

## 対数周波数領域における調波構造に着目した ピアノ多重音の自動採譜

東 真 梨 子<sup>†1</sup> 三 橋 渉<sup>†1</sup>

本研究は、実環境で演奏されたピアノ多重音を自動で楽譜データに変換することを目的としている。他の様々な楽器や歌声に対して高性能な自動採譜が実現できれば、音楽的な知識がなくても作曲・編曲を簡単にでき、曲名やアーティスト名の情報が無い場合の楽曲検索にも役立つだろう。また、音源分離の研究などと組み合わせることでより実用的なシステムとなるだろう。本研究では、音程や音符長の推定はもちろんのこと、音価（音符記号）の推定、伴奏・主旋律の推定まで行う。それに伴い、信号処理に確率・統計モデルを加えることで音価の推定と伴奏・主旋律の推定に音楽的知識を取り入れる。

### Automated Transcription for Polyphonic Piano Music with a focus on Harmonics in Log-Frequency Domain

MARIKO AZUMA<sup>†1</sup> and WATARU MITSUHASHI<sup>†1</sup>

The aim of this study is to transform polyphonic piano sound to musical scores in a real environment. Automated transcription of high quality for other instruments and singing voices can be applied to help composing music and identifying tune without information of song titles and singer names. In addition, this system can become useful when combined with the blind source separation technique. In this paper, We estimate musical note sequence and part of melody and accompaniment, not to mention pitch and note length. We therefore adopt musical knowledge by using probabilistic model in addition to technique of signal processing.

<sup>†1</sup> 電気通信大学  
The University of Electro-Communications

#### 1. はじめに

本研究では信号処理技術を用いて、音程を得るための基本周波数  $F_0$  や音符長 [s] を推定する。 $F_0$  においては対数周波数領域での調波構造に着目した手法で推定を行う。一般的に音程は 12 平均律で表現され、これを周波数で表すとき、その値は音程が高くなるにつれて指数関数的に上昇する。そのため、低周波数の音は高周波数の音に比べ、周波数の微少のずれで音程が違って聴こえてしまう。そこで、周波数領域において定  $Q$  フィルタバンク (Constant  $Q$  spectral transform<sup>1)2)</sup>) を用いる。また、音符長や音価の情報が含まれる隠れマルコフモデルを用いて、推定された音符長から音価を推定する。本研究では両手演奏を対象としており、音程遷移確率を取り入れることで主旋律・伴奏のパート分けも自動で行う。なお、本研究では休符は取り扱わない。

#### 2. 音符長 [s]

音符長 [s] を推定するには、音の発生時刻 (Onset 時間) と停止時刻を求める必要がある。まず全ての Onset 時間を求め、周波数情報より各音の停止時刻を推定する。本研究では休符を取り扱っていないため、この Onset 時間同士の間隔を求めることで音符長を得る。

##### 2.1 Onset 時間の検出

ピアノは音が発生する瞬間に振幅やエネルギーが増大するという特徴があるため、エネルギー情報より Onset 時間を推定する<sup>3)</sup>。

##### [Onset 時間検出のアルゴリズム]

観測波形を時間領域から時間-周波数領域のスペクトログラムに変換する。このスペクトログラムを 0 から 1 の値に収まるよう正規化し、実験的に閾値を設定する。閾値を基準にスペクトログラムの時間・周波数構造を '0' と '1' の二値行列に変換する (図 2 の  $D_2$  参照)。

