音楽情報科学　39－13
（2001. 2. 23）

採譜支援システムにおける要素技術

半田 伊吹, 武藤 誠, 日比 啓文, 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,hibi,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、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、そのようなシステムでは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している。本稿ではそのようなシステムを実装するにあたってどのような要素技術が必要かについて検証する。

キーワード　採譜、マン・マシンシステム、インターフェース

Required components for man-machine music transcription system

HANDA Ibuki, MUTO Makoto, HIBI Hirofumi,
SAKAI Shuichi and TANAKA Hidehiko

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

{handa,muto,hibi,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 exmamine what kind of components are required for the system.

key words　music transcription, man-machine system, interface

1. はじめに

演奏された楽曲に対して、本来記譜されていないものを楽譜にとることを採譜という。採譜の目的は、編曲してレパートリに入れ利用する場合と、学問的分析のための場合とがある。人間が採譜を行うには、訓練を受け技能を身につける必要がある。

音楽情報科学の扱う分野は広範に及ぶが、一部の技能を持った人間以外には困難な採譜を計算機によって行うというテーマを自動採譜とよぶ。自動採譜の最終出力は必ずしも楽譜そのものである必要はなく、その後の利便性からMIDIデータやSMDL[1]のような電子化された情報の方が都合がよい場合もある。

自動採譜を計算機上で行うことの目的は前述した人間による採譜の目的に加え、演奏内容に基づくデータベースの構築への応用などがある。ジャンルによっては楽譜が存在しない即興演奏が中心となる場合もあり、楽譜からではなく音響信号から記号化された演奏情報を抽出できることの意義は大きい。

また、更に別の目的がある。それは、人間の認知や判断といった頭脳の働きを理解しその機能を計算機で実現すること、つまり人工知能の研究である。応用ではなくその機能の実現そのものにも興味が持たれている。

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。そこで筆者らは、完全に計算機によって採譜をするのではなく、マン・マシンインターフェースを用いて人間も聴覚的、視覚的に得た情報を計算機に入力することによって、計算機の苦手な部分を補完する協調型の採譜システムを提案している[2]。

本稿ではまず第2.章で従来からの自動採譜システムの認識率向上の難しさを述べ、次に第3.章で認識率を高められると思われる協調型採譜システムについて述べる。そして第4.章でそのようなシステムが必要とする最小限の機能についての提案をし、最後に5.章でまとめる。

2. 自動採譜システム

従来行われている自動採譜の研究は、システム全体の構築を目指すものと、ある特定の問題を解決しようとするものとの二通りがある。

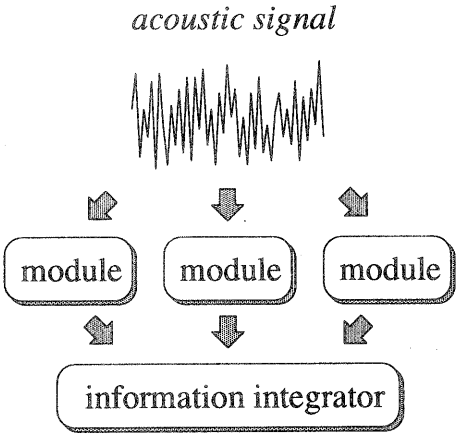

図1. 汎用的な利用を目指した採譜システムの構造

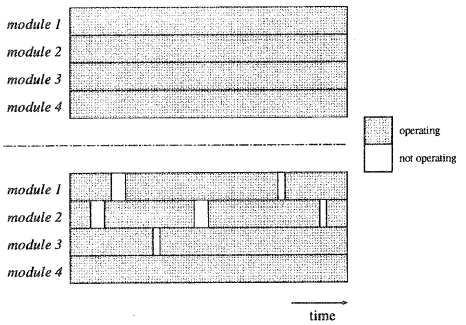

図2. モジュールやエージェントの稼働率

例えば柏野らの研究[3,4]では、複数の抽象度の異なるモジュールによる解析結果を統合して確率的に最尤な最終結果を得るOPTIMAを実装している。また、自動採譜ではなく更に認知の対象を拡大した聴覚的情景分析の研究として中谷らによる残差駆動型アーキテクチャ[5]がある。これも、それぞれが異なった機能を持った複数のエージェントとその統合方法について研究がなされている。これらの全体像を簡単に図示したものが図1である。一方、杉浦らによる文献[6]のように、システム全体の提案・実装ではなくて、自動採譜を行おうと時間周波数解析を行いうなりがあったときという特定の状況から不協和音程の検出を行うのに必要な音楽音響信号の処理方法についての研究例がある。

