

## 時間周波数平面上的倍音パターンと自己相関 に着目した採譜システム

深山幸穂<sup>†</sup> 片岡友徳<sup>††</sup> 田中大介<sup>††</sup> 増田玲往<sup>†</sup>

音楽信号をガボールウェーブレット変換により時間周波数平面に射影し、各楽器の倍音パターンを考慮した音高の推定と、信号強度の自己相関によるテンポと音符の長さの推定を特徴とする採譜システムを紹介する。

### A music transcription algorithm referring to harmonics patterns and correlations on the time-frequency plane

Yukio Fukayama<sup>†</sup>, Tomonori Kataoka<sup>††</sup>,  
Daisuke Tanaka<sup>††</sup> and Reo Masuda<sup>†</sup>

An algorithm for music transcription system that projects music signal onto the time-frequency plane applying Gabor wavelet transform shall be introduced. The system is also applying adaptive state estimation technique for actually played pitch names to cope with tones including considerable harmonics. In addition, the system identifies tempo and note lengths with referring autocorrelation of magnitude on the time-frequency plane.

### 1. はじめに

採譜システムでは一般に、マイク等で入力した音楽信号の強度を時間周波数平面に投影して演奏された音名や和音を推定し、その継続時間から音符や休符の長さを求めている。よく知られているように、楽器の音色は倍音（高調波）の構成で特徴付けられる。例えば、ヘ音記号が示す F 音の単音をフルート、ピアノ、トランペットで演奏した場合について、2 節に述べるガボールウェーブレット変換を用い、ある時点における周波数軸上の強度を図 1 に示す。本図では、半音ごとの白鍵と黒鍵を示す横軸に、左側より実際に演奏した F 音、大譜表中央の C 音を順に目盛っている。

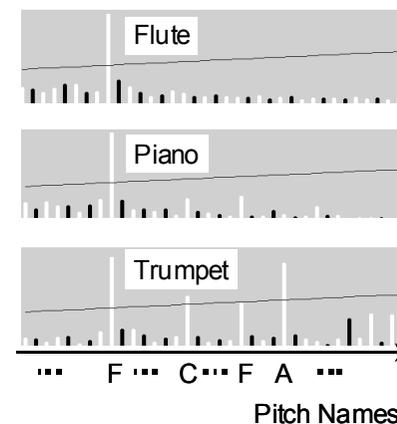


Fig.1 Harmonics of single tone in F

この例では、フルート、ピアノについては、図中に示した閾値により、実際の演奏

<sup>†</sup> 新居浜工業高等専門学校電子制御工学科  
Electronic Control Department, Niihama National College of Technology  
<sup>††</sup> 新居浜工業高等専門学校専攻科  
Advanced Study Courses, Niihama National College of Technology

音を推定できるが、トランペットでは実際に演奏していない倍音との区別が困難である。さらに、この楽器の倍音列が基音、2倍、3倍の高調波の順に、F, F, C, F, A, C, ... 音となっていることを考慮すれば、図1のトランペットの演奏音は3倍の高調波に相当し、公知のF0(基音)推定[5]の手法でも演奏音を検出できないと考えられる。また、最大の強度を示す音に着目する手法では、本質的に多重音の採譜に利用できないし、3節に述べるとおり、ピアノの低音弦では演奏した基音より2倍の高調波の強度が大きい場合があり、これに対処不可能である。

本報で紹介する採譜システムは、上述した背景の下、ピアノのように特徴のある倍音パターンを有する楽器の多重音の採譜への適用を主眼に開発を進め、音楽信号の強度をガボールウェーブレット変換により時間周波数平面に投影し、各楽器の倍音パターンを考慮した演奏音の推定と、信号強度の変化の自己相関に着目したテンポと音符の長さの判別を特徴としている。

## 2. 時間周波数平面への投影

本研究では、ガボールウェーブレット変換[6]により、時間領域の信号  $u(t)$  を時間周波数平面  $(a, b) = (f_0/f, b)$  上の複素量  $U(a, b)$  として投影する。なお、その絶対値  $|U(a, b)|$  は当該平面上における信号の強度(パワー密度)に相当する：

