

周波数基底チューニングに基づく演奏詳細解析に向けた スコアフォローイング

保利 武志[†] 中村 和幸^{†‡} 嵯峨山 茂樹^{†‡}

[†] 明治大学 大学院先端数理科学研究科

[‡] 明治大学 総合数理学部

1 はじめに

本稿では音楽音響信号に対し、和音単位で楽譜追跡する手法について述べる。多重音からなる音楽音響信号に対する音楽特徴量(音高, オンセット時刻, 音量, 音長)の抽出は音楽情報科学分野において極めて重要な要素技術であり, 特に演奏解析や演奏モデル獲得のためには高時間分解能な音楽特徴量の抽出(本研究ではこれを演奏詳細解析と定義する)が求められている。

我々はこれまでクラシック曲のピアノ演奏を対象とし, 楽譜情報を事前分布や制約として陽にモデル化した演奏詳細解析手法について検討してきた [1, 2]。これらはまず Score-based な MIDI 音源(楽譜演奏)を音響信号に変換し, パワースペクトルに対する Dynamic time warping (DTW) を用いた実演奏に対する楽譜追跡結果を事前情報として活用しており, 一定の高精度推定を実現する一方でそのパラメータに大きく依存しており, より精緻な事前情報が求められていた。また, 実録音データからの解析を想定する場合, 録音環境に基づくチューニングのずれに対する考慮も必要もあった。

図 1 は本研究における詳細解析のブロック図であり, 本稿では特に赤枠で囲われた音楽音響信号に対する音高周波数チューニングを備えた和音単位での楽譜追跡に焦点をあてる。

2 音高周波数チューニング

2.1 モデリング

録音環境や音源によっては数セント(半音 = 100 セント)チューニングがずれていることもあり, チューニング補正機能の有無は推定精度に大きく影響し得る。ここで, 一般にピアノにおける単音の周波数分布は, 基

“Comparison of Score Following Methods Based on Frequency Basis Vector Tuning Towards Precise Multi-Pitch Analysis”

by Takeshi HORI[†], Kazuyuki NAKAMURA[†] and Shigeki SAGAYAMA[‡],

[†]Graduate School of Advanced Mathematical Sciences and [‡]School of Interdisciplinary Mathematical Sciences, Meiji University, 164-8525, Tokyo, Japan

{hori, knaka, sagayama}@meiji.ac.jp

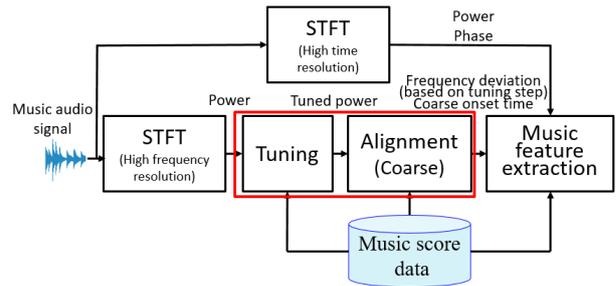


図 1: 詳細解析ブロック図. 高周波数分解能チューニング及び楽譜追跡情報を利用し詳細解析を行う

本周波数とその各倍音にピークが立つことが知られており, その周波数特性を考慮すれば, 音響信号に対し短時間フーリエ変換 (STFT) して得られたパワースペクトル系列(ただし本研究では A4 (= 440Hz) を基準とした対数周波数軸変換 (MIDI セント (c) 変換) を行った上で, 半音を任意の数だけ分割し三角窓を掛けている) に対し時間方向に総和をとって得られた対数周波数分布 $y(c)$ は, 最も低い音高(本研究では音高 A0 の基本周波数 $\mu_0 (= 27.5\text{Hz} = 2100c)$) 及び半音ごとにピークを持つような楕型の分布 $\mathcal{T}(c)$ とよく合致するはずである。本研究では次の 2 手法についてチューニング補正機能を比較検討した。

2.2 楕型形状最適化

セント軸においてピークは半音 (100 セント) 間隔で立つことから, 周波数分布 $y(c)$ の逆数を用いて,

$$\arg \min_d \sum_{m=0}^{87} \frac{1}{y(\mu_0 + 100m + d)} \quad (1)$$

として定式化することで, 差分 d が求められる。

2.3 混合正規分布 (GMM) 近似

楕型分布 $\mathcal{T}(c)$ を混合正規分布として近似すると,

$$\mathcal{T}(c) = v_c \sum_{m=0}^{87} \pi_m \mathcal{N}(\mu_0 + 100m + d, \sigma^2) \quad (2)$$

ただし, v_c は各セントにおけるエネルギー, π_m は第 m 倍音における混合係数であり, 分散 σ^2 は $2\sigma = 50\text{cent}$ とした. 以上より,

$$\arg \min_{\mathbf{v}, \pi, d} \sum_c |y(c) - \mathcal{T}(c)|^2 \quad (3)$$

とすることで差分が求められる.