前者のようなシステムの場合、それぞれのモジュールないしエージェントに持たせる機能はな

るべく汎用的であるように設計される。それは、極めて特殊な機能をもったモジュールやエージェントを用意しても、それをいつどのように作動させそこから得た結果をどう活用するかといった、全体を統合する段階での繁雑さを避けるためとも考えられるし、対象（ここでは音響信号）のなかから本質的なものだけを抜き出して捨象するという科学の大前提にのっとった方法ともとれる。このような思想で設計されたシステムでは、図2のように各モジュールないしエージェントは常に稼働しているか、かなり高い割合で稼働している。この理由は先に述べた通り、なるべく一般性が高い機能を持つように設計されているからである。

このように一般性を求めた場合、楽曲の演奏形態を問わず適用できるので全体としての適合率や再現率の平均は良い結果を得られることには間違いないのであるが、個々の曲に対しての認識率に満足のゆくものを得るのは非常に困難である。ある特定の演奏の仕方の曲に対しては一つの楽音の採りこぼしもなく誤認識もなく完全に採譜が可能な場合もあるであろうが、それは偶々である。一方、一般性を追求せずに、楽曲ごとの特徴、更に楽曲の小節ごとの特徴に着目して情報抽出を試みれば、適用範囲は著しく狭まるものの、その楽曲に対しては良質なシステムになる。著しく少数の特定の曲にしか適用できないのであれば学術的研究にはならないが、比較的よく現れる特徴というものをリストアップしてそれらに対処する処理系の集合体を確立すれば、実用的な採譜システムができると考えている。

このようなシステム設計段階における相違が認知においてどのような影響を生じさせるかについて具体例を挙げて検証してみる。例えばポピュラー音楽のある曲が、一つの小節の中ではベースの音高は一定で八分音符が8個並べられたリズムを刻んでいる、としよう。従来の一般性の高い方法によってベース音を検出しようとすると、小節が n 個あったとして、まず $8n$ 個の楽音の存在自体の発見をし、更に音高の同定をし、そのうえ音源同定をするならば、$8n$ 個の音が全て同じ音色であることまで認知しなくてはならない。この数段階に及ぶ全ての過程で誤りを避けなければ、完全に正しい結果は得られない。ところが、「一つの小節の中ではベースの音高は一定で八分音符が8個並べられたリズムを刻んでいる」という事実を知っていれば、そういう前提で信号を解析して音高だけを同定すればよい。こうした方が確度の高い結果が得られるであろうし、結果の情報量も1/8になっているので誤りを人間が訂正するのも容易である。このようにある特殊な機能しか持っていない処理系であっても、場合によっては非常に有効に利用できるのである。

3. 協調型採譜システム

一般化を目指した、稼働率が高くなるように設計したエージェントではなくて、様々な特殊な機能を持ったエージェントを用意しておき、それぞれを適切に使用することができるなら、確度の高い採譜ができるはずである。しかし、どうやったら適切に使用ができるのか、どうやったら情報統合ができるのかといった課題が残り、実現が困難となっている。

そこで、提案するマンマシン協調型採譜システムでは、エージェントの開始に関して人間が指令を与えるという構成のものを考えている。人間は聴取の対象となる楽曲を聞きながら、あるいはそれを時間周波数解析し画面上に表示し可視化したものを眺めながら、様々な特殊な問題に特化したエージェントを適宜利用して所望の結果を得るのである。

このような方針を採ると、準備するエージェントが無数必要になってしまうことが想像される。何から実装するべきかという問題が生じてしまうのである。そのことについて本稿では考え、必要最小限の機能を追求していくことにする。音楽的に訓練を受けていない人が計算機を使わないで採譜を行う場合、

- 聴取の対象となる楽曲を聴くことができるカセットテープデッキ
- 音高を決定する手助けとなる鍵盤楽器
- 採譜の結果を記載する五線譜
- 五線譜に音符を書き込む筆記用具

などが必要と考えられる。人間が採譜を行うのを支援するシステムの第一の目標としてまず目指すべきことは、これらの道具を駆使して採譜を行うのと最低でも同程度の負担で済むようにすることである。

