

スパース性を考慮した 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

近年, 音楽情報処理技術は飛躍的に進歩してきた。これにより, 楽曲の製作や編曲といった作業が特定の能力のある人のものだけでなく, パソコンとソフトさえあれば誰でも楽しめるようになり, 音楽を創造することに対する敷居の高さはかつてに比べて大幅に低くなった。しかしながら, 特定の能力のある人に頼らざるを得ない分野がまだまだ残されている。例えば, 耳にした音の高さを判別できる音感というものがそれに相当する。この能力を身につけるためにはかなりの経験や労力, 時間が必要になるが, 耳にただけの楽曲を再現しようとした時, あるいは楽譜に起すためには必ず要される能力である。特に和音のように同時刻に様々な高さの音が存在する信号を, 耳だけで解析するのは非常に難しく, そこには多重音解析の需要がある。

多重音の解析は単音の解析に比べて困難であり, これまでも数多くの手法が試されてきた [1, 2, 3]。Specmurt 法 [4] 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方法は, 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ったものであった [5, 6]。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ものであり, 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。

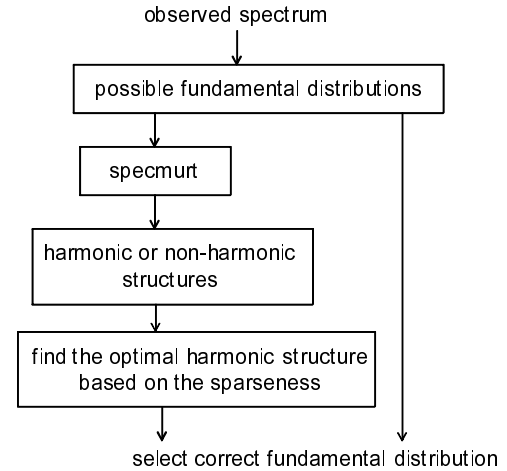


Fig. 1 Flowchart of Multi-pitch analysis with Specmurt based on the sparseness

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を Specmurt 法 [4] と呼ぶ。

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。Fig. 1 に手法の流れを示す。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. 2 はピアノ 3 和音の観測スペクトルである。ノイズがない場合, この図のように観測スペクトルのピークは基音 (基本周波数), もしくはその倍音 (高調波周波数) である。しかしこのスペクトルを見ただけでは, どのピークが

*Multi-pitch analysis with Specmurt based on the sparseness, by NISHIMURA, Daiki, NAKASHIKA, Toru, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe University)

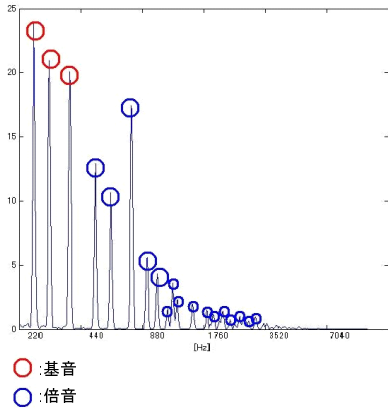


Fig. 2 Observed spectrum (piano triad)

