対数周波数逆畳み込みによる多重音の基本周波数解析

高橋　佳吾　　西本　卓也　　嵯峨山茂樹

東京大学大学院情報理工学系研究科
〒 113–0033 東京都文京区本郷 7–3–1
E-mail: {takahashi,nishi,sagayama}@hil.t.u-tokyo.ac.jp

あらまし　本報告では、多重音のモノラル音響信号の基本周波数を連続的な分布として出力する手法（Specmurt 法）を提案する。多重音を構成する各音が共通した調波構造パターン（高調波成分間の強度比パターン）のスペクトルを持つ場合、対数周波数軸上では、これらの互いの関係は、同一の倍音パターン形状を平行移動した関係となる。これは、多重音の基本周波数の分布と共通調波構造パターンとの対数周波数軸上の畳み込みと解釈でき、基本周波数分布を入力、共通調波構造パターンをインパルス応答とした線形系の出力と考えることができる。共通調波構造パターンを仮定して、対数周波数領域に対するフーリエ領域で除算を用いて逆畳み込みを行えば、基本周波数を連続分布として求めることができる。その結果を濃淡表示すれば、スペクトログラムに似た基本周波数分布表示が得られる。実験を通して、基本的な理論を検証し、実際の音楽信号に適用し、効果を確認した。
キーワード：多重音信号、多重ピッチ解析、対数周波数、逆畳み込み

Multi-Pitch Analysis Using Deconvolution of Log-Frequency Spectrum

Keigo TAKAHASHI, Takuya NISHIMOTO, and Shigeki SAGAYAMA

Graduate School of Information Science and Technology, University of Tokyo
7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–0033 Japan
E-mail: {takahashi,nishi,sagayama}@hil.t.u-tokyo.ac.jp

Abstract　This paper describes a method of transforming single-channel mixed sound signals into a continuous distribution representing a relative dominance of fundamental frequencies, which we call a “specmurt” method. Provided that each sound in a mixed sound has common harmonic structure pattern (a relative amplitude pattern of respective harmonic components), a harmonic structure of every single sound can be expressed as a parallel shift of the common pattern along a logarithmic frequency axis. Therefore, the spectrum of mixed sound can then be expressed as a convolution between the dominance distribution of fundamental frequencies and the common harmonic structure pattern. This relation can be regarded as a linear system with the dominance distribution of fundamental frequencies as an input, the common harmonic structure pattern as an impulse response, and the spectrum of mixed sound as an output. As considering in this way, the dominance distribution of fundamental frequencies can be obtained by deconvolution using a Fourier transformed harmonic structure pattern in the logarithmic frequency domain. As the experimental results, we validated our proposed method using musical signals and verified that it could successfully visualize a continuous distribution of the relative dominance of fundamental frequencies as a shading image, similar to ‘spectrogram’.
Keywords: multi-pitch signals, multi-pitch analysis, logarithmic frequency, deconvolution

1. はじめに

本報告では、楽器などの多重音のモノラル音響信号から、それらに含まれる基本周波数成分を解析し、視覚的に提示する信号処理手法を検討する。

音楽演奏などの音響信号から採譜したり演奏を解析するなどの場面で、信号中に含まれる基本周波数のみが分かりやすく表示される技術があれば、大変有用であろう。これはいわば、音響信号のピアノロール表示とも言え、理想的には合奏や合唱やピアノ演奏の音響信号から MIDI 信号へ変換・編集・楽譜変換などのフロントエンドとなる可能性がある。また、多重の音声の場合は、それらをスペクトルとして分離して認識したり、聴覚情景分析のツールとして使える可能性もある。

しかし、モノラル信号として与えられた多重音から、それを構成する基本周波数情報を抽出することは容易ではない。その理由は、信号には基本周波数成分のみでなく調波成分が複雑に重なり合い、通常のスペクトル解析の手法では基本周波数のみの情報に変換することが容易でないからである。

多重音の基本周波数推定に関してはさまざまな研究が行われて来た。最近の高精度な手法の例を挙げれば、調波構造を混合正規分布の重ね合わせとしてモデル化し、その重みを推定することで基本周波数を推定する方法 [1] [2] や、ハーモニッククラスタリング（調波構造の拘束を導入したクラスタリング手法）による基本周波数推定法 [3] [4] などがある。これらの手法は、優勢な基本周波数を抽出したり、音源数を与えて基本周波数を推定したり、音源数の推定を行ったりしているが、反復計算を用いて何らかの評価関数を最大化する手法であることから、局所解に陥る可能性を持ち、また相対的にエネルギーの小さい音は無視される傾向がある。