4. 基本的な機能

様々な特殊な機能を寄せ集めることで協調型採譜システムを設計するという方針を第3.章で述べたが、ここではまず最初に実装すべき基本的な機能について考えてみる。

〈4·1〉 範囲を指定した演奏機能　人間が採譜を行うときに、よく聞きとれなかった部分を繰り返し聞くことが容易にできると便利である。カ

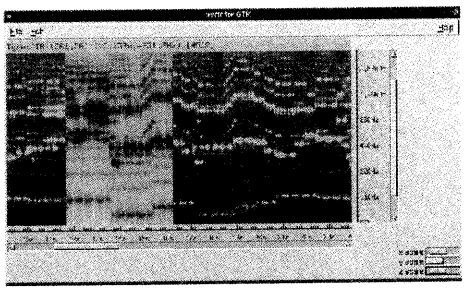

図 3. 演奏範囲の指定

セットテープやコンパクトディスクに録音された聴取の対象の特定の部分を繰り返し聞くには対象部分に対応する時刻やカウンタを覚えておく必要があり、操作が厄介である。そこで、図??のように音響信号を時間周波数解析した結果を画面上に表示して可視化し、ユーザはこれを参考に聴きたい部分を選択するようにすると便利である。時間領域を選択しておけば、あとはマウスやキーボードなどを一回叩くだけで、時間的に全くずれを生じさせずに所望の部分を繰り返し聴くことができる。

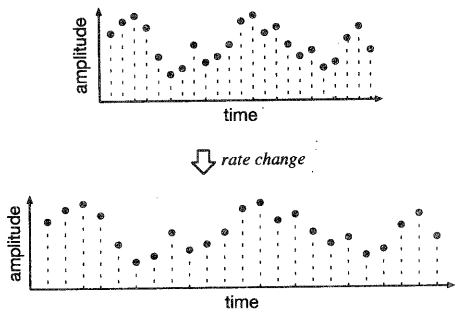

図 4. 人間による音楽の聴取

〈4·2〉 音響信号のストレッチ機能 所望の部分を繰り返し再生できるようになったことで作業が楽になったが、もっと欲をかくならば音程を変えずに再生速度を遅くできると便利である。これはストレッチという技術である。図4のように、単純に標本化の周波数を下げて再生した場合、全体としての演奏時間は延長されて楽曲がゆっくりと聞こえるようになることは容易に想像がつく。しかしこのような方法を採った場合、ピッチは低くなりフォルマントも保たれない。本来の音色とピッチを保ちつつゆっくりと演奏されたかのように加工するには工夫が必要である。

ピッチを変えずに演奏時間を延長するには、図5に示すように音響信号を数10ms程度の長さの塊に分解し、それぞれの塊の間に空白を挿入すればよい。それぞれの塊がある程度の長さを持っているため、人間が聴取した場合にはピッチをはっきりと同定できる。但しこのままだと音が途切れ途切れに聴こえ違和感が生じる。これを解決するために、塊の間の空白時間には前後の塊がつながって聴こえるような信号を前後両者の信号を用いて生成して、空白時間に挿入することが必要となる。

ここで原信号を $f(t)$ をストレッチする方法について述べる。まず $f(t)$ から以下の関数列 $g_n(t)$ を生成する。

$$g_n(t) = w_n(t)f(t) \quad \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

ここに関数列 $w_n(t)$ は

$$w_n(t) = \begin{cases} 0 \\ \quad (t < -T_e + nT_g, \\ \quad (n+1)T_g + T_e \leq t \text{ のとき}) \\ \frac{1}{2}\left\{1 + \cos\left(\frac{\pi(t-nT_g)}{T_e}\right)\right\} \\ \quad (nT_g - T_e \leq t < nT_g \text{のとき}) \\ 1 \\ \quad (nT_g \leq t < (n+1)T_g \text{のとき}) \\ \frac{1}{2}\left\{1 + \cos\left(\frac{\pi(t-(n+1)T_g)}{T_e}\right)\right\} \\ \quad ((n+1)T_g \leq t < (n+1)T_g + T_e \\ \quad \text{のとき}) \end{cases} \quad (2)$$

である。また T_g は塊の時間の長さ、T_e は塊の間の隙間の時間の長さである。

$g_n(t)$ を、全体の演奏時間が長くなるようにずらしながら

$$g(t) = \Sigma g_n\left(t - n(T_g + T_e)\right) \quad \cdots\cdots\cdots\cdots (3)$$

とつなぐことによって、ストレッチされた信号 $g(t)$ が得られる。このことを図示したのが6である。$w_n(t)$ の前側の曲線部分と後側の曲線部分はずらして足し合わせると1になるように設計されているので、マクロにみた $g(t)$ 振幅の連続性が保たれぶつぶつ途切れるような印象を与えない。なお、演奏時間は $(Te + Tg)/Tg$ 倍長くなる。

