

# ピアノ演奏コンピュータグラフィクス制作技術

巳波 弘佳, 古屋 晋一, 長田 典子

ピアノ演奏場面が含まれるアニメーションはこれまで数多く制作されてきたが、演奏する手元や指先は巧みな演出によって回避されてきた。それはアニメーターが作画することが不可能であるうえ、近年のコンピュータグラフィクス (CG) 技術を用いても困難であったからである。本稿では、ピアノ演奏 CG 制作に関する技術について、必要な技術課題とその解決例を紹介する。なお、実用的なピアノ演奏 CG 制作には、最適化問題としての定式化と効率的なアルゴリズムの設計というアプローチが有効であった。CG や音楽といった娯楽に関係する分野においても、オペレーションズ・リサーチの有効性が示されたと言える。

キーワード：ピアノ演奏, コンピュータグラフィクス, 最適化, アルゴリズム

## 1. はじめに

楽器演奏場面、特にピアノ演奏場面が含まれる映画やアニメーションはその創成期より数多く制作されてきている。しかし、俳優に演奏させることが困難であったり、作画に膨大な労力と正確性が必要なアニメーションでは実質的には制作不可能であったため、手元や指先を隠蔽するといった演出による解決策をとらざるをえなかった。

しかし近年、クラシック音楽がテーマのコミック「のだめカンタービレ」(二ノ宮知子原作) が人気を博し、テレビドラマ・テレビアニメ・実写映画も制作された。このように、楽器演奏そのものがストーリー上重要である映画やアニメーションが注目されたこともあり、演出だけで演奏場面を回避することには限界がきていた。そのため、楽器演奏場面を実際に作画することが緊急の課題となっていた。

楽器演奏場面の作成には 2 つのアプローチがある。1 つは、モーションキャプチャシステムを利用して演奏者の 3 次元的位置座標を記録し、それに基づいて CG (コンピュータグラフィクス) を生成するというものである。もう 1 つは、楽譜データだけから運指や手指・腕の動作を直接 CG として生成するものである。これらのアプローチによって楽器演奏 CG アニメーションを自動的・半自動的に生成することができれば、映画



図 1 マーカ

やドラマなどの制作における時間とコストの大幅な削減が可能となる。そのため、筆者の研究グループでは、楽器演奏 CG 自動生成システム、特にピアノ演奏 CG 自動生成システムの研究開発を進めてきた [1]。

本稿では、筆者らがかかわった、テレビアニメ「のだめカンタービレ」の巴里編 (第 10 話, 第 11 話) およびフィナーレ編 (全編) におけるピアノ演奏 CG 制作に用いた技術をはじめ、特にピアノ演奏 CG 制作に関する技術課題とその解決例の一部を紹介する。

## 2. モーションキャプチャを用いた CG 生成

モーションキャプチャは、人間などの動きをセンサを用いてコンピュータに 3 次元位置座標として取り込むシステムである。対象に取り付けたマーカ (図 1) という小球に赤外線を反射させて複数のカメラで撮影することにより、当該マーカに対応する点の 3 次元座標を決定する光学式が一般的に用いられている。

モーションキャプチャは、フレームごとに、フレーム内の各点の 3 次元座標値を記録する。ただし記録時点では、各点がどのマーカに対応するかはわからない。各点に対応するマーカ名を特定する処理を行うことに

