奏情報の系列が実時間に沿って並び，横軸にはガイド情報が持つ拍単位の時間が実時間に換算されて並んでいる．

4.1 DP マッチング

図 4 は探索平面の一部を表したものであり，各格子点は演奏情報とガイド情報の対応を表す．ある格子点 $Q(k,l)$ に対して，親候補の点を $P(i,j)$ とし，演奏情報とガイド情報の類似度 $S(k,l)$ を次の式で求める．

$$S(k.l) = max(sim(i,j,k,l) + S(i,j)) \qquad (1)$$

各格子点について $S(k,l)$ を最大にするような親 $P(i,j)$ を求めてゆき，最終的に曲の末尾を表す点（DP 平面の右上端の点）から順に親をたどることで一つの経路ができ，演奏情報とガイド情報の最も尤もらしい対応が得られる．ここで，式 (1) の $sim(i,j,k,l)$ は，2 点 $P(i,j)$，$Q(k,l)$ により値が決まるコスト関数である．

4.2 コスト関数の値を決める要素

$sim(i,j,k,l)$ は，次の 4 つのパラメータからなる．

- s_1:音高の一致度
- s_2:ガイド間のテンポと曲全体の平均テンポとの類似度
- s_3:ガイドの仮発音時刻と演奏の実発音時刻との類似度
- s_4:強拍である尤度

具体的には，まず s_1 については，演奏音とガイドを比較して音高が一致すれば単純に $s_1 = 1$ とし，s_2 は，$P(i,j)$ と $Q(k,l)$ を結ぶ直線と探索平面の対角線とのなす角が小さいほど値が大きくなるようにした．また s_3 は，演奏の発音時刻 t_P とガイドの仮発音時刻 t_G の差 $|t_P - t_G|$ が小さいほど大きな値をとるようにし，s_4 は，ガイドの比較対象となる演奏音の velocity が，その前後数音の velocity の平均値よりも大きいほど値が大きくなるようにした．ここで，ガイドが強拍に相当するかどうかは，ガイドに記述された拍単位の時刻を元に割り出される．

これら $s_n(i = 1,2,3,4)$ はそれぞれ最大値が 1 となるよう正規化した．そして，各 s_n の重み係数を w_n とし，

$$sim(i,j,k,l) = \sum_{n=1}^{4} w_n s_n$$

という式でコスト関数の値を定めた．ここで，$w_1 > w_2 > w_3 > w_4$ としてある．

ガイドとして与えるデータに関わる制約として，

- 強拍に相当する音符は必要最小限入れること
- 連続する同音の音符がある場合は省略せず記述する，あるいは，音価の異なる音符をガイドとして用いる

ことにした．これらの制約により，DP マッチングの段階では，ほぼエラーフリーとなっている．

5. HMM を利用した量子化処理

前章の処理により実演奏とガイドとの対応を取った後☆，隠れマルコフモデル（HMM）を用いて各ガイド間の音価列推定を行う．ここでは大槻らの示した演奏情報の発音時刻の間隔（IOI:Inter Onset Interval）から HMM を用いて意図された音価列を推定する手法[6]に基づき，その上で音長比に関するヒューリスティク

☆ 発音時刻が近接している演奏音は，50ms ごとにまとめて 1 つの和音として扱った．

スなども HMM に組み込む方法を検討した．

なお以下では，楽譜で表されるような整数関係にある音符の長さを「音価」と呼び，演奏における揺らぎのある音の長さを「音長」と呼ぶ．

5.1 HMM による音価列推定

音価列 $\Theta = (\theta_0, \theta_1, \ldots, \theta_T)$ を意図して演奏した結果，音長列 $\Phi = (\phi_0, \phi_1, \ldots, \phi_T)$ が観測される確率は，$P(\Phi|\Theta)$ で表される．逆に，観測された音長列 Φ が音価列 Θ を意図した結果である確率 $P(\Theta|\Phi)$ は，Bayes の定理より

$$P(\Theta|\Phi) = \frac{P(\Phi|\Theta)P(\Theta)}{P(\Phi)} \tag{2}$$

となる．ここで $P(\Phi)$ は Θ に依らない．よって式 (2) の分子が最大となるような Θ を求めることで，最も尤度の高い音価列 Θ^* を推定できる．すなわち，