以上のような方法で音響信号のピッチを変えることなく演奏時間を延長し人間にとって聴きとりやすくなる。しかし、図7のように塊を切りとる時間間隔と楽曲の拍位置を同期させずに切りとってしまった場合、違和感を覚える虞がある。これは対象となる楽曲の演奏形態に依存するのであるが、ポピュラー音楽のようにドラムセットを用

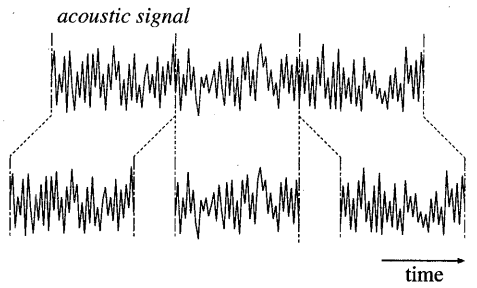

図 5.　ピッチ不変な演奏時間の延長

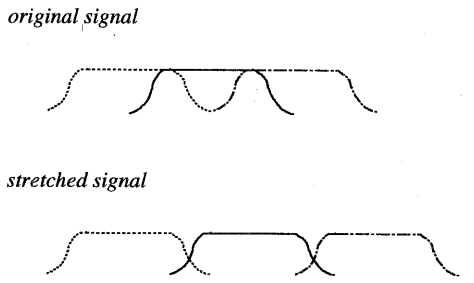

図 6.　分離した塊の滑らかな接続

いている場合にはその傾向が強い。それは、図7の左から2番目と3番目の拍のように、ドラムが鳴っている瞬間を切断してしまうので、その音が時間間隔をあけて2度聞こえてしまうことがあるからである。

このような現象を避けるためには、図8のように拍の周期と塊を切りとる周期とを同期させ、かつ切りとる時刻が拍の位置からずれるようにするとよい。

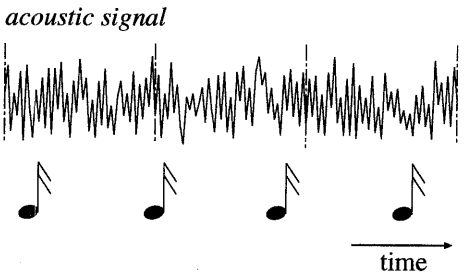

図 7.　切断時刻と拍位置の同期がとれていない状態

⟨4·3⟩　拍位置情報の表示　⟨4·2⟩節で述べた指定範囲の演奏を行う場合、時間周波数解析の結果表示の横軸の目盛は、秒やミリ秒などの絶対的な表示よりも、図9のように拍位置を表示した方

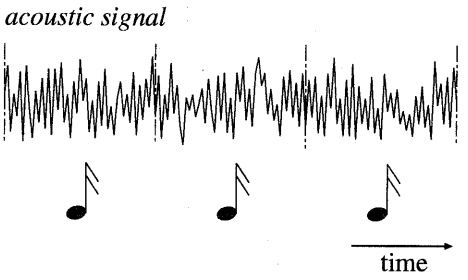

図 8.　切断時刻と拍位置の同期がとれている状態

が分かりやすいこともある。この図では四分音符だけを表示しているが、ビート構造は四分音符レベル、二分音符レベル、小節レベルといった階層構造を持っているので[7]、これらの情報も原信号から抽出して表示することが望ましい。

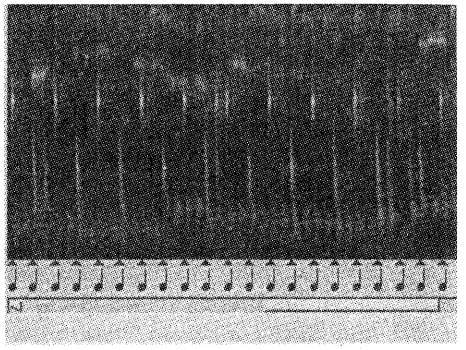

図 9.　拍位置による時間目盛

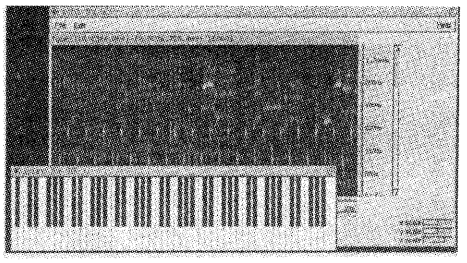

図 10.　計算機上の仮想鍵盤楽器

⟨4·4⟩　試聴のためのインターフェース　音響信号を時間周波数解析して可視化することは大切なことであり、⟨4·3⟩のように拍位置を表示することによって時間軸の意味がユーザにとって分かりやすくなる。しかし周波数方向の座標の意味を直感的にユーザに分かるようにすることは困難である。五線譜を表示することである程度分かるよ