みわ ひろよし

関西学院大学 理工学部 情報科学科  
〒669-1337 兵庫県三田市学園 2-1

ふるや しんいち

ハノーファー音楽演劇大学 音楽生理学・音楽家医学研究所  
ながた のりこ

関西学院大学 理工学部 人間システム工学科

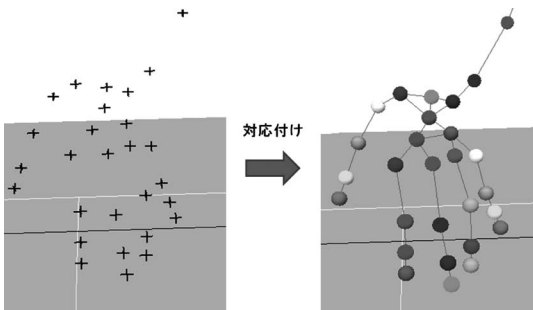


図2 点座標とマーカの対応

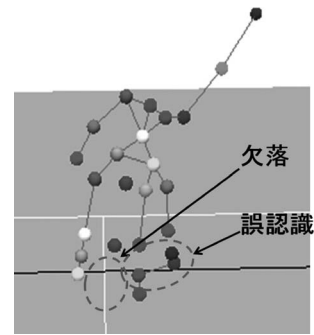


図3 誤認識と欠落

よって、モーションキャプチャデータは使用可能なものとなる(図2)。

モーションキャプチャは映画やCG制作において、現在では必要不可欠な手段であるため、ピアノを始め楽器演奏への適用を考えることは自然ではある。しかし実は、モーションキャプチャはピアノ演奏のデータ収録には向いてはいない。誤認識や欠落(図3)が多発することが理由の1つである。一般的なモーションキャプチャシステムには、記録した点位置に対応するマーカ名を推定する機能があるが、ピアノ演奏に特化したものではないため、対応づけを誤ること(誤認識)がある。実際、ピアノ演奏のデータ収録では、狭い領域内での手指の動きを取得することが必要であるが、きわめて高速な動きの際には、特に隣接したマーカとの誤認識が頻繁に発生する。密集した多数のマーカを正確に認識する必要のあるピアノ演奏の場合には誤認識は致命的である。また、正確性を高めるためにフレームレートを上げたとしても、通信速度やPC内部処理速度がボトルネックとなり、高性能のモーションキャプチャシステムであっても、実際にはマーカが存在する位置に対応する点座標が取得できず、欠落が多発する。また、手指が重なることでカメラの死角にマーカが入ることも多いうえ、そもそも手指の動きの激しさのためにマーカが外れることも少なくない。曲目によっては、1つのマーカにつき数百フレームにおいて欠落が生じる場合もある。このような誤認識や欠落をアニメーターの手作業で修復することは稼働量の観点からも非現実的であるため、実際のアニメーション制作の大きな壁の1つとなっていた。

もう1つの壁は、演奏動作データと音楽データの同期である。一般に、演奏動作データ収録と音楽データ収録はそれぞれ別に行われる。これは、モーションキャプチャで演奏動作データを収録する際には手指や楽器に付加されるマーカが演奏の妨げとなるため、高い品

質の演奏が必要な音楽データの収録を同時に行うことは不可能だからである。このため、両データ間のずれは不可避であり、楽器演奏CGを制作する際には、このずれを補正して両データを同期させる必要がある。これに対して、演奏動作と音楽双方の時間スケールを単純に伸縮させる方法では同期することはできない。同じ楽譜の演奏であっても、人間による演奏であるために、各音個別にずれが生じるからである。また、モーションキャプチャでは手指に付加したマーカが妨げとなるため、速いテンポの曲をそのままの速度で演奏することは困難である。商業作品である場合は、音質についても重要視されるため、高品質のピアノが用いられ、電子ピアノが使用されることはない。つまり、発音時刻が得られるMidi形式でデータが収録されることはなく、発音時刻や周波数情報を含まない無圧縮のWave形式データとして収録される。これらの理由から、Wave形式の音楽データにおいて音の発音時刻を検出し、発音時刻に合わせて動作データの延長・短縮を行うことにより、発音時刻と整合性のある演奏動作データの生成が必要である。

以上に挙げたように、モーションキャプチャシステムによってピアノ演奏データを収録してCG制作に用いることは実際には容易ではない。以下では、収録されたモーションキャプチャデータをCGとして使用可能なものにするための補正と、演奏動作データと音楽データの同期について述べる。

### 2.1 モーションキャプチャデータの補正

ピアノ演奏の収録のために、マーカは手指と腕につけられる。そのため、腕から指先にかけてのマーカの並びの位置関係は変わらず、またマーカ間の距離はほとんど変わらない。さらに、手指や腕には可動域や可動速度の制約もある。そこで、すべてのマーカ間の関係性と可動域・可動速度を表す手指腕モデルというものを考える。図4は、左手の手指腕モデルにおけるマー

