音楽情報科学 36－2
（2000．8．5）

音響信号からのメロディ検索と採譜システム

半田 伊吹, 武藤 誠, 坂井 修一, 田中 英彦

東京大学大学院工学系研究科
〒113-8656 東京都文京区本郷 7-3-1

{handa,muto,sakai,tanaka}@mtl.t.u-tokyo.ac.jp

あらまし:

従来から提案されている計算機による採譜システムは、処理全般を計算機に委ねるものが主流であった。しかし、そのようなシステムでは人間が容易に知り得るような情報も計算機では認識が困難である場合もあり、精度の高い採譜は実現しがたかった。

筆者らは認識率のより高いシステムを目指し、人と計算機がお互いに得意とする作業を分担し、協調して情報を補完しあう採譜システムを提案する。

また、そのシステムの実現にあたって有効と思われる、音響信号から旋律を探索する手法を実装し、予備的な実験と考察を行う。

キーワード: 採譜、音楽検索、マン・マシンシステム

Music transcription system with music retrieval method

HANDA Ibuki, MUTO Makoto,
SAKAI Shuichi and TANAKA Hidehiko

The University of Tokyo
7-3-1 Bunkyo-ku, Tokyo, 113-8656

{handa,muto,sakai,tanaka}@mtl.t.u-tokyo.ac.jp

ABSTRACT:

We think that complete transcription is difficult for a music transcription system wchich depends on only computational processes. So, we propose a man-machine system so that quality of transcription may inprove. The system contains man-machine interface, and human and machine co-operates in music transcription. We discuss abstract of the system and introduce music retrieval method which we regard as useful for the system.

KEY WORDS:

music transcription, music retrieval, man-machine system

1. はじめに

計算機による自動採譜処理に関する研究は、最初はその対象を単一音源による単旋律に限っていたが、次第に対象を広げ、単一音源の複数音を扱うようになった。更に複数音源による複数旋律の演奏を対象とした自動採譜処理が試みられるようになったが、対象が現実的な演奏に近付くに従って、それまでの単純な処理では十分な認識精度が得られないことが指摘されるようになった。認識精度の問題は、計算機の計算速度や記憶装置などの能力が向上すれば解決するというものではなく、自動採譜に有効な処理のアルゴリズム自体が解明されていないのである。

精度のよい自動採譜を実現するということは計算機科学や人工知能の研究としては大変意義のある大きなテーマであるが、一方採譜システムの応用範囲の広さを考えると、処理自体の仕組みには興味がないがすぐに利用したいという需要も大きく、精度の高い採譜システムの完成は急務であるとも言える。

筆者らは、完全に計算機によって採譜をするのではなく、計算機が処理を行う際に活用できる重要な情報であるが、計算機自身では抽出が困難なものをマン・マシンインターフェースを用いて人間が入力することによって、計算機の苦手な部分を補完する採譜システムを提案している[1]。

本稿では、そのような採譜システムの実現への適用を検討している、音響信号からの旋律探索手法の概要と予備的考察、およびその採譜システムへの適用の仕方について述べる。

2. 計算機と人間の協調による採譜

採譜処理というのは、結果から原因を推定するいわゆる逆問題の一つであるが、多くの逆問題と同様、容易に原因をつきとめることはできない。複数の音が同時に発音された場合には、同時にいくつの音が発せられたかすら計算機で判断するのは困難になる場合が多いのである。

このことは人間にも当てはまり、音だけを聞いて演奏に使われている楽器を全て挙げることを確実に誤りなくできるわけではない。しかし、その能力の程度は計算機に比して大変長けており、ある瞬間に発音されている楽器数をかなりの確度で言い当てることもできるし、ある特定の楽器が演奏に使われ出す時刻や使用が終る時刻も容易に分かる。このように人間は高い抽象度で音楽を理解する能力を持ち合わせている。このように高い能力を人間一般が有しているにもかかわらず、特別な訓練を受けた人以外が採譜を行おうとすると、なかなか思うようにいかない。うまくいかない理由は、先に挙げたようにどの楽器が使われているか分からない場合もないわけではないが、それ以上に着目したフレーズの進行の具合、つまり何度上行あるいは下行したのかを定量的に測ることができないからである。

音楽的な訓練を受けていない人が採譜を行うときのプロセスを考えてみる。音響信号を聞いた人は、図1に示すようにそこに含まれるフレーズを頭の中で「なんとなく」切り出すことができる。

それを構成する単音列において、ある単音とその次の単音の相対的な関係について上行、下行、平行といった定性的なことは分かるが、一方何度上行あるいは下行なのかという定量的なことが分からない。そこで、ピアノを弾きながら、あるいはDTM環境を整えた計算機上の打ち込みソフトでマウスを操作しながら、音高を知りたい単音の高さを探っていく。

