

打楽器音を考慮した音響信号と楽譜のアライメント

野口 綾子^{1,a)} 酒向 慎司^{1,b)} 北村 正^{1,c)}

概要: 本稿では、人間に合わせて計算機に伴奏を出力させることで合奏や演奏に関する支援を行うことを目標として、音響信号と楽譜のオンラインアライメントの精度の向上のためのモデルを提案する。本研究ではポピュラー音楽に焦点をあてており、その全体の演奏を取りまとめる重要な役割として打楽器がある。打楽器音は他の楽器音とは全く異なる響きを持ち、また打楽器音を含む和音のオンセットは他の楽器音のみからなる和音のオンセットよりもアライメントにおける重要度が高くなると予想されるため、他の楽器音とは区別した方が良いのではないかと考えられる。しかし従来法では、打楽器音を他の楽器音と同じ枠組みで扱っており、楽器種の違いを区別することができなかった。そこで本研究では、打楽器音の有無によってオンセットの重みを変えるよう従来モデルのアライメントアルゴリズムの拡張を行った。提案モデルによって楽器種を区別し、最適な重みづけを行った実験の結果、全ての許容誤差においてアライメント精度の向上と提案モデルの有効性を確認した。

SCRF Audio-to-Score Alignment Distinguishing Percussion Sounds from Other Instruments

Abstract: This paper aims to introduce a model which uses column alignment in scores in order to recognize onset positions in popular music. The objective is to play an accompaniment in real time, which is synchronized with the live performance. Because percussions are an important part of popular music, and as the previous model did not take it into consideration, we will focus on it in this research. Percussion sounds are radically different from the other ones composing popular performances, and thus they should be easily distinguished from the rest of the formation. Moreover, the main feature of this model is to distinguish the sounds of the instruments, as it greatly affects the onset. The effectiveness of the recognition will be increased toward an optimum by conducting experiments.

1. はじめに

音楽音響信号と楽譜のアライメント（スコアアライメント）とは、人の演奏の音響信号を入力として、楽譜上の演奏位置を認識する技術のことである [1, 2]。この技術は音楽演奏に関わる多くのアプリケーションに利用できる。実用例の1つとして、人間によって演奏された音楽に伴奏を同期させる自動伴奏というシステム [3] があり、現在でも盛んに研究されている [4]。スコアアライメントを用いたアプリケーションの実現における本質的な課題は、入力された信号に対して瞬時に演奏位置を求めることであり、より効果的に演奏位置とテンポを推定するためにアライメン

ト技術の向上が望まれる。

スコアアライメントに関する先行研究のひとつとして、MIDI 信号を対象とした単旋律および多声演奏に対して高精度で認識できる手法が提案されている [5]。しかし MIDI 信号を入力としたシステムは、使用楽器が限定されてしまうため自動伴奏などの応用を考えた場合に汎用的ではなく、MIDI 楽器に限らず歌声や管楽器など楽器種を問わない手法が望ましいと考えられる。近年では多くの楽器に対応できる音響信号を対象としたアライメント手法がいくつか提案されており、本研究でも同様の問題に取り組む。

音響信号に対応したスコアアライメントの先行研究として、動的時間伸縮法（Dynamic Time Warping; DTW または DP マッチング）[6] や隠れマルコフモデル（Hidden Markov Model; HMM）[7] を用いた手法が挙げられるが、これらはフレームレベルのマルコフ性、観測信号の独立性といった原理的な性質により、逐次的に信号が入力される実時