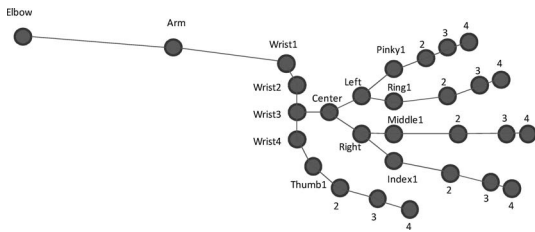


図4 手指腕モデル (マーカ名とマーカ間の関係性)

カ名とマーカ間の関係性のみを表したものである。

収録されたモーションキャプチャデータの補正は、フレームの集合と各フレーム内の点集合が与えられたとき、点集合が手指腕モデル上にあるという制約を満たし、マーカの移動量が小さくなるように、欠落した点集合を付加し、点集合の各点にマーカ名の対応を決定する最適化問題として扱うことができる。

この最適化問題を直接解いて最適解を求めることも可能であるが、シンプルなヒューリスティックアルゴリズムによっても高速に実用的な解を求めることができる。基本的な考え方としては、あるフレームまで点とマーカの対応が特定されているとき、次のフレームにおける各点に対応するマーカを、それ以前のマーカの動きから推定することを繰り返すというものである。各マーカごとに、推定対象フレームの直前の数フレームにおける動きの外挿によって得られる位置を中心とした小領域を当該マーカのとりうる領域とし、その領域内にある点を対応するマーカとして特定する、もしくはそのような点がなければ領域内に点を挿入してマーカを対応させる。動きの外挿においては、可動域や可動速度、鍵盤との距離などを考慮することによって、推定精度を高める。

このようにして各点に対応するマーカ名を特定する処理を行うことによって得られたモーションキャプチャデータは、人手によって正しく補正したものに比べて、94%から100%の精度であった。そのため、ほぼそのまま用いることができる。なお、シンプルなアルゴリズムであるため、実行時間は十分短い。

これにより、ピアニストによる演奏をモーションキャプチャによって収録したのち、上記アルゴリズムによって補正し、演奏動作データとしてすぐに使用可能な状態にすることができた。この演奏動作データに、アニメのキャラクタの手のモデルとの対応関係をセットすることによって、そのキャラクタの手による演奏シーンが短時間で生成できるようになった。図5に、実際に「のだめカンタービレ 巴里編」の第10話で使用さ



(C) ニノ宮知子・講談社/のだめカンタービレ2製作委員会

図5 のだめカンタービレ 巴里編 (第10話) のシーンから

れた例を挙げる。

このようにして、これまで現実的には不可能であった、CGによるリアルなピアノ演奏シーンの実現に成功した。

## 2.2 演奏動作と音楽の同期

演奏動作と音楽を同期させるためには、Wave形式の音楽データにおいて音の発音時刻を検出し、発音時刻に合わせて演奏動作の延長・短縮を行うことにより、発音時刻と整合性のある演奏動作データを生成する必要がある。

Wave形式のデータには、時刻ごとにすべての周波数が合成された値が格納されている。フーリエ変換によって周波数分解を行ったとしても、音には倍音と残音があり複数の周波数が重なっている。倍音とは、ある発音に含まれる、発音の周波数の整数倍の周波数をもつ音を言う。残音とは、発音後に響くことによって残る音を言う。ある時刻において複数の周波数を検出したとき、それが倍音か残音か区別することは難しい。実際、その時刻以前での発音の残音の信号強度が弱まったものか、その時刻で発音された弱い音そのものか、その時刻での発音の倍音かは、各時刻の周波数の値だけでは区別できない。単純な方法として、各時刻で最も信号強度の強い音をその時刻の発音とするものも考えられるが、強く弾かれた残音が、その直後の弱く弾かれた音よりも信号強度が強いことは頻繁にあるため、この方法は有効ではない。これらの理由により、Wave形式データを周波数分解して得られる、各時刻の周波数と信号強度の情報であるスペクトログラムを用いれば発音時刻の検出が容易にできるというわけではない。