$$\Theta^* = \arg\max_{\Theta} P(\Phi|\Theta)P(\Theta) \tag{3}$$

である．

本稿では，音価系列のモデルとして，短いひとまとまりの音価系列を単語として扱う「リズム単語モデル」[6] を用いている．各ガイド間に対して，リズム単語の集合 M_R の中から最もふさわしいリズム単語 R^* を選び出す，という処理を曲全体に渡って繰り返している．ガイド間すなわちリズム単語内では隣接する音価の遷移確率は 1 であるから，リズム単語 $R = (\theta_0, \theta_1, \ldots, \theta_T)$ の出現確率を $P_R(= P(\Theta))$ とし，音価 θ が音長 ϕ で演奏される確率密度を $L^{IOI}(\theta, \phi)$ とすると，式 (3) より，

$$R^* = \arg\max_{R \in M_R} P_R \prod_{t=0}^{T} L^{IOI}(\theta_t, \phi_t) \tag{4}$$

となる．☆

5.2 音価列推定に用いるヒューリスティクス

本稿ではさらに，次のようなヒューリスティクスを用いて式 (4) を拡張している．

(1) 局所的なテンポ変化の可能性
(2) 隣接する音長が有理数比になりやすいという傾向
(3) 拍打と velocity との関係

(1) については図 5 に示すように，変動テンポの場合は一定テンポの場合と異なり音長の生成確率密度 $L^{IOI}(\theta, \phi)$ の分布は時々刻々と変化する．本稿では，「星に願いを」の演奏データについて，特にフレージングに関わるテンポ変化の推移を 2 次関数で近似したモデル 2 つと，一定テンポの場合を合わせた計 3 個のフレーズモデル（この集合を M_p とする）を用意し，各 $P \in M_P$ を $L^{IOI}(\theta, \phi)$ に作用させて $L_P^{IOI}(\theta, \phi)$ を生成した．

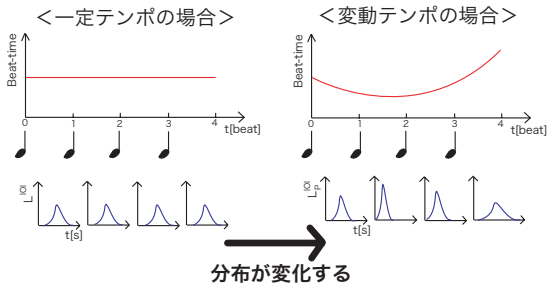

図 5 テンポ変化の有無

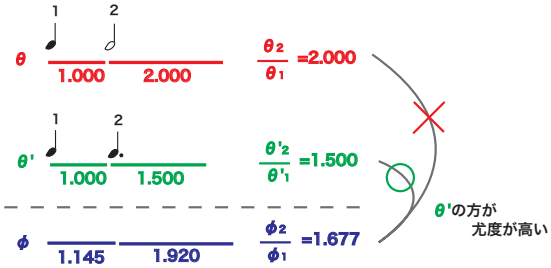

図 6 隣接 IOI の比

図 6 は，(2) について説明したものである．t 番目と $t+1$ 番目の音に対して，(2) を具体的に数値化したもの L^{ratio} は，隣接する音価の比と音長の比との差を用いて，

$$L^{ratio}(\theta_t, \theta_{t+1}, \phi_t, \phi_{t+1}) = f\left(\frac{\theta_{t+1}}{\theta_t} - \frac{\phi_{t+1}}{\phi_t}\right)$$

とすることにした．ここで f は分布関数であり，今回は分散 σ を適当に指定した正規分布関数を用いた．

(3) について，例えば 4/4 拍子の一般的なポップスでは奇数拍の velocity が大きくなりやすい，などの傾向がある．今回は 2 種類のベロシティモデル（この集合を M_v とする）を用意した．演奏音の velocity (v_t) と各モデル $V \in M_V$ の velocity とを比較して，両者が近いほど，velocity から見た仮定音価列の尤度 $L_V^{vel}(v_t)$ が高い，としている．ただし，音価列推定に用いる情報としての velocity 情報は，前述の (1)(2) に比べて優先順位が低いと考え，L_V^{vel} の値のばらつきは小さくした．