¹ 名古屋工業大学
Nagoya Institute of Technology

a) noguchi@mmsp.nitech.ac.jp

b) sako@mmsp.nitech.ac.jp

c) kitamura@mmsp.nitech.ac.jp

間アライメントにおいては必ずしも適さない．HMM を包含する条件付き確率場 (Conditional Random Field; CRF) をセグメントモデルへと拡張したセグメンタル条件付き確率場 (Segmental Conditional Random Field; SCRF) を用いたアライメント手法 [8] では、フレームレベルだけでなく 1 和音を 1 セグメントとするセグメントレベルの特徴量を導入することができる．それゆえ、音の発音に関する素性やテンポや音の継続長の変化など人の演奏の性質を表現することができ、高いアライメント精度が得られている．山本らは SCRF の枠組みをベースとして、演奏位置とテンポの推定を交互に繰り返す最適化アルゴリズムを提案しており、低計算量なアライメントを実現している [9]．文献 [9] では、300 ms の許容誤差において 75.2 % という精度を得ているが、これは実時間アライメントにおいて十分とは言えず、さらなる改善が望まれる．

従来法では、和音の定義を同時発音の音符全体としていたが、本来楽譜はパート単位で区別して書かれており、人間も選択的に音を拾っていると考えられる．そのため本研究では従来法とは異なる手法を検討する．ただし、問題とする音楽対象を広く扱うことは現実的ではないため、本研究では対象をポピュラー音楽に限定する．多くのポピュラー楽曲では打楽器が用いられており、打楽器音はその他の音高を持つ楽器の音とは全く異なる響きを持つ．それゆえ打楽器以外の奏者は多声演奏においても打楽器音を聞き分け、合わせて演奏することができる．スコアアライメントにおいても打楽器音は同様に重要な役割を持つと考えられるため、本研究ではポピュラー音楽における打楽器音の考慮を検討する．打楽器の演奏に重要なのはオンセットであり、打楽器音の含まれる和音は打楽器音を含まない和音よりもオンセットの重要度が高くなると考えられる．そこで本研究では楽譜情報を有効に活用することを検討する．本研究では楽譜情報を用いて和音を構成する楽器の種類によってオンセットの重要度を変更することで、アライメント精度の改善を図る．

2. セグメンタル条件付き確率場を用いたスコアアライメント

2.1 従来法における問題定義

従来研究では、音響信号と楽譜のアライメントは音響信号の各時刻に対して、楽譜上の演奏位置を推定する問題であると定義している．従来法では、SCRF に基づく演奏位置推定と線形動的システム (Linear Dynamical System; LDS) に基づくテンポ推定を組み合わせたモデルが提案されている．本手法では、音響信号と楽譜を実時間においてより高精度に認識するために、SCRF の枠組みを打楽器音の有無を考慮するように拡張する．図 1 に SCRF に基づく音響信号と楽譜のアライメントの基本構造を示す．従来法では図 1 のように、入力とする音響信号から特徴量を抽