そこで我々は、推定結果を一意に求めるいわゆる「ピッチ抽出」の考え方でなく、スペクトログラムの中で基本周波数成分以外を抑圧低減して表示し、いわば「音響的ピアノロール」を出力するような可能性を探った。このようなアプローチは、判定や抽出という過程を経ずに、“基本周波数らしさ” を視覚的に提示するため、音源数に制約のない、大小さまざまな音の混在をそのまま観察できることが期待できる。

第 2 章では本手法の原理について述べ、第 3 章で実音響信号に対して適用した結果について論じる。第 4 章では、実楽器音を対象に提案手法の効果を検証する。

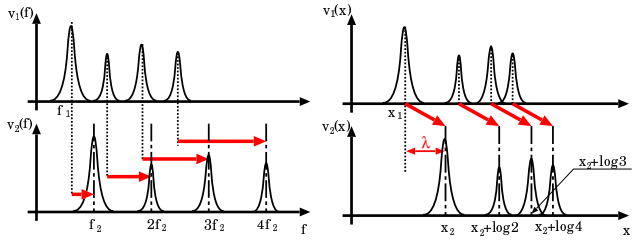
2. 対数周波数逆畳み込みによる基本周波数解析

2.1 対数周波数軸上での多重音スペクトル

まず、調和性を持つ単一音のスペクトルの場合を考えよう。基本周波数と第 2 倍音、第 3 倍音、第 4 倍音、...、第 n 倍音の周波数の関係は整数倍である。ここで周波数軸 f を対数周波数 $x = \log f$ に変換すると、これらの倍音は、x 軸では、対数基本周波数からそれぞれ $\log 2,\ \log 3,\ \log 4,\ \ldots,\ \log n$ だけ離れて位置し、それらの相対位置は基本周波数に依存しない。この関係を図 1 に示す。

次に、対象とする単一音の性質として、その基本周波数に依らず x 軸上の倍音の強度の比のパターンは一定であると仮定する。これを共通調波構造と呼び、対数基本周波数を原点に取って $h(x)$ と表すことにする。対数基本周波数が与えられれば、そこに $h(x)$ を異動することで、その単一音のスペクトルを表現できる。

異なる基本周波数の単一音が重畳した多重音の場合は、x 軸上にこれら異なる基本周波数の対数の位置ごとに共通調波構造



(a) 線形周波数軸上　(b) 対数周波数軸上

図 1　基本周波数と倍音の関係

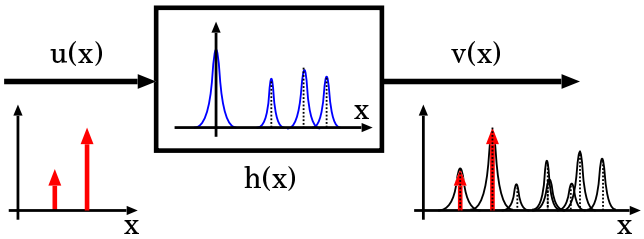


図 2　多重音生成のブロック線図表現 $u(x)$: 基本周波数分布、$h(x)$: 共通調波構造

$h(x)$ を配置すると、それらの和がそのスペクトルになる。構成音の強度（エネルギー）が異なるならば、それを係数として $h(x)$ に乗ずれば良い。

一般化し、多重音について、その構成音の基本周波数と強度の分布を表す関数を x 軸上で $u(x)$ と定義する。有限個数の多重音の場合は、$u(x)$ はデルタ関数の線形和であり、離散的な関数（正確には超関数）である。

以上から、対数基本周波数と強度の分布が $u(x)$ で、共通調波構造分布が $h(x)$ であるならば、多重音の対数周波数スペクトル $v(x)$ はこれらの畳み込み

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

として表現できる。この関係を図 2 に示す。

以上において、$u(x)$ は必ずしも離散関数である必要はなく、また $h(x)$ は調和関係である必要はなく、このような畳み込みの関係は一般的に成立する。また、x 軸を時間軸に見立てると、信号 $u(x)$ を、インパルス応答が $h(x)$ であるような線形フィルタに入力すると、その出力は $v(x)$ となると解釈することができ、線形系とのアナロジーを考えることができる。

