

音楽情景分析における 楽音認識と自動採譜

海野 由紀子 鈴木 啓高 斎藤 博昭 中西 正和
慶應義塾大学 理工学部 数理科学科

1. はじめに

音楽情報処理の分野でいう「自動採譜」とは、コンピュータが音楽の演奏を聞いてそこから楽譜(に相当する情報)を作り出すことをいう。これまでの音楽の分析・認識と自動採譜の研究は、サンブラや DAT に演奏情報(音の強弱や発音時間)をあらかじめ用意しておき、常に一定の演奏情報を扱うものがほとんどであった。しかし、人間の演奏は一定ではない。そこで本稿では人間の演奏から得られる音楽音響信号を入力として音楽の分析・認識と自動採譜を行う。音楽の分析・認識の内容は「楽音(音階を持つ音)の音高認識」、「拍子・小節認識」、「調性認識」を行うものとする。

2. 自動採譜システム

自動採譜のタスクは、楽音への群化、音楽的分析、楽譜出力の3つに分けることができる[1]。楽音の群化とは、主に基本周波数の認識をして、1つの楽音を1つの楽音として聞くことであり、音高(ピッチ)の認識がこれに当たる。楽音の群化により得られた音の羅列に対して、リズム(拍や拍子)の認識や調性(コード)認識など、音楽的な解析を伴った分析を続ける。これが音楽的分析で、楽譜の表記のために必要というだけでなく、信号処理部における誤認識を補償するための処理としても重要である。このような分析を行った上で、記譜に関する処理を施し、楽譜として出力する。以上が採譜処理の流れである。

3. 実装方法

電子楽器のアナログ出力端子(LINE OUT)を DAT-Link[4]に繋ぎ、演奏から得られる音楽音響信号をデジタル化(標本化、量子化)し処理を行う。

音高認識

デジタル化された音響信号の基本周波数(ピッチ)を抽出し音階に割り当てる。基本周波数の抽出にはスペクトル解析を用いる。図1にスペクトルの解析により基本周波数を求める手順を示す。短区間に分割するための窓関数はハミング窓を用いる。抽出した基本周波数は等分平均律による全音階を参照し[3]、音階上に同定される。本稿において処理対象とする演奏は単旋律のものとし、楽音の範囲は C3 do ~ C6 do とする。

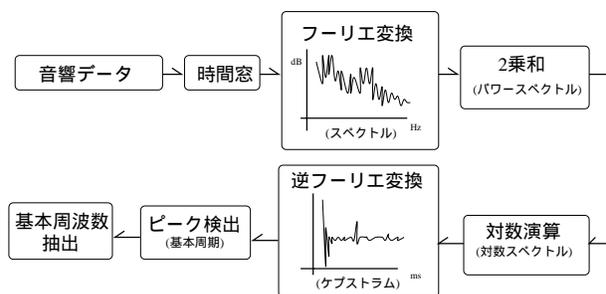


図1: スペクトル解析による基本周波数の抽出

拍節認識

次に、音高認識で得られた楽音を時間軸上に割り当てる作業を行う。演奏の各音の長さを認識するためには、固定の閾値でデータを振り分けていく必要がある。本稿においては、電子楽器のメトロノーム機能(以下クロックという)を閾値設定として使用する。閾値はクロックのテンポにより、以下の式に定める。閾値は1拍に入るデータ数となる。

$$Threshold = \frac{60 \times (Sampling\ Frequency)}{Tempo}$$

次に、得られた音長から小節線が引かれる場所を求める。この処理を行うにあたり、本稿においては拍節認識構造を提案し、この構造に基づき演奏が何拍子であるかを定める。拍節認識構造とは楽曲演奏(ドラムパターンなども含む)に対して、人間が持つ小節感覚を考慮し、演奏の各フレーズのまとまりに優先順位を付け、拍節の認識に繋げるものである。ここまでの処理により、楽音が時間軸上に並ぶこととなる。調性認識

音高認識により得られた音高列を用いて調性の認識を行う。本稿においては、吉野・安部の調性認識モデル[2]により提案された階層的全音階フレームを導入し評価を行う。階層的全音階フレームとは、長調のもの短調のもの2種類あり、入力メロディが全種の調それぞれに対してどの程度適合するかを数値化する評価システムである。適合性の判断は、全音階の主音、属音、3度音(以上が主和音構成音)、4度音(下属音)、6度音、2度音、7度音、非全音階音がこの順序で階層的に判断に関わる。つまり、ある調の解釈において、階層の上位の音高の出現頻度が高く、階層の最下位の音高(非全音階音)の出現頻度が低い程、メロディがその調に適合しているとして高い評価値を与える。本稿の調性認識処理は、演奏音が音高認識により順次音階上に同定される度に、全調の階層的全音