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

ここに、周波数  $f$  および時間推移  $b$  である。また、次式のマザーウェーブレット  $\psi(t)$  を基本周波数  $f_0 = 8.1758[\text{Hz}]$  として用い、式中の分散パラメータ  $\sigma$  は時間と周波数の分解能を調整する：

$$\psi(\tau) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ j2\pi f_0 \tau - \frac{1}{2} \left( \frac{\tau}{\sigma} \right)^2 \right\}. \quad (2)$$

また、当該平面の周波数軸は、MIDI標準[4]のノート番号  $l \in \mathbb{N}$  に対応する周波数  $f_l [\text{Hz}]$  は、国際標準ピッチ(1939)に基づく12音平均率の半音階[3]に従う：

$$f_l = 440 \times 2^{\frac{l-69}{12}}, \quad (3)$$

例えば、大譜表の中央のC音、および、その直ぐ上側の ... , A, A#, B, ... 音は、順に、 $f_{60} = 262, \dots, f_{69} = 440, f_{70} = 466, f_{71} = 494, \dots [\text{Hz}]$ , である。

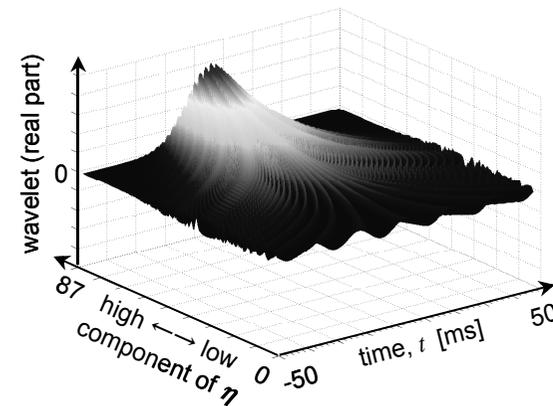


Fig.2 Gabor wavelet (real part)

いま、この時間周波数平面への投影を観測プロセスとみなし、次式の成分よりなる観測ベクトル  $\eta(k) = (\eta_0(k) \ \dots \ \eta_{p-1}(k))^T \in \mathbb{R}^p$  を定義する

$$\eta_i(k) = |U(2^{\frac{i+l_0}{12}}, k\Delta t)|, \quad (4)$$

ここに、信号の強度をノート番号  $l_0 \leq l \leq l_0 + p - 1$  の範囲で時点  $t = k\Delta t$ , ただし、サンプリング間隔  $\Delta t$  にて評価する。

図 2 に (1) 式において信号  $u(t)$  との内積を求めるウェーブレット  $\psi(t/a)/\sqrt{a}$ , ( $b=0$ ) の実部を, 観測ベクトルの成分  $\eta_i$  に対応させて表示した. ここに,  $a=2^{-\frac{i+l_0}{12}}$ ,  $l_0=21$  すなわち  $f_{i+21}=440 \times 2^{\frac{i-48}{12}}$  [Hz], ( $0 \leq i \leq 87$ ) と選んでおり, ピアノの 88 鍵盤の音高に対するウェーブレットの概形が把握できる.

### 3. 楽器の倍音パターンを考慮した演奏音の推定

1 節で述べたように, 本報では特徴のある倍音パターンを有する楽器の多重音の採譜に対処するため, 以下, 状態推定の手法を適用している. まず, 時点  $k$  において状態ベクトル  $x(k) = (x_0(k) \cdots x_{n-1}(k))^T \in \mathbf{R}^n$  を考え, その成分は当該楽器の音域に属する音名  $\cdots, A, A\#, B, C, C\#, D, \cdots$  において, それらの音が実際に演奏されている確からしさとする. さらに, この状態ベクトルは零平均, 共分散行列  $V(k) \in \mathbf{R}^n$  の白色雑音  $v(k) \in \mathbf{R}^n$  で駆動されている仮定する. このとき, 次式の状態方程式を記述できる:

$$x(k+1) = F(\theta_1)x(k) + v(k), \quad (5)$$

ここに, 状態遷移行列  $F(\theta_1) \in \mathbf{R}^{n \times n}$  は音名や和音の変化の遷移確率を表し, 調性を示すパラメータ  $\theta_1$  に依存する. ただし,  $F$  の詳細が不明な場合の扱いは後述する.