音楽データの発音時刻を検出する問題について述べる。入力として Wave 形式の音楽データと、楽譜データがある。ここでの楽譜データとは、総音数・各音の周波数・同時音数（一つの和音に含まれる音の数）を保持しているものとする。なお、発音時刻検出のためには、発音される順序さえ正しければよいので、音の長さの情報は必要ではない。楽譜の始まりから発音する順序で音符番号  $i$  を割り当てる。このとき、楽譜において同時に発音すべき音は一組とし、音符番号  $i$  は同一とする。そして音符番号  $i$  の最大値を総音数  $n$  とする。音符番号  $i$  において同時に発音すべき音の数を同時音数  $H_i$  とする。音符番号  $i$  での各音符が示す音階の周波数の集合を同時音集合  $F_i = \{f_0, f_1, \dots, f_{H_i}\}$  とする。これらの情報からなるものを楽譜データとする。

楽譜データにおける音符番号  $i$  の発音時刻  $t_i$  とは、Wave 形式の音楽データにおいて、基準となる時刻からの経過時間とする。

発音時刻検出問題とは、Wave 形式の音楽データと楽譜データを入力として、楽譜データの各音符番号  $i$  の発音時刻を決定する問題である。ここでは、これを以下のように扱う。まず、Wave データからフーリエ変換によってスペクトログラムを作成する。スペクトログラムにおいて時刻  $t$  において検出された周波数の集合を  $A(t)$  とする。 $A(t)$  のうち、音符番号  $i$  の同時音集合  $F_i$  すべての音に相当する周波数の信号強度を足し合わせた値を  $\widetilde{S}_i(t)$  とする。すべての音符番号についてこの値を求める。ただし、楽器の音は、電子楽器の場合を除いて音階が示す周波数と完全に一致するとは限らないため、誤差として一定の幅を許容するものとする。 $F_i$  に含まれるすべての音について、各音の  $k$  倍音 ( $k = 1, 2, \dots$ ) の周波数が  $A(t)$  に含まれているならば、それらの信号強度を  $k$  で割った値を  $\widetilde{S}_i(t)$  に加える。このようにして得られた値を  $S_i(t)$  とする。

次に、同時音集合  $F_i$  の時刻ごとの発音可能性について述べる。 $w_i(t)$  を、 $\max\{dS_i(t)/dt, 0\}$  と定義する。信号強度が急増している時刻で発音している可能性が高いため、 $w_i(t)$  が大きければ発音可能性が高い。逆に信号強度が減少しているならば、発音時刻後の残音部分である可能性が高く、その場合は  $w_i(t)$  が 0 となっている。さらに、ノイズについても考慮する。ここでのノイズとは、注目している音のほかの音の発音可能性を意味する。これは、発音時刻を検出するにあたり、ほかの音が発音している可能性が高い時刻には注目している音が発音している可能性が低くなることを考慮するためである。ノイズ  $N_i(t)$  は、 $S_i(t)$  の算出に使用した周波

数以外のすべての周波数の信号強度の総和に係数をかけたものとする。同時音集合  $F_i$  の時刻ごとの発音可能性を表す重みを、 $W_i(t) = w_i(t) - N_i(t)$  と定義する。このようにして、楽譜データに含まれるすべての同時音集合  $\{F_i | i = 1, 2, \dots, n\}$  に対して、それぞれの各時刻での発音可能性を表す重み集合  $\{W_i(t) | i = 1, 2, \dots, n\}$  と、ノイズ集合  $\{N_i(t) | i = 1, 2, \dots, n\}$  が得られる。重みが大きくノイズが小さければ、そこで発音している可能性が高い。そこで、楽譜データにおける音符番号の順序で、重みの総和が最大となる発音時刻系列を求めるというアプローチを採った。時間を離散化して動的計画法に基づいたアルゴリズムによって、Wave 形式データから発音時刻を決定することができる。

