

スパース性基準による F0 周波数選択を用いた Specmurt による多重音解析*

☆西村大樹, 中鹿亘, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

多重音の解析は単音の解析に比べて困難であり, これまでにも数多くの手法が試されてきた. 本研究で用いる Specmurt[1] は多重音解析手法の一つである. 従来の Specmurt 法を用いた多重音解析のアプローチは, 共通調波構造と呼ばれるモデルを自ら反復的に生成することで疑似的な楽器情報を得て, それをもとに基本周波数分布を求めるというものであった [2]. これらの手法は, ある音色に関して, 全ての音の周波数成分間のパワー比が一定であるという前提のもとに共通調波構造をモデル化しているが, 厳密には周波数成分間のパワー比は音階によって異なるため, 完全に正しい共通調波構造を得ることはできないと考えられる. そこで我々は共通調波構造をモデル化しないで基本周波数分布を求める, 新しい多重音解析手法を提案する.

2 Specmurt 法の概要

共通調波構造を $h(x)$, 基本周波数分布を $u(x)$ とする時, 多重音スペクトル $v(x)$ は

$$v(x) = h(x) * u(x) \quad (1)$$

と表すことができる. ここで, 共通調波構造とは全ての音の周波数成分間のパワー比が基本周波数にかかわらず普遍的に一定である調波構造パターンのことであり, 基本周波数分布とは基本周波数がどの値にどれだけの成分を持つかを表したものである.

また, 共通調波構造 $h(x)$ が既知であるなら, 基本周波数分布 $u(x)$ は

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

のように, $v(x)$ と $h(x)$ の逆畳み込みで求められる.

ただし, 線形周波数スケールでは基本周波数が $\Delta\omega$ 変化すると n 次の高調波周波数は $n\Delta\omega$ 変化してしまい, 式 (1) は成立しない. そこで, これら各関数を対数周波数スケールで扱うことにする. 対数周波数スケールでは基本周波数が Δx 変化すると全ての高調波周波数も Δx 変化し, 式 (1) が成立する.

以上のように, 対数周波数スケールで逆畳み込みすることにより, 基本周波数分布を求める方法は Specmurt 法 [1] と呼ばれている.

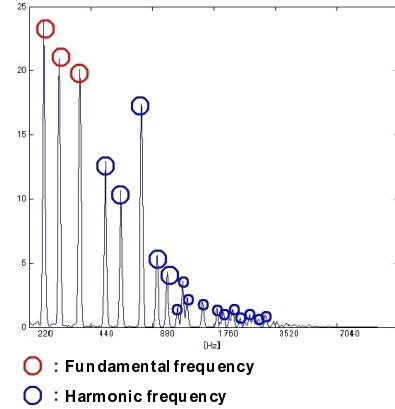


Fig. 1 Observed spectrum (piano triad)

3 提案手法

提案手法では, 共通調波構造 $h(x)$ をモデル化しないで基本周波数分布 $u(x)$ を求める. 以下に具体的な処理手順を示す.

3.1 基本周波数分布として考えられる候補

観測スペクトルを元に解の基本周波数分布として考えられる候補をいくつか求める. Fig. 1 はピアノ 3 和音の観測信号スペクトルである. ノイズがない場合, この図のように観測信号のピークは基本周波数, もしくは高調波周波数である. しかしこのスペクトルをみただけでは, どのピークが基本周波数でどのピークが高調波周波数に相当するのかわからない. そこである程度の大きさをもつピークの組み合わせを考え, それらを解の候補とする. その際, ピークは全てインパルスとして処理する. 解が単音から n 和音までの可能性を考えると候補の組み合わせパターン数は $\sum_{k=1}^n n C_k$ になる. この中から解の基本周波数分布に相当するピークの組み合わせを探していく.

3.2 最適な調波構造の決定

式 (1) を変形し

$$h(x) = u(x)^{-1} * v(x) \quad (3)$$

とする. 観測信号のピークをもとに得られた λ 個の解の候補 $u_i(x)$ をそれぞれ $u(x)$ として式 (3) に代入し, 各 $u_i(x)$ に対応する $h_i(x)$ を求める. これより得られた $h_i(x)$ は $u_i(x)$ と同じ λ 個存在するが, その中から最適な共通調波構造 $\hat{h}(x)$ を一つだけ見つける.

