

ウェーブレット変換を用いた採譜システム

深山幸穂, 日野祐志, 伊藤里美
 新居浜工業高等専門学校

マイクロフォンで受信した音楽信号より音符を表示する採譜システムにおいて、まず、直交ウェーブレット変換によるブレイクポイント検出を用いて音符の長さを把握し、次に、各ブレイクポイント間にガボールウェーブレット変換を用い、楽器の種類による倍音構成を観測課程として扱った適応状態推定により和音を推定する構成を特徴とするアルゴリズムを提案した。

A Music Transcription System Applying Wavelet Transform

Yukio Fukayama, Yuji Hino and Satomi Ito
 Niihama National College of technology

An algorithm for music transcription system that listens to sounds and displays notes of the corresponding tones has been proposed. The algorithm features a two-stage processing that observes note length at first, and then, pitch names. The former stage effectively detects break points caused by starting of next note or frequency hop at pitch name changing with dyadic wavelet transforms. The latter stage analyzes component of each pitch name on the interval of the adjoining break points with Gabor wavelet. The algorithm is applied adaptive state estimation technique to cope with tones including considerable harmonics.

1. はじめに

本研究による採譜システムは図1に示す通り、マイク等で入力した音楽信号から、音の長さや音名、和音を推定し、音符や休符に表す。

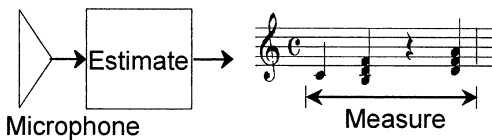


Fig.1 Music Transcription

本報で紹介するシステムにおいては、基準ピッチ、 $A = 442[\text{Hz}]$ 、とし、12音平均率により各音名間に次式の関係式を用いている。

$$f_{l+1} = 2^{\frac{1}{12}} f_l \cong 1.0594631 f_l \text{ (chromatic interval),}$$

$$f_{l+12} = 2 f_l \text{ (octave) for } l \in \mathbf{Z}. \quad (1)$$

本報では、音符の長さの把握法と、楽器の種類による倍音構成の相違を考慮し和音の推定法を中心に述べる。

2. アルゴリズムの構成

本研究では、初期の段階において、ウェーブレットパケットを用いる採譜システムの検討を行った。これは、ウェーブレットパケットに、例えば、C, D, E音を順に入力し、図2に示すとおり、パケットの各成分が音名に応じて時間変化することを用いていた[1]。

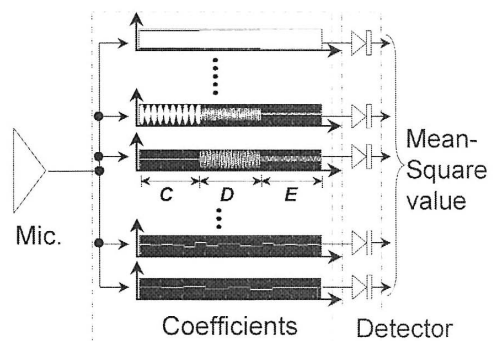


Fig.2 Decomposition by Wavelet Packet

このとき、ウェーブレットパケットの各成分の実効値（二乗平均値の平方根）について、最大の実効値から音の大きさと長さを知り、各実効値と最大の実効値との比率のパターンから直交射影を用いて音名の推定を行った。本例では、ウェーブレットパケットとしてDaubechiesの3次[5]を用い、サンプリング周波数、8192[Hz]、二乗平均を求める時間間隔、1/128[s]、としている。

しかしながら、図2の構成による採譜システムを筆者らは1段階処理と称しており、この方法では二乗平均値を求める時間幅に起因して、音符の長さの識別の精度が低下する問題が生じる。例えば、図3に示すとおり、C音とD音を3：1の長さで入力しても、識別された音の長さは4：2となる場合が生じる。

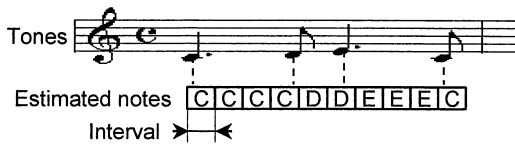


Fig.3 One-stage Processing

以上のような音の長さの識別に問題に対処するため、本研究では図4に示す2段階処理でアルゴリズムを構成することにした。この方式では、まず、2進(dyadic)ウェーブレット変換による音楽信号や、その時間微分の不連続点(break point)として、音高や和音の変化を検出して音符の長さを把握する。次に、音楽信号の2つの不連続点に挟まれた部分に連続ウェーブレット変換を適用して音名や和音の推定を行う構成である。

