

計算機と人間の協調による採譜システム

半田 伊吹, 木下 智義, 武藤 誠, 坂井 修一, 田中 英彦

{handa,kino,muto,sakai,tanaka}@mt1.t.u-tokyo.ac.jp

東京大学大学院 工学系研究科*

1 はじめに

計算機の音楽の世界への応用は様々な試みがなされており、演奏された楽曲の音響信号から楽譜を生成するいわゆる自動採譜に関する研究もなされている。音楽情景分析の処理モデルである OPTIMA[1][2] では、入力を多角的に解析し、その結果を統合するということにより最尤推定像を出力できるという特徴を備えているが、その認識精度は実用上十分とは言えず、改善が課題となっている。

計算機による音楽認知が難しい理由は、単に計算量の問題だけではなく、計算機の性能が向上すれば解決されるというものではない。音響信号に対してどのようにアプローチすれば精度のよい認識ができるかが明かにされていないのである。

本稿では、採譜作業を全て計算機に委ねずに人間と協調して採譜を行うシステムを提案し、その全体像とインターフェースについての考察を行う。

2 計算機と人間の協調による採譜

OPTIMA のように従来研究のなされている楽曲認識システムは、全ての処理を計算機が行っている。しかし、人間にとっては比較的容易に知り得るような情報であっても計算機では抽出困難なものに関しては、計算機で不十分な認識処理を行って精度を下げるよりは、人間が計算機を支援して精度の高い認識が可能となったほうが実用的といえよう。

そこで、採譜システムを新たな視点から構築しようと考え計算機と人間の協調による採譜システムを提案する。

提案するシステムの概要を図1に示す。

まず対象となる音楽音響信号は、人間の耳によって聴取され、計算機には符合化されたデータとして入力される。

楽曲を聴いた人間は、そこに含まれる情報を意識的あるいは無意識的に見出し、ときには情動を揺さぶられることもあるし、はっきり認識しつつも感性には影響をあまり受けない場合もある。楽曲の特徴として

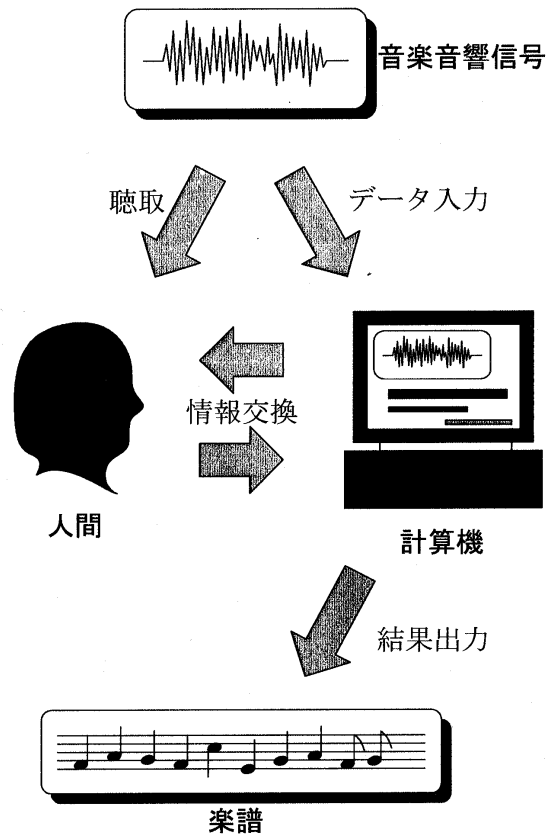


図1: 提案する採譜システムの概要

物理的特徴 音楽音響信号の周波数スペクトルの特徴、すなわち音楽音響信号を物理的な音として捉えたときの特徴

音楽理論的特徴 楽曲を楽譜として表現したときに、楽譜上に現れる特徴

感性的特徴 人間が楽曲を聴いたときに受ける印象や感想

の3種類を挙げることができる [3]。これらは明確な区分があるわけではないが、この分け方に従うと計算機は物理的特徴を正確に認識することを得意とし、人間は音楽理論的特徴を計算機よりはるかに的確に捉えられると言えるであろう。そこで、計算機と人間はお互いの得た知識をやりとりし、最終的に得た知識をもとに計算機が採譜結果を出力する。