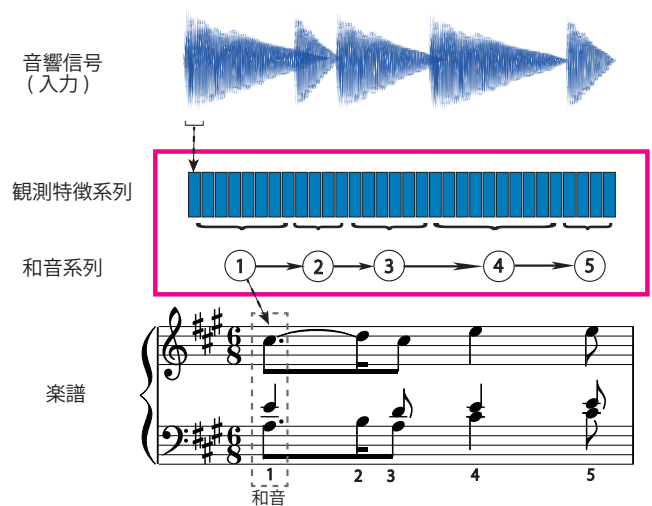


図 1 SCRF に基づく音響信号と楽譜のアライメントの基本構造

出し、フレームレベルの系列である観測特徴系列にセグメンテーションを行うことで和音系列に置換する．その後、SCRF によって演奏位置を推定している．

従来研究において定式化されている SCRF の確率モデルについて述べる．ここで、入力信号から得られる特徴量の系列を $\mathbf{o} = \{\mathbf{o}_t\}_t^T$ とし、 t はフレーム番号、 T は総フレーム数とする．一和音に相当するフレーム群を一つのセグメントとして、セグメント番号を n 、総セグメント数を N とするとき、セグメントの系列を $\mathbf{q} = \{q_n\}_n^N$ 、 $q_n = (t_n^s, t_n^e, s_n)$ とする． t_n^s は n 番目のセグメントの先頭、 t_n^e は最後尾、 s_n はセグメントに対応する拍位置を示す．よって、最適なセグメント位置を決定する問題は以下のように表される．

$$\hat{\mathbf{q}} = \arg \max_{\mathbf{q} \in \mathcal{Q}} p(\mathbf{q}|\mathbf{o}), \quad (1)$$

ここで \mathcal{Q} は考えうるあらゆるセグメンテーションの集合とする．セグメント表現に基づけば SCRF による条件付き確率は、次のように表される．

$$p(\mathbf{q}|\mathbf{o}) = \frac{1}{Z(\mathbf{o})} \Psi(q_1) \prod_{n=2}^N \Psi(q_{n-1}, q_n) \prod_{n=1}^N \Phi(q_n, \mathbf{o}), \quad (2)$$

$Z(\mathbf{o})$ は正規化行列、 $\Psi(q_n, q_{n-1})$ は遷移関数を示す． $\Phi(q_n, \mathbf{o})$ は観測関数である．式 (1) は Viterbi アルゴリズムによって解くことができるが、入力系列全体が必要となるためオンライン処理には適さない．そこで従来研究では、SCRF における演奏位置推定処理で一定の遅延を許容し、その過程で得られたテンポから LDS によって現在の演奏位置を予測することによってオンライン処理による演奏位置推定手法が提案されている．

2.2 演奏の遷移モデル

従来法では、楽譜に基づく演奏は、演奏位置が時間経過にもなると変化していく過程と捉えられている．ここで

は、従来法における演奏の遷移モデルを和音、テンポと音の継続長の変化の2つに分けて説明する。

まず和音のモデル化では、奏者は楽譜という明確な指示に基づいて演奏すると仮定すると、和音は $0, 1, 2, \dots$ と遷移していくと推測できる。しかし、人間の演奏は確定的ではなく、楽譜の誤読、または演奏誤りによって、局所的な音符の挿入、脱落、置換が発生することがある。また、練習時の演奏を想定すると演奏の途中楽譜の先頭から弾き直したり、大きく弾き飛ばしたりすることも考えられる。不確実性を持った演奏に対して効率的に楽譜追跡を行う手法として $\alpha\beta\gamma$ 法 [11] が提案されており、楽譜上の任意の和音間の遷移を考えることで、音符の脱落誤り、弾き飛ばし、弾き直しを見通しよくモデル化している。 $\alpha\beta\gamma$ 法を用いて演奏の性質ごとに遷移確率を与えることで、高速に演奏位置を探索することができる。

2.3 セグメンタル条件付き確率場による演奏モデル

本研究では、従来法に基づき、SCRFF の演奏遷移モデルと観測モデルを拡張する。従来法では、式 (2) に演奏遷移モデルと観測モデルを統合して以下のように表している。

$$\begin{aligned}
 p(\mathbf{q}, \mathbf{r} | \mathbf{o}) = & \frac{1}{Z(\mathbf{o})} \psi^d(s_1, d_1, r_1) \phi^s(q_1, \mathbf{x}_{1:t_1^e}) \phi^c(q_1, \mathbf{o}_{1:t_1^e}) \\
 & \cdot \phi^f(q_1, c_{1:t_1^e}) \phi^a(q_1, a_{1:t_1^e}) \\
 & \cdot \prod_{n=2}^N \psi^t(s_n, s_{n-1}) \psi^d(s_n, d_n, r_n) \psi^r(r_n, r_{n-1}) \\
 & \cdot \prod_{n=2}^N \phi^s(q_n, \mathbf{x}_{n:t_n^e}^s) \phi^c(q_n, \mathbf{v}_{n:t_n^e}^c) \\
 & \cdot \phi^f(q_n, c_{n:t_n^e}^f) \psi^a(q_n, a_{n:t_n^e}^a) \quad (3)
 \end{aligned}$$