2.2 逆畳み込みによる基本周波数解析

前節の議論から、共通調波構造 $h(x)$ が存在して既知ならば、基本周波数推定の問題は (1) 式の逆問題として定式化できる。すなわち、基本周波数の分布は、多重音から観測した対数周波数スペクトル $v(x)$ に対して、共通調波構造 $h(x)$ を逆畳み込み (deconvolution)

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

することにより復元できる。$h^{-1}(x)$ は、$h(x)$ をフィルタと見なした場合の逆フィルタのインパルス応答に相当するが、このような逆畳み込みは、フーリエ領域での除算としても行える。

$u(x), v(x), h(x)$ の（逆）フーリエ変換をそれぞれ $U(y), V(y), H(y)$ とすると、(2) 式は、

$$U(y) = \frac{V(y)}{H(y)} \tag{3}$$

と書ける。このような y 領域からフーリエ変換によって $u(x)$

を求めることができる。なお、この y 領域表現を次節で述べる理由により specmurt と呼ぶ。

以上は $h(x)$ が全構成音について共通で、かつ既知であると仮定していたが、これらの仮定が成り立たない場合でも、基本周波数成分のみを強調する効果は期待できる。これを、後の章で実験検証する。

2.3 ケプストラム法と本方法の対比

音声信号処理の分野において、基本周波数抽出にしばしば利用されるケプストラムと本方法の間に興味深い関係がある。信号のスペクトルは、周波数を横軸に、スペクトル強度を縦軸に取って表現されるが、それより

- 縦軸を対数変換し逆フーリエ変換すると cepstrum
- 横軸を対数変換し逆フーリエ変換すると specmurt

が得られるという関係がある。

スペクトラム（spectrum）の縦軸を対数変換し逆フーリエ変換して得られるものを、最初の 4 文字を逆転してケプストラム（cepstrum）と呼ぶのであるから、我々は、スペクトラム（spectrum）の横軸を対数変換し逆フーリエ変換して得られるものを、最後の 4 文字を逆転してスペクマート（specmurt）と呼ぶことにする。また、スペクマートを用いて多重音を解析するこの手法を、ケプストラム法に習ってスペクマート法と呼ぶことにする。

両手法とも基本周波数推定あるいは解析に用いられるが、ケプストラム法は非線形性のために、原理的には多重音の解析に適用できない。スペクマート法（本方法）は、調波構造が共通で既知であると仮定する制約はあるが、多重音の基本周波数が解析できる。

2.4 定 Q フィルタを用いた基本周波数解析手順

提案手法では、スペクトルを対数周波数軸上で観測する。このため、対数周波数軸上で等しい窓幅を持つ分析フィルタを使用することにする。これは、音楽で用いられている音階との整合性もよい。そこで、対数周波数スペクトルを抽出するための分析フィルタとして、中心周波数が指数関数的に変化する定 Q フィルタバンクを利用する。ピッチ変動に対し振幅変動が小さい場合、調波信号を解析する上で定 Q フィルタが頑健性に優れている [5]。そこで、分析中心周波数が指数関数的に変化する定 Q フィルタバンクを利用する。これは、人間の聴覚器官とも類似しており、生理物理学的な興味も持たれているものである [8]。

以上のスペクトル解析法を用いて、本分析手順は次のようにまとめられる。仮定した共通調波構造 $h(x)$ のフーリエ変換を $H(y)$ とすると、短時間フレームごとに

(1) 多重音信号を定 Q フィルタバンク分析して $v(x)$ を得る。
(2) それをフーリエ変換して $V(y)$ を得る。
(3) $V(y)/H(y)$ を逆フーリエ変換して $u(x)$ を得る。

を実行する。

3. 実 験 検 証

本手法（specmurt 法）の理論を検証するため、モノラル音楽音響信号の基本周波数解析を行った。解析条件は表 1 の通りである。共通調波構造 $h(x)$ としては、線スペクトル形状の倍音構造を仮定し、各調波成分の強度比は周波数に反比例する（すなわち $1/f$ 特性のエネルギー分布）とした。

3.1 MIDI 音源を対象にした実験検証

まず理論検証のため MIDI 音源を対象とした動作確認を行った。図 3 は、RWC 研究用音楽データベースのクラシック音楽データベース [7] に付属している “J. S. Bach: 音楽の捧げ物（BWV1079）から「6 声のリチェルカーレ」” の MIDI データの一部を使用して行った実験結果例である。この処理により高調波成分が低減されて基本周波数成分が観察しやすくなり、いわば「音響信号ピアノロール表示」が得られている。参照のために図 3(c) に示した原 MIDI 信号のピアノロールとの対応が観察できる。