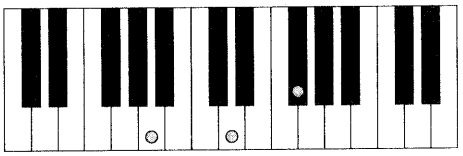

図 11. 鍵盤と時間周波数解析結果との間の情報交換

うになるであろうが、楽音は調波構造を有しており、ある周波数のパワーが強いからといってそこから五線譜上での位置の情報を得ても、パワーが強くなっているのが基音によってではなく倍音によってであった場合には、ほとんど意味をなさなくなる。

視覚によってでは直観的に伝わりにくい周波数方向の情報は、音そのものによって聴覚で理解するのが望ましいと思われる。そこで、図10のように計算機上に仮想の鍵盤楽器を置くことによって、いつでも音高を参照できるようにする。但し、それだけでは計算機の横に本物の鍵盤楽器を置くのと変わりがないので、図11のように計算機による解析結果を鍵盤上に表示させ理解を助けたり、逆に鍵盤を叩く位置の情報から何らかの情報を導きだして時間周波数解析の結果上に可視化することも可能になる。

(4·5) 和音の認識 小節によっては、和音を特定できると採譜の助けとなる場合がある。OPTIMA[3, 4]などでは、抽出した和音の情報を巧み利用して採譜面の精度をあげるような試みがなされている。つまり、和音情報はどの時刻でも必ず検出するようになっている。

しかし、和音情報が有用な小節とそうでない小節があるはずであり、人間が「和音情報を知りたい」と思ったときだけそれを得るということにしたほうが、情報抽出処理系の信頼性も高くなるはずである。OPTIMA では図12のように周波数成分から楽音一つ一つをクラスタリングし和音情報を得ようとしているが、和音の演奏の仕方まで人間側が計算機に教えることができるならば、図13のように周波数成分からいきなり和音を同定できる場合もあるであろう。

5. まとめ

本稿ではマンマシン協調型採譜システムの設計の思想と、それを実現するにあたって最小限必要な信号処理技術やインターフェース技術について述べた。第3.章に挙げた

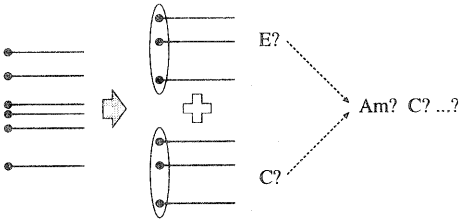

図 12. 人間による音楽の聴取

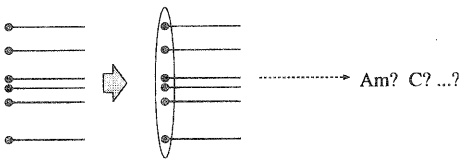

図 13. 人間による音楽の聴取

- 聴取の対象となる楽曲を聴くことができるカセットテープデッキ
- 音高を決定する手助けとなる鍵盤楽器
- 採譜の結果を記載する五線譜
- 五線譜に音符を書き込む筆記用具

という計算機を使わない採譜に必要な道具に計算機がとって替わる道筋を検証したが、後者2点についての検証が十分でなかったため、今後はこれらの点について研究を進める予定である。

文 献

(1) ISO/IEC DIS 10743, "Standard Music Description Language (SMDL)", 1995

(2) 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

(3) 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

(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 における和音の認識」, 電子情報通信学会論文誌, Vol. J79-D-II, No. 11, pp. 1762–1770, 1996

(5) 中谷 智広, 後藤 真孝, 川端 豪, 奥野 博: 「残差駆動型アーキテクチャの提案と音響ストリーム分離への応用」, 人工知能学会誌, Vol. 12, No. 1, pp. 111–119, 1997

(6) 杉浦 勇樹, 阪口 豊: 「うなりを利用した不協和音程の検出」, 情報処理学会音楽情報科学研究会, 2000-MUS-36, pp. 1–6, 2000

(7) 後藤 真孝: 「音楽音響信号を対象としたリアルタイムビートトラッキングに関する研究」, PhD thesis, 早稲田大学, 1998