したがって、スコアアライメントの問題は式 (3) の最大化に帰着される。ここで、 ϕ^s はパワースペクトル特徴、 ϕ^c はクロマ特徴、 ϕ^f はオンセット関数特徴、 ϕ^a はオンセット特徴の観測モデルを表す。また、 ψ^t は楽譜位置、 ψ^d は継続長、 ψ^r はテンポの演奏遷移モデルを表す。

ただし、必要に応じて特徴量の選択を行えば、それに伴って条件付き確率が変わることに注意する。このモデルは、テンポを含まないセグメントモデルとテンポモデルの統合モデルとして解釈することができる。つまり、テンポ以外の潜在変数を定数とすれば式 (3) はテンポを含まない通常のセグメントモデルとして解釈され、逆に、テンポを定数とみなせば式 (3) は線形動的システム (Linear Dynamical System; LDS) によるテンポモデルと解釈することもできる (詳細は文献 [9] を参照)。このような従来研究の提案モデルの二つの解釈は、効率の良い推論アルゴリズムを導く。

3. 打楽器を考慮した演奏モデル

3.1 合奏観点から見た打楽器の重要性

音響信号において、打楽器音は重要な役割を持つ。すなわち打楽器はその演奏される音によってを統率し曲をまとめる。打楽器音を含む多声演奏では、打楽器によって演奏されるリズムやテンポ、ノリに合わせて他のパート奏者の演奏が行われることでまとまりのある演奏が可能となる。また、打楽器音は音高の種類が他のパートに比べて明らかに少なく、打楽器音の響きは他の楽器の音色とは全く異なる。それゆえ、様々な楽器の音が重なり合う多声演奏においても、他のパートの奏者は打楽器音を見失わずに聞き取ることができる。このことは、音響信号と楽譜のアライメントにおいても同じである。つまり、アライメントに打楽器音の区別を取り入れることでアライメント精度の向上が見込まれる。

しかし、既に述べたように、打楽器音の多くはその他の楽器の音色とは異なるため、従来の音響信号のアライメントでは扱われていなかった。例えば打楽器には、その他の楽器においては最も重要な特徴量となる音高の種類が明らかに少ない。これは、従来のアライメント手法では対応できないことを示している。さらに、打楽器音を含む音響信号を打楽器音に対応していないアライメントの枠組みで行うと、打楽器音はその性質のためにノイズと判断され、アライメントの精度を低下させる恐れがある。従って、音響信号と楽譜のアライメントにおいて打楽器音とその他のパートは、それぞれ区別される必要がある。音響信号と楽譜のアライメントに打楽器の区別を取り入れる方法として、ビートトラック [13] を用いることが考えられるが、実時間アライメントにおいて計算量の増加は望ましくないため、本研究では採用しない。

本研究では各パートの音高の系列を得るための楽譜情報として、MIDI を用いる。MIDI はパート毎にそれぞれ発音位置や音高、さらに継続長など様々な情報が書かれており、打楽器のパートも例外ではない。これにより、複雑な処理を用いることなく、簡単に打楽器とそれ以外のパートの楽譜情報を正確に区別して扱うことができる。このような背景から本研究では、スコアアライメントにおいても楽器の特性を考慮する。

3.2 打楽器音を考慮した系列ラベリング問題としての音響信号と楽譜のアライメント

第2章において、音響信号と楽譜のアライメントの枠組みを系列ラベリング問題として SCRFF を用いて定式化を行った。ここで、従来法を拡張し打楽器音とその他の楽器を区別するための二つの観測モデルを定義する。