表 1 多重音の基本周波数解析の実験条件

処理データ	サンプリング	16(kHz)
	フレーム長	64(msec) [1024 点]
	フレームシフト	32(msec) [512 点]
分析フィルタ	形状	Gabor 関数
	分散 Q 値	6.03% [∼100(cent)] 8.35% [∼140(cent)]
	分解能	20(cent)
調波構造モデル	形状	整数倍音の線スペクトル
	高調波の強度比	周波数に反比例 ($1/f$ 特性)
	考慮した次数	第 14 次高調波まで

表 2 実験に用いた RWC 研究用音楽データベースのモノラル音楽音響信号

楽器名	DISC	ファイル名	使用音名
ピアノ	1	011PFNOF.WAV	A4 B4 C#5
トランペット	6	211TRNOF.WAV	D#5 E5 A5
アルト（女声）	11	461ALA1F.WAV	F4 G4 A4 B4 C#5

表 3 単音のモノラル音楽音響信号から合成した多重音

	ピアノ, トランペット	アルト（女声）
二重音	A4,B4	F4,G4
三重音	A4,B4,C#5	F4,G4,A4
四重音	A4,B4,C#5,D#5	F4,G4,A4,B4
五重音	A4,B4,C#5,D#5,F5	F4,G4,A4,B4,C#5
オクターブ	A4,A5	B3,B4
半音	A4,A#4	A4,A#4

3.2 単一実楽器音の多重音を対象とした実験検証

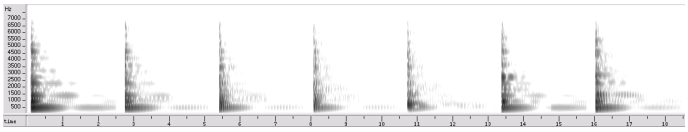
実楽器の場合でも、同一楽器音ならば、調波構造 $h(x)$ が比較的類似しており共通に扱える可能性がある。そこで、まず同一楽器音の多重音の基本周波数解析実験を行った。

実験では、RWC 研究用音楽データベースの楽器音データベース [6] から、ピアノ、トランペット、アルト（女声）の単音データを使用し、これを人工的に合成した多重音を用いた。表 2 に、このデータベースの中から使用した音響信号のファイル名および音名を示す。表 3 にこれを用いて人工的に作成した多重音の構成を具体的に示す。

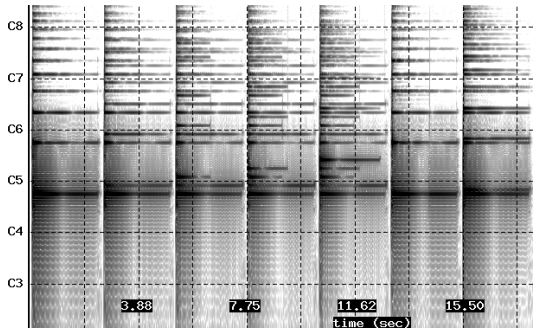
図 4 は本手法との比較のために、ピアノにおける通常のスペクトログラムと、対数周波数スペクトル、そして提案手法による基本周波数解析結果を示したものである。また、図 5 は、表 3 に示すようなその他の楽器における多重音の解析を行った結果を示したものである。これらの図では、濃淡で基本周波数の分布 $u(x)$ の推定結果を示す。図 4 において、高調波成分のスペクトル強度が低減され、相対的に基本周波数が強調されている様子を観察できる。

3.3 実楽器による音楽演奏を対象とした実験検証

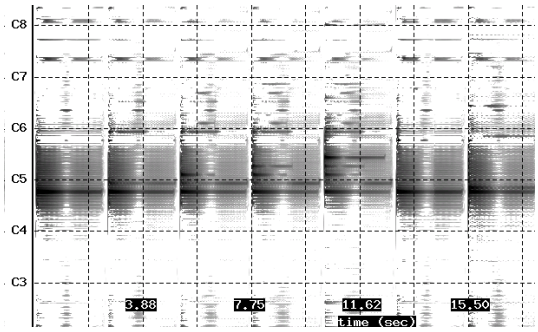
異種の実楽器音が混在する音楽信号について本手法の有効性を調べるため、RWC 研究用音楽データベースのクラシック音