一方, 本節では 2 節に述べた  $\eta(k)$  の成分を変数変換した観測ベクトル  $y(k) = (y_0(k) \cdots y_{p-1}(k))^T \in \mathbf{R}^p$ , ただし,  $y_i = \log \eta_i - \bar{y}_i$  を考える. これは, により, 信号  $u(t)$  のダイナミックレンジが広い場合や後述する閾値処理への便宜による. このとき, 次式の状態方程式が記述できる:

$$y(k) = H(\theta_2)x(k) + w(k), \quad (6)$$

ここに, 観測雑音  $w(k) \in \mathbf{R}^p$  は零平均, 共分散行列  $W(k) \in \mathbf{R}^{p \times p}$  で上述の

$v(k)$  とは互いに独立な白色雑音であり, 観測行列  $H(\theta_2) \in \mathbf{R}^{p \times n}$  は楽器の種類を示すパラメータ  $\theta_2$  に依存する. 状態ベクトル  $x$  の各成分に対応する単音を順に鳴らし, このときの観測ベクトル  $y$  の各成分の値を得るとき, 観測行列  $H$  の各成分は, 両ベクトルの成分間の相互相関である. 図 3 にグランドピアノについて, 観測行列の一例を示す. 前述したように, 低音弦では対角線上の成分で示される基音より 2 倍の高調波の強度が大きいことが特徴的である.

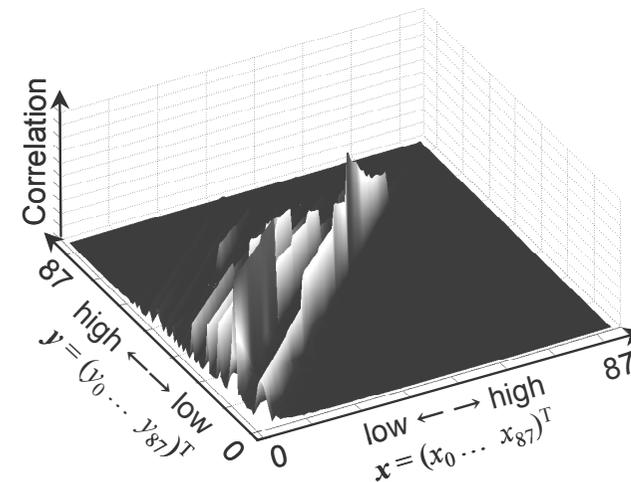


Fig.3 Observation Matrix (Piano)

いま, 信号  $u(t)$  の時間周波数平面上への射影から観測ベクトル  $y(i)$ , ( $i=0,1,\dots,l$ ) を得たとき, 実際に演奏した音の情報を含む状態ベクトル  $x(k)$  について, その推定値  $\hat{x}(k|l)$  は適当な初期値  $\hat{x}(0|-1) = x_0$  を仮定することにより, 以下に述べるカルマンフィルタ[2] を用いて  $k=1,2,\dots$  と逐次に求めることができる:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k|k-1)\tilde{\mathbf{y}}(k|k-1), \quad (7)$$

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{F}(\theta_1)\hat{\mathbf{x}}(k-1|k-1), \quad (8)$$

ここに、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。

$$\tilde{\mathbf{y}}(k|k-1) = \mathbf{y}(k) - \mathbf{H}(\theta_2)\hat{\mathbf{x}}(k|k-1). \quad (9)$$

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：

$$\begin{aligned} \mathbf{K}(k|k-1) &= \mathbf{P}(k|k-1)\mathbf{H}^T(\theta_2) \\ &\times [\mathbf{H}(\theta_2)\mathbf{P}(k|k-1)\mathbf{H}^T(\theta_2) + \mathbf{W}(k)]^{-1}. \end{aligned} \quad (10)$$

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。

$$\mathbf{P}(k|k-1) = \mathbf{F}(\theta_1)\mathbf{P}(k-1|k-1)\mathbf{F}^T(\theta_1) + \mathbf{V}(k), \quad (11)$$

$$\mathbf{P}(k|k) = [\mathbf{I}_n - \mathbf{K}(k|k-1)\mathbf{H}(\theta_2)]\mathbf{P}(k|k-1). \quad (12)$$

