Music Score Alignment and Computer Accompaniment

楽譜アラインメントとコンピュータ自動伴奏

Roger B. Dannenberg, Christopher Raphael

音楽の音を音楽記述と関連付けることで，これらのシステムは，人間の演奏に追随する飽きの来ない、表情豊かな伴奏を生成するだけでなく，時にはそこから学習もする.

現在，コンピュータでアクセス可能な形式の音楽が，量の上でも種類の上でも爆発的に増大しているのを私たちは目の当たりにしている．音楽データは，主に 2 種類に分かれる：サンプリングされた音響信号ファイル（例えばコンパクトディスクに収録されたたり，Web 上に散在するもの），および，音楽の記号表現である，後者の基本は，音符ごとに音高（pitch），開始時刻（onset），持続長がついた音符のリストである．音楽音響信号と音楽記号表現の対応は，話し声の音響信号とテキストの関係と同様である．いずれの場合も，音響信号は，演奏や発声の多彩に富んだ表現力豊かなニュアンスを表すことができる．もっとも，人のリスナー以外の何物もこれを「理解する」のはたやすくない．一方，テキストも音楽記号表現も，高次の「単語」をけちって格納でき，その認識もたやすくできる．

本稿では，機械リスニング（machine listening）の一種である，楽譜マッチング（music score matching），楽譜追跡（score following），楽譜アラインメント（score alignment）に焦点を当てる．我々は，同じ音楽に対する記号表現と演奏音響信号との間の対応関係を探究し，音響信号中の適切な音楽「イベント」，通常は音符，の開始時刻を同定する．このマッチング問題は，通常「オフライン」と「オンライン」の 2 種類の異なるバージョンに分かれる．

オフラインマッチングは，楽曲の演奏全体を用いて音響信号と記号表現である楽譜との対応関係を推定する．オフライン問題では，対応関係を確定中に，「未来を先読みする」ことが許される．それゆえに，オフラインマッチングは，演奏への見出し付けとみなすことができ，録音へのランダムアクセスが可能となる．そのような見出しを用いて，リスナーは，音楽作品中の任意の位置（例えば 48 小節（measure）の 2 拍目）から始めて，可視化された楽譜表現を音響信号とリンクさせたり，アニメーションを録音済み（prerecorded）音響信号と調整させたりできる．

オフライン楽譜マッチングは，新しくて魅力的なアプリケーションに直ちに応用することができる．例えば，音楽のディジタル編集や後処理では，音響信号中の特定の音符の位置を様々な方法で調整し，バランスを変え，微調整する必要がある．楽譜マッチングにより音響エンジニアはそれらの位置を自動的に同定できるので，その作業が劇的に単純化されることになる．もう 1 つの例は，音楽音の合成に関連している．音楽音合成の品質は，一般的に様々な条件下（音高，ダイナミックレンジ，アーティキュレーションなど）で音響サンプルをどれだけ集めるかに依存している．楽譜マッチングの使用により，この時間のかかるデータ収集を自動化することができる．違った方向としては，楽譜マッチングによりタイミングやテンポに関する量的情報を得ることもできるので，音楽表現に関する研究や理解の一助となる．楽譜マッチングは，音響信号を音符に基づいて分割することで，

他の多くの音楽属性（ダイナミクス，ビブラート，音色など）についての研究の鍵ともなり得る．この探究は，さらに，未来の合成音楽に大きく貢献することになろう．

オフライン楽譜マッチングの実時間版の親類は，楽譜追跡と呼ばれ，音響信号をそれが得られた順にオンラインで処理する．そのため，「先読み」は可能ではない．楽譜追跡の目的は，楽譜に記されている音楽イベントを高精度かつ低遅延で同定することである．私たちの心に最も響くアプリケーションは，自動伴奏（accompaniment）システムである．これは，生のソロ演奏者の演奏に追随した柔軟な伴奏を生成する．生の演奏者を「聞く」ことは，いかなる解法において必要な要素技術である．他のアプリケーションとしては，視聴覚機材の音楽演奏への自動調整（例えばオペラの字幕（supertitle）や照明効果など）や，楽譜に基づいた実時間の音響品質向上（音高補正や自動ページめくりなど）などがある．

