

打楽器音を対象にした自動採譜システム

7Q-2

後藤 真孝 田淵 仁浩 村岡 洋一

早稲田大学 理工学部

1. はじめに

本稿では、複数の打楽器によって演奏された音楽を対象にした自動採譜システムについて述べる。従来おこなわれてきた自動採譜の研究は単音から和音へ、単一楽器から複数楽器へと対象とする音楽の制約を徐々に減らす方向で進められている[1]-[7]。しかし、これらの研究はピアノなどの楽音(musical tone)を発する楽器のみを対象にしたものであり、打楽器などの無音(unpitched sound)を発する楽器は対象にしていなかった。

一方、ポップスやロックなどの音楽では打楽器が重要な役割を演じている。ところが従来の自動採譜をこれらの音楽に適用すると、打楽器音がノイズとなって楽音の認識を妨げるために採譜が困難になる。したがって、これらの音楽を自動採譜するためには打楽器音に対応した自動採譜を実現することが本質的に重要である。そこで打楽器音が含まれているこれらの音楽を自動採譜するための第一段階として、複数の打楽器のみによって演奏された音楽を対象にした自動採譜システムを実現した。

2. システムの特徴

本システムは複数の打楽器音が混合した音響信号を入力し、打楽器の種類とその発音時刻・強度をあらわすシンボルを出力する。具体的な入力信号はポップスやロックなどの音楽で多く使用されるドラムスの演奏音とし、楽音は含まれていないものとする。このドラムスは、Bass Drum (BD)、Snare Drum (SD)、Low Tom (LT)、Middle Tom (MT)、High Tom (HT)、Hi-hat Close (HC)、Hi-hat Open (HO)、Ride cymbal (RI)、Crash cymbal (CR) の9種類の打楽器で構成され、個々の音は事前に計算機に登録しておく。このうちHCとHOは同一の打楽器であるが、演奏音が異なるために分けてある。以上9種類の打楽器はErich von HornbostelとCurt Sachsの体系的楽器分類によると[8]、膜鳴楽器(BD、SD、LT、MT、HT)と体鳴楽器(HC、HO、RI、CR)の二つに分けられる。前者の5つの音は中低域の周波数成分が多く、後者の4つの音はすべてシンバル系の音なので高域の周波数成分が多い。

採譜結果として出力するシンボルには標準MIDIファイル(SMF)を用いる。SMFは現在広く使用されている演奏データのフォーマットなので、本システムが出力したSMFを用いて自動演奏することにより、採譜結果を音で確認することができる利点を持つ。

本システムは入力信号をA/D変換した後、周波数解析をおこない時間周波数マップを得る。時間周波数マップとは、横軸に時間 t 、縦軸に周波数 f をとり、各時刻において存在する周波数成分の強度を濃淡で示した図である。次にこの時間周波数マップを解析し、複数の打楽器がいつどのくらいの強さで鳴っているかを認識する(楽器同定/音源分離)。最後にこれらの認識結果をSMFにシンボル化して出力する。

3. 周波数解析

16bit、44.1kHzでサンプリングされた入力信号に対し、4096点の時間窓を100点ずつシフトさせてFFTをおこなう。FFTの窓関数にはハニング窓を使用し、周波数成分の存在する場所に対して複素スペクトル内挿法を適用する。ただし、これは従来の複素スペクトル内挿法[9]を新たにハニング窓に対応するように補正した方法で、式は付録に示す。こうしてFFTの結果である複素スペクトルから精度良く周波数を求め、時間周波数マップを得る。

4. 楽器同定/音源分離

打楽器音はこれまで自動採譜の対象にされてきた楽音とは違い、衝撃音が多いために複雑なスペクトル構造を持つ。そのため基音と倍音が明確に分けられず、音階上に周波数同定できない。このような特徴を持つために、従来の楽音を対象にした手法では楽器同定/音源分離をおこなうことができない。

そこで、人間の音楽認識過程を考慮した認識手法を考案した。本手法は以下のような特徴を持つ。

1. 低い音/高い音の各発音時刻の候補を求めてから、その時刻に発音した音の楽器同定をおこなう。

これは、人間が音を聞く際に楽器の識別ができなくても、低い音あるいは高い音が鳴ったという認識だけは誰にでも瞬時にできることに相当する。

また、これはすべての時刻の音に対して楽器同定をおこなう詳細探索にくらべ計算効率が良い(粗探索の機構)。

2. 各楽器ごとにテンプレートパターンとその特徴をあらわす重み関数を用意し、後述する距離尺度を用いてテンプレートマッチングをおこなう。

人間が採譜する場合にも、特定の音に意識的に注意を向けなければ認識するのが難しい。これは、認識したい楽器ごとに聴覚の受容感度の周波数特性を変化させて注意を向けていると考えられる。そこで、このような仕組みを重み関数により実現する(選択的注意の機構)。