(a) 線形周波数軸上 （表 3 に示した様々な多重音）



(b) 対数周波数軸上 （表 3 に示した様々な多重音）



(c) 基本周波数解析結果

図 4　本方法によるピアノの多重音の解析結果例: 横軸は時間（フレーム番号）、縦軸は音名（対数周波数）を示す

楽データベースを対象として、実音楽信号の基本周波数解析実験を行った。図 6 は、このデータベースの Disc3 に収録されている “J. S. Bach: 音楽の捧げ物（BWV1079）から「6 声のリチェルカーレ」” の一部分（3.1 節で実験に用いた音楽と同一曲の同一部分）についての実験結果例である。MIDI 信号の場合ほど顕著ではないが、異種の楽器音が混在する実際の音楽信号についても、基本周波数以外の調波成分の低減の効果が見られる。

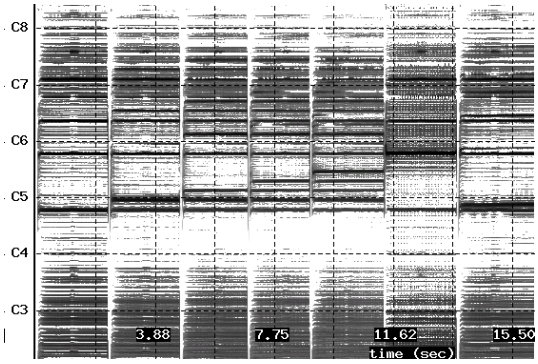
4. おわりに

本報告では、対数周波数軸で多重音のスペクトルを考えることにより、共通調波構造を仮定して多重音のモノラル音響信号の基本周波数分布を解析する手法（specmurt 法）を提案した。検証のため、MIDI 音源、単一楽器の多重音、異種楽器の合奏を対象にした実験を行った。

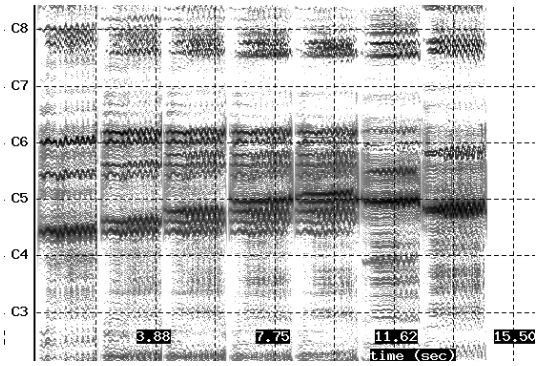
今後は、逆フィルタに代えてマッチトフィルタ（matched filter）の利用や、複数種類楽器の多重音の基本周波数解析方法への拡張を検討する予定である。また、解析結果の表示方法についても、カラー表示にするなど、情報のより豊富な表示方法を検討する。

謝　　辞

本原稿を作成するに当たって、同研究室の武田晴登、亀岡弘和の両氏に多大な協力を頂いた。ここに記して感謝する。



(a) トランペットの多重音



(b) アルト（女声）の多重音

図 5　実楽器音の多重音を用いた基本周波数分析実験結果例 （表 3 の多重音構成）

文　　献

[1] 後藤真孝, “実世界の音楽音響信号を対象としたメロディーとベースの音高推定,” 情報処理学会研究報告, SIGMUS31-16, pp.91-98, 1999.

[2] M. Goto, “A Predominent-F_0 Estimation Method for CD Recordings: MAP Estimation Using EM Algorithm for Adaptive Tone Models,” Proceedings of the 2001 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2001, pp. V-3365–3368, 2001.

[3] 亀岡弘和, 西本卓也, 篠田浩一, 嵯峨山茂樹: “ハーモニッククラスタリングによる多重音の基本周波数推定アルゴリズム,” 情報処理学会研究報告, SIGMUS50-5, pp.37-43 ,2003.

