

平均律音楽の純正律化—ピアノソロでの試み—

松村 崇志^{1,a)} 井上 真郷^{1,b)}

概要: 平均律, 純正律は音律 (音楽で使われる音の周波数比の決め方) の一種で, 代表的な楽器はそれぞれピアノ, バイオリンである. 何れも基準音 (ピアノでは 88 鍵盤の中央近くのラの周波数を 440Hz としたりする) に対して相対的な周波数比により各音の周波数を定め, 平均律の場合は「半音上がる毎に 2 の $1/12$ 乗倍」と, 無理数を用いて決める. 一方純正律は小さな正の整数の比を用いるため, 和音 (異なる高さの音を複数同時に鳴らす) の構成音が調和してうなりのない美しいものとなる. しかし, 純正律はハ長調などの調毎に周波数比率が異なるため, ピアノなどの演奏中に再チューニングが出来ない楽器では, 転調が出来ないという欠点がある. 本研究では matching pursuit とピアノの特性を利用して, 一つのピアノで演奏された楽曲の録音データを純正律で演奏されたような楽曲に変換する手法を提案する. これにより平均律楽器の表現の幅が広がる事が期待できる. 結果, 時間-周波数スペクトルを期待通りに変換でき, 音楽経験の乏しい人でも違いを聞き分けることが出来た.

キーワード: matching pursuit, クロマベクトル, ピアノ音楽, 平均律, 純正律

Translation of Equal Temperament Music into Pure Intonation music: A Trial for Piano Solo Plays

MATSUMURA TAKASHI^{1,a)} INOUE MASATO^{1,b)}

Abstract: Both 12-tone equal temperament and pure intonation are musical temperaments, the representative examples of their instruments are pianos and violins, respectively. In both temperaments, the frequency of the standard tone A near the center of a keyboard is set to 440Hz. In the equal temperament, the frequency ratio of every neighboring semitones is set to 2 to the power $1/12$. On the other hand, in the pure intonation, the frequency ratio is set to the fraction of small positive integers. According to this feature, chords in pure intonation are usually more beautiful than ones in equal temperament. However, pure intonation has disadvantages; the frequencies of tones depend on the tonality of a music and just-in-time re-tuning at every modulation is not a realistic idea for fixed-tone instruments, e.g., pianos. In this report, we suggest a method to translate the recorded sound data of a piano solo play (of course in equal temperament) into ones in the pure intonation. In this method, we utilize the matching pursuit algorithm and some characteristics of pianos. As a result, the change of time-frequency spectrum was as expected. Moreover, the change of hearing impression was somewhat obvious even for musical novices.

Keywords: matching pursuit, chroma vector, piano music, equal temperament, pure intonation

1. 背景・目的

1.1 表記

ピアノの各鍵盤に対応する音は音名+オクターブ数で表記する. 音名は C, C#, D, D#, E, F, F#, G, G#, A, A#, B, (ド, ド#, レ, レ#, ミ, ファ, ファ#, ソ, ソ#, ラ, ラ#, シ) の 12

¹ 早稲田大学大学院 先進理工学研究科 電気・情報生命専攻
Department of Electrical Engineering and Bioscience, Graduate School of Advanced Science and Engineering, Waseda University, Shinjuku-ku, Tokyo, 169-8555, Japan

a) w-only.matsumura@akane.waseda.jp

b) masato.inoue@eb.waseda.ac.jp

個で、高さを表すオクターブ数は0から7の整数で定義される。オクターブはCから始まってBまで同じ値をとり、例えばA4はオクターブ4のA(ラ)の音を意味する。本研究の対象となるピアノ音楽は図1のように88個の鍵盤のうちA2からG#6の48鍵盤を用いて演奏されたものとする。また、本研究での平均律は十二平均律を指す。

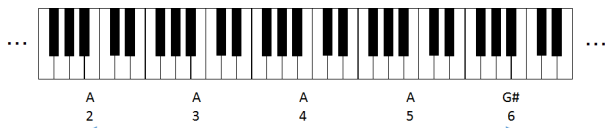


図1 用いる鍵盤

$\mathbb{R}, \mathbb{C}, \mathbb{Z}$ は各々実数体, 複素数体, 整数集合, \mathbb{R}^N は N 次元の実ベクトル空間, \mathbb{Z}^+ は正の整数の集合を表すとする。また $\Re()$ は括弧内の複素ベクトルの実部, ベクトル \mathbf{a}, \mathbf{b} の内積は $\mathbf{a} \cdot \mathbf{b}$, $\|\mathbf{a}\|_1, \|\mathbf{a}\|_2$ はそれぞれ \mathbf{a} の L1, L2 ノルムを表すとする。更に周波数 (Hz) は小数点第二位以下を四捨五入して表記する。

