

# SVMによる頑健なピアノ演奏の音高推定

総合研究大学院大学 複合科学研究科 統計科学専攻

今村 武史

## 1 まえがき

本研究では、SVM (Support Vector Machine) を用いて、頑健にピアノ演奏の音高推定を行う方法を提案する。近年、インターネット上などで大量の音楽データが利用できるようになり、楽曲の検索や音楽データからの情報抽出の技術が注目されている。ピアノ演奏の音高推定はその技術の一つであり、様々な先行研究が報告されている。その一つにSVMに基づく方法 (Poliner and Ellis 2007, Costantini, Perfetti and Todisco 2009) があり、音楽に関する知見を利用することなく、高い精度で音高推定を行えるが、低音域、高音域では十分な学習データが得られず推定精度が低くなるという問題があった。また各音高についてSVMを構成するため、学習のための計算量が大きいという問題があった。本研究では定Qスペクトルを用い、音高間で共有できる学習データを作成することでこれらの問題を解決することを試みた。

## 2 SVMによる音高推定

### 2.1 先行研究

先行研究Poliner and Ellis 2007では特徴量として短時間フーリエ変換によるスペクトルを使用しており、音高毎に作成した学習データセットによって各音高の有無を識別するSVMの学習を行い、One vs All識別器を構成する (Scholkopf and Smola 2001)。またCostantini, Perfetti and Todisco 2009では定Qスペクトルを特徴量としているが、学習データやSVMはPolinerらと同様に音高毎に準備している。

### 2.2 提案法

短時間フーリエ変換によるスペクトルの場合、音高が変われば各高調波を示すスペクトルピークの間隔が変化してしまう。一方、定Qスペクトルの場合は音高が変わるとスペクトルピークの全体的な位置は変化するが、各スペクトルピーク間の相対的な位置関係は変化しない(図1)。従って、定Qスペクトルを音高に応じて周波数方向にシフトすることで、各スペクトルピークの位置を一致させることができ、シフト後の定Qスペクトルを異なる音高間で学習データとして共有することが可能となる。

これらの学習データによって学習を行ったSVMに、推定対象の音高に対応した周波数方向シフトを施した採譜用楽曲の定Qスペクトルを与えることで、異なる音高の有無の推定を単一のSVMで行うことができる。ただし、大きく離れた音程の音高ではスペクトル包絡の形状や、観測できる高調波数が変化してしまうので、各オクターブ毎に学習データとSVMを準備する。

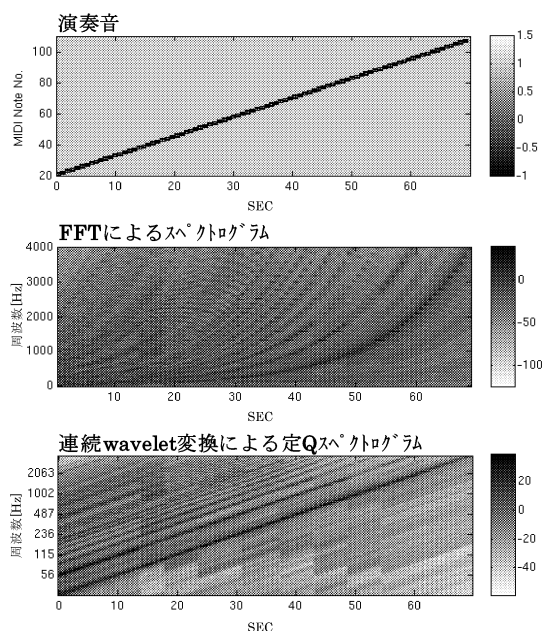


図1. 通常のスペクトログラムと定Qスペクトログラム

### 3 実験

提案法と従来法 (Poliner and Ellis 2007) の比較実験を行った。ただし従来法ではSVMによる音高推定の後、更にHMMによる時間方向の平滑化を行って採譜精度の向上を図っている。ここではSVMによる音高推定の性能を比較する。

#### 3.1 実験データ

実験データはMIDIデータ (Classical Piano Midi Pageより入手) から作成したデータ (合成音) と、自動演奏ピアノによる演奏データ (Real Piano, Automatic Piano Transcription (Lab ROSA)より入手) を用いた。従来法では窓長1024点、シフト数80点の短時間フーリエ変換によってスペクトルを求め、各スペクトルピークの高さを揃えるための正規化を施したものを特徴量とした。提案法ではGabor関数をマザーウェーブレットとする連続Wavelet変換によって構成した1/48octフィルタバンクにより求めた定Qスペクトル (Q=34.6) を、各フレームのノルムで正規化して特徴量とした。

#### 3.2 実験結果

図2に合成音に対する各手法の音高別推定精度を示す。提案法によって低音域、高音域の精度が改善された。同様の傾向はReal Pianoでも確認された。各データセットに対する提案法、従来法の平均音高推定精度を表1に示す。Real Pianoでは提案法の性能は若干低くなっていたが、合成音では大幅に性能を向上させることができ、結果としてテストデータ全体の推定精度を向上させることができたことが分かる。また学習時間についてはSVMの数が減ったことで、従来法では全音高で1週間程度かかっていた学習時間を2日程度に短縮することができた。

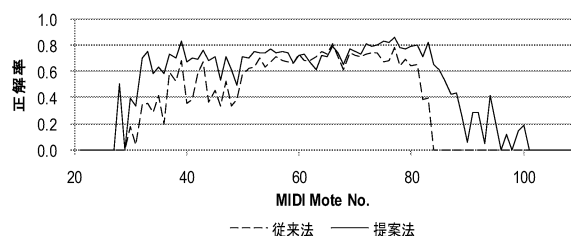


図2. 合成音に対する音高別正解率

表1. 全テストデータに対する正解率

	従来法	提案法
合成音 (14 曲)	61.46%	71.48%
Real Piano (10 曲)	51.17%	47.78%
テストデータ全体	57.17%	61.61%

### 4 まとめ

異なる音高間で学習データセット、SVMを共有し、音高推定精度の改善、および計算時間の短縮を実現することができた。今後はフレーム間の依存関係にも着目して更に推定精度の向上を目指していきたい。

### 参考文献

- G. Poliner and D. Ellis (2007). A discriminative model for polyphonic piano transcription, *EURASIP Journal on Advances in Signal Processing*,8, 1–9.
- G. Costantini, R. Perfetti and M. Todisco (2009). Event based transcription system for polyphonic piano music, *Signal Processing*,89,1798–1811
- B. Scholkopf and A. J. Smola(2001). Learning with kernels, The MIT Press.
- Classical Piano Midi Page, <http://www.piano-midi.de/>
- Automatic Piano Transcription, LabROSA, <http://labrosa.ee.columbia.edu/projects/piano/>