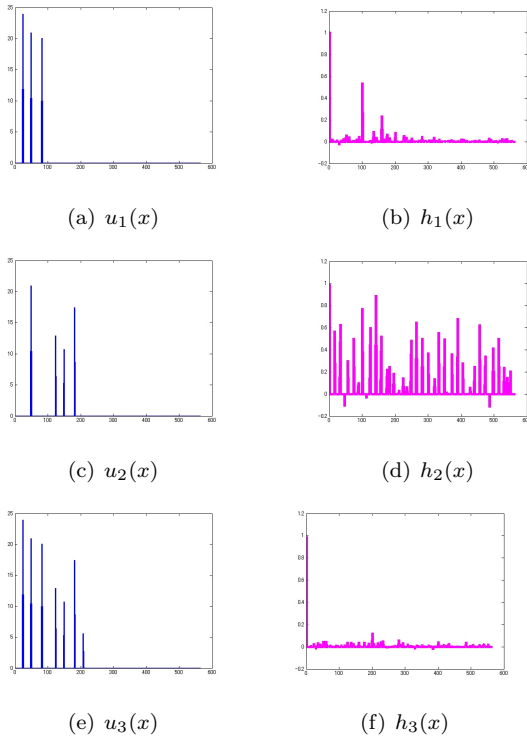


Fig. 3 Examples of $u(x)$ and $h(x)$

基音でどのピークが倍音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 $\sum_{k=1}^n C_k$ 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．観測スペクトルのピークをもとに得られた解の候補を $u(x)$ として式 (3) に代入し，それぞれの

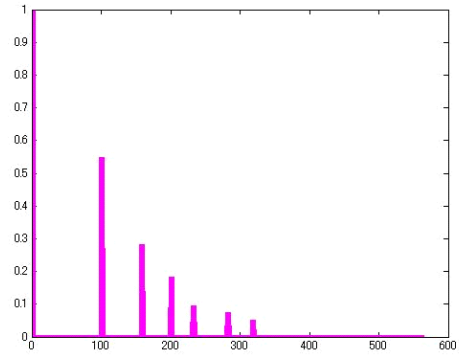


Fig. 4 Example of harmonic structure

ピークの組み合わせに対応する $h(x)$ を求める．これより得られた $h(x)$ は $u(x)$ と同じ $\sum_{k=1}^n C_k$ 個存在するが，その中から最適な $h(x)$ を一つだけ見つける．最適な $h(x)$ 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持たない $u(x)$ に対応する調波構造のことである．

Fig. 3 の左側は観測スペクトルのピークを元に得られた $u(x)$ の例を示しており，右側はそれに対応する $h(x)$ である．Fig. 3(a) はピークの組み合わせに全ての基本周波数を持ち，高調波周波数をひとつも持たない $u_1(x)$ を示しているが，それに対応する $h_1(x)$ (Fig. 3(b)) はピアノ A3 の調波構造の例である Fig. 4 によく似た形をしている．一方， $h_2(x)$ (Fig. 3(d)) や $h_3(x)$ (Fig. 3(f)) は，ピークの組み合わせに基本周波数が足りない，あるいは組み合わせに高調波周波数を持つような $u_2(x)$ (Fig. 3(c))， $u_3(x)$ (Fig. 3(e)) から生成されたものである． $h_1(x)$ に比べると $h_2(x)$ は Fig. 4 には無い箇所に大きなピークを数多く持っており， $h_3(x)$ はピークのあるべきところにピークを持っていないことがわかる．ここで， $h_3(x)$ のように高調波周波数を持たない構造を非調波構造と呼ぶことにする．

3.2.1 非調波構造の棄却

Fig. 4 はピアノの単音 (A3) の調波構造であるが，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，それに対応する $u(x)$ は解の候補から外れる．

3.2.2 スパース性の考慮

理想的な調波構造は Fig. 4 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が，Fig. 3(d) の $h_2(x)$ のような最適でない調波構造は本来ピーク

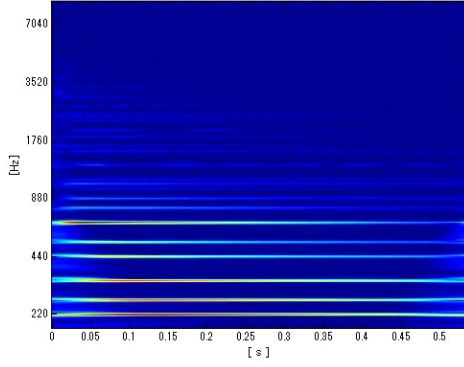


Fig. 5 Observed spectrum used in the experiments

のない場所に大きなピーク (ノイズ) を持つ．よって棄却されなかった各 $h_j(x)$ のスパース性を調べ，最もスパースなものを最適な調波構造 $h_{\hat{j}}(x)$ とする．しかし，実際は最適な調波構造でも Fig. 3(b) の $h_1(x)$ のように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，スパース性の計算には L_0 ノルムではなく，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{j} は以下のように