1.2 平均律・純正律

楽器が発する音の周波数の決め方の代表例に平均律・純正律がある。この特徴をピアノを例に比較する。

平均律楽器の代表であるピアノは、まずA4を一般的に440Hzと決め、他の鍵盤の周波数は「半音上がる毎に $2^{1/12}$ 倍」という無理数比を用いた法則で決める。これは「1オクターブ上がると周波数が2倍になる」という性質から、対数軸上でオクターブを均等に分割したものである。

これに対し純正律では、ハ長調などの調ごとに異なる整数比で他の鍵盤の周波数を決める。ここで調とは曲の雰囲気などを決めるものであり、一部の例外を除き曲はある時刻において必ず1つの調に属する。全部で24種類の調が存在し、曲中で調が変わる事を転調と呼ぶ。

具体的に、ハ長調のとき平均律と純正律のそれぞれの場合のC4, D4, E4, F4, G4, A4, B4の周波数を表1で示す。

音	C4	D4	E4	F4	G4	A4	B4
平均律	261.6	293.7	329.6	349.2	392.0	440.0	493.9
純正律	264.0	297.0	330.0	352.0	396.0	440.0	495.0

表1 ハ長調の場合のC4, D4, E4, F4, G4, A4, B4の周波数 (Hz)

いま、ハ長調において、C4, E4, G4の鍵盤を同時に鳴らすことで得られる和音について考える。ここで、鍵盤に対応する音はその鍵盤が持つ周波数以外に n 倍 ($n = 1, 2, \dots$) の周波数成分(倍音)と半分の $1/2$ 倍音を含む。まず平均律の場合のC4, E4, G4の和音に含まれる周波数成分は表2のようになる。

音	1/2 倍音	基音	2 倍音	3 倍音	...	n 倍音
C4	130.8	261.6	523.3	784.9	...	261.6 · n
E4	164.8	329.6	659.3	988.9	...	329.6 · n
G4	192.0	392.0	784.0	1176.0	...	392.0 · n

表2 C4, E4, G4の和音に含まれる周波数成分(平均律)

例えばC4の3倍音とG4の2倍音に注目すると、このような近い周波数を持つ成分がうなりを生み出してしまい、和音が不協和音となってしまう。一方純正律の場合のC4, E4, G4の和音は、周波数比率がC4:E4:G4=264.0:330.0:396.0(=4:5:6)という整数比であり、和音に含まれる周波数成分は表3のようになる。

音	1/2 倍音	基音	2 倍音	3 倍音	...	n 倍音
C4	132.0	264.0	528.0	792.0	...	264.0 · n
E4	165.0	330.0	660.0	990.0	...	330.0 · n
G4	198.0	396.0	792.0	1098.0	...	396.0 · n

表3 C4, E4, G4の和音に含まれる周波数成分(純正律)

このように純正律は倍音成分の周波数が一致してハモるため、和音が美しく鳴る。ただし、純正律は調によって比率の整数値が異なるため、転調する曲に対応する事が出来ない。そのため、ピアノのように演奏中すぐにチューニングができない楽器は妥協案として平均律を用いて調律される。例えばバイオリンなどは奏者によって音の高さを微調整することで純正律を再現できる楽器である。

本研究では、平均律楽器の代表例であるピアノのみで演奏された音楽の録音データを、純正律で演奏されたかのように変換することを目的とする。平均律にも独自の美しさがあり、純正律は一部の和音は却って汚い響きになるなどの欠点もあるが、平均律でしか演奏できなかった楽器の表現の幅が広がることを期待できる。

2. 手法

入力音楽波形が持つ平均律周波数の振幅・位相の時系列情報を計算し、それぞれの平均律周波数を、入力音楽波形が属する調によって一意に対応する純正律の周波数に置き換え、波形を再構成し、純正律の出力音楽波形を得る。ただし本研究では、曲中の調はあらかじめ与えられているものとした。手法の大まかな流れを図に示す。