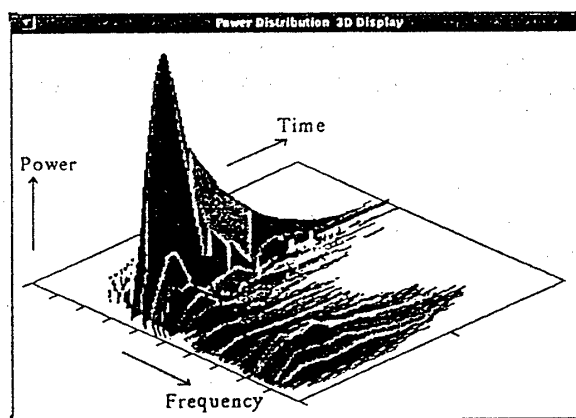


図1: Snare Drumのパワー分布形状(テンプレートパターン)

まず、周波数解析の結果得られた時間周波数マップから周波数成分のパワー分布の時間変化(パワー分布形状)を得る。

1. の発音時刻候補の決定は、このパワー分布形状の時間軸方向の一次微分が連続して正の値である座標 (t, f) を求めることによりおこなう。この座標が低域に密集している時刻を低い音の発音時刻候補とし、高域に密集している時刻を高い音の発音時刻候補とする。

次に、低い音、高い音の各発音時刻候補の音に対して、前者には膜鳴楽器、後者には体鳴楽器の同定をおこなう。各楽器の同定は、事前に登録した打楽器音のパワー分布形状(テンプレートパターン)と入力音のパワー分布形状(入力パターン)との距離を、考案した距離尺度で求めることによりおこなう。ただし距離を求める前処理として、正確な距離が得られるように両者の音量を合わせておく必要がある。打楽器の発音強度は合わせる前の両者の音量差から求める。

複数の打楽器が同時に鳴っているときには、入力パターン中に複数のテンプレートパターンが含まれている。そこで、個々の打楽器を同定する音源分離を実現するために、以下のような距離尺度を導入する。まず、各時刻、周波数における入力パターン $IP(t, f)$ とテンプレートパターン $TP(t, f)$ の差 $DIF(t, f)$ を $IP > TP$ ならば 0、 $IP < TP$ ならば正と定義する。そして、この差 DIF を時間-周波数平面上で重み関数をかけて積分することにより距離を求める。これは、単にパターン間のユークリッド距離をとるのではなく、入力パターン中にテンプレートパターンが含まれていれば小さい距離になる距離尺度である。そのため、同時に鳴っている他の打楽器音の成分があっても、単独の打楽器音の場合とほぼ同様な距離が得られる。また重み関数は、時間-周波数平面上の各楽器固有の領域におけるパターン間の差 DIF を強調し、2. の選択的注意の機構を実現するものである。

こうして求めた距離を閾値処理し、楽器同定/音源分離をおこなう。その際、HO だけは演奏上の性質から発音/消音時刻をそれぞれ求めておく必要がある。そして、その発音中は同一楽器音である HC を無視するように排他処理をおこなう。

5. 実験結果

市販のサンプラーをドラムスの音源として演奏音を入力し、SMF の出力結果を得る実験をおこなった。図 2 のドラム譜の演奏音を自動採譜した結果を図 3、図 4 に示す。この例と同程度の難易度のいくつかのドラムパターンで実験した結果、誤認識なく採譜できた。次に、2 つの打楽器が同時に発音するすべての組合せを実験した結果、実際の演奏上はほとんどない体鳴楽器同士の組合せを誤認識することがあったが、その他は問題なく採譜できた。また、3 つの膜鳴楽器が同時に発音するすべての組合せを実験した結果、誤認識なく採譜できた。

6. おわりに

以上、打楽器音を対象とした自動採譜システムについて述べた。本認識手法が有効に機能し、複数の打楽器によって演奏された音楽を自動採譜できることが確認できた。今後は実験結果で述べた誤認識を減らすとともに、入力する打楽器音の制約を減らしていく予定である。また楽音、打楽器音が混合した音楽の自動採譜システムを実現することを考えている。

参考文献

[1] 上野 勝彦, 中村 泰明, 井口 征士: 音楽情報の入力法 — 採譜 —, 信学技報, PRL78-21 (1978).
 [2] 新原 高水, 今井 正和, 井口 征士: 歌唱の自動採譜, 計測自動制御学会論文集, Vol.20, No.10, pp.940-945 (1984).
 [3] 阿久津 達也, 小池 汎平, 山内 宗, 吉田 実, 田中 英彦: ICOTone on PSI - Act II, 第 35 回情報学大会, 5F7-7 (1987).

