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。

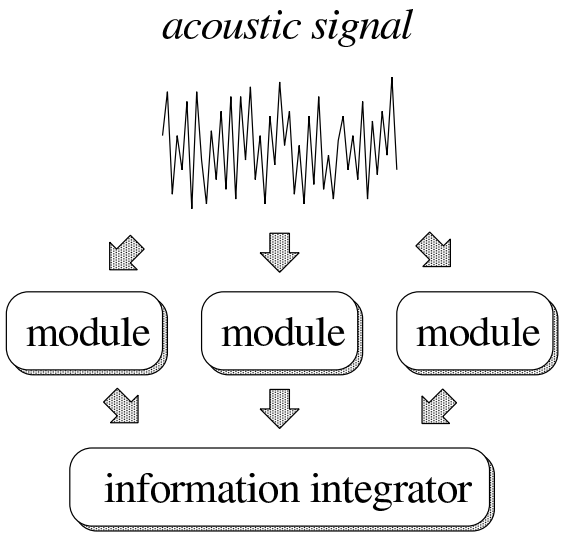

図 1. 汎用的な利用を目指した採譜システムの構造

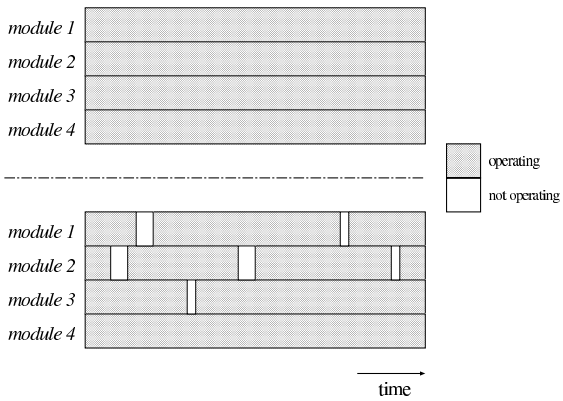

図 2. モジュールやエージェントの稼働率

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

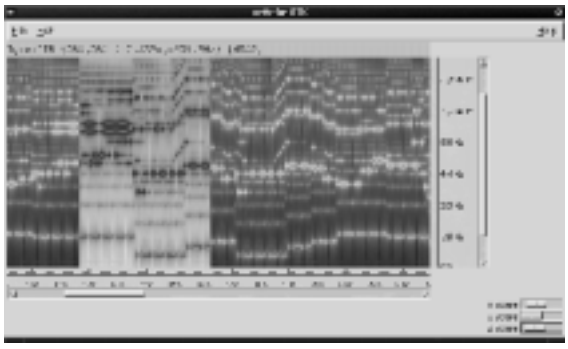

図 3.　演奏範囲の指定

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。

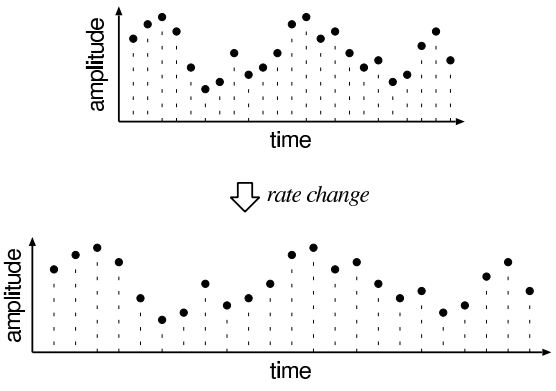

図 4.　人間による音楽の聴取

⟨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。

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

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

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

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

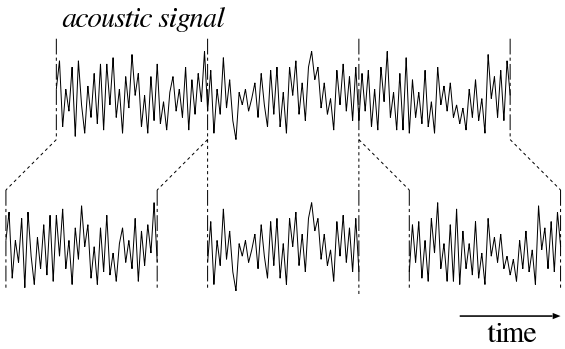

図 5.　ピッチ不変な演奏時間の延長

original signal