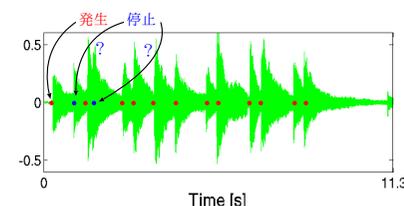


図 1 Onset 時間と停止時間の例

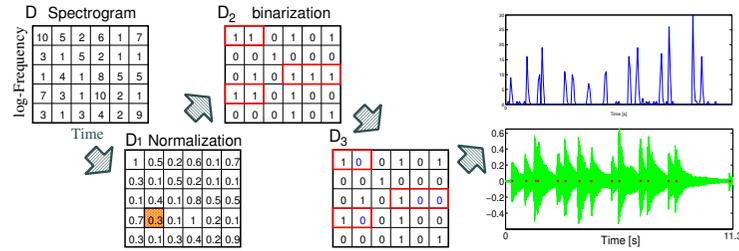


図 2 onset 時間検出の例と実装結果

次に、図 2 の  $D_2$  の赤線で囲まれている部分のように '1' が連続する場合、先頭の値を除いて全て '0' に変換する。この処理を行う理由は、ここでは音が発生した瞬間の情報のみを必要としているからである。最後に、各時間ごとに周波数軸の成分を加算し、実験的に設定した閾値を越えるピーク値が示す時間を検出する。これが Onset 時間となる (図 2 の右に示す波形中の赤点を参照)。

### 3. 基本周波数 F0

基本周波数 F0 を推定し、最も近い 12 平均律を求めることで音程を決定する。現在、F0 推定に関する様々な手法が研究されているが、和音などのような多重音に対する推定は単音に比べて困難である。また、実環境における音源を用いた F0 推定は、雑音や反響の影響により理想の波形とは異なるため容易ではない。

本研究では、対数周波数領域における調波構造の特性を活かした手法により F0 を推定する。

#### 3.1 調波構造

楽器音のほとんどは調波構造を持ち、基本周波数 F0 とその整数倍に位置する倍音成分で成り立っている。F0 は音程を、倍音成分は楽器の音色を形成している。周波数領域における倍音成分は必ず F0 成分の整数倍の位置に存在するため、倍音と F0 の位置関係は F0 の値に依存する。これに対し対数周波数領域では、倍音と F0 の関係は式 (1) で表され、F0 の値に依存しないことが確認できる (図 3)。

$$h(x, \mathbf{a}) = \sum_{n=0}^N a_n \delta(x - \log n) \quad (1)$$

ここで、 $x$  は対数周波数領域、 $h(x)$  は調波構造、 $N$  は調波の数、 $\mathbf{a} = \{a_1, \dots, a_N\}$  は第  $n$  調波の強度比を表し、 $\delta(x)$  は  $x$  に関するデルタ関数である。

#### 3.2 Specmurt 分析

前章の特性を活かした F0 推定法が、嵯峨山らの提案した Specmurt 分析である<sup>4)5)6)</sup>。ここでは、各楽器によって調波構造が共通であると仮定し、これを共通調波構造と呼ぶ。対数周波数領域  $x$  において観測スペクトルを  $v(x)$ 、共通調波構造を  $h(x)$ 、基本周波数 F0 分布を  $u(x)$  とするとき、 $u(x)$  は  $v(x)$  と  $h(x)$  の逆畳み込みによって得られる。

$$u(x) = h^{-1}(x) * v(x) \quad (2)$$

しかし実際には共通調波構造  $h(x)$  は未知であるため、 $h(x)$  を仮定する必要がある。実環境でのピアノ音のスペクトルを観測し調波構造を比較したところ、同じピアノであっても音源の位置・音程・演奏時間などの条件によって異なる調波構造を示すことが確認された。そのため、調波の強度比  $\mathbf{a}$  を式 (3) のように仮定し実装を行う。ここで、調波の数  $N$  は 10 に設定した。

$$\mathbf{a} = \left\{1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{N}\right\} \quad (3)$$