*"Intelligent music transcription system with human assistance"
Ibuki Handa, Tomoyoshi Kinoshita, Makoto Muto, Shuichi Sakai, Hidehiko Tanaka
Graduate School of Engineering, the University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan



図 2: 譜例

3 インターフェース

ここでは計算機と人間がどのようなやりとりをどのようにするかについて検討する。

まず大前提となることは、人間の行うべき作業が極端に複雑になってはいけないということである。人間だけで採譜を行う場合よりも面倒なものになっては本末顛倒である。そこで、システムには、

- 人間のみが採譜を行う場合より、より多くの情報が得られる
- 人間のみが採譜を行う場合より少ない作業量で採譜が行える

といった仕様が要求される。ここでは特に後者を満たすことを考えることとする。

例えば図2のような楽曲を認識する際に、計算機上では図3のように時間、音高で区切った空間を埋める問題として処理しているとする。このとき、計算機がマトリックスを示してあとは人間が穴を埋めてくれれば計算機への負担は少ないが、人間にとっての負担は大きい。音楽の専門的知識がない人間でも容易にできることといえば、音が鳴っているか否かの判断や音高の変化のおおまかな判断くらいであろう。そこで、図4に示すように、計算機は発音時刻の候補を示し、人間はそこに音の有無および音高について一つ前の発音とのおおまかな相対関係だけを入力するようなシステムが要求仕様を満たすと考えられる。計算機側が担う発音時刻の特定は、決して容易な問題ではないが、ビートトラッキング問題として成果も得られている[4]。人間が入力する情報のうち、単音が鳴っているの否かの情報は非常に重要である。このように計算機にとっては大変有用な情報を人間が簡単に入力できるようなインターフェースが望まれる。

4 おわりに

本稿では、計算機だけでは困難な採譜処理に人間の助けを加えて構築する採譜システムを提案した。この手法により認識精度が従来より高い採譜システムができるものと考えられる。

ところが、完全に計算機に処理を委ねたシステムに比べ、認知科学や計算機科学の観点からは消極的でつまらないものという印象を与えかねない。しかし、どの

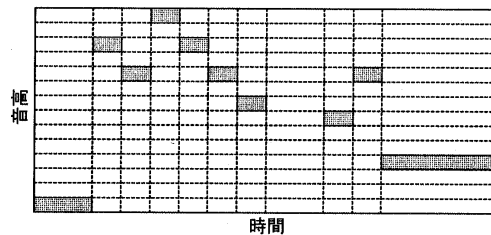


図 3: 抽象化された楽曲

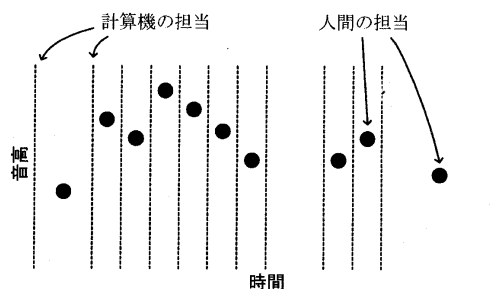


図 4: 計算機の出力情報と人間の入力情報

ような情報を得ると認識精度が向上するのかという見識が、本システムの実装を行う上で得られるであろうから、結果として学問的にも意義のある研究であると考えられる。

参考文献

- [1] 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-DII, No.11, pp. 1751-1761, 1996
- [2] 柏野邦夫, 木下智義, 中臺一博, 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-DII, No.11, pp. 1762-1770, 1996
- [3] 武藤誠, 木下智義, 半田伊吹, 坂井修一, 田中英彦: 「音楽音響信号からの楽曲の感性的特徴の抽出」, 情報処理学会平成 11 年後期全国大会講演論文集 (2), 4G-6, pp. 2-11-2-12, 1999
- [4] 後藤真孝, 村岡洋一: 「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システム—コード変化検出による打楽器音を含まない音楽への対応—」, 電子情報通信学会論文誌, Vol.J81-DII, No.2, pp. 117-127, 1998