おそらく，楽譜追跡というチャレンジを最も直接的に説明する方法は，単純な例を用いることであろう．単旋律（同時に一つの音しか演奏されない，monophonic）楽器が少雨数の音符しか含まない既知の楽譜から，例えばABCD のように，演奏されているとしよう（図1）．音高の値を 50 ミリ秒程度ごとに音響信号から推定する．音高推定は正確ではないが，音符ごとに多数の推定結果が存在し，次のような結果が得られる：

AAAAAAAAaABABABBBCBBCBcBBBBBdBBBddCDCCCCCcCCCCDDBDdDDcDBDDDDDD

ここで，小文字は楽譜に記された音より 1 オクターブ低い音を表す．

楽譜マッチングは，この系列を ABCD という音符系列に対応した領域に分割できる必要がある．図 1 に示した 1 つのアプローチでは，「スコアモデル」（Score Model）と名付けられた，A，B，C，D の 4 つの可能な状態からなる単純な状態グラフを使用する．各ステップでは，モデルは，同一状態に留まり続けるか，次の状態に進む．この例では 62 の音高推定値が入力されるので，システムはグラフ中の長さ 62 の経路すべてを考慮する必要がある．各経路は，実際のデータ系列とどの程度ぴったりとマッチングするかによって，スコア付けされる．各記号に対して，正確に一致した場合はペナルティ無し，オクターブ間違いや隣接する音高の誤りには小さいペナルティを与え，それ以外の場合には大きなペナルティを与える．このようにして，各系列には全ペナルティ（発生したペナルティの合計値）が与えられ，最適な状態系列はこのペナルティを最小化する系列として定義される．（読者の中には動的計画法（dynamic programming）に詳しい方もおられよう．動的計画法を用いると，すべての経路を列挙しなくても最適解を見つけることができる．）楽譜マッチングの最適状態系列は，例えば次のようなものである：

AAAAAAAAAAAABBBBBBBBBBBBBBBBBBBCCCCCCCCCCCCCCDDDDDDDDDDDDDDDD

これは明らかにおもちゃの例であるが，基本的なアイディアは音楽という領域で遭遇する問題の大部分をカバーするように，拡張し一般化することができる．例えば，多重奏（polyphonic）音楽は和音（chord）系列として表現することができ，音符が始まるか終る時刻ごとに新しいコードが出現する．音符長の変動に関する確率分布は，各音符にいくつかのグラフ状態を割り当て，遷移確率を含めることで表現できる．

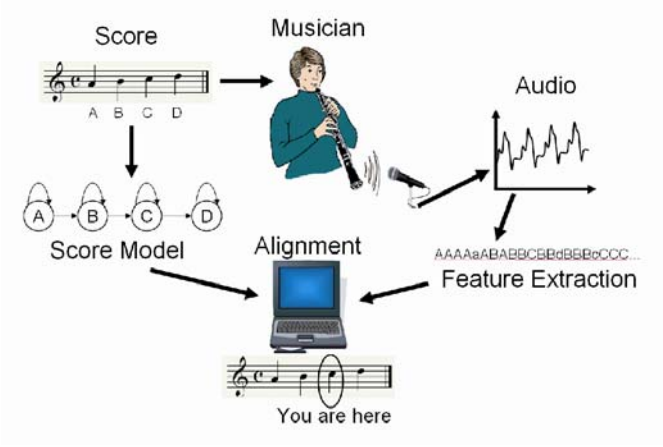

図 1．音高推定の記号表現を得るために，音響信号を分析する（「特徴抽出」，feature extraction）．これらの推定値は記号表現された楽譜から導出されたスコアモデルの状態系列に対応する．次に，アラインメント（alignment）処理が音響信号の時間位置を楽譜上の位置と関連付ける．

状態グラフのアイディアを多重奏音楽に直接拡張するには，音高推定値を音高推定値の集合に置き換えることになろう．グラフの音符系列は和音系列に置き換えられよう．残念ながら，多重音高推定は信頼できない（活発な研究分野である）ため，今日の楽譜アラインメントシステムは信号中の周波数成分の分布に基づく表現，別称は信号のスペクトル（spectrum），を用いている．スペクトルが使用されるのは，スペクトルが音響信号から容易に計算でき，楽譜から合理的に予測できるからである．

