

5G-5

単旋律の生演奏を収録した CDからの自動採譜システム

村瀬 樹太郎 鈴木 啓高 斎藤 博昭 中西 正和
慶應義塾大学大学院 理工学研究科 計算機科学専攻

1. はじめに

音楽を聴いてそれを楽譜としてあらわすことは、音楽的知識や聴音能力が必要で、一般的な人にとって簡単なことではない。そこで、計算機を用いて採譜を自動的に行うことができれば、特別な能力のない人でも採譜ができるようになる。

アナログ信号は、対象とする音楽音響信号以外の信号を多く含んでいるため処理が困難で、これまでの主な研究はMIDIデータを対象としていた。本研究ではアナログで得た音響データから計算機を用いて自動的に楽譜を作成することを目的とする。

2. 自動採譜システム処理の流れ

自動採譜システムの処理は、入力音楽信号の(1)楽音への群化、(2)音楽的解析、(3)楽譜出力の3つの段階に分けることができる[1]。

(1) 楽音への群化

人間が知覚的に一つの音とみなす楽音を音響信号においてカテゴリ化することであり、処理としては基本周波数とその継続時間の認識が主なものである。

(2) 音楽的解析

テンポ、拍子の認識、楽音の音階への割り当て、調の決定などの音楽的解析を伴った処理である。

(3) 楽譜出力

音符の位置決めなどの処理が必要である。

3. 生演奏の特徴

生演奏では人間が演奏するため、MIDIのように機械的に合成された音による演奏と大きく異なる特徴がある。

まず、その1つはスペクトルに現れる。音は基本的に基本周波数とその倍音(Harmonic)で構成されている。人間が演奏する場合、機械のように一定の音を出し続けられないので、楽器にもよって異なるがピッチの揺らぎが生じてしまう。また、その音を録音する際に雑音の混入は避けられない。これらのことから、音は必ずしも基本周波数とその倍音成分だけで構成されているとは言えない。

また、リズムにも決定的な特徴が現れる。人間の演奏では表情をつけるために、意識的なリズムやテンポにおける微妙な表現やゆらぎが入り、本来同じ音符は同じ長さであるが、実際は音の長さにはばらつきが生じてしまう。

これらが自動採譜の処理、つまり音の認識を困難にする大きな原因であると考えられる。

4. 本研究の手法

4.1 ピッチの抽出方法

音声信号は短区間において周期的であるという性質を利用する。そのため本研究においても音声データを標本化、量子化をしてデジタル化し、ハミング窓を用いて短区間に分割して分析を行う。これらに対してケプストラム法を使う。これは、フーリエ変換し、対数を取り、逆フーリエ変換を行って得られた結果を、低ケフレンシー部と高ケフレンシー部に分けてピッチを抽出する方法である。

その抽出したピッチを平均律表[2]を用いて決められた音名に変換する。

4.2 拍の認識方法

拍を認識するための音の長さは、ピッチを抽出するために用いた短区間のフレームのずらした数とする。すべての音の長さの中から、ある閾値以上で一番短いものを基準とし、それを八分音符と決め、他の音はその基準とした音の長さから最も近い整数倍の値にして、音符を割り当てる。

4.3 採譜方法

いままでに得られた結果から Music TeX Version 5.01 を用いて楽譜出力する。本研究では最低限の楽譜を目的としているため、音符と休符のみを出力し、小節線、スラー、連符などは出力をしない。

4.4 評価方法

本研究では重み付けを用いた次のような評価法を考案する。

4.4.1 パラメーター

正解譜の音一つに対し、以下の3つの設定されたパラメーターについて評価する。

- ピッチの同定

完全にピッチの認識が行われたもの、正しい認識とそうでないもの両方が出力されたもの、全く認識できていないもの、の3つに分類する。

- 音の長さ

本来、楽譜に表示されている音符より人間は多少短く演奏する。つまり、音が短く認識された場合に休符によって補われていれば正しい認識であり、元の音より長く認識してしまうことは正しい認識ではないと言える。

- 音符の個数

1つの音に対し1つの音が認識されることが望ましいが、1つの音を複数の音として認識してしまう場合と、2つ以上の同じ音を1つとして認識してしまう場合が考えられる。

4.4.2 重み付け

人間がある楽譜とある楽譜を比べる際、それぞれの楽譜から音楽的知識をもとに曲をイメージする。よって、先のパラメータを以下のように需要度が高いと思われ順に、1つの音に対し最高で10pのポイントとなるようにポイントを与える。

- 1 ピッチ R (5p or 3p or 0p)
- 2 音の長さ L (3p or 2p or 0p)
- 3 音符の数 (拍) M (2p or 0p)

一曲全体に対するピッチ認識率 Pa [%]、一曲全体に対する楽譜を含めた認識率 Sa [%] は、 i 番目の音の R 、 L 、 M 、をそれぞれ R_i 、 L_i 、 M_i 、総単音数を N すると、以下のように定義できる。

$$Pa = \frac{\sum_i R_i}{N * 5} * 100 \quad (1)$$

$$Sa = \frac{\sum_i (R_i + L_i + M_i)}{N * 10} * 100 \quad (2)$$

5. 実験

5.1 実験条件

市販されている単音のみの CD から DAT を通して A/D 変換 (サンプリング周波数 44100Hz で標本化、16bit で量子化) を行う。

5.2 実験結果および認識率

Music TeX を用いて得られた採譜結果は図1にようになる。

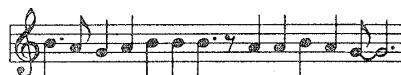


図1: 実験1の結果

表1: 認識率一覧 (%)

	N	Pa	Sa
実験1	12	100.0	98.3
実験2a	28	97.9	66.1
実験2b	28	100.0	93.9
実験3	31	98.7	87.1
実験4	25	96.0	90.0

総単音数 N : 音響信号中の総単音数
 ピッチ認識率 Pa [%] : ピッチ認識率
 全体認識率 Sa [%] : 楽譜としての認識率

6. おわりに

本研究で扱った7つの曲のうち、中音域の曲に関してはかなりのピッチ認識率が得られた。実験を通して本研究で提案した評価方法の有効性が認められたと言える。

楽譜にするにあたって、ビートトラッキングなどいくつかのシステムを組み合わせることによってより良い楽譜の作成ができると考えられる。また、本研究で提案した重み付けによる評価方法を、パラメーターの変更などにより、さらに結果を正しく反映するように改善を図ることが必要である。

参考文献

- [1] 長嶋洋一, 橋本周司, 平賀護, 平田圭二. コンピューターと音楽の世界. 共立出版. 1998.
- [2] Tim Kientzle. A Programmer's GUIDE TO SOUND. Addison Wesley Developers press. 1997.