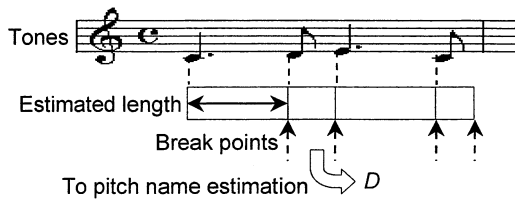


Fig.4 Two-stage Processing

3. 音符の長さの把握

3.1 2進ウェーブレット変換の適用

新たに開始する音や、音高の変化でもたらされる、音楽信号(時間関数)もしくは、その1、2階程度の時間微分の不連続点は、その時間間隔より音符の長さの把握に応用できる。このような不連続点の検出には、2進ウェーブレット変換が好適と知られている[2]。

2進ウェーブレット変換、 $X(a,b)$ 、および、対応する時間関数、 $x(t)$ 、について、次の関係が知られている。

$$X(a,b) = \int_{-\infty}^{\infty} x(t)\phi_{a,b}^* dt, \quad (2)$$

$$\phi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \phi\left(\frac{t-b}{a}\right), \quad (3)$$

ここに、 $\phi(t)$ はアドミッシブル条件を満たすマザーウェーブレットであり、2進ウェーブレット変換であれば、さらに次式の関係がある。

$$(a,b) = (2^{-j}, m2^{-j}) \quad (j,m \in \mathbf{Z}). \quad (4)$$

いま、マイクロフォンで受信した音楽信号、 $\eta(t)$ 、が楽音、 $s(t)$ 、と雑音、 $n(t)$ 、より成っているとす。

$$\eta(t) = s(t) + n(t), \quad (5)$$

ここに、 $s(t)$ は時刻、 $t = t_i$ 、において不連続点を持ち、 $n(t)$ は周波数領域においてパワースペクトラムが有界な2次モーメントを持つと仮定する。

このとき、2進ウェーブレット変換、 S および N 、と対応する時間関数、 $s(t)$ および $n(t)$ 、には次の関係式が知られている[2]。

$$\begin{aligned} \inf |S(2^{-j}, m2^{-j})| \\ \geq 2^{\frac{j}{32}} \beta (|s(t_i+0) - s(t_i-0)|), \end{aligned} \quad (6)$$

$$\max E |N(2^{-j}, m2^{-j})|^2 \leq 2^{\frac{31j}{16}} c, \quad (7)$$

ここに、下限および最大値は不連続点が次の閉区間に存在するとして評価されている。

$$t_i - 2^{-(j+1)} \leq m2^{-j} \leq t_i + 2^{-(j+1)}. \quad (8)$$

さらに、関数、 β 、はマザーウェーブレット、 $\phi(t)$ 、に依存し、定数、 c 、は雑音、 $n(t)$ 、の

ワースペクトラムに応じて与えられる。

従って、(6)式より、 S の下限が、(7)式より、 N の最大値が抑えられることより、適切にマザーウェーブレットを選択すれば、2進ウェーブレット変換により、音楽信号の不連続点が検出できることになる。

3.2 マザーウェーブレットの選定

不連続点の検出に好適なマザーウェーブレットの例として、図5に Coiflet の5次 [5] の Coiflet を用い、Detail-3成分の変化を示す。この例では、C major の典型的なカデンツ(Tonic, Sub-dominant, Tonic, Dominant, Tonic)を計算機より与えており、楽器のダイナミクスによる過渡現象(波形のなまし効果)を模擬するため、時定数、 $T=0.05$ [s]、の一次遅れを通して

本例より、和音の切替りによる時間微分の不連続点が、Detail-3成分のピークとして良好に検出できることを確認できる。



Fig.5 Simulation of Break Points Detection (Coif 5th, Detail3, $T=0.05$ [s])

Table 1 Performance in Break Point Detection




Wavelet	Result at B.P.	Peak/Average		
		T=0	0.01	0.05
coif(5th)		398	143	106
db(6th)		34.6	10.9	7.5
sym(8th)		98.8	35.5	26.3