以上のパラメータを用いて

$$L_t = L_P^{IOI}(\theta_t, \phi_t) \cdot L^{ratio}(\theta_t, \theta_{t+1}, \phi_t, \phi_{t+1}) \cdot L_V^{vel}(v_t)$$

☆ 本稿では音長の分布 $L^{IOI}(\theta, \phi)$ については正規分布 $N(\mu, \sigma^2)$ で近似し，$\mu = \theta$，$\sigma = 0.05\mu + 0.011$(秒単位)[6] を用いた．

表 1 対応探索実験（ガイド数の括弧内の値は,（ガイド数）÷（音数）を表す．正解率の単位は [%]）

実験対象曲		最多ガイド			2 拍ごとのガイド			1 小節ごとのガイド		
曲目	音数	ガイド数	誤り数	正解率	ガイド数	誤り数	正解率	ガイド数	誤り数	正解率
星に願いを	480	107(0.223)	0	100	52(0.108)	0	100	29(0.060)	0	100
美女と野獣	544	118(0.217)	0	100	59(0.108)	0	100	30(0.055)	0	100
いつか夢で	274	92(0.336)	0	100	–	–	—	34(0.124)	0	100
ノクターン	193	57(0.295)	0	100	17(0.088)	0	100	9(0.046)	6	33

$$R^* = \arg \max_{R \in M_R, P \in M_P, V \in M_V} P_R P_P P_V \prod_{t=0}^{T} L_t \tag{5}$$

として，最も尤度の高いリズム単語 R^* を推定している．ここで P_P，P_V はそれぞれ，フレーズモデル P，ベロシティモデル V の出現確率である．

6. 実験と評価

本章では，ヤマハサイレントピアノ用データとして販売されている楽曲 3 曲（ディズニーシリーズ）と手打ち込みで作成したノクターン，計 4 曲を用いて，対応探索実験および音価列推定実験を行った．なお,「いつか夢で」が 3 ／ 4 拍子，それ以外は 4 ／ 4 拍子である．今回用いた曲の全長はいずれも 60 秒前後である．

6.1 対応探索実験

表 1 は実験対象となる 4 曲それぞれについてガイド数を変えて正解率を求めたものである．上から 3 曲については，ガイド数を約 1 小節につき 1 個に抑えても正解率 100 %となった．対象とした曲に限れば，演奏された音とその約 1 割のガイドとの対応は完全に取れたことになる.「ノクターン」を対象とした場合に誤りが認められたが，この理由は，ガイドに用いたのと同じ音高が演奏中に多数存在したためである．

今回色々な条件でガイド情報を与える際，曲中でメロディの音高にばらつきがある（同音高を持つ音符が近接していない）部分や，テンポ変化の少ない部分では，ガイド数をさらに減らしても対応が取れる場合がある，ということが分かった．

対応探索の段階でのエラーを極力無くすため，ガイド情報の与え方として，以下のような指針が挙げられる．

- 同一音高を持つ音符が近接している場合は，意図的に特異な音高を持つ音符をガイドとして与える．
- 単音のみならず和音をガイドとして用いる．
- 演奏音の疎密度やテンポ変化の程度に応じてガイド情報の量を変える．

表 2 音価列推定実験（各値は適切なリズム単語を選び出した割合を表す．単位は [%]）

曲目	L^{IOI} のみ	$L^{IOI} + \alpha$
星に願いを	64.29	100.00
美女と野獣	93.10	93.10
いつか夢で	96.97	100.00
ノクターン	100.00	100.00