モデル1 打楽器補助アライメント

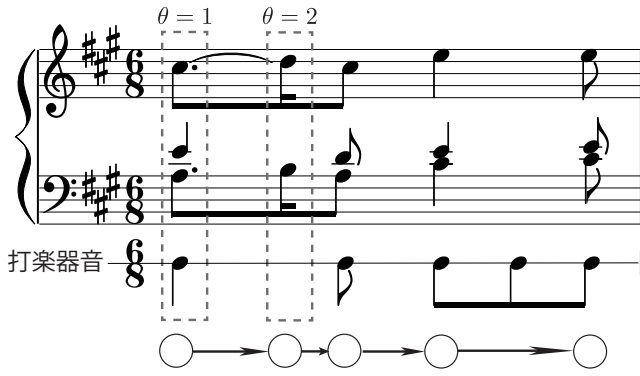


図 2 提案モデルにおける楽譜の和音表現

モデル 2 オンセットの重要度を考慮したアライメント

モデル 1 では、打楽器のみからなる和音は考慮しない。つまりこのモデルでは、打楽器以外の楽譜のアライメントを行うために打楽器の楽譜を補助的に用いる。モデル 2 では、モデル 1 に加えて楽器の種類によってオンセットの重要度を変えることでアライメント精度の向上を図る。モデル 2 はモデル 1 の拡張モデルであり、用いる和音表現は同一のものとなる。図 2 に、提案モデルにおける楽譜の和音表現を示す。従来モデルでは、各パートの同時発音の音の集合を多声演奏における和音の定義としていたが、本研究では図 2 に示すように打楽器音の有無によって和音を 2 種類の和音タイプに再定義する。それぞれのモデルに用いる和音タイプ θ は以下の 2 種類である。

- $\theta = 1$ 打楽器音とそれ以外のパートの音を両方含む和音
- $\theta = 2$ 打楽器音を含まない和音

本研究では、全体の楽譜は音高のある楽譜系列 $x_o = \{b_j, p_j, l_j\}_{j=1}^{J_o}$ と打楽器の楽譜系列 $x_d = \{b_j, l_j\}_{j=1}^{J_d}$ から構成されるものであり、この二つの楽譜系列を統合した和音系列 $s = \{s_t, \theta_t\}_{t=1}^T$ と音響信号の対応を考えている。 b は楽譜上の時刻、 p_j は音名、 l_j は拍を単位とする音符長を示す。

ここで、和音タイプとみなさない打楽器音のみからなる和音では音高は存在しないが、直前の和音が持続していると考え音高は直前の和音のものとすることに注意する。

3.3 打楽器の有無を考慮した観測モデル

3.3.1 音高と音色の観測モデル

第 2 章でも述べたように従来法では、楽譜系列は必ず音高のある和音系列としていた。しかし本研究では、音高のない音も考慮するため、観測モデルを再考する必要がある。再定義した観測モデルを以下に示す。

$$\tau^s(q_n, \mathbf{x}_{t_n^s:t_n^e}) = \begin{cases} 1 & (p_n \text{ が未定義}) \\ \phi^s(q_n, \mathbf{x}_{t_n^s:t_n^e}) & (\text{otherwise}) \end{cases} \quad (4)$$

$$\tau^c(q_n, \mathbf{v}_{t_n^c:t_n^e}) = \begin{cases} 1 & (p_n \text{ が未定義}) \\ \phi^c(q_n, \mathbf{v}_{t_n^c:t_n^e}) & (\text{otherwise}) \end{cases} \quad (5)$$

ここで、 x_t は音響信号の連続ウェーブレット変換、 v_t はクロマベクトルを示す。

3.3.2 発音に関する観測モデル

打楽器を考慮する際、音響信号から打楽器のオンセットを正確に得ることが望ましい。しかし、多声演奏などの複雑な音響信号から打楽器のみのオンセットを得ることは容易ではなく、実時間性が重要なオンラインアライメントにおいては計算時間の増加が問題となる。