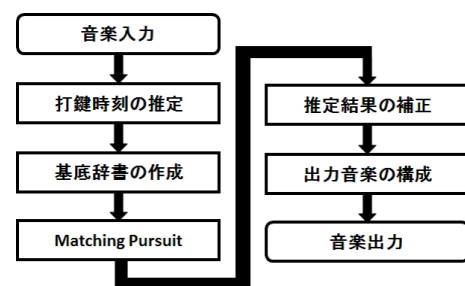


図2 フローチャート

2.1 打鍵時刻の推定

まず曲中で音が新しく発生した区間、つまりピアノが鍵盤を叩いた打鍵時刻を窓番号単位で検出する手法 [1] を示す。入力波形ベクトル \mathbf{x} を K 個の窓に区切った波形ベクトル $\mathbf{x}^k (k = 1, 2, \dots, K)$ について、パワーの変化の尺度 α_p と時間-周波数スペクトルの変化の尺度 α_f を以下の式で定義する。

$$\alpha_p \equiv \frac{P(k+1) - P(k)}{P(k)}$$

$$\alpha_f \equiv \frac{\sum_f p(k+1, f) - p(k, f)}{P(k)}$$

このどちらかが閾値を超えた窓番号 k を打鍵時刻と認識することにする。

$$\alpha_p > \beta_1 \text{ or } \alpha_f > \beta_2$$

ここで $P(k)$ は \mathbf{x}^k のパワー、 $p(k, f)$ は \mathbf{x}^k の離散 Fourier 変換の周波数 f のパワースペクトルを表す。閾値を超えた L 個の窓を打鍵時刻 (窓番号) の集合 $\mathbf{D} \equiv (d^1, d^2, \dots, d^L)$ と定義する。

2.2 基底辞書の作成

次に 2.3 で解説する matching pursuit で用いる周波数の辞書 \mathbf{S} をどのように決めるかを述べる。ここで用意したい周波数の辞書とは、例えば C4, E4, G4 の和音が鳴っている音楽については C4, E4, G4 の鍵盤から発生する周波数を指す。しかし、音名に加え音の高さも推定することは困難であることから、音名のみを推定し、その音名を持つすべての鍵盤から発生する周波数を列挙することにする。具体的には、2.1 で求めた打鍵時刻 (窓番号) $\mathbf{D} \equiv \{d^1, d^2, \dots, d^L\} (d^l \in \mathbb{Z}^+)$ について、 d^l の波形から計算されたクロマベクトル $\mathbf{V} \equiv \{\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^L\} (\mathbf{v}^l \in \mathbb{R}^{12})$ によってその打鍵時刻 d^l で鳴らされた音 $\mathbf{G} \equiv \{\mathbf{g}^1, \mathbf{g}^2, \dots, \mathbf{g}^L\} (\mathbf{g}^l \in \mathbf{X}^{P_l}, \mathbf{X} \equiv \{1, 2, \dots, 12\})$ を推定し、その音を持つすべての鍵盤から発生する周波数の集合 $\mathbf{S} \equiv \{\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^L\} (\mathbf{s}^l \in \mathbb{R}^{M_l})$ を列挙する。ここで P_l, M_l は打鍵時刻 d^l で鳴らされた音名の個数およびその音名を持つすべての鍵盤から発生する周波数の個数である。特に複数の音が鳴らされたと推定された場合は、解析のために \mathbf{s}^l の要素の一部を統合する処理を行う。

48 鍵盤すべての周波数ではなく、ある程度用意する周波数を限定するのは、対象の音楽を演奏しているピアノに調律のずれがある場合など近い周波数を持つ成分に間違えられてしまうことがあるためである。特に対応する純正律が異なる周波数成分に間違えられると大きく結果が変わってしまう。

1. クロマベクトルによる音名決定

各次元が入力波形ベクトル $\mathbf{x} \in \mathbb{R}^N$ における 12 半

音の成分を表す 12 次元の特徴ベクトル [2] であり、 $\mathbf{v} \equiv (v_1, v_2, \dots, v_{12})$ と定義する。計算手順を以下に示す。