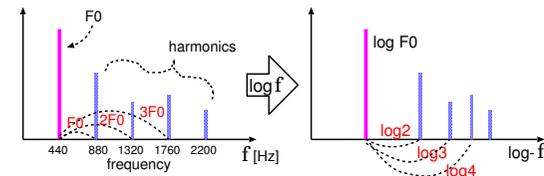


図 3 単音の調波構造の例

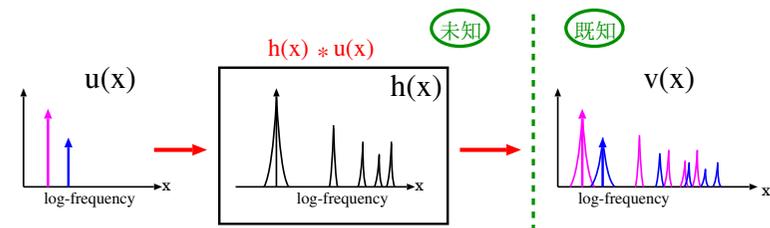


図 4 Specmurt 分析の例

### 3.3 閾値処理による F0 推定

前章でも述べたように、実環境における調波構造は同じ楽器であっても必ずしも共通になるとは限らない。そのため、調波構造が共通であることを仮定とする Specmurt 分析のみに依存すると、倍音成分が残ることがある。そこで次の閾値処理を加える。

Specmurt 分析後、エネルギーの値が実験的に設定した閾値  $T_1$  を越える周波数  $p$  を、低周波数から順に検出する。この時のエネルギーの値を基準とし、新たな閾値  $T_2$  を設定する。 $p$  から  $\log n$  ( $n = 1, \dots, N$ ) 離れた成分のうち、閾値  $T_2$  を下回る成分のみを削除する。これを繰り返し、最後に残ったピークの示す値が F0 となる。 $T_2$  を設定する理由は、図 5 の例のようにオクターブ違いの音が同時に演奏されたとき、低い音の倍音成分と高い音の F0 成分がちょうど重なるため、低い方の倍音を削除する際に高い音の F0 成分を同時に削除してしまうのを防ぐためである。F0 推定を行った結果を図 6 に示す。

### 3.4 音程と音符長の推定

前章までの F0 推定処理だけでは、誤推定は完全には無視できない。そこで、得られた F0 を Onset 間隔ごとに処理することで最終的な音程を決定する。また、同時に各音符の停止時間も推定し、音符長も得る。

図 7 の表 1 は、前章までの処理で推定された各時間ごとの F0 の例を表したものである。以下の処理は Onset 間隔ごとに行う。この Onset 間隔の長さを  $L$  とおく。まず、同一音程