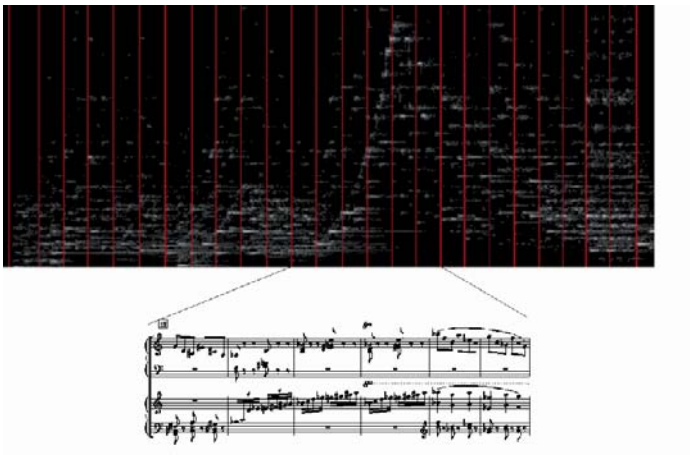

図 2. Gustav Holst の惑星（The Planets）より「水星」の冒頭部のスペクトログラム．縦線は自動楽譜アラインメントによって推定された小節区切りの位置を表す．添付の音響ファイルには小節線の位置にクリック音が加えられている。

コードからスペクトルを予測するのに，音楽音響学に関する基本的な知識を使用することができる．和音中の各音符が生成するスペクトルは，調波構造（harmonics）の集合で近似できる．音高の集まりは，このような単音符モデルを重ね合わせとしてモデル化する．次に，ペナルティ関数は，予測スペクトルと観測スペクトルとの距離に基づいて定義するか，あるいは，和音構成音の仮説が与えられた際のデータスペクトルの確率モデルに基づいて定義する．スペクトル予測の別の方法として，楽譜を音楽シンセサイザーで演奏し，得られた音響信号からスペクトルを計算する方法がある．この方法も，同一楽曲の 2 つの異なる演奏の音響信号間のアラインメントに通じる可能性を秘めている．

楽譜から得られるスペクトルと音楽音響信号から得られるスペクトルを比較する際に用いられる一般的な方法は，詳細なスペクトルを 12 要素から成るベクトルに圧縮することである．このベクトルの各要素は 12 の音高のクラス（例えば，C，C#，D）の 1 つに関連するスペクトルの全エネルギーを表す．これらのいわゆる「クロマベクトル」（chroma vector）間のユークリッド距離は，楽譜アラインメントの頑健な距離尺度である[6]．

これらの本質的なアイディアを武器に，研究者は実際の音響信号を扱える楽譜マッチング器を開発してきた．例えば，図 2 はオーケストラ演奏のスペクトログラムで，認識された小節区切り（measure）が縦線で示されている．この音響信号の認識は，前述の「状態」の概念を，時間変化するテンポを含むように拡張することで，達成している[8]．対応する音響ファイルには，小節区切りを表すクリック音が加えられており，xavier.informatics.indiana.edu/~craphael/acm で聞くことができる．

コンピュータ自動伴奏

コンピュータ自動伴奏は，今日までにオンライン楽譜追跡の大成功を収めた代表であり，伴奏システムはいくつかの分野で有望となっている．まず第 1 は，ソロ演奏者の伴奏という伝統的なシナリオであり，生の演奏者が柔軟かつ繊細な（飽きの来ない）伴奏で演奏したい場合である．このような伝統的な分野では，人の伴奏者では強硬な競争となるのに対して，自動伴奏システムでは人の伴奏者をいくつかの点で凌駕している．例えば，自動伴奏システムはほとんど無制限の技術的能力を備え，任意の速い音符と複雑なリズムの連繋が可能である．そのような妙技は，コンスタンツ大学（Universität Konstanz）の Jan Beran とケンブリッジ大学（Cambridge University）の Nick Collins によって自動伴奏システムのために明示的に作曲された最近の音楽において探求されている．（デモは xavier.informatics.indiana.edu/~craphael/music_plus_one/）．恐らく，自動伴奏システ