Tonal Music Recognition and Automatic Music Transcription
in Auditory Scene Analysis

Yukiko UNNO Hiroataka SUZUKI Hiroaki SAITO
Masakazu NAKANISHI

Department of Computer Science, Faculty of Science and
Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku,
Yokohama, Kanagawa 223, Japan

Tel: 045-563-1141(ex.3765)

(Email: {uni, hiro, hxs, czl}@nak.ics.keio.ac.jp)

階フレームを通り各調の重み付けを行うものとする．最終的に重み値(評価値)が最高であった調を，入力演奏の調とする．図2に階層的全音階フレームを用いた調性認識処理の概念図を示す．

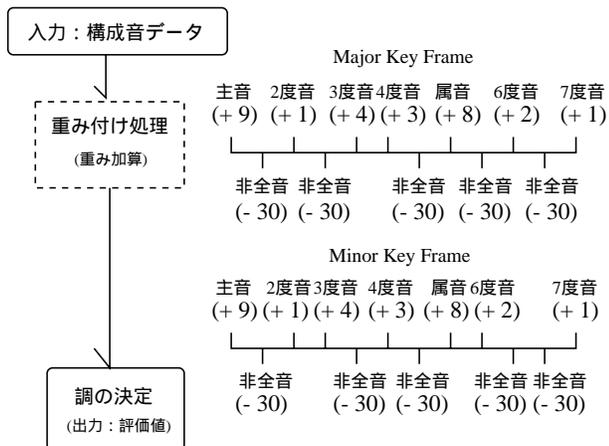


図2: 全音階フレームによる調性認識処理の流れ
自動採譜処理

音響データ全体に対してそれぞれの処理が終了したら，調性認識処理で評価値が最高であった調を演奏の調性とし，音高認識や拍節認識で得られた結果を譜面として出力する．譜面出力には MusicTeX を用いる．

4. 実験

YAMAHA W-7 シンセサイザを演奏楽器として使用した．演奏の内容を以下の表1と楽譜を図3に示す．

表1: 実験のサンプル

標本化	48 kHz
量子化	16 bit
演奏曲	ふるさと
拍子	3 拍子
調性	へ長調
音色	クラリネット (音源 no.G072)
テンポ	60 (bpm)
クロック	1 拍毎に発音
クロック音量	レベル 3



図3: 「ふるさと」の楽譜

結果

上記の条件の下で得られた採譜結果を以下に示す．



図4: 採譜結果

拍節認識により，演奏は3拍子と認識された．また，調性認識においては「へ長調」が評価値127で最も高く，続いて「変イ長調」(評価値69)，ト長調(評価値21)の順で調性名が得られた．

5. 結論

採譜結果が演奏通りの楽譜であるかを，図3と図4を比較して考える．連続音の表記法は異なるが，図3の楽譜(実験用に与えられた楽譜)に忠実な演奏と図4の楽譜(採譜結果)に忠実な演奏は同じ演奏となる．よって，演奏通りの採譜結果が得られたと言える．また，調性認識の結果も「へ長調」という正しい調性名が得られた．

6. おわりに

音楽情報処理の分野において，それぞれの認識システムが独立して研究されているなか，本稿では「音楽認識」のインテグレーションシステムを提案することができた．また，「入力を一定の演奏情報に限定しない」という本来の目的も達成でき，人間の演奏に必ず存在するずれや不定な発音時間にもある程度対応できた．しかし「認識率」と「処理の対象とする音楽音響信号の制約条件」の関係は未だ反比例にあり，今後も処理の対象とする音楽音響信号の範囲を広げること，リアルタイム性を意識することなど，より実用的な自動採譜システムの構築にはより一層の思案が必要である．

参考文献

- [1] 片寄 晴弘，“自動採譜(概論)”，bit 別冊コンピュータと音楽の世界，pp.74-88，1998.
- [2] 吉野 蔵，阿部 純一．“調性認識：メロディの調を解釈する計算モデル”，bit 別冊コンピュータと音楽の世界，共立出版株式会社，pp.117-131，1998.
- [3] Tim Kientzle，“A Programmer's GUIDE TO SOUND”，ADDISON-WESLEY，1997.
- [4] DAT-Link <http://www.tc.com/>