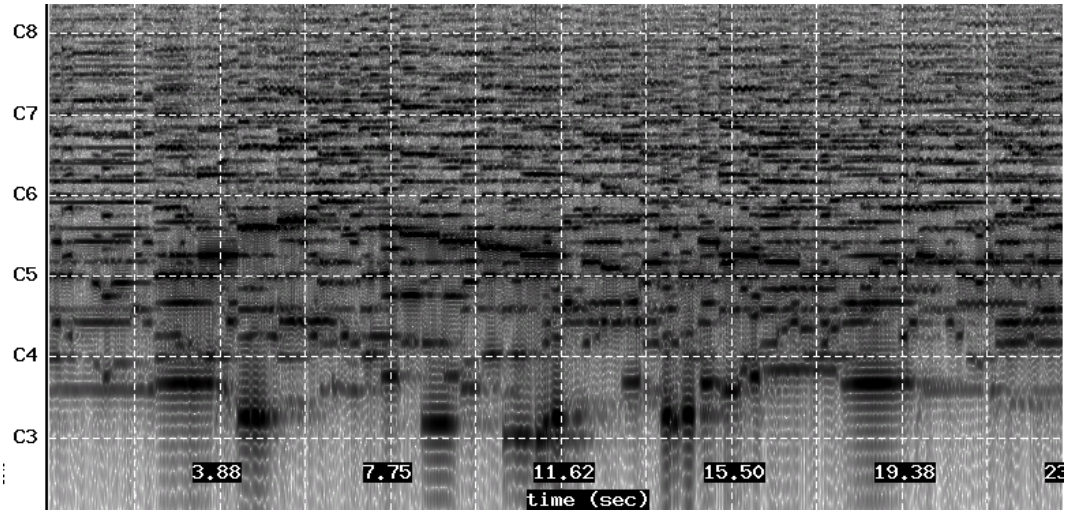
[4] 亀岡弘和, 西本卓也, 嵯峨山茂樹: “ハーモニッククラスタリングによる多重音信号音高抽出における音源数とオクターブ位置推定,” 情報処理学会研究報告, 2003-MUS-51, pp. 29-34, Aug 2003.

[5] 西 一樹, 安藤 繁: “ピッチ変動と振幅変動に頑健性を有する調波信号抽出フィルタ,” 電気学会論文誌, Vol.119-E, No.4, pp.215-220, 1999.

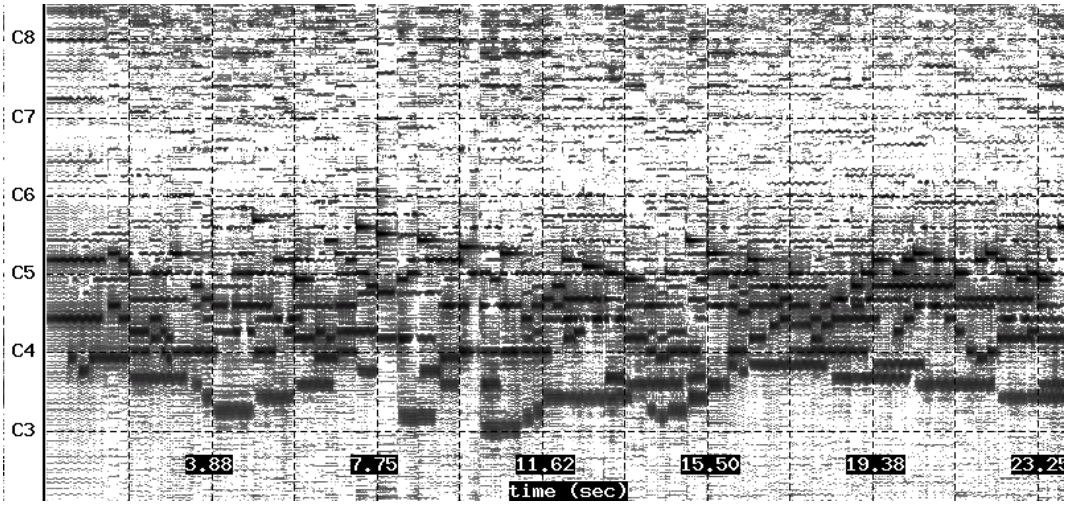
[6] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: “RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース,” 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-45-4, Vol.2002, No.40, pp.19-26, May 2002.

[7] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: “RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース,” 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-44-5, Vol.2002, No.14, pp.25-32, February 2002.