- (1) \mathbf{x} の離散 Fourier 変換のパワースペクトル $p(f_i) (f_i \equiv f^* i/N, i = 0, 1, 2, \dots, N/2)$ を求める。ただし f^* はサンプリング周波数である。
- (2) $p(f_i)$ の周波数軸を、単位 Hz の周波数 f_i から単位 cent の周波数 g_i に変換し、 $p(g_i)$ とする。

$$g_i \equiv 1200 \log_2 \frac{f_i}{440 \times 2^{-4}}$$

- (3) \mathbf{v} の各次元 v_c を得る。ここで $c \equiv 1, 2, \dots, 12$ の整数値で 12 半音に対応する。

$$v_c \equiv \sum_{h=O_L}^{O_H} \sum_{i=0}^{N/2} H_{c,h}(g_i) p(g_i)$$

ただし、 O_L, O_H はオクターブの範囲を表し、今回はそれぞれ 2.5 とする。また H は

$$H_{c,h}(g_i) \equiv \begin{cases} \frac{1 + \cos 2\pi \frac{F_{c,h} - g_i}{200}}{2} & (|F_{c,h} - g_i| \leq 100) \\ 0 & (|F_{c,h} - g_i| > 100) \end{cases}$$

$$F_{c,h} \equiv 1200h + 100(c - 1)$$

という Hanning 状の窓で定義する。さらに $F_{c,h}$ は平均律の半音は 100cent に、つまり 1 オクターブは 1200cent に相当することを考慮して、音名 c 、オクターブ数 h の周波数を cent 単位で表すものである。

- (4) v_c をその総和 $\sum_{c=1}^{12} v_c$ で割ることで $\|\mathbf{v}\|_1 = 1$ となるように正規化する。

この手順で打鍵時刻 d^l の波形から \mathbf{v}^l を得る。 \mathbf{v}^l の各要素 v_c^l について閾値 β_3 を越えた音 c は鳴っていると判定する。

$$v_c^l > \beta_3$$

打鍵時刻 d^l について、閾値を超えた音名 $\{c_1, c_2, \dots, c_{P_l}\}$ を $\mathbf{g}^l \equiv \{c_1, c_2, \dots, c_{P_l}\}$ とし、 $\mathbf{G} \equiv \{\mathbf{g}^1, \mathbf{g}^2, \dots, \mathbf{g}^L\}$ を得る。以降、窓番号 $k (d^l \leq k < d^{l+1}, l = 1, 2, \dots, L - 1)$ の区間では $\mathbf{g}^l \in \mathbf{G}$ の音が継続して鳴っているものとする。また、1 窓分のサンプルだけで推定すると誤推定することがあるので、打鍵時刻の周辺でやや長めに切り取った波形で計算した。

2. 辞書の作成 (列挙)

打鍵時刻 d^l で鳴らされた音 \mathbf{g}^l についてその音を持つすべての鍵盤から発生する周波数を \mathbf{s}^l にすべて列挙する。例えば C の音のみ鳴っていると推定された場合、C の鍵盤は 48 鍵盤の内 C3 から C6 の全部で 4 つがあり、これらの鍵盤によって発生する周波数の一覧は、各鍵盤の倍音数を 6 としたとき、図 4 のように 15 個となる (重複により 24 個から減らされる)。

音	1/2 倍音	基音	2 倍音	3 倍音	4 倍音	5 倍音
C3	65.4	130.8	261.6	392.4	523.3	654.1
C4	130.8	261.6	523.3	784.9	1046.6	1308.0
C5	261.6	523.3	1046.6	1569.8	2093.2	2616.5
C6	523.3	1046.6	2093.2	3139.6	4186.4	5233.0

表 4 C3 から C6 の鍵盤によって発生する周波数 (Hz) 一覧

3. 辞書の作成 (要素の結合)

打鍵時刻 d^l で鳴らされた音 g^l の要素が複数ある場合は和音が鳴っていると推定し、特に対応する純正律の周波数が一致するような値に近い周波数 f_x, f_y は、その平均値 $f_z=(f_x + f_y)/2$ という 1 つの周波数にまとめる処理を行う。例えば C,E,G の音が鳴らされた ($g^l = \{C,E,G\}$) と推定された場合、各音の周波数を全て列挙すると、C の音には $f_x=1308.1(\text{Hz})$, E の音には $f_y=1318.5(\text{Hz})$ の成分が含まれ、どちらも純正律周波数は $1320.0(\text{Hz})$ に対応する。そこで 2 つの代わりに、平均の $f_z=(1308.1 + 1318.5)/2 = 1313.3(\text{Hz})$ を辞書に登録する。これは、対応する純正律の周波数が一致するような値に近い周波数成分は 2.3 で解説する matching pursuit を適用した際に f_x の成分が誤って f_y に計算されてしまう恐れがあるためであり、2 つの周波数の中間値である周波数でまとめて計算することとする。