そこで本研究では、打楽器のみのオンセットではなく、和音全体のオンセットと楽譜の対応を考える。打楽器も含む和音全体のオンセットを得るには、従来法に用いられている Spectral Flux [12] が有効である。音符のオンセット検出に関する研究として、パワーの変化に着目する手法 [14] も考えられるが、実際に演奏された音響信号の振幅が楽譜からの推定振幅変化に対応しなかった場合精度が著しく低下してしまうという問題点を踏まえ、本研究では採用しない。Spectral Flux は、パワースペクトルの各周波数成分に対して、時間方向の差分をとったもので、ピアノなど音の立ち上がりが顕著な楽器はもちろんのこと、バイオリンやクラリネットなどの音の立ち上がりが顕著ではない楽器音の場合にも頑健な特徴量となることが知られている。さらに、Spectral Flux は打楽器音のような音高のない楽器音の立ち上がりの検出に対しても有効であることが知られている。

打楽器には明確な音高がないため、打楽器音を含む和音の場合はオンセットがより重要になると考えられる。このようなことから、本研究ではオンセットに関する観測モデルに、和音タイプを定義し、その和音タイプによって重み付けを行う。和音のオンセットと楽譜を対応付ける際には、和音タイプによってオンセットの重要度が変わることには注意する。オンセットに関する観測モデルはオンセット関数特徴とオンセット特徴であり、それぞれ次のように再定義する。

$$\tau^f(q_n, c_{t_n^f:t_n^e}) = w(\theta_n) \phi^f(q_n, c_{t_n^f:t_n^e}) \quad (6)$$

$$\tau^a(q_n, a_{t_n^a:t_n^e}) = w(\theta_n) \phi^a(q_n, a_{t_n^a:t_n^e}) \quad (7)$$

ここで、 c_t は時刻 t に対する Spectral Flux、 a_t はオンセット結果である。

$w(\theta_n)$ は素性タイプによる重みであり、次式のように定める。

$$w(\theta_n) = \begin{cases} \nu_1 & (\theta_n = 1) \\ \nu_2 & (\theta_n = 2) \end{cases} \quad (8)$$

表 2 パラメータ設定 .

パラメータ	
クロマ特徴に対する重み	$\lambda^c = 0.1$
オンセット特徴に対する重み (1/3)	$\lambda_0^a = -0.3, \lambda_1^a = -0.01$
オンセット特徴に対する重み (2/3)	$\mu_0^a = -0.3, \mu_1^a = -0.01$
オンセット特徴に対する重み (3/3)	$\mu_h^a = -0.15h (h \geq 2)$
状態遷移確率	$\alpha = 1, \beta = 0, \gamma = 0$

本研究では ν によってそれぞれの和音タイプに重みを考慮する .

3.4 セグメンタル条件付き確率場による打楽器の発音位置を考慮したスコアアライメント

本節では、これまでに述べた観測モデルを第 2 章で述べた SCRF の枠組みに統合する . 再定義を行った御行に関する観測関数式 (4) と式 (5), オンセットに関する観測関数式 (6) と式 (7) を式 (3) に統合すると、以下の式 (9) が得られる .

$$\begin{aligned}
 p(\mathbf{q}, \mathbf{r} | \mathbf{o}) = & \frac{1}{Z(\mathbf{o})} \psi^d(s_1, d_1, r_1) \tau^s(q_1, \mathbf{x}_{1:t_1^e}) \tau^c(q_1, \mathbf{o}_{1:t_1^e}) \\
 & \cdot \tau^f(q_1, c_{1:t_1^e}) \tau^a(q_1, a_{1:t_1^e}) \\
 & \cdot \prod_{n=2}^N \psi^t(s_n, s_{n-1}) \psi^d(s_n, d_n, r_n) \psi^f(r_n, r_{n-1}) \\
 & \cdot \prod_{n=2}^N \tau^s(q_n, \mathbf{x}_{t_n^s:t_n^e}) \tau^c(q_n, \mathbf{v}_{t_n^s:t_n^e}) \\
 & \cdot \tau^f(q_n, c_{t_n^s:t_n^e}) \tau^a(q_n, a_{t_n^s:t_n^e}) \quad (9)
 \end{aligned}$$

これにより、打楽器音の有無を考慮したスコアアライメントが可能となる .