このような状況を想定して、図2に示すような進行の度合を定量的に測ることができない人が採譜を行うのを計算機が手助けすることを本システムでは目指している。

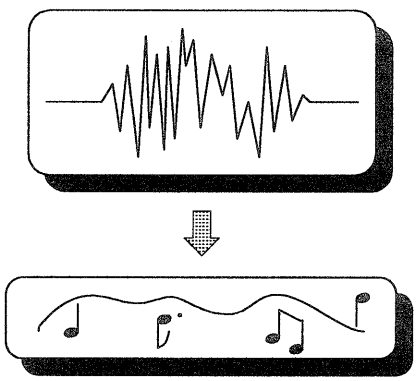

図 1. 人間による音楽の聴取

3. 音響信号からのメロディ検索

〈3·1〉 演奏内容からの検索 近年のマルチメディア技術の発展に伴い、文書だけでなく画像データや音声、音楽データなどもデータベースとして利用できるようになってきた。音声、音楽検索の手法は、書誌情報による検索と内容検索の2つに大きく分類で

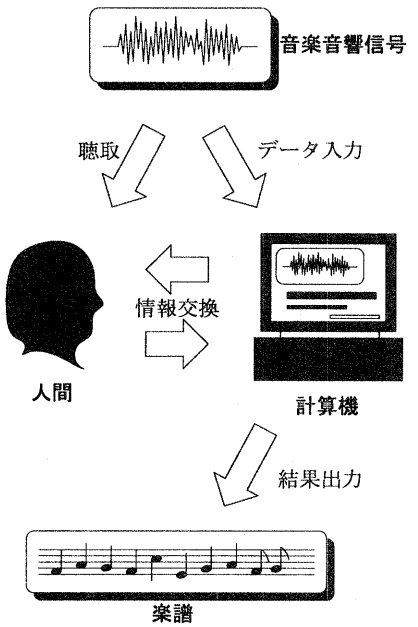

図 2. 採譜システムの概要

きる。

音楽データに対する書誌情報による検索は、曲名や作曲者、歌唱者など、楽曲の持つ書誌的な情報をキーにして検索するもので、技術的には図書のデータベースなどの検索と同等である。

一方、内容検索は音楽の演奏内容を検索のキーとするものである。この場合、書誌情報が分からない場合でも検索できたり、多様な検索が可能となるなどの期待も大きい。

音響信号の内容を検索する研究として、柏野らの研究[2]があるが、これは時間的に長い音響信号の中に、それより短いキーとほとんど同じ信号が含まれるかどうか、含まれる場合はどこかを探索する技術である。音楽音響信号を対象とした場合、キーには頭に浮かんだ旋律だけではなく伴奏も忠実に含ませなければならない。また、キーも音響信号である必要があるので、鼻歌などを入力とする場合には音響信号を合成する必要もある。このため、この技術を音楽の内容検索に適用するのは困難である。

これに対し、柳瀬らの研究[3]では、予めパートごとに情報が区分されており、かつ演奏情報を高度に捨象してあるMIDIデータを用いて内容検索を試みている。この方法によると、大規模なデータベースから、1つのパートの部分的な旋律をキーにして検索することができる。

しかし、この方法では音響信号に対応したMIDIデータのように符合化されたデータを持たせなくてはならず、自動採譜が実用的な段階に達していない現在では手作業で行うことになる。

本研究では、図3のように音響信号そのものからメロディをキーとして検索する手法を提案し、それを逆に採譜システムに応用することを考えている。

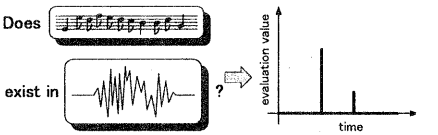

図 3. 音響信号からのメロディ探索

⟨3·2⟩ 時間–周波数解析 音響信号からのメロディ検索において、後の処理を容易にするために時間–周波数解析を行う。解析の方法はFFTやフィルタバンクを用いることが多いが、FFTでは周波数軸上のサンプル点が線形に並ぶため、音楽音響信号の解析には不向きである。そこで、中心周波数を等比数列に並べたフィルタバンク[4]を用いることにする。この手法によって解析されたものは、図?? に示すような、パワが時間と周波数の関数として表される。

⟨3·3⟩ スペクトログラムからのメロディ検出 従来の採譜システムでは、スペクトログラムから鳴っていると思われる音を抽出するというアプローチをとっている場合が多いが、ここでは逆に、鳴っている音を仮定してその音が鳴っている確度を調べるというアプローチを採る。