ムはいつの日か，ジャズなどの即興的音楽にも居場所を見つけることになろう．

ほとんどのコンピュータ自動伴奏システムは，図 3 に示される線に従って，4 つの明確に独立した構成要素からできている．すなわち，音響特徴抽出，楽譜追跡，伴奏制御，音楽合成である．入力は，演奏者と伴奏システムの双方のための機械可読形式の楽譜であり，出力は実時間の音楽伴奏である．特徴抽出部では，入力音響信号に対して信号処理を行い，データ（音高推定やフレーム単位のスペクトルなど）のストリームを生成する．楽譜追跡部では，この系列を楽譜とマッチングを取り，ソロ演奏者の現在の位置の推定をさらに展開させる．これは，図 1 の楽譜アラインメントのリアルタイム版である．楽譜追跡部は，ソロ演奏者の演奏について実況コメント（例えば「今ここです.」）を行い，遅延時間が変わりながらソロ演奏のイベント時刻推定を提供する．

伴奏制御は自動伴奏システムの脳に当たり，楽譜中の情報や，音楽的な制約，過去の演奏から抽出された知識などを統合し，伴奏音符を演奏すべき時刻を決める．4 番目の要素，音楽合成，は実際の伴奏音が生成する．生成では，音符ベースのサウンドシンセサイザーが使われる場合もあれば，録音済み音響信号をその場で再合成する場合もある．

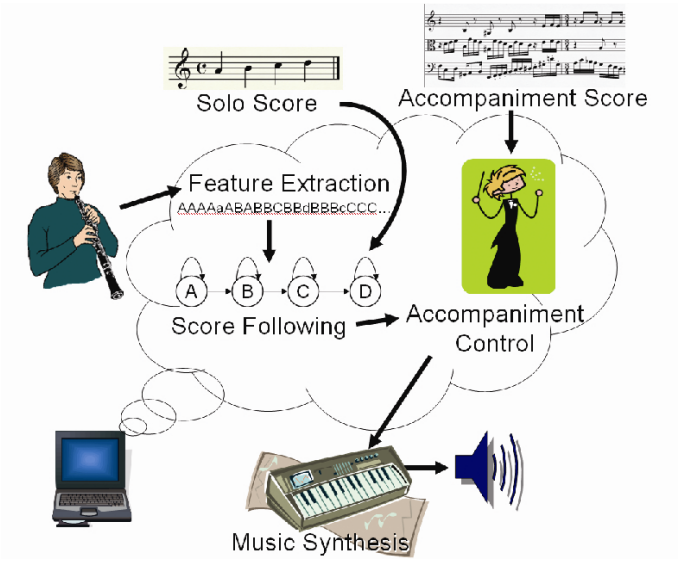

図 3. コンピュータ自動伴奏システムの構造

音符が聞えるまで待って反応することで，同期を達成しようとするような自動伴奏システム（あるいは個人）では，常に遅れる．というのは，すべての音楽イベントは遅延を伴って検出されるからである．もっとうまいやり方は，検出された音符の開始時刻だけでなく他の情報を基に，音楽がこの後どのように展開するかを予測して，音楽の各パートを調整することである．音楽予測に関する本質的なチャレンジ問題では，楽譜，楽譜追跡器の出力，ソロ演奏者や人が演奏した伴奏からの過去の学習例などを含む，種々の知識の統合が必要である．1980 年代に開発された初期の自動伴奏システムは，この予測を導くのに，人が書いたルールを使っていた[1,3]．最近では，不完全情報からの自動学習と予測というチャレンジに，信念（belief）ネットワークによる取り組みが始まっている[9]．

初期の自動伴奏システムは，スペクトルデータではなく，音符系列にマッチングさせ，データ速度と計算量を低減していた[1,3,12]．隠れマルコフモデル（Hidden Markov model）が音響信号のスペクトル表現とともによく用いられ，リハーサルデータから学習し，より正確な追跡とより少ない遅延の応答を可能にしている．他のアプローチ－Probabilistic Vocalist Tracker [4]－では，楽譜上の位置を連続確率密度関数（continuous probability density function）として表現することで，離散状態問題を回避した（デモは www.cs.cmu.edu/~music/video/CANGIO.MOV）．コンピュータ自動伴奏は伴奏の反応性と音楽性を重視しているのに対して，オフライン楽譜マッチングの主な目標は，アラインメントの正確性であり，短時間音響信号フレームと，マッチングに特化した特徴量によって達成可能である[7]．