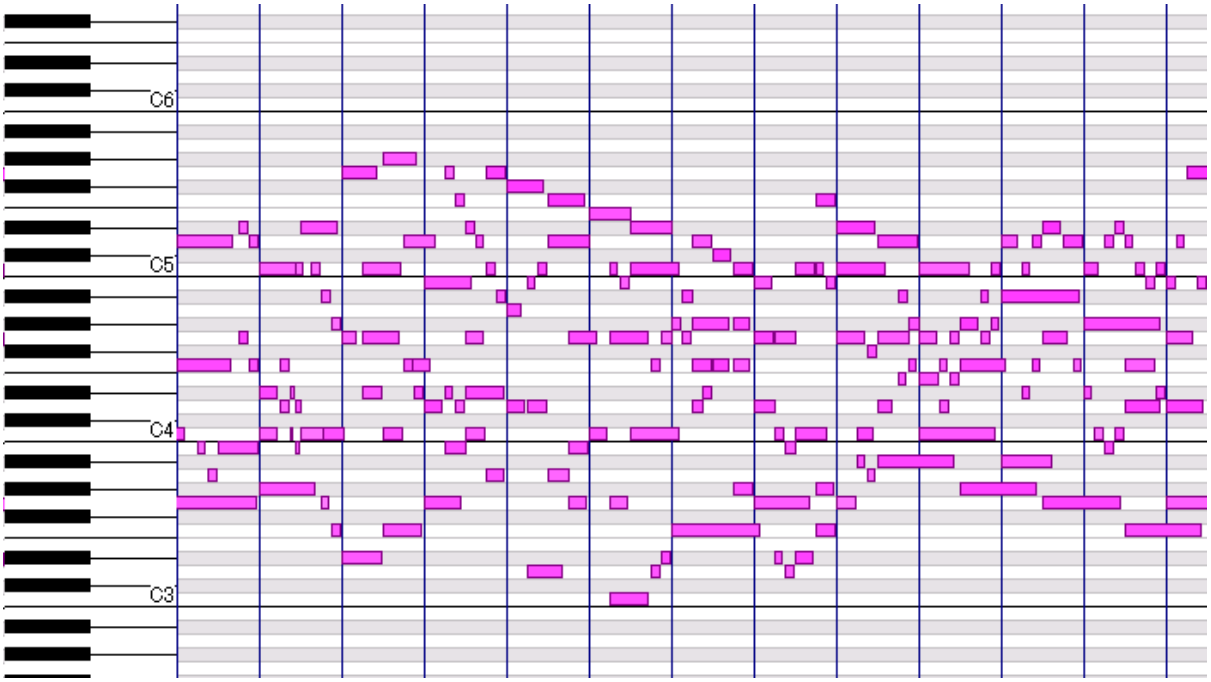
[8] 淀川, 東倉, 中根, “視聴覚の認識科学,” 電子情報通信学会, pp. 99-128, 1998.