演奏動作データにおける打鍵時刻の検出は、モーションキャプチャ収録の際、鍵盤にマークを付加し、マークが一定距離以上、一定時間以上沈むところを打鍵時刻とすることでできる。

モーションキャプチャによって得られた演奏動作データを打鍵時刻で区切った区間に分ける。対応する音の発音時刻の区間が打鍵時刻の区間より長ければ、等間隔にフレームを挿入し、フレーム間を線形補間によって補間する。逆に短かければ、等間隔にフレームを抜き取る。このようにして発音時刻と打鍵時刻が同期し、Wave 形式の演奏と、別にモーションキャプチャによって収録された演奏動作を同期させることができる。

上記の方法による同期方法について、実際の演奏データに基づいて性能評価を行ったところ、90%以上の割合で、人間が検知できない誤差範囲で発音時刻を検出した。これは、きわめて高い精度で同期ができることを意味する。

### 3. 楽譜データからの CG 自動生成

次に、楽譜データからのピアノ演奏 CG 自動生成技術について述べる。これは、楽譜を読み込み、運指を決定し、それに基づいて手指腕の骨格を表すボーンモデルの CG を生成するものである (図 6)。

#### 3.1 運指決定

適切な運指は自明に定まるものではない。実際、ピアノの学習の初心者にとって運指は自明ではないため、多くの教則本には運指指示も同時に記載されている。運指の決定については、通常は演奏者が経験的に、もしくは試行錯誤することで決定するものであるが、人間の介在を排除するためには、運指の自動決定法が必要である。

運指決定に関してはこれまでも研究はあり、ピアノ



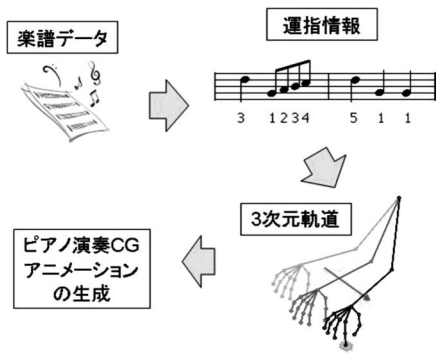


図6 楽譜データからのCG自動生成

演奏では隠れマルコフモデルに基づいて確率論的手法を用いて運指を決定するもの [2], 鍵盤楽器ではなく管楽器であるフルートの運指 [3] や、弦楽器であるギターの運指 [4] に関する研究などが挙げられる。ここでは、実際のシステム開発を優先し、短期間で小さい開発稼働で実用的な性能を実現するという観点から検討した。

一般的なピアノ演奏者は、弾きにくさが最小、つまりもっとも弾きやすい運指を選択するとみなした。指と鍵盤（音番号）の組を考え、それらの間での手指の移動しにくさ（弾きにくさ）をコストとする。つまり、音番号  $p$  の鍵盤を指  $q$  で弾いている状態から、音番号  $p'$  の鍵盤を指  $q'$  で弾いている状態への遷移に対して、コスト  $c(p, p'; q, q')$  が与えられているものとする。楽譜データによって音番号の系列  $s_1, s_2, \dots, s_n$  が与えられたとき、各音番号を弾く指  $f_1, f_2, \dots, f_n$  が、 $\sum_{i=1}^{n-1} c(s_i, s_{i+1}; f_i, f_{i+1})$  が最小となるように決定することを、最適運指決定と定義した。これは手指への負担が小さくなるような運指が実際にも選ばれやすいという経験に基づくものである。この最適化問題は、コストが与えられれば、音の数  $n$  に関する線形オーダの計算量で容易に求めることができる。

さまざまな曲での評価実験の結果、このようなシンプルな方法であっても、弾きやすさという観点からはどのような曲であっても妥当な運指を決定することができることがわかった。ただし、初心者向けの曲や比較的簡単な曲であれば、一般的に用いられる運指とほぼ同じものを出力できるが、中級・上級レベルの曲になると異なってくる。これは、高いレベルの曲では、単純な弾きやすさ以外の要因が関係していることを示唆している。