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

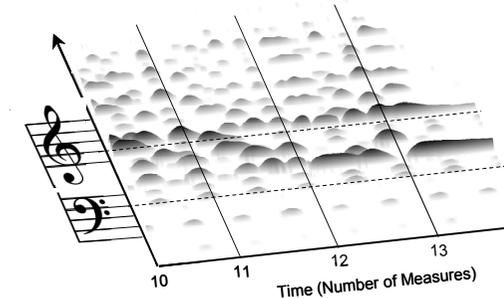
以上の議論において、観測雑音が小さく、共分散行列  $\mathbf{W}(k) \approx \mathbf{O}_{p \times p}$  と近似できれば、状態推定は次式の擬似逆行列に帰着し、状態遷移行列  $\mathbf{F}(\theta_1)$  と無関係になる。これは、前述したように  $\mathbf{F}(\theta_1)$  の詳細が不明な場合には好都合である。

$$\hat{\mathbf{x}}(k|k) = \mathbf{H}^T(\theta_2)[\mathbf{H}(\theta_2)\mathbf{H}^T(\theta_2)]^{-1}\mathbf{y}(k). \quad (13)$$

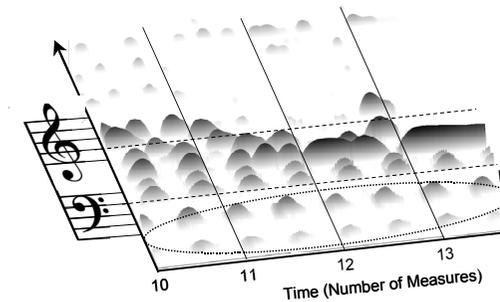
以上に述べた状態推定の手法の採譜システムにおける有効性を確認するため、図 4 には、ドビュッシー作曲「子供の領分」より「ゴリウォークのケーキウォーク」の第 10-13 小節をピアノで演奏し、これを 2 節に述べた手法で得た時間周波数平面上の観測ベクトル  $\mathbf{y}$  を上段 (a) に、さらに 3 節に述べた手法で推定した状態ベクトル  $\hat{\mathbf{x}}$  を中段 (b) に、また下段 (c) には演奏に用いた譜面を示している。状態推定により、(a) で顕著な倍音成分が (b) では抑制され、さらに、(a) では把握が困難な左手のバス

の左手のバスの動きが、(b) では楕円 B 部に示すように認識できる。

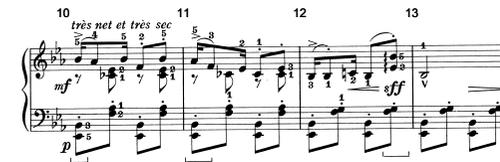
(a) Projection on the time-frequency plane:  $\{\mathbf{y}(k)\}$



(b) Estimated music:  $\{\hat{\mathbf{x}}(k)\}$



(c) Actually played music



10th-13th measure, Golliwog's Cakewalk  
C. Debussy: Children's Corner, 1908  
Fig.4 Estimation of actually played music 1

#### 4. テンポと音符の長さの推定

本報の採譜システムのテンポと音符の長さの推定法を、図 5 下段 (b) に示すドイツ民謡「かえるの歌」の筆者らによる簡単な編曲を用いて説明する。同図上段 (a) は下段の曲をピアノで演奏した信号から本報のシステムで採譜した結果である。

(a) Estimated music: (Tempo)=111.2



(b) Actually played music: (tempo)=112

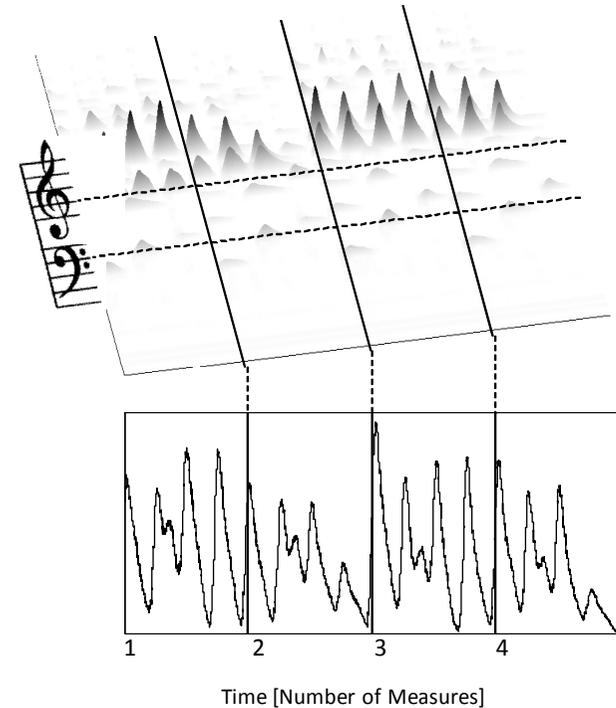


1st-4th measure, German folk song: Froshgesang (arranged)

