

# 電子楽器の音色制御パラメータの自動最適化に関する研究

新富康生<sup>1</sup> 藤井慶<sup>2</sup>

**概要:** 現代の音楽制作に欠かせない楽器の1つにシンセサイザがある。この楽器は設定できるパラメータが多く、特に初学者にはパラメータの操作が困難という問題がある。この問題を解決するため、著者らはこの問題を組み合わせ最適化問題の一種と捉えて、自動でパラメータを設定するシンセサイザシステムの開発を目指している。本報告では、最適化に用いる特徴量の検証を重点的に行い、実験で得た類似度の主観評価値と相関の高い特徴量を用いて最適化を行う。3か所のMFCCを組み合わせた特徴量が最も高い精度を出し、入力音と近い音を出力することができたが、様々な楽器音に対する最適化の検討、評価実験の実施などの課題が挙げられる。

**キーワード:** 音楽情報-すべて、電子楽器・演奏インタフェース、感性情報処理

## Study of Automatic Optimization of Synthesizer's Parameters

KOKI SHINTOMI<sup>†1</sup> KEI FUJII<sup>†2</sup>

**Abstract:** Synthesizer is one of the most well-known electronic instruments. This instrument has many parameters, and it is difficult to operate them. To solve this problem, we aim to develop a synthesizer system that automatically sets parameters. In this paper, I focus on the verification of the features to be used for optimization and perform optimization using features. Features that combine MFCC from three locations were found to have the strongest correlation and output sounds close to the input sounds, but there are still issues to be addressed, such as optimization for various instrumental sounds and evaluation experiments.

**Keywords:** Music Information, Electronic Musical Instruments / Performance Interfaces, Intuitive Information Processing

### 1. はじめに

現代の音楽制作に欠かせない楽器の1つにシンセサイザがある。この楽器は音の合成、加工により新しい音色を作り上げていく楽器であり、1台で様々な音色を作り出すことができる。しかし、その柔軟性が故に設定できるパラメータが多く、理想の音色を作り出すまでに操作に慣れる必要があり、多くの時間を要することとなる。特に初学者にはパラメータの操作が困難という問題がある。そこで、本問題を組み合わせ最適化問題の一種と捉え、目標の音色を最も模倣できるパラメータ値を自動探索する研究が行われている[1, 2]。最適化は、目標の音色、シンセサイザの出力音色の双方からそれぞれ特徴量を求めてその距離を最小化するようにパラメータ値を補正することで行われる。特徴量は最適化精度を左右する要であり、先行研究[3, 4]等でMFCCや倍音などの特徴量が試みられているが、決定打となるような特徴量は未だ見つかっておらず、フレーム長の影響や複数種の特徴量の組み合わせなど、更なる調査が期待されている。

そこで本研究では、音色見本を提示すると自動でパラメータを設定するシンセサイザシステムの開発を最終目標とし、その中の特徴量の検証に重点を置き、相関の強い最適

な特徴量を見つけることを一番の目的としている。ただし、シンセサイザは既存の楽音の模倣だけでなく、様々な音色を作り出せるという特徴があり、ユーザはそのとき想定している音色をそのまま出力することだけを望んでいるわけではない。そこで本システムでは必ずしも目標音と全く同じ音を出力することではなく、目標音近傍の音を出力することを試みる。目標音近傍の音を自動的に生成するためには目標音と出力音の2音の差を正確に表現できる特徴量が必要となり、相関の強い特徴量を見つけるため、特徴量単体や複数の特徴量を組み合わせた特徴量、フレーム長とフレーム数に対する相関係数の関係について調査する。その後、距離を単調減少に減らす反復処理によって最適解を求める貪欲法を用い音色制御パラメータを最適化する。

以下、2節では検証した特徴量、3, 4節では特徴量の検証と聴覚実験、5節では最適化を用いたパラメータの探索、6, 7節で考察とまとめを述べる。

### 2. 特徴量

本研究では先行研究[5]で最も強い相関を得たMFCCに加えて、次の7つの特徴量を追加検証する。

- FFT 特徴量

<sup>1</sup> 熊本高等専門学校電子情報システム工学専攻  
National Institute of Technology, Kumamoto College  
<sup>2</sup> 熊本高等専門学校セキュリティグループ

National Institute of Technology, Kumamoto College Security Group

- ・ DFT 特徴量
  - ・ 倍音成分を用いた特徴量
  - ・ 隣接する特徴量データの差分特徴量
  - ・ ZCR を組み合わせた特徴量
  - ・ 単体で用いた特徴量, 差分特徴量, ZCR 特徴量の 3 つの特徴量を組み合わせた特徴量
  - ・ 1 音内 3 か所から特徴量抽出し組み合わせた特徴量
- これらの特徴量のユークリッド距離を 2 音の音色差とする.

2.1 節で特徴量の距離計算に用いたユークリッド距離, 2.2 節から 2.10 節で各特徴量について述べる. 以降から MFCC, FFT, DFT, 倍音成分を用いた特徴量をまとめて特徴量単体という.

## 2.1 ユークリッド距離

2 音の距離を求める際に用いる. 点  $\vec{p}$  および  $\vec{q}$  が  $n$  次元ユークリッド空間内の 2 点とすれば, ユークリッド距離は次式(1)で求められる.

$$d(\vec{p}, \vec{q}) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

ここでの  $p_i$  は点  $\vec{p}$  の  $i$  次元目の値を表す.