スコア追跡のアプリケーション

楽譜追跡とコンピュータ自動伴奏により，作曲者はコンピュータの超絶技巧能力を活用し，生演奏の音に対してディジタル音響信号処理を適用することができるようになった．また，コンピュータ自

動伴奏はいくつかの特筆すべき音楽教育システムにおいても使用されている．

The Piano Tutor. 楽譜追跡は，生徒と人の指導者との間のインタラクションをモデル化した Piano Tutor [2]と呼ばれる知的ピアノ指導システムの基礎となっている．このシステムは約 100 曲のレパートリーを持ち，生徒に応じて選ぶことができる．生徒が音楽を演奏すると，実時間楽譜追跡が生徒の演奏を追跡し，ディスプレイ上の楽譜を更新することで自動的にページをめくる．演奏終了後，システムは楽譜アラインメントを使用して，弾き飛ばした音符と誤って弾いた音符を同定する．生徒の演奏テンポは自動的に推定され，タイミングの誤りが同定される．この分析に基づき，エキスパートシステムによる案内により，「よく見て．ここで音符を見逃しましたよ．」という声が発声され，同時に，楽譜上の問題領域が四角で囲まれる．楽譜追跡によって，Piano Tutor は比較的制約のないピアノ演奏を評価できるようになり，その結果，単純なインタフェースになり，かつ，（私見だが）生徒にとってより楽しい経験となろう．（ www.cs.cmu.edu/~music/video/pianotutor.mov）

SmartMusic. SmartMusic（smartmusic.com） は，一般のパソコン向けの商用ソフトウエア製品である．コンピュータ自動伴奏を音楽教育用ツールとして使い，管楽器，ボーカル，初級の弦楽器演奏者のために 30,000 曲以上の音楽に対して伴奏を有している．従来の音楽科の生徒は，一人で練習するものの，効率的に練習するための技量や動機付けが不足する場合が往々にしてあった．SmartMusic は，生徒が最初から最後まで楽曲を習得できるように挑戦させる．伴奏付きで練習することで，ソロパートからは見逃しがちなリズムとハーモニーの文脈で自分自身の音楽を聞くことができる．また，伴奏によって音程の基準が与えられるので，生徒は，何時音程を外したのか，何時修正動作を取る必要があるかを，よりよく聞き知ることができるようになる．伴奏付き練習した生徒についてのごく小規模な研究によると（music.utsa.edu/tdml/conf-VI/VI-Repp/VI-Repp.html)，伴奏付きで練習することで演奏技量と自信が向上し，動機付けがより高まったと感じ，練習時間も増大している．

Music Plus One. Music Plus One は，主として西洋芸術音楽を対象としたシステムであり，事前録音の音響信号からフェイズボコーダ（phase vocoder）を使用して再合成することで伴奏を作成する．したがって，元の演奏の豊かな音のパレットと音楽性を保ったまま，ソロ演奏者との同期が可能になる．システムは，隠れマルコフモデルを使って楽譜追跡をし，伴奏制御部を含む信念ネットワークを使ってパート間の調整を行う．このネットワークは，楽譜追跡器，楽譜そのもの，過去のソロ演奏と人の演奏した伴奏に基づいて，今後の伴奏音の開始時刻を予測する．これらの予測により，フェイズボコーダが継ぎ目なく接続した音断片の列（a trail of bread-crumbs）が生成される．

まとめ

オンライン，オフライン楽譜マッチングアルゴリズムの能力向上によって，ディジタル音響編集，音響データベースの構築，実時間演奏品質向上，自動伴奏システムにおいて，多くの新しいアプリケーションが実現可能になった．音楽はほとんどすべての仮定が破られるほど変化に富んでいる一方で，関連するアルゴリズムは既存の音楽分野だけでなく新しい挑戦的な音楽分野においても信頼性高く機能し，多くの新しい可能性を約束している．