表1に類似の検討結果纏めた。ここでは、図5に紹介したカデンツについて、不連続点の検出性能を、Coiflet の5次、Daubechies の6次、Symmetric の8次、について、すべて Detail-3成分を比較している。いずれのケースにおいても、楽器のダイナミクスを模擬する時定数、 T 、を変化させながら、不連続点に対応するピークの高さと、当該信号の2乗平均値の平方根との比を示している。この表の比較から、不連続点の検出には、Coiflet の5次の Detail-3成分が利用できることが分かる。

さらに、Coiflet の5次の Detail-3成分の実用性を確認するため、図6にキーボード(CASIO CTK-130)のバイオリン音を与えた例を示す。演奏者が参照した譜面に応じて、音符の長さを把握でき、良好な結果と考えられる。



Fig.6 Break Point Detection from Keyboard (Vivaldi: Violin Concerto in E major, "La primavera")

4. 和音の推定

4.1 ガボールウェーブレットの適用

隣接する不連続点の間から切出された信号、 $u(t)$ 、のウェーブレット変換、 $U(a,b)$ 、をガボールウェーブレット、 $\psi(t)$ 、を用いて、時間周波数平面、 $(a,b)=(\omega_0/2\pi f_1, b)$ 、への射影として求める。:

$$U(a,b) = \int_{-\infty}^{\infty} u(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a} \right) dt, \quad (9)$$

$$\psi(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ j\omega_0 t - \frac{1}{2} \left(\frac{t}{\sigma} \right)^2 \right\}, \quad (10)$$

ここに、周波数、 $\dots, f_0, f_1, f_2, \dots$ は音名、 $\dots, A, A\#, B, C, C\#, D, \dots$ に対応し、通常、 $f_0=442$ [Hz]、として、(1)式の関係がある。

観測ベクトル、 $\mathbf{y} = (y_0 \dots y_{n-1})^T \in \mathbf{R}^n$ 、はガボールウェーブレット変換の各成分の絶対値、 $|U_k|$ 、を次のように、最大値を1に規格化して与える。

$$y_k = \frac{|U_k|}{\max |U_k|}, \quad (11)$$

$$U_k = U(\frac{\omega_0}{2\pi f_k}, b), \quad (k = 0, \dots, n-1). \quad (12)$$

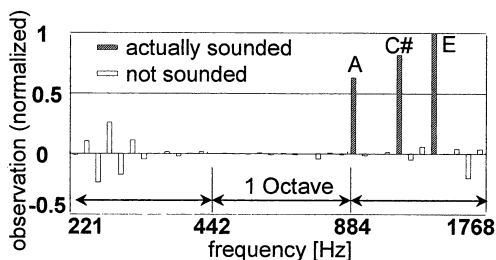


Fig.7 Components of Observation Vector

図7は観測ベクトルの成分、 y_0, \dots, y_{n-1} 、をA-C#-Eの和音(tonic in A major)を与えた場合を示しており、例えば、0.5程度の適切な閾値の設定により、ガボールウェーブレット変換を用いて、和音が推定できることが分かる。

Table 2 Observation Performance (Flute tone)

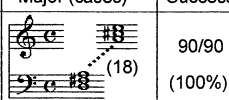
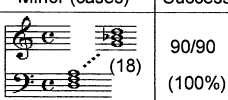
Major (cases)	Success	Minor (cases)	Success
 (18)	90/90 (100%)	 (18)	90/90 (100%)

表2はキーボード(CASIO CTK-130)のフルート音を用い、majorおよびminor、それぞれ18種類の三和音を5回づつ与えた場合において、ガボールウェーブレット変換により間違いなく和音が推定できることを確認した例である。

4.2 倍音成分の強い楽器への対処

よく知られているように、フルートの音は正弦波に近く、含まれている倍音成分も小さいが、例えば、トランペットでは倍音成分を多く含む。

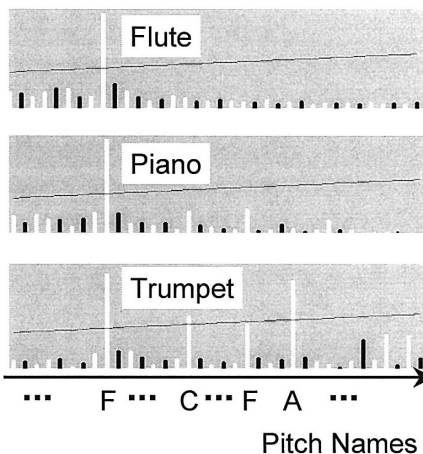


Fig.8 Harmonics of single tone in F

図8は前節に述べた機能に、フルート、ピアノ、トランペットのFの単音を入力した例を示す。この例から、フルート、ピアノにおいては、図中に示した閾値により、実際に演奏している音を推定できるが、トランペットでは倍音成分により、演奏していない音を区別できない。