$$\hat{j} = \operatorname{argmin}_j \sum_{i=1}^N |h_j(x_i)| \quad (4)$$

$$\hat{j} = \operatorname{argmin}_j \sum_{i=1}^N h_j(x_i)^2 \quad (5)$$

と計算される．また， L_0 ノルムの時は以下のように \hat{j} を決定する．

$$\hat{j} = \operatorname{argmin}_j \sum_{i=1}^n \delta(h_j(x_i)) \quad (6)$$

ただし

$$\delta(h_j(x_i)) = \begin{cases} 1 & (h_j(x_i) \neq 0) \\ 0 & (h_j(x_i) = 0) \end{cases}$$

である．

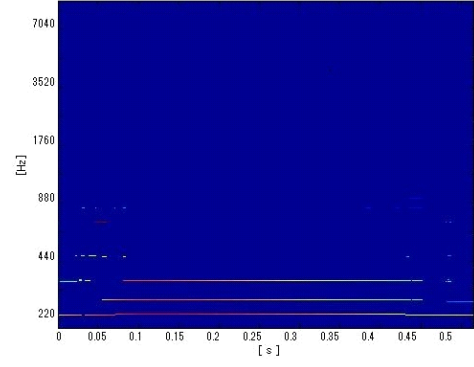
3.2.3 解の選択

以上の方法で最もスパースな $h(x)$ ，すなわち最適な調波構造が決定される．最後に，最適な調波構造に対応する $u(x)$ を解の候補から探し出し，解とする．

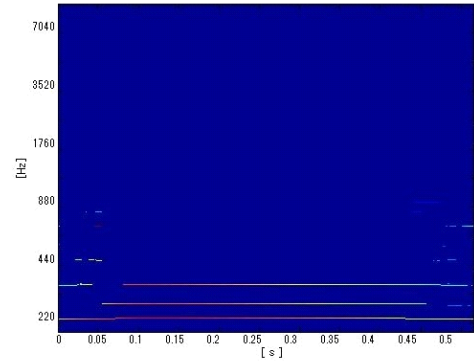
これらの手順を全てのフレームに適用することで，入力信号全体の基本周波数分布を求める．

4 実験

提案手法により実際に多重音解析が可能か，約 0.5 秒のピアノ 3 和音 (A3, C4, E4) による実験を行った．



(a) L_1 norm



(b) L_2 norm

Fig. 6 Results using our proposed method

本実験では周波数分析に Wavelet 変換を用いた．Fig. 5 は実験に使用したデータの Wavelet 変換された観測スペクトルである．横軸が時間，縦軸が対数周波数になっている．このスペクトルの下 3 本が基本周波数に相当し，残りは全て高調波周波数である．

Fig. 6 に実験結果を示す．Fig. 6(a) はスパース性を考慮する際に L_1 ノルムを用いた結果であり，Fig. 6(b) は L_2 ノルムを用いた結果である．また，結果の比較のために L_0 ノルムを用いた結果を Fig. 7 に示している．

4.1 考察

実験結果は L_1 ノルム， L_2 ノルムのどちらも多くのフレームで正しく 3 和音の解析ができており，提案手法の有効性が示せた．しかし両方とも前半の音の立ち上がりや後半の減衰部分での解析に失敗している．Fig. 8(a) はフレーム中盤の解析に成功したスペクトルであり，Fig. 8(b) はフレーム前半の解析に失敗したスペクトルである．横軸が対数周波数，縦軸が強度を示している．この図からわかるように，解析に失敗したスペクトルは成功したものに比べて，各周波数ピーク間の谷の部分の部分が比較的大きな値を持っている．