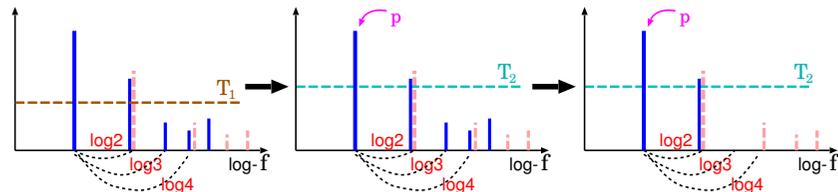


図 5 閾値処理による F0 推定の例

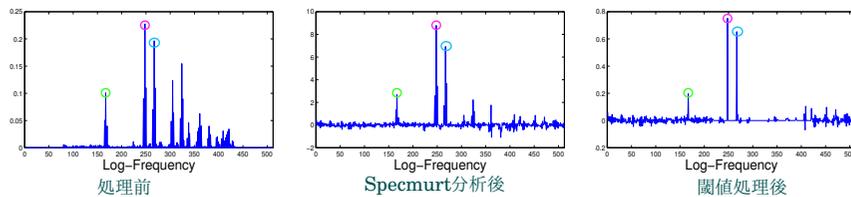


図 6 F0 推定結果の一例

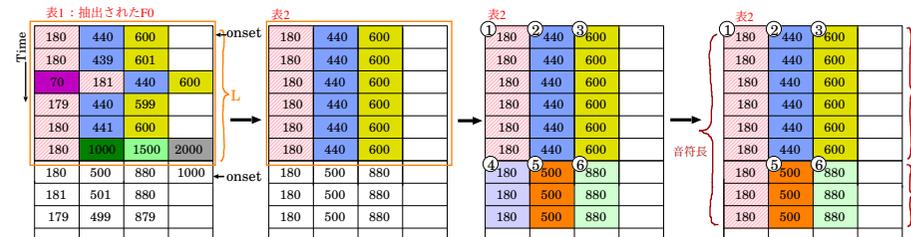


図 7 音程と音符長の推定

と考えられる F0 ごとに分類する。 $L$  に対してある程度の長さを持つ音程のみを実際に演奏されているものとみなし、それ以外は削除する。ここで、本研究では休符を扱わないため、採用された音程の長さが全て  $L$  となるように補間する (図 7 の表 2)。この処理を各 Onset 間隔ごとに行いラベル付けをする。次に、図 7 のラベル 1 とラベル 4 のように同じ音程が隣り合う場合、連続して同じ音程が演奏されたのか、それとも継続している音であるのかを判断する必要があるため次の処理を行う。

まず処理を行う前の表 1 に戻り、隣り合うラベルの後番 (図 7 の例の場合はラベル 4) の Onset 時間の一つ前の時間に同じ音程があるかを検索する。もし同じ音程が発見された場合、前番の音からつながっていると見なし、ラベルを統一する。このようにして、各音符の音程と音符長が同時に求まる。

## 4. 主旋律・伴奏の分類

主旋律と伴奏のパート分けを自動で行う。本研究では伴奏と主旋律を、音程遷移確率と音程差の情報により区別する。具体的には、次のような処理を行う。

### 【音程差】

鍵盤上で片手で弾くことのできる音程差の限界を設定する必要がある。これにより、同時に演奏されている音符同士が明らかに離れていた場合、音程差の情報だけで区別することが可能となる。本研究では、片手で可能な音程差を 14 半音に設定した。

### 【音程遷移】

同時に演奏された音程が、前記で設定した音程差より小さい場合、一つ前に演奏された音

程との遷移確率によってパートを決定する (図 8). 音程遷移確率については 4.1 章で説明する.

[ 削 除 ]

図 4 のように, 音程差の情報により 3 つ以上のパートが存在する (2 つの手で弾くことが不可能な) 場合, 前の音符との音程遷移確率がどちらのパートにおいても低い音符を削除する.

[ 移 動 ]

F0 推定での倍音の誤推定と考えられる場合, 音程遷移が最も高くなるオクターブ差の音程へと移動する. 前記で述べた [削除] の例の中にもこのような倍音の誤推定が含まれることがある. その場合, 削除するのではなく移動を採用するが, やはり倍音誤りであるかどうかを判断するのは困難である. 現時点ではこの処理は行っていないが, 現在どのようにプログラムに取り入れるかを考慮している.

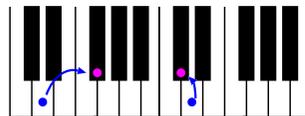


図 8 音程遷移による区別

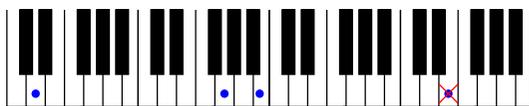


図 9 削除の例

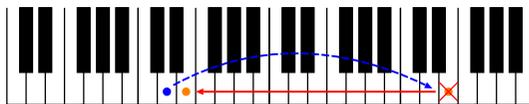


図 10 移動の例

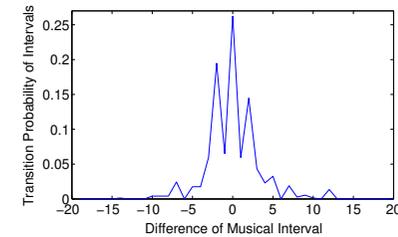


図 11 音程遷移確率の統計結果

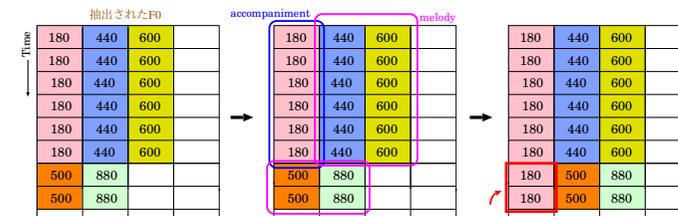


図 12 パートの補間

4.1 音程遷移確率

本研究で用いる音程遷移確率は, 連続する 2 音符の音程差に対して統計を取ったものである. 実際に 20 曲の楽譜から 2,314 遷移の統計をとった結果を図 11 に示す. 横軸は半音を '1' としたときの音程差, 縦軸は遷移確率を表している.