この問題に対処するため、以下、状態推定の手法を適用する。まず、隣接した不連続点(音高や和音の変化点)の間を、 k 番目の状態として考える。各状態は確率変数ベクトル、 $\mathbf{v} \in \mathbf{R}^n$ 、で駆動され、その統計量は零平均、共分散行列、 $\mathbf{V}(k) \in \mathbf{R}^n$ 、で与えられ、i.i.d.(independent, identically distributed)特性に従うと仮定する。

いま、状態間には逐次に、次式のシステム方程式が成立すると考えられる。

$$\mathbf{x}(k+1) = \mathbf{F}(\theta_1)\mathbf{x}(k) + \mathbf{v}(k), \quad (13)$$

ここに、 $\mathbf{x} = (x_0 \dots x_{n-1})^T \in \mathbf{R}^n$ は状態ベクトルであり、その成分は音名、 $\dots, A, A\#, B, C, C\#, D, \dots$ において実際に演奏されている楽音が存在する確からしさを表す。さらに、状態遷移行列、 \mathbf{F} 、は音高や和音の変化の遷移確率

を表し、調性を示すパラメータ、 θ_1 、に依存する。ただし、調性が不明な場合、 F として零行列を与えても、以下の議論は成立する。

観測ベクトル、 $\mathbf{y} = (y_0 \ \dots \ y_{n-1})^T \in \mathbf{R}^n$ 、は(11)、(12)式で与えられ、状態ベクトルとは次式の観測方程式で関係づけられる

$$\mathbf{y}(k) = \mathbf{H}(\theta_2)\mathbf{x}(k) + \mathbf{w}(k), \quad (14)$$

ここに、観測雑音、 $\mathbf{w} \in \mathbf{R}^n$ 、は零平均、共分散行列、 $\mathbf{W}(k) \in \mathbf{R}^n$ 、のi.i.d. 特性と仮定し、観測行列、 \mathbf{H} 、は楽器の種類を示すパラメータ、 θ_2 、に依存する。観測行列の各成分は、図8に示すように、状態ベクトルの各成分を単音ずつ順に鳴らした際の観測ベクトルの値を、楽器ごとに整理して与える。

4.3 状態推定の応用

求める状態ベクトルの推定値、 $\hat{\mathbf{x}}$ 、は適当な初期値、 $\hat{\mathbf{x}}(0|-1) = \mathbf{x}_0$ 、を仮定することにより、次式のカルマンフィルタ[6]を用いて、 $k=1, 2, \dots$ と逐次に求められる

$$\begin{aligned} \hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k|k-1)\tilde{\mathbf{y}}(k|k-1), \\ \hat{\mathbf{x}}(k|k-1) &= \mathbf{F}(\theta_1)\hat{\mathbf{x}}(k-1|k-1), \end{aligned} \quad (15)$$

ここに、

$$\hat{\mathbf{x}}(k|m) = (\hat{x}_0(k|m) \ \dots \ \hat{x}_{n-1}(k|m))^T \in \mathbf{R}^n$$

観測ベクトル、 $\mathbf{y}(0), \dots, \mathbf{y}(m)$ 、に係わる σ -集合体に基づいて得られる状態ベクトル、 $\mathbf{x}(k)$ 、の推定値である。

このとき、時点、 k 、におけるイノベーション、 $\tilde{\mathbf{y}}(k|k-1) \in \mathbf{R}^n$ は次式である。

$$\begin{aligned} \tilde{\mathbf{y}}(k|k-1) &= \mathbf{y}(k) - \mathbf{H}(\theta_2)\hat{\mathbf{x}}(k|k-1) \end{aligned} \quad (16)$$

さらに、カルマンゲイン、 $\mathbf{K} \in \mathbf{R}^{n \times n}$ 、はイノベーションの直交条件より次式となる。：

$$\begin{aligned} \mathbf{K}(k|k-1) &= \mathbf{P}(k|k-1)\mathbf{H}^T(k|k-1) \\ &\times [\mathbf{H}(\theta_2)\mathbf{P}(k|k-1)\mathbf{H}^T(\theta_2) + \mathbf{W}(k)]^{-1} \end{aligned} \quad (17)$$