以上の様手順を踏むことで最終的に打鍵時刻 $d^l \in \mathbf{D}$ ごとに周波数の辞書 $s^l \in \mathbf{S}$ を得られる。以後、 s^l の要素数を M_l とする。

2.3 matching pursuit

matching pursuit とは、入力ベクトル $\mathbf{x} \in \mathbb{R}^N$ を M 個の基底ベクトル $\phi_m \in \mathbb{R}^N (m = 1, 2, \dots, M)$ の線形和で近似的に表現するための手法 [3][4] である。解である線形和の係数を $\mathbf{w} \equiv (w_1, w_2, \dots, w_M), w_m \in \mathbb{R}$ とする。

$$\mathbf{x} \simeq \sum_{m=1}^M w_m \phi_m$$

1 ステップごとに基底を 1 つ選択し、残差ベクトル \mathbf{r} を更新しながら \mathbf{x} を分解する。以下に matching pursuit のアルゴリズムを示す。ここではベクトルの添字は右上は反復回数 i , 右下は要素番号 m を示すことにする。

(1) 以下の初期化を行う。

$$\mathbf{r}^{(0)} \equiv \mathbf{x}, \quad \mathbf{w}^{(0)} \equiv \mathbf{0}$$

(2) 以下の更新を行う。

$$\begin{aligned} \hat{m} &\equiv \underset{m}{\operatorname{argmax}} |\mathbf{r}^{(i)} \cdot \phi_m| \\ \hat{w} &\equiv \mathbf{r}^{(i)} \cdot \phi_{\hat{m}} \\ \mathbf{r}^{(i+1)} &\equiv \mathbf{r}^{(i)} - \hat{w} \phi_{\hat{m}} \\ w_m^{(i+1)} &\equiv w_m^{(i)} + \begin{cases} \hat{w} & (m = \hat{m}) \\ 0 & (m \neq \hat{m}) \end{cases} \end{aligned}$$

(3) 以下の条件を満たしていれば $\mathbf{w}^{(i+1)}$ を解とし、そうでなければ i を 1 増やして (2) へ戻る。

$$1 - \frac{\|\mathbf{r}^{(i+1)}\|_2}{\|\mathbf{r}^{(i)}\|_2} < \beta_4$$

本研究では基底を音波形に拡張したものを扱う。入力をサンプリング周波数 f^* の音波形ベクトル $\mathbf{x} \in \mathbb{R}^N$ とし、基底は周波数集合 $\mathbf{F} \equiv \{f_1, f_2, \dots, f_M\} (f_m \in \mathbb{R})$ の各要素 f_m に対応した M 個のベクトル $\phi_m \in \mathbb{C}^N$ とする。

$$\begin{aligned} \mathbf{x} &\simeq \sum_{m=1}^M \Re(w_m \phi_m) \\ \phi_m &\equiv \frac{1}{\sqrt{N}} \left[e^{2\pi j (f_m / f^*) n} \right]_{n=0}^{N-1} \end{aligned}$$

ここで大きさだけでなく位相も考慮できるようにするために解である線形和の係数の各要素を $w_m \in \mathbb{C}$ とする。前述のアルゴリズムの (2) の更新部分を以下の様に変更する。ただし $W(\mathbf{r}^{(i)})$ は窓関数 W にかけた状態の $\mathbf{r}^{(i)}$ を、 $I(\hat{w})$ は窓関数による影響を無くするために大きさ $|\hat{w}|$ を $N/(W$ の面積) 倍した \hat{w} を表す。

$$\begin{aligned} [\hat{m}, \hat{\theta}] &\equiv \underset{[m, \theta]}{\operatorname{argmax}} (W(\mathbf{r}^{(i)}) \cdot \Re(e^{j\theta} \phi_m)) \quad (\theta \equiv (-\pi, +\pi]) \\ \hat{w} &\equiv \left\{ W(\mathbf{r}^{(i)}) \cdot \Re(e^{j\hat{\theta}} \phi_{\hat{m}}) \right\} e^{j\hat{\theta}} \\ \mathbf{r}^{(i+1)} &\equiv \mathbf{r}^{(i)} - \Re(I(\hat{w}) \phi_{\hat{m}}) \\ w_m^{(i+1)} &\equiv w_m^{(i)} + \begin{cases} I(\hat{w}) & (m = \hat{m}) \\ 0 & (m \neq \hat{m}) \end{cases} \end{aligned}$$