4.2 補 間

パート分け処理後, ある時間に伴奏か主旋律のどちらか一方しか存在していないことが起こり得る. 例えば, 伴奏などで全音符のように長い音符が演奏されたときに起こりやすい. 休符を扱っていないため, 各時間には必ず伴奏と主旋律が存在しなければならない. ここで, パートが抜けている場合は図 12 のように, 足りないパート部分を前の時間の音符を引き延ばすことにより補う.

5. 音価推定

推定された音符長 [s] から, 隠れマルコフモデル (Hidden Markov Model ; HMM) の

Viterbi 検索を用いた手法<sup>7)8)</sup>により音価(音符記号)を推定する。この手法はリズムが与えられていない状況での音楽演奏を対象とし、曲中でのテンポの揺らぎにも対応が可能である。この確率モデルは、状態遷移確率と出力確率の2種類の統計を用いる。なお、本研究では図13に示す8種類の音価を対象としている。

### 5.1 状態遷移確率

状態を音価とし、各音価から各音価へと遷移する確率が状態遷移確率である。実際に、様々なジャンルの楽譜100曲から9,232遷移の統計をとった。遷移確率の統計をとると同時に、各状態が一番最初に演奏される確率である初期状態確率も求めている。

### 5.2 出力確率分布

ここでは、各音価が実際に演奏されたときの音符長の統計結果を出力確率として用いる。実際に、人によって演奏された914音符から統計をとった。この統計結果を正規分布に変換して実装に用いている。また、統計量がまだ少ないため、特に演奏時間が短い音価は隣接する分布同士がほとんど重なってしまうということが見受けられた。ここで、統計結果よりテンポの平均をもとに各音価の平均音符長を求め、この平均音符長が分布の中心となるように平行移動を行った(図14)。

