重みつきノルム基準による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

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，耳にし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，これまでに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 [5, 6, 7]．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

$$v(x) = h(x) * u(x) \tag{1}$$

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．

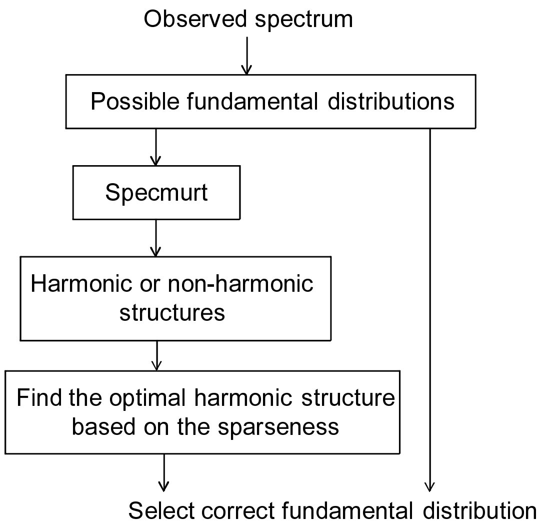


Fig. 1　Flowchart of weighted norm criteria of F0-frequencies selection for specmurt-based multi-pitch analysis.

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

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

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 [4] と呼ばれている．

3　提案手法

提案手法では，共通調波構造 $h(x)$ をモデル化せず，重みつきノルムを利用して基本周波数分布 $u(x)$ を求める．Fig. 1 は提案手法の流れである．以下に具体的な処理手順を示す．

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

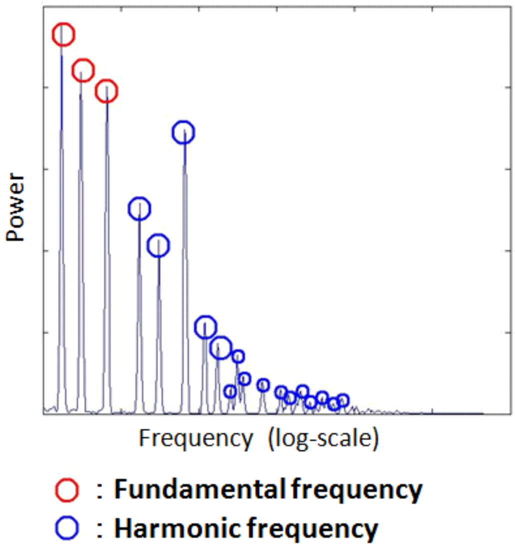


Fig. 2　Observed spectrum (piano triad)

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見ただけでは，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 λ は $\sum_{l=1}^{M} {}_MC_l$ 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

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

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．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．

3.3　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. 3 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．

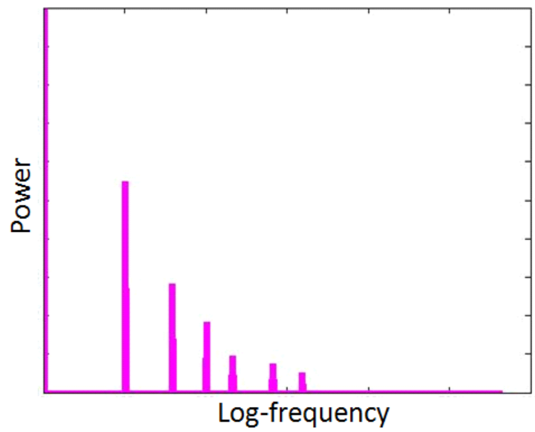


Fig. 3　Example of harmonic structure (piano A3)

3.4　スパース性の考慮

理想的な調波構造は Fig. 3 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他の $h_i(x)$ に比べて基本周波数と高調波周波数以外にピークを持たず，高調波成分に大きなピークを持つと考えられるため，スパース性を

$$Sparseness(i) = \alpha La(i) - (1-\alpha) Lb(i) \tag{4}$$

と定義する．α は重みである．L_1 ノルムを第一項，または第二項に用いる場合，

$$La(i) = \sum_{x=1}^{X} \{1 - \sum_{j=1}^{N} \delta(\Omega_j - x)\} |h_i(x)| \tag{5}$$

$$Lb(i) = \sum_{x=1}^{X} \sum_{j=1}^{N} \delta(\Omega_j - x) |h_i(x)| \tag{6}$$

となる．δ はクロネッカーのデルタである．第一項の $La(i)$ は高調波成分を除く $h_i(x)$ のスパース性を，第

二項の $Lb(i)$ は高調波成分の和を計算している．同様に L_2 ノルムを式 (4) に用いる場合，

$$La(i) = \sum_{x=1}^{X} \{1 - \sum_{j=1}^{N} \delta(\Omega_j - x)\} h_i(x)^2 \quad (7)$$

$$Lb(i) = \sum_{x=1}^{X} \sum_{j=1}^{N} \delta(\Omega_j - x)\, h_i(x)^2. \quad (8)$$

となる．$\hat{h}(x) = h_{\hat{i}}(x)$ であるとすると，$\hat{i}$ は以下のように

$$\tilde{i} = \operatorname*{argmin}_i Sparseness(i). \quad (9)$$

と計算される．

3.5 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

4.1 実験環境

提案手法の有効性を確かめるため，RWC データベース [8] の「RWC-MDB-J-2001 No. 9: Crescent Serenade」をテストデータ (Fig. 4(a)) として使用した．MIDI データをピアノ，アコースティックギター，バイオリンで演奏させ，演奏区間は 13 秒とした．本実験では周波数分析に Wavelet 変換を用いた．3.1 節で述べたパラメータ M は 7 とし，3.4 節で用いたパラメータ N は 6 に設定して実験を行った．