また、時間に伴う音の変化を追うために、 \mathbf{x} をそれぞれ K 個の窓関数で区切った $\mathbf{x}^k (k = 1, 2, \dots, K)$ に置き換える。窓番号 $k (d^l \leq k < d^{l+1}, l = 1, 2, \dots, L-1)$ では \mathbf{F} を 2.2 で用意した M_l 個の周波数集合 $s^l \in \mathbf{S}$ とする。窓番号 k の matching pursuit の結果を \mathbf{w}^k とし最終的には $\mathbf{W} = (\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^K)$ という時系列情報が得られる。

2.4 推定結果の補正

ある打鍵時刻 d^l から d^{l+1} における、2.2 で述べたような 2 つの周波数を統合し平均をとった周波数成分 f_z の振幅 $|w_z^k| (d^l \leq k < d^{l+1})$ はうなりが起きているため、ピアノ音楽らしい減衰をしない (2.3 では右下の添字を要素番号としたがここでは便宜上 $|w_a^k|$ は matching pursuit で計算した周波数 f_a の窓番号 k の係数を表す)。この事を実際に MIDI(譜面) データを基に作成したピアノ音楽で確認する。あるピアノの C4,E4,G4 の和音波形について、これまでの手順を踏んで、C の音の $f_x=1308.1(\text{Hz})$ と E の音の $f_y=1318.5(\text{Hz})$ の成分を、その平均である $f_z=(1308.1 + 1318.5)/2=1313.3(\text{Hz})$ を代わりに辞書に登録して、matching pursuit で得られた $f_z=1313.3(\text{Hz})$ の振幅 $|w_z^k|$ を図 3 に示す。

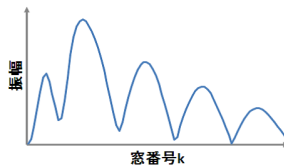


図3 得られた $|w_z^k|$

いま, 対応する純正律周波数が等しい倍音周波数 f_x, f_y を各々含む2つの鍵盤を同時に叩いて鳴らすことを考える. この場合, 本来の研究目的でもあるように, 純正律で調律された鍵盤で演奏された音楽に変換するならば, matching pursuit によって周波数成分 f_x, f_y のそれぞれの大きさ $|w_x^k|, |w_y^k| (k = d^l, d^l + 1, \dots, d^{l+1})$ を求める必要がある. しかし, 本研究ではこの2つをまとめて $f_z = (f_x + f_y)/2$ という1つの周波数成分として $|w_z^k|$ を求めており, この一つの計算結果から $|w_x^k|, |w_y^k|$ を求めることは不可能である. そこで得られた $|w_z^k|$ に対し自然な減衰をするような補正処理を行う. 本研究では単純にピークの前を直線で結ぶ方法 [7] を取った. この場合, 図3は以下の図4のように補正される. 図 (AfterRevise)

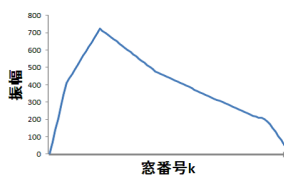


図4 補正後の $|w_z^k|$

2.5 出力音楽の構成

得られた K 個の窓の平均律周波数成分の情報 $\mathbf{W} = (w^1, w^2, \dots, w^K)$ をもとに純正律の音楽波形ベクトル \mathbf{y} を構成する. 入力波形ベクトル \mathbf{x} と同様に \mathbf{y} を K 個に区切った窓 $\mathbf{y}^k (k = 1, 2, \dots, K)$ に, それぞれ \mathbf{x}^k に対応させた純正律周波数に対応した波形ベクトル $\phi'_m \in \mathbb{C}^N$ を代入する. M^k は窓番号 $k (d^l \leq k < d^{l+1})$ の基底の要素数を示す.

$$\mathbf{y}^k \equiv \sum_{m=1}^{M_l} \Re(w_m^k \phi'_m) + \mathbf{r}^k$$

$$\phi'_m \equiv \frac{1}{\sqrt{N}} \left[e^{2\pi j (f'_m / f^*) n} \right]_{n=0}^{N-1}$$

ここで f'_m は f_m に一意に対応する純正律周波数である. また, \mathbf{r}^k は窓番号 k における matching pursuit 終了後の残差信号である. これを足し合わせるの, matching pursuit の結果である \mathbf{W} の情報のみで構成するとピアノの特徴であるハンマー音などがなくなりピアノらしさが失われた音楽になってしまうためである. matching pursuit を終了したときの残差信号 \mathbf{r}^k には音程を表す平均律周波数成分が

除去されハンマー音などピアノらしさを表す成分が残っていることを利用し, 最後に加えることにした. K 個の窓について \mathbf{y}^k を計算し, 得られた $\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^K$ から純正律の音楽波形ベクトル \mathbf{y} を得る [5].