存在を仮定した音が鳴っているということの確度として、SN比の適用を考える。SN比を論じるときには通常、どれが信号でどれが雑音であるということが明確にされているが、ここでは存在仮定のたてかたによって、信号とみなされる成分と雑音とみなされる成分が変わってくる。あくまでも便宜的なものと考えて頂きたい。

さて、ある単音を人間が聴取しようとするとき、その音の認識を妨げるのはその単音の周波数成分と重なるような周波数成分を持つ音である。つまり、多くの楽器のがそうであるように周波数成分が基音とその倍音で構成されているとするとき、目的の音の周波数成

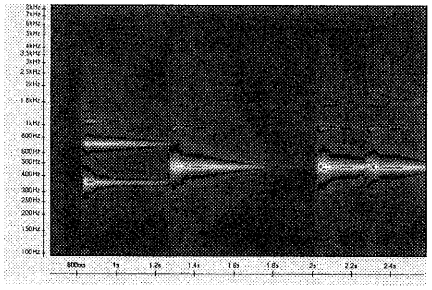

図 4. スペクトログラム

分 (基音や倍音) の整数分の 1 の周波数の成分があると邪魔になるのである。

ここで、特殊な場合を除いて倍数がある程度大きくなると倍音のパワは小さくなることを勘案して、信号を構成するのは基音に対して 6 倍までの正整数倍の周波数と仮定すると、雑音となる周波数成分は表 1の下線付きの周波数となる。表中にあるにもかかわらず下線が引かれていないものは、ある存在を仮定した音の基音ないし倍音と同じ周波数になっている成分であり、信号でもあり雑音でもあるといった矛盾を生じさせないために信号としてのみ扱うものである。

表 1. ノイズとみなす周波数

着目する周波数	周波数に乗ずる係数				
	1/2	1/3	1/4	1/5	1/6
f_0	$\underline{\frac{1}{2}f_0}$	$\underline{\frac{1}{3}f_0}$	$\underline{\frac{1}{4}f_0}$	$\underline{\frac{1}{5}f_0}$	$\underline{\frac{1}{6}f_0}$
$2f_0$	f_0	$\underline{\frac{2}{3}f_0}$	$\underline{\frac{1}{2}f_0}$	$\underline{\frac{2}{5}f_0}$	$\underline{\frac{1}{3}f_0}$
$3f_0$	$\underline{\frac{3}{2}f_0}$	f_0	$\underline{\frac{3}{4}f_0}$	$\underline{\frac{3}{5}f_0}$	$\underline{\frac{1}{2}f_0}$
$4f_0$	$2f_0$	$\underline{\frac{4}{3}f_0}$	f_0	$\underline{\frac{4}{5}f_0}$	$\underline{\frac{2}{3}f_0}$
$5f_0$	$\underline{\frac{5}{2}f_0}$	$\underline{\frac{5}{3}f_0}$	$\underline{\frac{5}{4}f_0}$	f_0	$\underline{\frac{5}{6}f_0}$
$6f_0$	$3f_0$	$2f_0$	$\underline{\frac{3}{2}f_0}$	$\underline{\frac{6}{5}f_0}$	f_0

これによると、基音 f_0に対して信号を構成する周波数 F_sは

$$F_s = \{f_0, 2f_0, 3f_0, 4f_0, 5f_0, 6f_0\} \quad \cdots\cdot(1)$$

となり、雑音を構成する周波数は F_nは

$$F_n = \left\{\frac{1}{6}f_0, \frac{1}{5}f_0, \cdots, \frac{5}{2}f_0\right\} \quad \cdots\cdots\cdot(2)$$

である。

ここでは単音 Nを代表する値として、音高と継続時間だけを用いることにする。

$$N = (h, d) \quad \cdots\cdots\cdots\cdots\cdots\cdot(3)$$

ただし、hは MIDI ノート番号のような音高を表す値、dはその音の継続時間である。

時刻 t、周波数 fのパワを $P(t, f)$ し、単音が開始されたと仮定する時刻を t_0とすると、信号相当周波数のパワの和は

$$P_s = \sum_{i=0}^{d-1} \sum_{j \in F_s} P(i + t_0, j) \quad \cdots\cdots\cdots\cdot(4)$$

となり、雑音相当周波数のパワーの和は

$$P_n = \sum_{i=0}^{d-1} \sum_{j \in F_n} P(i + t_0, j) \quad \cdots\cdots\cdot(5)$$