Fig.5 Estimation of actually played music2

本例では、テンポ (四分音符) = 112/分 で演奏したところ、採譜システムによる推定値は (四分音符) = 111.2/分 で、ほぼ正確であり、推定された音符の長さは概ね妥当であるが、休符の検出に関して不正確な箇所がある。この問題については、今後、閾値設定の改良を行いたい。

以下、図 5 の結果に至る処理の概要を説明する。まず、ピアノ演奏による音楽信号を 2 節に述べた手法で時間周波数平面に投影すると図 6 の上段が得られる。これを周波数軸方向に加算すると、すべての音高をトータルした信号強度の時間変化が同図の下段に得られる。

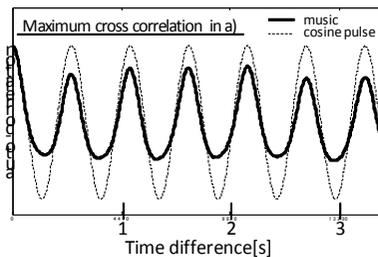


1st-4th measure, German folk song: Froshgesang (arranged)

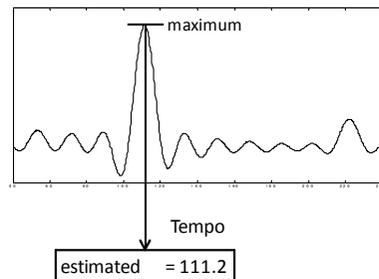
Fig.6 Projection on the time-frequency plane  
and its summation on the time axis

さらに、図 6 下段の信号の自己相関を求め、これを横軸に時間差をとって図 7 上段 (a) の太線に示す。同図では、細線に示すように、テンポに相当する周波数の余弦波信号を重ね、太線の信号と相関値を求める。同図下段 (b) では、余弦波信号の周波数に 60 を乗じてテンポに換算し、これを横軸にとって両信号の相関値の変化を把握する。

(a) Auto-correlation of music magnitude and superimposed cosine pulse



(b) Searching for maximum correlation in (a)



1st-4th measure, German folk song: Froshgesang (arranged)

Fig.7 Tempo estimation with correlation

このとき、当該相関値を最大とす余弦波の周波数から、入力した楽曲のテンポを推定できる。推定したテンポの逆数から、リズムの周期が求められるので、時間周波数平面で当該幅の区間ごとに、対象となる音高における音楽信号の成分の有無を閾値で判別すれば、多声部を有する音楽信号の採譜を行える。

## 5. まとめ

音楽信号をガボールウェーブレット変換により時間周波数平面に射影し、各楽器の倍音パターンを考慮した音高の推定と、信号強度の自己相関によるテンポと音符の長さの推定を特徴とする採譜システムを設計・製作し、ピアノ演奏による多声部の音楽信号の採譜に適用し、その実用性を確認した。

## 参考文献

- [1] Y. Fukayama et al.: Identification of instruments and keys for Music on the time-frequency plane; Proceedings of 41st international symposium on stochastic systems theory and its applications, pp.19-24, ISCIE, Kobe JP (2009)
- [2] L. Van Tree: Detection, estimation, and modulation theory; Wiley, New York NY (1967)
- [3] M. Kennedy: Oxford dictionary of music; Oxford, New York NY (1985)
- [4] MIDI manufacturers association: <http://www.midi.org/aboutmidi/tutorials.php>
- [5] A. Klapuri and Manuel Davy, editors: Signal processing methods for music transcription, Springer, New York NY (2006)
- [6] G. Goto et al: RWC music database: Music genre database and musical instrument sound database; Proceedings of the 4th international conference on music information retrieval, pp.229-230, ISMIR, Baltimore MD (2003)