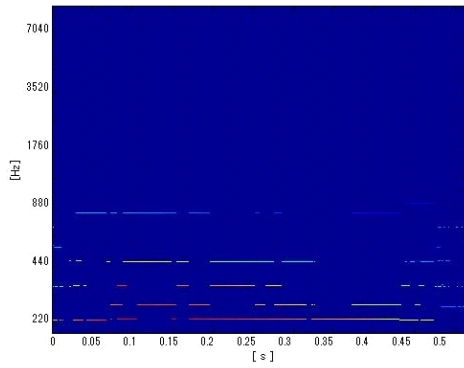


Fig. 7 Result using L_0 norm

実験の手順の中で観測スペクトルから解の候補である $u(x)$ を生成する際、ピークは全てインパルスとして処理していたため、Fig. 8(b) のようなピーク間の谷の部分の値が十分小さくない場合は、Fig. 8(a) のような谷の値が十分小さい場合に比べて大きな誤差が発生したのだと考えられる。

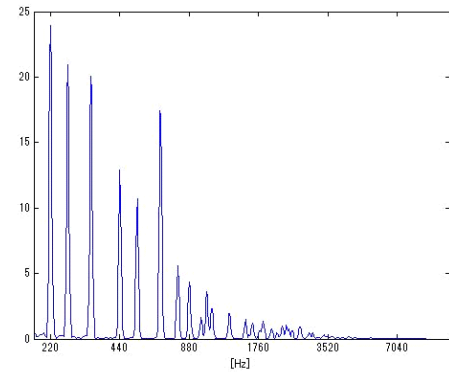
一方、 L_0 ノルムでの解析はほとんどのフレームで失敗している。このことから、 L_0 ノルムのような値の有無だけに注目して評価するものより、各値の大きさを考慮できる L_1, L_2 ノルムの方が本手法では有効であることがわかる。

5 おわりに

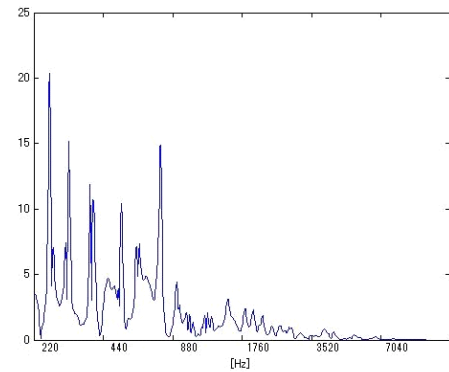
本稿では共通調波構造をモデル化しないで、スパース性を考慮した Specmurt による多重音解析の有効性を示した。この手法は音色の学習を必要とせず、また和音数などといった知識も用いなくて多重音の解析ができる。今後は様々な楽器に対する有効性や楽曲レベルでの有効性を示せるようにすることが課題である。また、今回解析に失敗したスペクトルにおいても対処できるような手法や、時間軸方法の情報を考慮に入れた解析方法を検討していくことも視野に入れて研究を進めていきたい。

参考文献

- [1] 柏野邦夫他, “音楽情景分析の処理モデル OPTIMA における和音の認識,” 電子情報通信学会論文誌, vol. J-79-D- , no. 11, pp. 1762–1770, 1996.
- [2] M. Goto, “A Predominant-F0 Estimation Method for CD Recordings: MAP Estimation using EM Algorithm for Adaptive Tone Models,” Proc. IEEE 26th International Conference on Acoustics, Speech and Signal Processing (ICASSP2001), pp. V-3365–3368, 2001.



(a) Observed spectrum (0.25 s)



(b) Observed spectrum (0.05 s)

Fig. 8 Comparison spectrum

- [3] H. Kameoka, *et al.*, “Extraction of Multiple Fundamental Frequencies from Polyphonic Music Using Harmonic Clustering,” Proc. ICA 2004, pp.533-536, 2004.
- [4] 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.
- [5] 亀岡弘和他, “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書, 2004-MUS, vol. 84, pp. 41–48, 2004.
- [6] 斉藤翔一郎他, “凸射影法に基づく Specmurt 分析の共通調波構造推定,” 情報処理学会研究報告書, 2006-MUS, vol. 45, pp. 13–18, 2006.