となる。この 2 つの量から信号とみなされる成分と雑音とみなされる成分の比をとりたいが、一般的な SN 比は値域が有界でないので、値域が $[0, 1]$ となる指標を

$$\alpha = \begin{cases} \frac{2}{\pi} \arctan\left(\frac{P_s}{P_n}\right) & (P_N > 0) \\ 0 & (P_S = P_N = 0) \\ 1 & (P_S > 0, P_N = 0) \end{cases} \quad (6)$$

のように定義すると、αは存在を仮定した単音が鳴っている可能性に応じて 0 から 1 の値をとる。大きいほど可能性が高い。

ここまでは単音の存在の評価の仕方の話であるが、メロディの探索にあたっては、それを構成する単音全てについて、時間的な連続性を考慮にいれつつαを求め、積をとることにする。つまり、旋律が

$$F = (N_1, N_2, \cdots, N_m) \quad \cdots\cdots\cdots\cdot(7)$$

のように m 個の単音列で表されているならば、i 番目の単音に対してのαを求め、

$$\beta = \prod_{i=1}^{m} \alpha_i \quad \cdots\cdots\cdots\cdots\cdots\cdot(8)$$

とする。

この評価量βによって、単音ではなくてメロディが対象とした音響信号に含まれているかどうかが判断できる。

MIDI 楽器のオーボエ、ピアノ、ベース、ドラムの音色によって演奏された 1 分程度の演奏に対して、それに含まれる 2 小節強のメロディを検索した結果を図 5に示す。図中の横軸はスペクトログラムの時間分解能を基準にしており、ここでは 1024/44100 秒である。また、縦軸は式 (8) で表される評価量である。図には急峻なピークが見出せるが、これは検索をかけた旋律と全く同じ部分の開始時刻である。うまく検索ができたことを示している。

同一楽曲で演奏する楽器を変えた場合も幾つか実験を行なったが、ピークの高さにはばらつきがでるものの、正確に検索を行えている。また、同一楽曲で検索キーを含むパートを除いたものや楽曲自体全く別のものにして検索をかけたところ、ピークは現れず、誤検出する恐れも少ないことが分かった。

ここに示したのは予備実験であるので、値がいくつ以上のときに検出されたとするかの閾値の設定については現在検討中である。

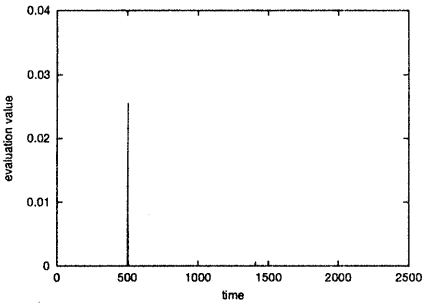

図 5. 正しい旋律を含む楽曲に対する探索結果

〈3·4〉 検索手法の採譜への応用 音響信号から旋律を探索する手法の概要とその予備的な実験結果を示したが、ここではこの手法を採譜に応用することを考える。

計算機上で存在すると思われる旋律を複数生成し、それぞれの存在可能性の評価量 (8) を求め、最も値の大きいものを正解とするというアプローチをとる。この方法が有効であるというためには、正解の旋律とそれ以外の旋律とで評価量に著しく差が生じなくてはならない。

先に述べた実験結果では、全く異なる旋律に対してはピークが現れないことが確かめられた。しかし、かなり似ているが少し異なる、というような旋律に対しても評価量が正解に対してのそれより小さくなることが望まれる。

このことを確かめるために、旋律を構成する単音のうち1つだけを本来の音高からずらしたものを手動で 80 種類用意し、評価を行った。ピークが検出される時刻での評価量の分布は図6のようになった。横軸に意味はなく、80 種類の旋律が与える評価量が分かりやすいように横方向に散りばめただけである。線で示されているのは、正解をキーとして探索したときのピークでの評価量である。

この結果から窺えることは、旋律中の1音を変えただけでも評価量が大分小さくなることが多いが、場合によってはあまり変わらない、あるいは大きくなってしまうこともあるということである。採譜システムへの適用をする場合には、正解以外をキーとしたときは必ず評価量が正解のときより小さくなる必要があるので、このままの適用はできない。

評価量が小さくならないような旋律について大雑把であるが傾向を調べたところ、以下の2点が言えそうである。

- 本来の音からのずれが完全8度であるような場合
- 本来の音からずらした結果、他のパートが奏でる音と合致してしまった場合

これらは確かに納得のゆくことである。このような場合の誤検出を回避するために、音色の情報を検出評価に適用することを検討している。オクターブの誤りや他のパートの演奏を誤認してしまう場合、その音の音色は他の音の音色と著しく異なり、誤りであることを知ることができると考えている。