(a) 解析前の対数周波数スペクトログラム （本文中の $v(x)$）

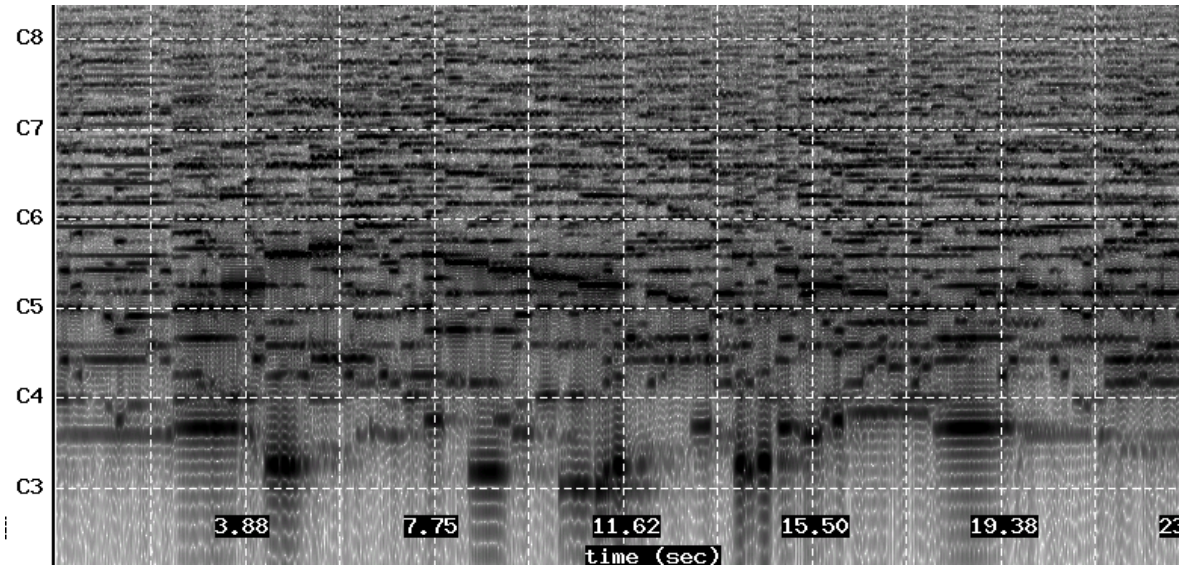


(b) 本方法による基本周波数分布解析結果の濃淡表示 （本文中の $u(x)$）

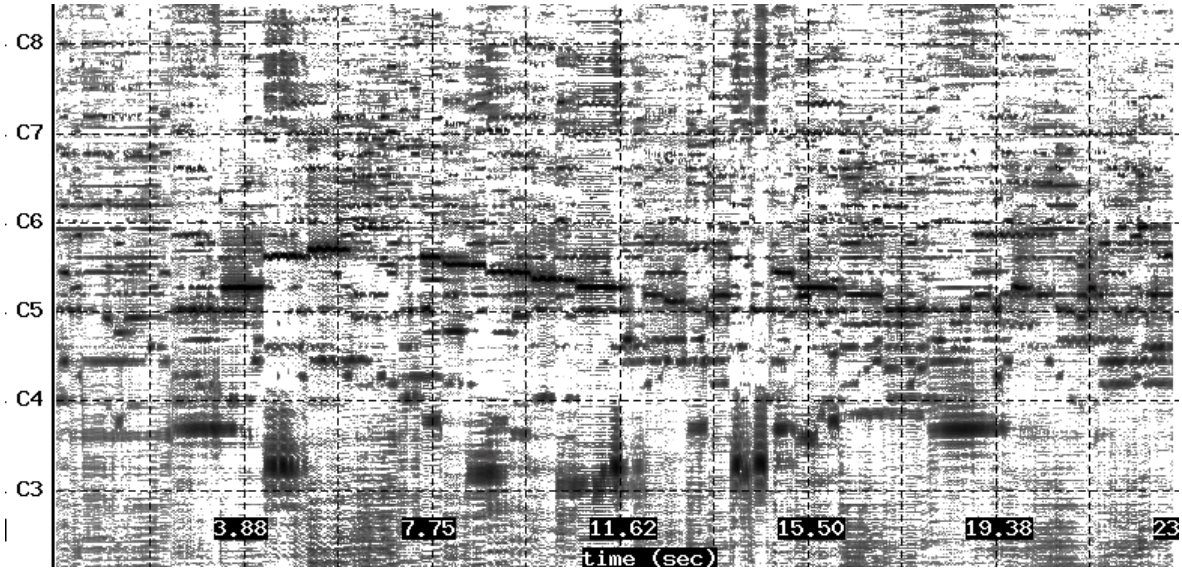


(c) MIDI 信号のピアノロール表示

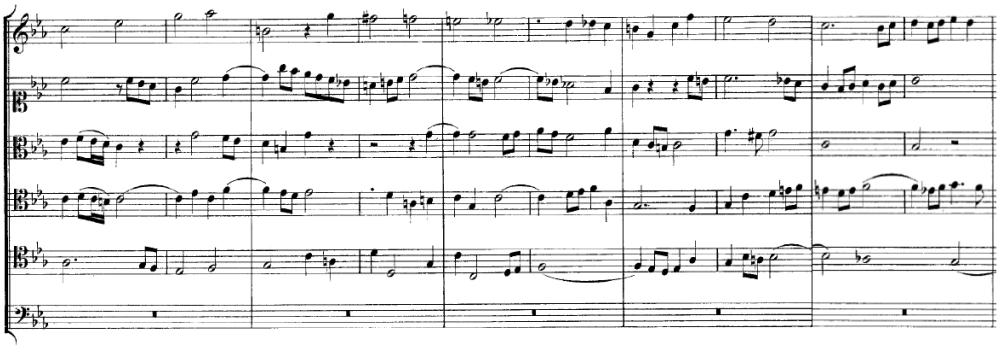
図 3　MIDI 音源による “J. S. Bach: 音楽の捧げ物 （BWV1079） から「6 声のリチェルカーレ」” の一部分の音響信号解析結果例 （横軸は時間、縦軸は音名で表した周波数を表す）



(a) 本方法適用前の対数周波数スペクトログラム （本文中の $v(x)$）



(b) 本方法による基本周波数分布解析結果の濃淡表示 （本文中の $u(x)$）



(c) 対応する楽譜 （第 19 小節から第 23 小節。楽譜の位置と (a), (b) の表示がほぼ対応）

図 6　実楽器合奏による音楽演奏 “J. S. Bach: 音楽の捧げ物 (BWV1079) より「6 声のリチェルカーレ」” の一部分の音響信号解析結果例 （横軸は時間、縦軸は音名で表した周波数を表す）