*Sparseness criteria of F0-frequency selection for Specmurt-based multi-pitch analysis, by NISHIMURA, Daiki, NAKASHIKA, Toru, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe University)

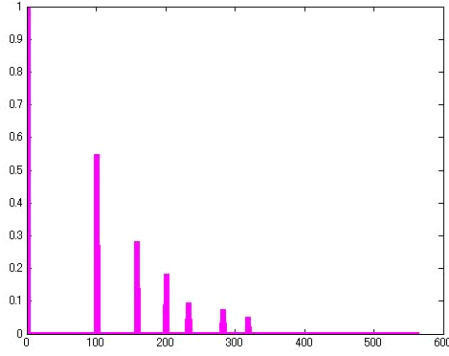


Fig. 2 Example of harmonic structure (piano single tone)

最適な調波構造 $\hat{h}(x)$ とは、ピークの組み合わせに全ての基本周波数を持ち、かつ高調波周波数を含まない基本周波数分布 $\hat{u}(x)$ に対応する調波構造のことである。

3.2.1 非調波構造の棄却

$h_i(x)$ の中に高調波周波数成分を持たないものが存在する。そのような構造をここでは非調波構造と呼ぶことにする。調波構造を持つ音響信号を解析の対象としているため、非調波構造は最適な調波構造 $\hat{h}(x)$ ではないとし、この段階で棄却される。

Fig. 2 はピアノの単音 (A3) の調波構造であるが、調波構造は音色や音階が変わっても各高調波周波数のパワーが変わるだけでピークの出現する箇所は変わらないことが知られている。このことから音色や音階によらず高調波周波数の位置はわかるため、そこにピークを持たない非調波構造は簡単に見つけ出せる。見つけた非調波構造は棄却され、最適な調波構造の候補から外れる。

3.2.2 スパース性の考慮

理想的な調波構造は Fig. 2 のように基本周波数と高調波周波数以外にピークを持たない。一方、得られた $h_i(x)$ の中で最適でない調波構造は本来ピークのない場所に大きなピーク (ノイズ) を持つ。そこで前段階で棄却されなかった各 $h_i(x)$ のスパース性を調べる。先験情報より、理想的な調波構造は基本周波数と高調波周波数以外にピークを持たずスパースであると考えられるため、 $h_i(x)$ の中から最もスパースなものを最適な調波構造 $\hat{h}(x)$ とする。しかし、実際は最適な調波構造でも本来ピークの無い場所に小さなノイズを持っているため、各値の大きさを考慮できる L_1 ノルム、もしくは L_2 ノルムを適用する。 $\hat{h}(x) = h_i(x)$ であるとする、 $\hat{i}$ は以下のように

$$\hat{i} = \operatorname{argmin}_i \sum_{x=1}^X |h_i(x)| \quad (4)$$

$$\hat{i} = \operatorname{argmin}_i \sum_{x=1}^X h_i(x)^2 \quad (5)$$

と計算される。

3.2.3 解の選択

以上の方法で最適な調波構造 $\hat{h}(x)$ が決定される。最後に、 $\hat{h}(x)$ に対応する $\hat{u}(x)$ を解の候補から探し出し、解とする。

これらの手順を全てのフレームに適用することで、入力信号全体の基本周波数分布を求める。

4 実験

提案手法の有効性を確かめるため、RWC データベース [3] の「RWC-MDB-J-2001 No. 9: Crescent Serenade」をテストデータとして使用した。MIDI データをピアノで演奏させ、演奏区間は 13 秒とした。Table 1 に従来の Specmurt 法による解析結果と提案手法による解析結果を示す。提案手法による解析から最も高い正解率を得られたことがわかる。

Table 1 Accuracy results

従来手法	提案手法 (L_1)	提案手法 (L_2)
89.2%	91.6%	92.5%

5 おわりに

本稿では共通調波構造をモデル化しないで、スパース性を考慮した Specmurt による多重音解析の有効性を示した。この手法は音色の学習を必要とせず、また和音数などといった知識も用いないで多重音の解析ができる。今後は、オクターブの和音に対処できるような手法や、調波構造を持たない音色に対する解析を検討していくことも視野に入れて研究を進めていきたい。

参考文献

- [1] 高橋佳吾他, “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書, 2003-MUS, vol. 127, pp. 61–66, 2003.
- [2] S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto and S. Sagayama, “Specmurt Analysis of Polyphonic Music Signals,” IEEE Trans. on ASLP, vol. 16, no. 3, pp. 639–650, 2008.
- [3] <http://staff.aist.go.jp/m.goto/RWC-MDB/>