当然周波数成分の重なりが生じ、それぞれの単音の音色の持つ特徴量がそのままの形で保たれているとは限らないが、木下らの研究[5]に周波数成分の重なり適応処理を用いた音源同定処理に関しての考察があり、応用を検討している。

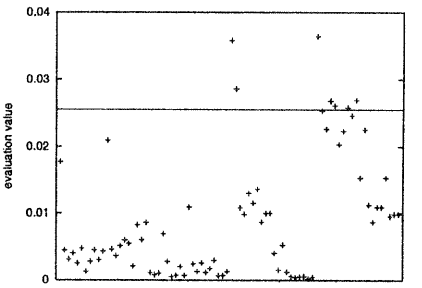

図 6. 正解とは異なるキーに対する評価量

このような音響信号からの旋律探索手法を確立した後、採譜システムへの応用を行おうとしているが、その方法は以下の通りである。まず、計算機が大量の旋律候補を生成し、それぞれの存在可能性の評価量を求め最大のものを正解とする。但し、例え時間的に短くても旋律の刻むリズムは何通りもあるし、リズムが決定されても進行の仕方が自由なので、全て生成するのは不可能である。そこで、生成する旋律の数を減らすために図7に示すように人間がインターフェースを介して入力した情報を活用することを考えている。音楽的な訓練を受けていない人間は、2つの音の音

高の高低を定量的に判断することが苦手であるが、どちらが高いか低いかを定性的に判断する能力は先天的に兼ね備えている。ここで提案するマン・マシン協調システムは、そのような人間の能力を巧みに計算機への助けとし、逆に定量的判断が苦手な人間を計算機が手助けすることにより、精度の高い採譜を行おうとするものである。

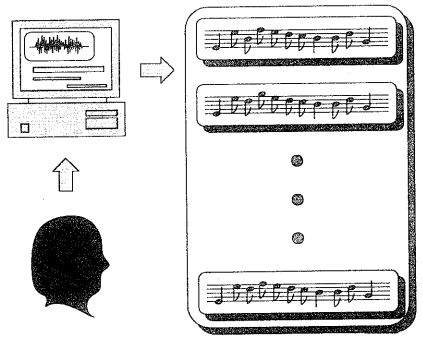

図 7. 人間の与えた情報によるメロディ候補の削減

4. まとめ

本稿では、マン・マシン協調による採譜システムと、それを実現する際に有効と思われる音響信号からの旋律探索手法について、概要を述べた。

人間が計算機に与えられる情報としては本稿ではごく概念的なことしか述べなかったが、今後はインターフェースの詳細な設計をする必要がある。また、旋律探索手法について有効性がある程度見えてきたが、採譜システムへの適用については手法自体に更なる検討が必要である。

なお、本稿では検索を1つの旋律を単位として行っていたが、人間側からの情報入力が容易であれば、和音のような複数音が同時に鳴っているものを一気にまとめて検索することも可能と考えている。例えば、図8のようなCとAmの計算機による聴取を、3音まとめて行えるようにすることを考えている。

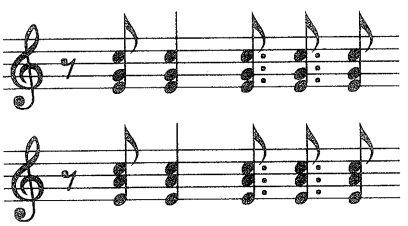

図 8. 和音をまとめての検索

文　献

(1) 半田 伊吹, 木下 智義, 武藤 誠, 坂井 修一, 田中 英彦: 「マン・マシン協調による採譜システム」, 情報処理学会音楽情報科学研究会, 99-MUS-34, pp. 21–26, 2000

(2) 柏野 邦夫, ガビン スミス, 村瀬 洋: 「ヒストグラム特徴を用いた音響信号の高速探索法 — 時系列アクティブ探索法」, 電子情報通信学会論文誌, Vol. J82-D-II, No. 9, pp. 1365–1373, 1999

(3) 柳瀬 隆史, 高須 淳宏, 安達 淳: 「音楽検索における自動インデクシング報」, 情報処理学会研究報告, 98-DBS-116(2)-42, pp. 117–124, 1998

(4) 柏野 邦夫, 中臺一博, 木下 智義, 田中 英彦: 「音楽情景分析の処理モデル OPTIMA における単音の認識」, 電子情報通信学会論文誌, Vol. J79-D-II, No. 11, pp. 1751–1761, 1996

(5) 木下 智義, 坂井 修一, 田中 英彦: 「周波数成分の重なり適応処理を用いた複数楽器の音源同定処理」, 電子情報通信学会論文誌, Vol. J83-D-II, No.4, pp. 1073–1081, 2000