また、推定共分散行列、 \mathbf{P} 、は次のとおり逐

次に求められる。

$$\begin{aligned} \mathbf{P}(k|k-1) &= \mathbf{F}(\theta_1)\mathbf{P}(k-1|k-1)\mathbf{F}^T(\theta_1) \\ &+ \mathbf{V}(k), \end{aligned} \quad (18)$$

$$\begin{aligned} \mathbf{P}(k|k) &= [\mathbf{I}_n - \mathbf{K}(k|k-1)\mathbf{H}(\theta_2)]\mathbf{P}(k|k-1), \end{aligned} \quad (19)$$

ここに、適当な初期条件、 $\mathbf{P}(0|-1) = \mathbf{P}_0$ 、を仮定する。

4.4 状態推定の実施例

図9に鍵盤型ハーモニカ(YAMAHA ピアニカ P-32D)を用い、単音、C、を入力した場合について状態推定の例を示す。本例では、図9の下段に示すとおり、 $\mathbf{y} = (y_0 \ \dots \ y_{n-1})^T$ には倍音成分を含むが、状態推定により、正しい推定値、 $\mathbf{x} = (x_0 \ \dots \ x_{n-1})$ 、が得られている。

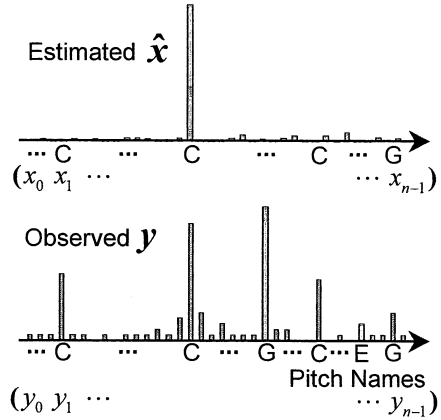


Fig.9 State estimation in a single tone case

さらに、図10は、同一の鍵盤ハーモニカを用いて、音名、E, G, Cの三和音を入力した場合を示す。図9の場合と同様に、正しい推定値が得られている。なお、図9、図10の例では共に(13)式の状態遷移行列、 \mathbf{F} 、として零行列を与えた。

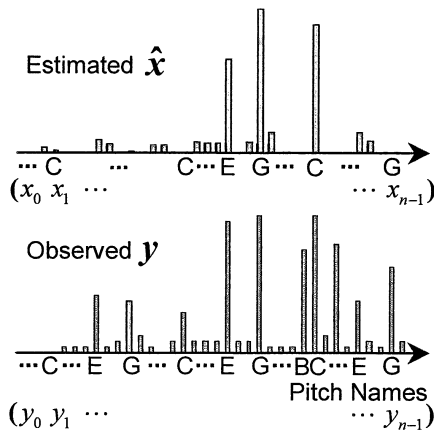


Fig. 10 State estimation in a chord case

5. 結論

次の特徴を持つ採譜アルゴリズムを提案した。

- 1) 音符の長さの把握に，Coiflet の 5 次の Detail-3 成分に着目した 2 進ウェーブレット変換が不連続成分が効果的である。
- 2) 和音の判別にガボールウェーブレット変換が効果的である。
- 3) 倍音を含む楽器による和音の推定にカルマンフィルタを用いた状態推定が効果的である。

参考文献

- [1] Y. Fukayama: A study of transcript algorithm for multi-voice music applying wavelet packet; Proc. of 34th Intl. Symposium on Stochastic Systems Theory and Its Applications, pp.48-53, ISCIE, Fukuoka JP, 2002.
- [2] Z. Xie and T. Suzuki: Introduction to Wavelet and Stochastic Processes (in Japanese); Uchida Rokakuho, Tokyo JP, 2002.
- [3] G. Strang and T. Nguen: Wavelet and Filter Banks; Wellesley-Cambridge, Wellesley MA, 1996.
- [4] M. V. Wickerhauser: Adapted Wavelet Analysis from Theory to Software; IEEE, Piscataway NJ, 1994.

[5] I. Daubechies: Ten Lectures on Wavelet; SIAM, Philadelphia PA, 1992.

[6] T. Katayama: Kalman Filter and Its Application (in Japanese); Asakura, Tokyo JP, 1983.

[7] N. Kennedy: The Oxford Dictionary of Music; Oxford Univ., Oxford UK, 1983.