上記の方法では、指と鍵盤の組と、その組の間の遷移すべてについてコストの値を与えなければならない

ため、膨大な組合せとなり、コストの決定は容易ではない。これに対して、逆問題を解くというアプローチ [3] もあるが、ここでは、筋肉負荷の観点からコスト決定法を検討した。

コストは指と鍵盤の組から組への遷移のしにくさを表すものであるため、演奏者の腕の筋肉負荷と関係しており、筋肉負荷が小さいほど弾きやすいと考えられる。これまでに報告されているピアノ打鍵時の手、指および上肢や体幹のバイオメカニクスの研究では、熟練者と初心者の打鍵時における力発揮の違い [5] などが調べられているが、運指と関連筋活動の関係を調べた研究はないため、まずプロのピアニストの筋活動量を表面筋電位法によって測定することにより、筋活動量と弾きやすさの間の相関を調べることにした。

浅指屈筋は、上腕骨の内側上顆、尺骨と橈骨の上端部の前面から起こる筋で、手根部では4本の腱に分かれて、第2~5指の中節骨の底につき、手首を曲げ、第2~5指の中節を曲げる作用がある。総指伸筋は、上腕骨の外側上顆から起こり、4腱に分かれて第2~5指の指背を長く走ったあと、末節骨につき、第2~5指を伸ばし、また手首も伸ばす作用がある [6]。打鍵動作では、まず伸筋の活動によって指が持ち上がる。次に下降運動が始まるのとはほぼ同時に伸筋の活動が止まり、代わりに屈筋が活動する。鍵盤の下方への移動が停止するとほぼ同時に2回目の伸筋活動が現れ、再度指が上昇し始める。そこで、ピアノ打鍵動作時の浅指屈筋、総指伸筋の活動を表面筋電位法により計測した (図7)。

複数の熟練ピアニストを対象に、2段階の音量（大きい・小さい）と2段階のテンポ（速い・ゆっくり）、そして指と鍵盤の組と組の間のさまざまな遷移の組合せについて筋電位を計測した。その結果、弾きやすさと筋電位の間には強い相関があることがわかった。したがって、筋電位の大小関係をコストの大小関係に反映させることによって、運指決定問題におけるコストを決定することができる。また、同じ遷移であっても、音量とテンポの組合せによって筋電位が異なることも

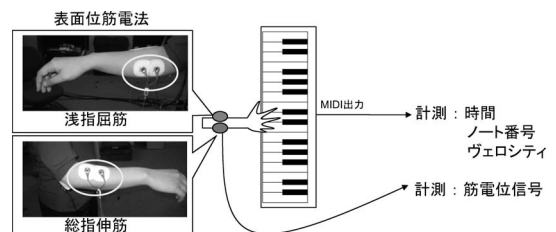


図7 筋電位測定

わかった。これは、筋電位測定結果から作ったコストを利用することで、曲調によって異なる運指を生成できることを意味する。このようにして得られたコストを用いて前記の運指決定法により運指を決定すると、ピアノ演奏者にとって自然な運指、もしくは代替可能な運指を得ることができた。

### 3.2 手指腕動作生成

運指に基づいて手指の骨格の動きを決定する方法について述べる。人間の手指は関節と筋が複雑に組み合わさった構造をしており、ロボットアームの制御に用いるような単純なモデルでは表現できない。ここでは、手指腕を表す構造として、棒状の剛体を関節で接続し、関節における回転には筋肉の作用を反映した負荷がかかるとしたモデル（ボーンモデル）を構成した。

指が鍵盤を打鍵する瞬間をキーフレームとし、まずキーフレームにおける手指腕の形状を決定する。基本的な考え方としては、直前のキーフレームからの変化が小さく、かつボーンモデルにおける負荷が小さくなるような形状を決定する問題を最適化問題として定式化し、準ニュートン法に基づいたアルゴリズムにより、手指腕の形状を決定することにした。