4.2 実験結果

Table 1，2，3 は式 (4) の重み α を 0.0 から 1.0 まで変化させてピアノ，ギター，バイオリンを提案手法で解析した結果である．表中の (L_n, L_m) は式 (4) の第一項に L_n ノルムを，第二項に L_m ノルムを使用したことを示している．正解率の計算には以下の式を用いた．

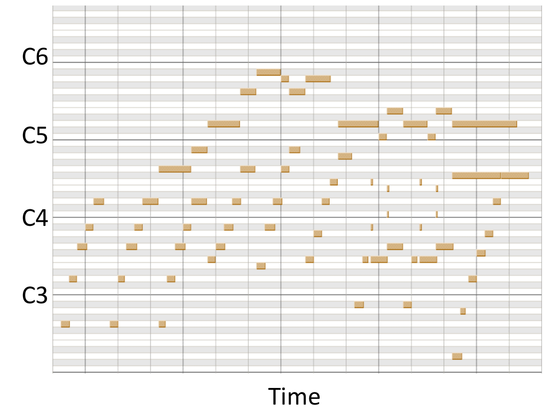
$$Accuracy(\%) = \frac{N_{all} - (N_{ins} + N_{del})}{N_{all}} \times 100 \quad (10)$$

N_{all}，N_{ins}，N_{del} はそれぞれ全ノート数，挿入誤り数，削除誤り数である．

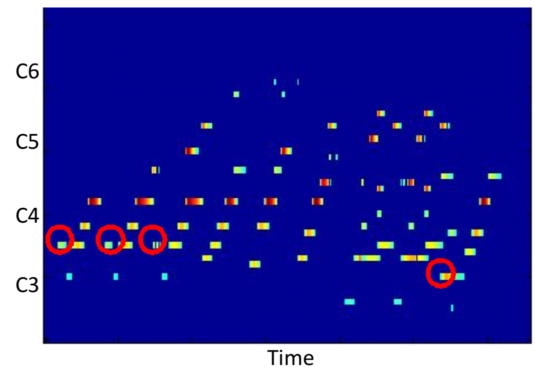
Table 1，2，3 より，ピアノにおいては (L_2, L_1) もしくは (L_2, L_2) を用いたものが，ギターにおいては (L_2, L_1) を用いたものが，バイオリンでは (L_1, L_1) を用いたものが最も高い正解率を示した．また，重み α はピアノは大きなものが，バイオリンは小さなものが高い正解率を示した．以上より，最適なノルムと重みパラメータは楽器によって変化することがわかる．

Fig. 4(b) は提案手法による解析結果の一例である．ほとんどのノートは正確に解析できているが，一部にオクターブ誤りが起こっている．

Table 4 は従来の specmurt による多重音解析 [5] と提案手法を比較した結果を示している．各楽器において提案手法の正解率が従来手法を上回っていることがわかる．



(a)



(b)

Fig. 4 (a) Piano-roll of test MIDI data. (b) An example of analysis result. Red circles indicate some mistaken notes.

5 おわりに

本稿では共通調波構造をモデル化しないで，重みつきノルムによる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や人の歌声による和音 (コーラス) の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．

Table 1　Accuracy results for piano [%]

α	(L_1, L_1)	(L_1, L_2)	(L_2, L_1)	(L_2, L_2)
0.0	77.5	77.3	87.3	75.7
0.1	83.3	83.8	89.5	84.3
0.2	85.4	83.3	90.8	89.2
0.3	87.0	84.7	90.6	89.3
0.4	88.8	86.3	89.4	89.2
0.5	88.2	88.0	89.9	91.1
0.6	89.2	90.4	91.4	92.0
0.7	90.1	89.1	91.8	92.9
0.8	89.8	90.4	92.6	92.6
0.9	**90.7**	90.1	**92.7**	**92.7**
1.0	90.6	**90.6**	92.4	92.4

Table 2　Accuracy results for guitar [%]

α	(L_1, L_1)	(L_1, L_2)	(L_2, L_1)	(L_2, L_2)
0.0	76.5	77.5	78.4	78.3
0.1	77.8	**78.4**	78.3	77.4
0.2	76.7	75.8	77.0	75.9
0.3	77.6	75.9	78.5	77.6
0.4	**78.4**	76.3	**79.7**	78.4
0.5	78.1	75.7	79.0	77.8
0.6	77.5	76.3	**79.7**	**78.8**
0.7	77.4	78.4	79.1	77.8
0.8	77.6	78.2	79.5	77.7
0.9	77.0	77.3	77.7	77.9
1.0	77.0	77.0	78.0	78.0

Table 3　Accuracy results for violin [%]

α	(L_1, L_1)	(L_1, L_2)	(L_2, L_1)	(L_2, L_2)
0.0	71.4	69.2	**69.6**	**68.6**
0.1	**71.7**	65.4	68.4	66.7
0.2	71.1	69.5	63.1	64.5
0.3	70.6	**69.6**	59.3	67.7
0.4	63.3	68.4	58.8	64.9
0.5	57.5	66.7	57.3	62.6
0.6	53.9	63.6	58.3	61.9
0.7	52.9	56.2	59.5	60.2
0.8	55.0	52.3	58.3	57.2
0.9	57.7	54.8	59.3	56.2
1.0	57.7	57.7	60.0	60.0

Table 4　Comparison of the proposed method to a conventional method.

	proposed method	conventional method
Piano	92.7%	89.2%
Guitar	79.7%	74.3%
Violin	71.7%	65.0%

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-II, 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.

[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.

[7] 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.

[8] http://staff.aist.go.jp/m.goto/RWC-MDB/