3. 評価実験

3.1 条件

サンプリング周波数 f^* は 44100(Hz), 窓長は 2048(サンプル), 窓のシフト幅は 256(サンプル) とする. 窓関数は Hamming 窓を用いた. 閾値 $\beta_1, \beta_2, \beta_3, \beta_4$ は後述の MIDI 音楽データを複数用意して, 実験により $\beta_1 = 0.25, \beta_2 = 70, \beta_3 = 0.25, \beta_4 = 10^{-5}$ と決めた. 以降時刻 (t) をサンプル数の単位で表すことにする (1秒が 44100 サンプルに相当).

3.2 sin 波音楽データ (Input1.wav)

C4, E4, G4 の和音の周波数を持つ, 定常的な sin 波音楽データに提案手法を適用し, Output1.wav を得た. matching pursuit によりうなりを消すことができたか確認する. 以下の表5の周波数を持ち, 各々の振幅と初期位相が等しい sin 波を 18 個混ぜ合わせたものである.

音	1/2 倍音	基音	2 倍音	3 倍音	4 倍音	5 倍音
C4	130.8	261.6	523.3	784.9	1046.5	1308.1
E4	164.8	329.6	659.3	988.9	1318.5	1648.1
G4	196.0	392.0	784.0	1176.0	1568.0	1960.0

表5 sin 波音楽データに含まれる周波数成分 (Hz)

入力波形 \mathbf{x} と出力波形 \mathbf{y} の概形, および適当な窓番号 k における \mathbf{x}^k と \mathbf{y}^k の離散 Fourier 変換の 700(Hz) から 1400(Hz) 周辺のパワースペクトルを示す. ただし, $\mathbf{D} = \{d^1 = 0\}$, $\mathbf{G} = \{g^1 = \{C, E, G\}\}$ とあらかじめ与えた.

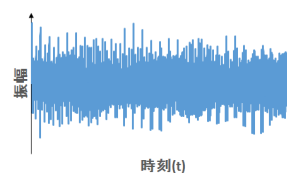


図5 \mathbf{x} の概形

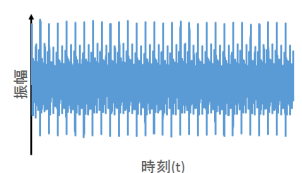


図6 \mathbf{y} の概形

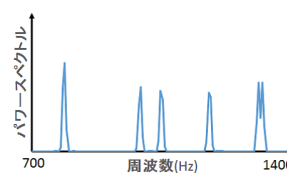


図7 \mathbf{x}^k のパワースペクトル

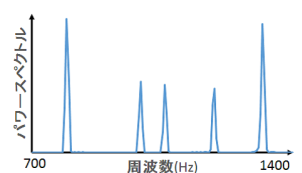


図8 \mathbf{y}^k のパワースペクトル

図5,6 から変換の前後でうなりが消えていることがわかる. また, 図7の5つのスペクトルの山が一番左が 784.9(Hz)

と 784.0(Hz), 一番右が 1308.1(Hz),1318.5(Hz) の周波数成分に対応する. 図 8 と比較すると 1308.1(Hz),1318.5(Hz) の成分が一つにまとまっていることが確認でき, また 784.9(Hz) と 784.0(Hz) 周辺のパワースペクトルについても実際はうなりを持つためパワースペクトルは周期的に動いているが, 理論通り 1308.1(Hz),1318.5(Hz) の成分と同程度のパワーを取るようになった. うなりを発生しない真ん中の 3 つの山は, 988.9(Hz),1046.5(Hz),1176.0(Hz) に対応し, 大きさを保ったまま変換できたことがわかる.

3.3 MIDI 音楽データ (Input2.wav)

次に, 自作の MIDI データにピアノのサウンドフォントファイルをあてることで作成した MIDI 音楽データに提案手法を適用し, Output2.wav を得た. ハ長調で使われる主音である, C4,D4,E4,F4,G4,A4,B4,C5 という 8 つの単音と C4,E4,G4 と F4,A4,C5 と G4,B4,D5 の 3 つの和音の計 11 音を順番に鳴らしたものである. サウンドフォントは [8] の HP のフリーのものを用了. まず, D, G の推定結果を正解とともに表 6 に示す. ただし D については窓番号の代わりにその窓の中央の時刻 (t) を掲載した.