次に、各関節点についてキーフレーム間の軌道を決定する。単純に、キーフレーム間の形状変化を線形補間することで比較的自然的な滑らかな動きが生成できたが、力学的・解剖学的な制約を考慮していないため、物理法則にそぐわない動きが生じ、違和感を抱かせてしまう箇所も出てきてしまう。そこで、手指腕の負荷が小さくなるように、かつ物理法則に反しないように軌道を決定するようにした。そのために、鍵盤を指定した強さで弾くことを、重力加速度も考慮したうえでその鍵盤へ指定した速度で到達するようにボーンモデルを動かすことととらえ、その際のボーンモデルにおける負荷の総和が最小となるようにボーンモデルの動きを決定する最適化問題として定式化した。この最適化問題は多峰性関数であり、一般的に解くことが困難であるため、可能な動きを限定することと、ヒューリスティックな初期解から始めてランダム多スタート局所探索によって解を探索することにした。その結果、このアルゴリズムによって、自然なピアノ演奏動作 CG が実現できるようになった。なお、鍵盤への指の到達速度を考慮できるため、演奏の強弱を反映できる。モーションキャプチャを利用した CG 制作では、演奏の強弱を変更する場合は、再度演奏しなおさなければならないが、この方法では速度指定を変更するだけで容易に新たな CG を制作できる。

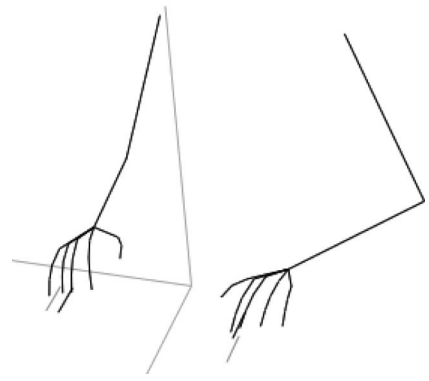


図8 CGアニメーションの1シーン

図8に、制作したCGアニメーションの1シーンを例として挙げる。

## 4. まとめ

本稿では、特にピアノ演奏CG制作のために必要な技術課題とその解決例を紹介した。ここで述べた技術はすべて、実システムへの適用が目的であったが、最適化問題としての定式化と効率的なアルゴリズムの設計というアプローチがたいへん有効であった。実際、ここで述べた技術の一部は、テレビアニメ「のだめカンタービレ」シリーズにも用いられ、大きな成功を収めた本作品に貢献することができた。CGや音楽といった娯楽に関係する分野においても、オペレーションズリサーチのアプローチの有効性が示されたと言えよう。

## 謝辞

ピアノ演奏CGの研究開発に多大な協力をいただいた、釘本望美様、藤村武史様、平田純也様、そしてピアノニストとしてご多忙のところ、快くご協力いただきました古屋絵理様に感謝します。また、ピアノ演奏CGの制作にかかわる機会をいただきました、株式会社トリロジー・フューチャー・スタジオ代表取締役社長の平正昭様に感謝いたします。

## 参考文献

- [1] N. Kugimoto, K. Takai, R. Miyazono, K. Omori, T. Fujimura, S. Furuya, H. Katayose, H. Miwa and N. Nagata, "CG Animation for Piano Performance," ACM SIGGRAPH 2009, Aug. 3-7, 2009.
- [2] 米林, 亀岡, 嵯峨山, "手の自然な動きを考慮した隠れ変数付き隠れマルコフモデルに基づくピアノ運指決定," 情報処理学会研究報告, 音楽情報科学, Vol. 2007, **81**, 179-184, 2007.
- [3] 澤井, 黒木, 松井, "フルートの運指最適化と逆最適化を用いたパラメータチューニング," オペレーションズリサーチ学会誌, **53**(1), 39-46, 2008.

- [4] A. Radisavljevic, P. Driessen, "Path Difference Learning for Guitar Fingering Problem," *Proc. of the International Computer Music Conference (ICMC)*, 2004.
- [5] S. Furuya, R. Osu and H. Kinoshita, "Effective Utilization of Gravity during Arm Downswing in Keystroke by Expert Pianists," *Neuroscience*, 822–831, 2009.
- [6] 加藤征, 福島統, 国府田稔, 「新解剖学」, 日本医事新報社, 2002.