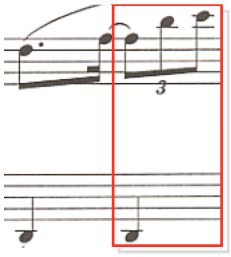

図 7 「美女と野獣」で誤認識があった箇所

6.2 音価列推定実験

表 2 は，課題曲に対して各小節の 1 拍目にあたるガイドを与え DP マッチングを行った後，ガイド間の音価列推定を行った結果である．ただし,「ノクターン」については 1 小節ごとのガイドではマッチングの正解率が 100 %とならなかったので，2 拍ごとのガイドを用いて実験を行っている． 実験では，5 章で述べた手法のうち，音長の生成確率 L^{IOI} のみ用いた方法と他の工夫を加えた方法を用い，いずれが有効であるか比較した．その結果，音価推定の正解率は 4 曲全てについて後者の方が前者と同等以上の値が得られた．特に「星に願いを」については大きな改善が見られた．この曲に関してさらに精緻に調べたところ，今回用いた工夫のうちどの 1 つを欠いても音価推定率が減少することがわかった．

「美女と野獣」では，図 7 の四角で囲った箇所について，3 つの音符を順に 8 分，4 分，8 分音符とする誤認識があった．各音の IOI は順に 193，302，161（単位は [ms]）であった．この誤りに対処するには，記譜上の慣例を取り入れていかなければならない．現状では，やむを得ない誤認識であると受けとめている．

今回の実験では明らかにならなかったことだが，フレーズモデルは音長生成確率密度の分布自体を操作

するので，一定テンポで演奏された部分を変動テンポであると解釈した場合，かえって誤った音価列を推定する可能性もある．これを避ける手段として，曲中にテンポ変化が認められる箇所については，ガイド中に「フレーズモデルを適用する」という情報を入れることが有効であると考える．

また音価，音長の比を考慮した値 L^{ratio} は隣接する 2 音についてのみ依存するものとしたが，数音の足し合わせを考慮することでさらに高い効果が得られる可能性がある．

7. おわりに

パフォーマンスレンダリング，音楽認知・知覚に関する研究領域において，情緒を含んだ演奏の deviation データベースの作成が強く求められている．その要請に応えるものとして，本論文では，演奏表情を含んだ SMF 形式の演奏データに対し，ガイドとなる少数の音列情報を与え，DP と HMM を用いて効率的に deviation データを得る手法について述べてきた．比較的演奏表情が豊かなデータを対象に対し，提案手法が有効であることを検証した．

今後はデータベースの作成に力を入れつつ，その過程を通じてモデルパラメータのチューニングを行っていきたい．今回は MIDI データのみを対象としたが，実信号への対応できるようにシステムの拡張を行う予定である．また，タクトスを自動付与する機構についての検討も進める予定である．

参考文献

1) 平賀, 平田, 片寄: 蓮根，目指せ世界一のピアニスト，情報処理, Vol.43, No.2. pp. 136–141 (2002).
2) http://shouchan.ei.tuat.ac.jp/~Rencon/
3) 片寄, 平賀, 平田, 野池, 橋田: ICAD-Rencon ー報告と課題ー, 情報処理学会音楽情報処理科学研究報告, No.47-14, pp. 79–83 (2002).
4) P. Desain and H. Honing: The Quantization of Musical Time: A Connectionnist Approach", MIT press, Computer Music Journal, Vol. 13, No. 3 pp. 56–66 (1989).
5) 片寄, 井口: 知的採譜システム，人工知能学会誌, Vol.5, No.1, pp. 59–66 (1990).
6) 大槻, 中井, 下平, 嵯峨山: 隠れマルコフモデルによる音楽リズムの認識，情報処理学会論文誌, Vol.43, No.2, pp. 245–255, (2002).
7) 浜中, 後藤, 麻生, 大津: 発音時刻の楽譜情報の位置を確率モデルにより推定するクォンタイズ手法, 情報処理学会論文誌, Vol.43, No.2, pp. 234–244, (2002).
8) Werner Goebl, Roberto Bresin: Are computer-controlled pianos a reliable tool in music performance research? Recording and reproduction precision of a Yamaha Disklavier grand piano, MOSART workshop, Barcelona, Nov. pp.15–17 (2001)
9) 高見啓史，片寄晴弘，井口征士：ピアノ演奏における演奏情報の抽出，電子情報通信学会論文誌,Vol.J72-D2 No.6, pp. 917–926 (1989).