4. 実験

4.1 実験条件

本研究では、RWC 研究用音楽データベース [10] よりポピュラー音楽の音響信号を用いて打楽器音を考慮した音響信号と楽譜のアライメントの評価を行った . それらの録音物を用いることで打楽器音の有用性の評価はもちろん、テンポ変動、多種類の楽器に対する頑健性を評価することができる . MIDI データはボーカルパートがソフトウェア音源であり、さらにハンドラベリングされている部分では若干のズレが生じてしまうため、より正確な評価を行うために、MIDI データから音響信号を合成して実験に用いる . その際には、iTunes を使用した .

まず、本研究における提案モデルであるモデル 1 と、楽器の種類を考慮していない従来法のモデルの比較実験を行った . その後、モデル 1 とモデル 2 の比較実験を行った .

評価基準には、アライメント結果から得られるオンセット検出の再現率、適合率、 F -値の三つを用いる ([12] の方法に則る) . F -値は適合率と再現率の調和平均であり、そ

れらのトレードオフを考慮した評価指標となり、数値が高いほどアライメント性能が優れていることを示す . なお、オンセット選出結果のオンセット位置が許容誤差以内であれば正解とする . 許容誤差は 100 ms, 300 ms, 500 ms それぞれで評価を行う .

パラメータは従来法において予備実験で決定されたもの表 2 を用いる . 文献 [9] の実験条件と同様に、本研究においても弾き飛ばしなどは起こらないとしていることに注意する . [12] で示されているように、実際にはオンラインであっても遷移確率の調整によって容易に対応できる .

4.2 実験結果

表 1 はそれぞれのモデルの実験結果を示している . モデル 1 では、従来モデルに比べ F -値はすべての許容誤差で向上した . つまり、音響信号と楽譜のアライメントに打楽器を考慮することは効果的であることがわかる . さらに、依然音高のあるオンセットは重要であるが、音高のない打楽器のオンセットを補助的に用いることも重要であることがわかった .

モデル 2 についてもモデル 1 の結果からさらに F -値が全ての許容誤差において向上したが、従来モデルとモデル 1 の比較に比べてその数値向上の程度は小さかった . 楽譜情報を区別することだけでなく、最適な重みを決定することもアライメント精度を向上させると考えたが、実際は精度の改善は微々たるものであった . つまり、このモデルは現段階では有効ではないと考えられる . しかし、正しく重み調整をすることで今後アライメント精度をさらに向上させられる可能性があるといえる . さらに、曲ごとに重みは異なると考えられるため、より細かくパラメータの変更を検討する必要がある .

5. 結論

本論文では、和音に打楽器音が含まれるかどうかによってオンセットの重要度を変化させ、適切な重みを与えることで、ポピュラー音楽における音響信号と楽譜のアライメントの精度の向上を図った . そのため、従来法では各パートの同時発音の音の集合としていた和音を、打楽器音とそれ以外のパートの音の両方を含む和音と打楽器音を含まない和音の 2 種類にタイプ分けし、最適な重みを与えた . このように従来法に和音タイプに応じた重みの変化を加えることにより、オンセットの重要度を予測される演奏位置に応じて変化させることを可能とした .