多くの音楽家は練習をより楽しく有益なものにするために自動伴奏システムを使うが，技術を広範囲に受け入れてもらうためには，研究者達はいくつものチャレンジを，特に，いかに音楽的美意識のより深淵な理解をどう捉えていくか，乗り越えなければならない．そのためには，音楽学と計算機科学など，これまでは互いに限定的にしか意識されてこなかった複数の分野の専門知識が必要となる．

コンピュータ自動伴奏は，何百万もの音楽家や音楽愛好家の生活に触れるような商用のアプリケーションを見つけて行くはずである．音楽家はこれらの機械で置き換えられるのであろうか．我々は，これは主たる効果とはならないと考えている．そうではなく，自動伴奏システムによって，音楽がより多くの人々により近づき易いものとなり，音楽作成

についての鑑賞力と愛が育まれるであろう．究極の効果は，リスナーと参加者両方から，「全人（all-human）の」音楽への要求が高まることであろう．

文献

1. Baird, B., Blevins, D., and Zahler, N. Artificial intelligence and music: Implementing an interactive computer performer. *Computer Music Journal* 17, 2 (Summer 1993), 73.79.
2. Dannenberg, R., Sanchez, M., Joseph, A., Capell, P., Joseph, R., and Saul, R. A computer-based multimedia tutor for beginning piano students. *Interface Journal of New Music Research* 19, 2.3 (1993), 155.173.
3. Dannenberg, R. Real-time scheduling and computer accompaniment. In *Current Research in Computer Music*, M. Mathews and J. Pierce, Eds. MIT Press, Cambridge, MA, 1989, 225.261.
4. Grubb, L. and Dannenberg, R. A stochastic method of tracking a vocal performer. In *Proceedings of the 1997 International Computer Music Conference* (Thessaloniki, Greece, Sept. 25.30). International Computer Music Association, San Francisco, 1997, 301.308.
5. Hu, N., Dannenberg, R., and Tzanetakis, G. Polyphonic audio matching and alignment for music retrieval. In *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, Oct. 19.22). IEEE Computer Society Press, Piscataway, NJ, 2003, 185.188.
6. Loscos, A., Cano, P., and Bonada, J. Score-performance matching using hidden Markov models. In *Proceedings of the 1999 International Computer Music Conference* (Beijing, Oct. 22.27). International Computer Music Association, San Francisco, 1999, 441.444.
7. Orio, N. and Schwarz, D. Alignment of monophonic and polyphonic music to a score. In *Proceedings of the 2001 International Computer Music Conference* (Havana, Sept. 17.22). International Computer Music Association, San Francisco, 2001, 155.158.
8. Raphael, C. A hybrid graphical model for aligning polyphonic audio with musical scores. In *Proceedings of the Fifth International Conference on Music Information Retrieval* (Barcelona, Spain, Oct. 10.14). Audiovisual Institute Pompeu Fabra University, Barcelona, Spain, 2004, 387.394.
9. Raphael, C. A Bayesian network for real-time musical accompaniment. In *Proceedings of Neural Information Processing Systems* (NIPS) 14 (Vancouver, B.C., Dec. 3.8). MIT Press, Cambridge, MA, 2001.
10. Raphael, C. Automatic segmentation of acoustic musical signals using hidden Markov models. *IEEE Transactions on PAMI 21*, 4 (Apr. 1999), 360.370.
11. Schwarz, D., Cont, A., and Schnell, N. From Boulez to ballads: Training IRCAM's score follower. In *Proceedings of the 2005 International Computer Music Conference* (Barcelona, Spain, Sept. 5.9). International Computer Music Association, San Francisco, 2005.
12. Vercoe, B. The synthetic performer in the context of live performance. In *Proceedings of the International Computer Music Conference 1984* (Paris, Oct. 19.23, 1984). International Computer Music Association, San Francisco, 1985, 199.200.

Roger B. Dannenberg（roger.dannenberg@cs.cmu.edu）はペンシルベニア州ピッツバーグ市カーネギーメロン大学の計算機科学と芸術の准研究教授である．

Christopher Raphael（craphael@indiana.edu）は，インディアナ州ブルーミントン市インディアナ大学の情報学部准教授と音楽学部客員教授である．

訳：藤原弘将（京都大学・情報学研究科）