正解		推定結果	
打鍵時刻	音名	D	G
22050	C	21248	C, Cs
33075	D	32512	C, D
		33280	C, D
44100	E	44032	D, E
55125	F	54784	E, F, Fs
66150	G	65536	F, G
77175	A	76544	G, A
88200	B	87808	B, A
99225	C	98816	C, B
110250	C, E, G	109568	C, E, G
132300	F, A, C	131584	F, A, C
154350	G, B, D	153600	G, B, D

表 6 MIDI 音楽データの D, G の推定結果

余分な音が若干含まれていたり, 打鍵時刻にずれが生じてはいるものの, 元の音楽が大きく音を失うことはなかった. 特に C4,E4,G4 の和音の部分に注目する. この和音の立ち上がり付近の窓 k における波形 x^k と y^k の離散 Fourier 変換のパワースペクトルの, 周波数 1300(Hz) から 1350(Hz) 周辺を実験の前後で観察すると図 9, 図 10 のようになった.



図 9 x^k のパワースペクトル

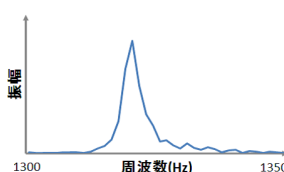


図 10 y^k のパワースペクトル

不協和音の原因である 1308.1(Hz) と 1318.5(Hz) の 2 つの成分が 1320(Hz) という 1 つの成分にまとめられ, 協和音が持つようなハモリの成分となっていることがわかる.

3.4 実音楽データ (Input3.wav)

ピアノ音色を持つ電子キーボードによってメヌエット (J.S. バッハ) の冒頭 8 小節を自ら演奏した実音楽データに提案手法を適用し Output3.wav を得た. MIDI 音楽データと同様, 元の音楽から大きく音を失うことなく変換できた.

4. 考察・今後の課題

提案手法によって sin 波で構成された人工音楽およびピアノで演奏された実音楽でスペクトルを期待通りに変化させることが出来た. 特に sin 波で構成された人工音楽では, 音楽経験のない素人でもはっきり違いを聴き取る事が出来た. また MIDI データによって再現したピアノ音楽とピアノの音色を持つ電子キーボードによって演奏したピアノ音楽に対して, 若干の打鍵時刻や音名の推定にずれはあったものの, 大きく入力音楽を損なうことなく変換ができた.

今後の課題としては, 楽曲の波形データから演奏しているピアノの調律のずれを取得方法ピアノの個体差に対応するための閾値設定の方法などがあげられる. 今回, 入力音楽を演奏しているピアノは理論上の周波数を持つように調律された状態であると仮定したが, 実際には, 調律のずれやインハーモニシティ [9] という, 倍音の周波数が理論値の整数倍より若干高くなる現象によって周波数が理論値からずれることがある.

本研究で用いた入力・結果音楽ファイルは下記の URL から聞くことが出来るので参考にさせていただきたい.

URL:http://ux.getuploader.com/SIGMUS101_MatsumuraTakashi/

参考文献

- [1] 亀岡 弘和, 篠田 浩一, 嵯峨山 茂樹 : 周波数領域の DP マッチングによる自然楽器演奏の和音ピッチ推定 (2002)
- [2] 後藤 真考 : リアルタイム音楽情景記述システム : サビ区間検出手法 (2002)
- [3] 中静 真 : スパース信号表現とその音声・画像処理への応用 (チュートリアル講演)(2009)
- [4] 松野 高道, 浅井 健史, 岩田 卓哉, 本谷 秀堅 : 医用画像中の臓器レジストレーションのための画像辞書生成 (2010)
- [5] 中村 尚五 : ビギナーズデジタルフーリエ変換, 東京電機大学出版局
- [6] 純正律音階と平均律音階 (<http://www.gabacho-net.jp/whims/whim0010.html>) (2013.11.14).
- [7] MathWorks, DSP System Toolbox, 包絡線検出 (<http://www.mathworks.co.jp/jp/help/dsp/examples/envelope-detection.html>) (2013.11.14).
- [8] SoundFonts site (<http://www.sf2midi.com/>) (2013.11.05).
- [9] ピアノの仕組みを知ろう (<http://piano.s20.xrea.com/mecha/doc03.html>) (2013.11.14).