実験では、二つの提案モデルを従来法のモデルと比較した . モデル 1 では、打楽器の楽譜を補助的に用いて打楽器以外の楽譜のアライメントを行い、モデル 2 ではその最適な重みづけを行った . 両モデルにおいてアライメント結果から得られる F -値の向上が見られたため、提案法は有効であることが示唆された . しかし、モデル 2 ではアライメン

表 1 打楽器補助アライメントにおける提案法と従来法の比較

Measure	100 ms			300 ms			500 ms		
	F-値	適合率	再現率	F-値	適合率	再現率	F-値	適合率	再現率
従来モデル	54.6	54.1	55.0	75.2	74.6	75.9	78.5	78.0	79.2
モデル 1	59.5	59.1	60.0	81.5	80.9	82.2	84.8	84.1	85.5
モデル 2	60.3	60.1	60.5	81.8	81.3	82.3	85.4	85.1	85.7

ト精度の大きな向上は見られなかったため、更なる改良が望まれる。

今後の展望として、曲ごとに打楽器音の有無や重要度は違うと考えられるため、楽譜から最適な和音タイプによる重みを推定することなどが考えられる。

参考文献

- [1] Dannenberg, Roger B and Raphael, Christopher: Music score alignment and computer accompaniment, *Communications of the ACM*, vol. 49, no. 8, pp. 38–43 (2006).
- [2] Montecchio, Nicola and Cont, Arshia: A unified approach to real time audio-to-score and audio-to-audio alignment using sequential montecarlo inference techniques, *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference*, pp. 193–196 (2011).
- [3] J. Bloch and R. Dannenberg: Real-time computer accompaniment of keyboard performances, in *Proc. of the International Computer Music Conference (ICMC)*, pp. 279–289 (1985).
- [4] 武田晴登, 西本卓也, 嵯峨山茂樹: HMM による MIDI 演奏の楽譜追跡と自動伴奏, *情報学音楽情報科学研報, MUS-66*, pp. 109–116 (2006).
- [5] D. Schwarz and N. Orio and N. Schnell: Robust polyphonic MIDI score following with hidden Markov models, in *Proc. of the International Computer Music Conference (ICMC)*, pp. 442–445, (2004).
- [6] Hu, Ning and Dannenberg, Roger B and Tzanetakis, George: Polyphonic audio matching and alignment for music retrieval, *Computer Science Department*, pp. 521 (2003).
- [7] Orio, Nicola and Déchelle, François: Score following using spectral analysis and hidden Markov models, in *Proc. of the International Computer Music Conference*, vol. 1001, no. 27, pp. 1708–1710 (2001).
- [8] Joder, Cyril and Essid, Slim and Richard, Gaël: A conditional random field framework for robust and scalable audio-to-score matching, *Audio, Speech, and Language Processing, IEEE Transactions*, vol. 19, no. 8, pp. 2385–2397 (2011).
- [9] R. Yamamoto and S. Sako and T. Kitamura: Robust on-line algorithm for real-time audio-to-score alignment based on a delayed decision anticipation framework, in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, pp. 191–195 (2013).
- [10] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, *情報処理学会論文誌*, vol. 45, no. 3, pp. 728–738 (2004).
- [11] 中村栄太, 武田晴登, 山本龍一, 齋藤康之, 酒向慎司, 嵯峨山茂樹: 任意箇所への弾き直し・弾き飛ばしを含む演奏に追従可能な楽譜追跡と自動伴奏, *情報処理学会論文誌*, vol. 54, no. 4, pp. 1338–1349 (2013).
- [12] S. Dixon: Onset detection revisited, in *Proc. of Inter-*

- national Conference on Digital Audio Effects (DAFx)*, pp. 133–137 (2006).
- [13] 後藤真孝, 村岡洋一: 音楽音響信号に対するビートトラックシステム, *情報処理学会研究報告*, vol. 94, no. 71, pp. 49–56 (1994).
 - [14] 三浦雄文, 赤羽歩, 佐藤誠, 津田貴生, 井上誠喜, 小宮山撰: 音響信号の振幅を用いるオーケストラ楽曲の楽譜追跡, *The Journal of The Institute of Image Information and Television Engineers*, vol. 61, no. 7, pp. 997–1005 (2007).