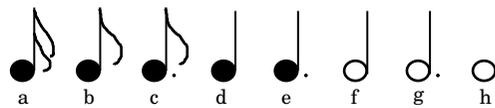


図13 音価の種類

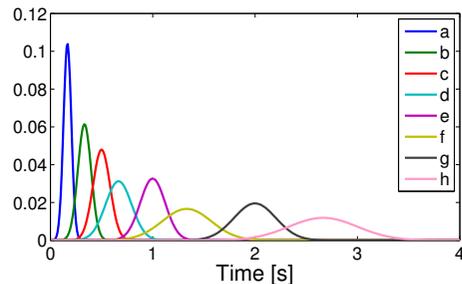


図14 出力確率分布

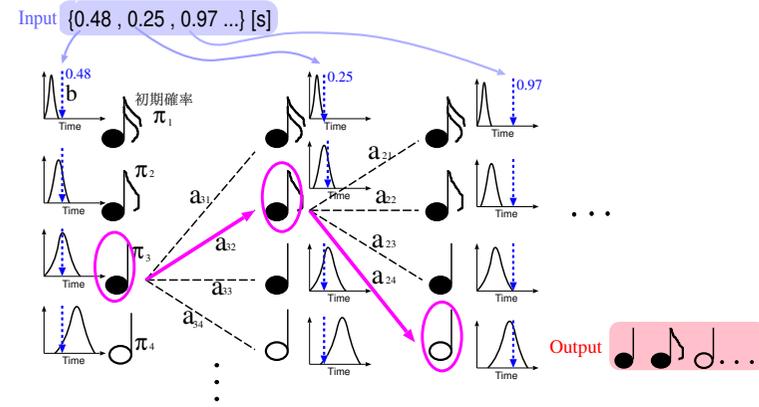


図15 音価推定の例

### 5.3 Viterbi 検索による音価列の推定

音符長 [s] の列を入力とし、入力値に対する各音価の出力確率と状態遷移確率の積が最大となる状態(音価)に遷移していくことで音価の列を得る(図15)。

## 6. 実験

### 6.1 設定

広さが10畳ほどの音楽室において両手で演奏されたピアノ音源を対象に Matlab 上で実装を行った。音源はサンプリング周波数 22.05 [kHz] で標準化し、Constant Q spectral transform におけるシフト長は 1024 点 (0.046 [s]) とした。

### 6.2 実験結果と考察

実装結果を楽譜に表したものを図16に示す。青色の音符は音価の誤推定、桃色の音符は F0 の誤推定であることを示している。

音価の誤推定の理由は、正しい音価と誤りの音価の出力確率分布が広い範囲で重なっていること、前の音符からの状態遷移確率が誤りの音価の方が高いことによるものと考えられる。音程の誤推定については、誤って倍音を検出してしまったことによるものが見られた。音符の欠落については、同時にオクターブ差の音程が演奏されており、倍音処理の際に高い音程の F0 も同時に抑制されてしまったことによるものである。

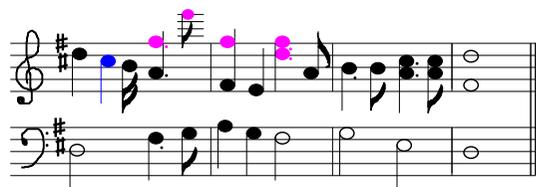


図 16 採譜結果



図 17 元の楽譜

## 7. まとめと課題

本稿では、雑音や反響の影響がある実環境での両手ピアノ演奏に対する自動採譜結果を示した。まだ倍音処理による音程の誤推定が多く見られるため、解決すべき課題が多く残されている。一番の課題は F0 推定法を改良し、精度を上げることである。音程遷移確率や音価推定の出力確率においては、統計量を増やす必要がある。また、曲のジャンル別の統計をとることも視野にいれている。

(平成 23 年 1 月 14 日受付)

(平成 23 年 2 月 11 日採録)

## 参 考 文 献

- 1) BrownJudith C.: Calculation of a constant Q spectral transform, *The Journal of the Acoustical Society of America*, Vol.89, pp.425–434 (1991).
- 2) BrownJudith C. and PucketteMiller S.: An efficient algorithm for the calculation of constant Q transform, *The Journal of the Acoustical Society of America*, Vol.92, pp.2698–2701 (1992).
- 3) Giovanni Costantini, Renzo Perfetti and Massimiliano Todisco: Event based transcription system for polyphonic piano music, *Signal Processing*, Vol.89, pp.1798–1811 (2009).
- 4) 高橋佳吾, 西本卓也, 嵯峨山茂樹: 対数周波数逆畳み込みによる多重音の基本周波数解析, 情報処理学会研究報告, 2003-MUS-53-13, pp.61–66 (2003).
- 5) Shigeki Sagayama, Keigo Takahashi, Hirokazu kameyama and Takayuki Nishimoto: Specmurt Analysis: A Piano-Roll-Visualization of Polyphonic Music Signals by Deconvolution of log-frequency Spectrum, *In SAPA2004* (2004).
- 6) 亀岡弘和, 斎藤翔一郎, 西本卓也, 嵯峨山茂樹: Specmurt における準最適共通調波構造パターンの反復推定による多声音楽信号の可視化と MIDI 変換, 情報処理学会研究報告, 2004-MUS-56-7, pp.41–48 (2004).
- 7) 斎藤直樹, 中井 満, 下平 博, 嵯峨山茂樹: 隠れマルコフモデルによる音楽演奏からの音符列の推定, 情報処理学会研究報告 [音楽情報科学] (1999).
- 8) 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: 発音時刻の楽譜上の位置を確率モデルにより推定するクオンタイズ手法, 情報処理学会論文誌 (2002).