[4] 吉田 実, 下山 健, 小池 汎平, 田中 英彦: ICOTone — 音楽情報認識ユニット minoru —, 第 37 回情報学大会, 5F7-2 (1988).
 [5] 片寄 晴弘, 井口 征士: 知的探譜システム, 人工知能学会誌, Vol.5, No.1, pp.59-66 (1988).
 [6] 柏野 邦夫, 田中 英彦: 音源分離同定システムについての考察, 第 43 回情報学大会, 7C-1 (1991).
 [7] 長東 哲郎, 才藤 直樹, 井口 征士: 異種楽器を対象とした探譜システム, 1992 年信学春大会, D-499 (1992).
 [8] ダイアグラムグループ 編, 香川 達夫 監修: 楽器, 株式会社マール社 (1992).
 [9] 原 祐一郎, 井口 征士: 複素スペクトルを用いた周波数同定, 計測自動制御学会論文集, Vol.19, No.9, pp.718-723 (1983).

付録

従来の複素スペクトル内挿法は方形窓を使用した FFT の結果を用いるもので、ハンギング窓には対応していない。そこで、この複素スペクトル内挿法の結果をハンギング窓に対応するように補正する式を示す。

$$z = 3(f - m) - 1$$

$$f' = m + z$$

$$a' = 6z(z - 1) \frac{\sin \pi(f - m) a}{\sin \pi z}$$

これは、ハンギング窓を使用した FFT の結果から従来の複素スペクトル内挿法で求めた周波数を f 、振幅を a 、 $m = [f]$ ($[\cdot]$ はガウス記号) としたときに、補正後の周波数 f' と振幅 a' を求める式である。



図 2: 入力した演奏の楽譜例

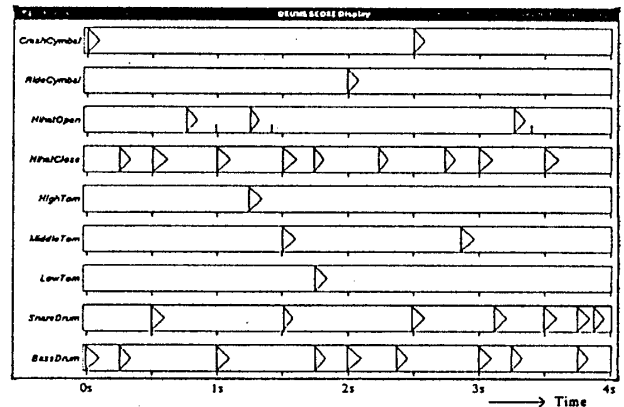


図 3: 自動採譜結果例 (画面表示)

HEADEF	Format=0	Mtrks=1	Division=48	[Time]	[NoteNum]	[GateTime]	[Velocity]
Track start				167:	42(F#2)	1	61
Tempo = 500000us per MIDI quarter-note				168:	36(C2)	1	67
Time Signature = 4/4				169:	41(F2)	1	77
MIDI clocks in a metronome click = 24				192:	36(C2)	1	85
32nd-notes in a MIDI quarter-note = 8				192:	51(D#3)	1	78
Text = "Automatic Transcription"				215:	42(F#2)	1	58
				228:	36(C2)	1	70
				239:	38(D2)	1	80
				0:	36(C2)	1	82
				0:	49(C#3)	1	82
				24:	36(C2)	1	70
				24:	42(F#2)	1	68
				48:	38(D2)	1	80
				48:	42(F#2)	1	92
				73:	46(A#2)	22	78
				96:	36(C2)	1	80
				96:	42(F#2)	1	82
				119:	48(C3)	1	81
				120:	46(A#2)	16	62
				144:	45(A2)	1	83
				144:	42(F#2)	1	86
				145:	38(D2)	1	59
							Track end

図 4: 自動採譜結果例 (SMF)