3 楽譜とのマッチング

3.1 定式化

ある音楽音響信号がピアノの楽譜に基づき演奏された結果であるとするならば, 各時刻のパワースペクトルは楽譜に記載された音高情報に基づき生成されたものであると仮定できる. また各単音に対しそれぞれの(対数)周波数分布をモデル化可能ならば, 和音表現は単音の重ね合わせ, すなわち単音周波数分布の混合として近似できる(ただし厳密にはパワースペクトルの加法性は成り立たない).

従って, 四分音符や八分音符といった音価に対応する音長を継続長確率, 音高遷移(パターン)を遷移確率としてモデル化するならば, 音楽音響信号は楽譜上の和音に対応した混合分布を出力として持つ隠れセミマルコフモデル(HSMM)として近似できる.

3.2 観測系列の近似

短時間フーリエ変換(STFT)により得られるパワースペクトルを直接観測系列とした場合, 周波数分布の次元が高すぎるために次元の呪いの影響を過度に受けるため, 適切に次元削減を行う必要がある. 音楽音響信号において最も簡易かつ代表的な手法としては12次元のピッチクラスに集約したChroma vector[3]が考えられるが, オクターブ情報を極力保存することを考えると, 一般的なピアノにおける最大出現音高数88次元を確保することが望ましい. 本研究ではパワースペクトルに対し, 半音単位でピークを持つ三角窓を施した88次元のパワースペクトル系列を適用しChroma vectorと合わせて比較評価した.

4 評価実験

4.1 実験条件

チューニング機能の有無に対する楽譜追跡精度及び, 従来楽譜追跡に使用されてきたDTWとHSMMによる比較評価を目的とし, それぞれの条件について正解オンセット時刻に対する推定精度評価を行った. 実験は楽譜演奏としてClassical Archivesから, また実演奏としてInternational Piano-e-Competition database

表 1: 楽譜追跡 (オンセット時刻) 推定精度 (%)

	DTW	HSMM
No-tuned chroma	32.5	40.8
Tuned chroma	68.6	74.2
No-tuned 88dim	31.7	36.5
Tuned 88dim	70.5	75.4

より, 合計10曲, それぞれ曲冒頭の45–60秒程度に対し, +75セントの補正を加えた上で, サンプリングレート44.1kHzでWAVE変換後, 窓幅4096, ハーフオーバーラップでSTFTを行った. チューニングテストには半音を4, 8, 16分割した分解能で式(1)と式(3)に基づき比較したところいずれも同等の性能を確認したため, 計算速度に勝る楕円形状最適化を適用した.

4.2 実験結果

表1に比較評価した実験結果を示す. "No-tuned", "Tuned"はそれぞれチューニング補正の有無を表し, "chroma"と"88dim"は観測系列で, それぞれChroma vectorと88次の三角窓を掛けたものを示す. 比較実験の結果チューニングの有無は推定精度に大きく依存し, またDTWよりHSMMのほうが高い正答率を示したが, 各時刻のパワー系列のピーク検出等に基づき継続長を推定し初期パラメータとすることでDTW, HSMMはいずれも精度向上を期待できる.

5 おわりに

本稿では演奏詳細解析に向けて, 音楽音響信号と楽譜とのチューニング機能を有する和音単位での楽譜追跡手法について述べた. 評価実験の結果, チューニング補正機能及びHSMMの有効性が示された. 今後は本稿の結果に基づき詳細解析を行い, 実演奏解析データベースを構築し, 自動演奏表情付与学習の実現を目指したい. なお, 本研究はJSPS科研費17H00749の助成を受けて行われた.

参考文献

- [1] 保利武志 *et al.*, "多重解像度NMFに基づく音響信号演奏詳細解析," 情報処理学会研究報告音楽情報科学 (MUS), vol. 2018, no. 16, pp. 1–6, 2018.
- [2] T. Hori *et al.*, "Multiresolutional hierarchical bayesian NMF for detailed audio analysis of music performances," in *Proceedings of APSIPA ASC*, 2018.
- [3] T. Fujishima, "Real-time chord recognition of musical sound: A system using common lisp music," *Proceedings of ICMC*, pp. 464–467, 1999.