## 2.2 メル周波数ケプストラム (MFCC) 特徴量

人間には「周波数の低い音に対して敏感で, 周波数の高い音に対して鈍感である」という聴覚特性がある. MFCC は, 音の周波数に対しこの性質を考慮したメル尺度に変換し離散コサイン変換をかけた多次元ベクトルの特徴量であり, 音声認識や楽器の識別等によく使用される特徴量である. 本研究では実験に用いる各音から LibROSA の `librosa.feature.mfcc()` を用いて MFCC を算出し, 13 次元を特徴量として使用する.

## 2.3 ゼロクロッシングレート (ZCR) 特徴量

時間波形の符号が切り替わる回数を特徴量とする. 一般的には ZCR の値が大きいくほど耳障りなノイズ音になる. 本研究では 1 秒間あたりのゼロ交差回数を特徴量とする.

## 2.4 FFT 特徴量

MFCC は人の音高知覚特性を鑑みて FFT スペクトルを要約するものだが, MFCC を求める際に算出される FFT の値が特徴量として効果的であると考え, 検証対象とする. FFT とは離散フーリエ変換を計算機上で高速に計算するアルゴリズムのことをいい, 主にデジタル信号を周波数解析する際に用いる. 本研究では実験に用いる各音に窓をかけ, 窓処理により切り出した波形を Numpy の `fft` パッケージを用いて求める. 求めた値を正規化し, 特徴量として各周波数の周波数強度を得る.

## 2.5 DFT 特徴量

FFT はフレーム長が 2 のべき乗でなければならないが, DFT はそのような制限がない分自由にフレーム長を決められ, FFT と比べ実験試料の音高の周期長に対応したフレーム長で処理できる. 本研究では式(2)を用いて算出する.

$$F(t) = \sum_{x=0}^{N-1} f(x) e^{-i\frac{2\pi tx}{N}} \quad (2)$$

## 2.6 倍音成分を用いた特徴量

音には音高に相当する基本周波数の他にその整数倍の周波数の振動がいくつも生じやすく, これが音色を特徴づける一因になっている. この整数倍の振動のことを倍振動といい, それぞれの音のことを倍音という. 本研究では各音の  $\sin$  波と  $\cos$  波の内積を求め, 二乗和平方根を求める. 基本周波数をもとに任意の倍数分繰り返し行うことにより倍音成分を算出する.

## 2.7 隣接する特徴量データの差分特徴量

各ベクトル内で隣接する値との差を算出したものを差分特徴量とする. 特徴量算出後の特徴ベクトルの  $n$  次元目の特徴量データ  $x(n)$  と 1 つ前の特徴量データ  $x(n-1)$  の差を特徴量とすると, 式(3)で算出される.

$$y(n) = x(n) - x(n-1) \quad (3)$$

## 2.8 ZCR を組み合わせた特徴量

先行研究[5]の検証結果より, MFCC の次に強い相関が見られた ZCR を組み合わせることで, より強い相関が得られる可能性が示唆された. そこで, 各特徴量単体と ZCR 特徴量をそれぞれ組み合わせる. 組み合わせ方について以下に述べる.

組み合わせる際それぞれの特徴量に重み付けを行い, 重みパターンごとに相関係数を求める. それぞれの特徴量のユークリッド距離を式(1)で求め, 標準化した値をそれぞれ  $F_1, F_2$  とする. ここでの  $F_1$  は各特徴量単体,  $F_2$  は ZCR 特徴量となる.  $F_1, F_2$  に掛ける重みを  $W_1, W_2$  とおくと, 2 つの重みは式(4)の関係にあり, それぞれの重みは 0.1 刻みで更新する.

$$W_1 + W_2 = 1.0 \quad (\text{ただし } 0 \leq W_1 \leq 1, 0 \leq W_2 \leq 1) \quad (4)$$

次の式(5)で重み付き距離  $D_w$  を算出する.

$$D_w = F_1 W_1 + F_2 W_2 \quad (5)$$

## 2.9 3 つの特徴量を組み合わせた特徴量

各特徴量単体とそれぞれの差分特徴量, ZCR 特徴量の 3 つを組み合わせることでより高い相関が得られると考え, 検証を行う.

それぞれ 3 つの特徴量のユークリッド距離を式(1)で求め, 標準化した値をそれぞれ  $F_1, F_2, F_3$  とする. ここでの  $F_1$  は各特

微量単体,  $F_2$ は $F_1$ の差分特徴量,  $F_3$ は ZCR 特徴量となる.  
 $F_1, F_2, F_3$ に掛ける重みを $W_1, W_2, W_3$ と置くと, 3 つの重みは式(6)の関係にあり, それぞれの重みは 0.1 刻みで更新する.

$$W_1 + W_2 + W_3 = 1.0$$

(ただし,  $0 \leq W_1 \leq 1, 0 \leq W_2 \leq 1, 0 \leq W_3 \leq 1$ ) (6)

次の式(7)で重み付きデータ $D_w$ を算出する.

$$D_w = F_1W_1 + F_2W_2 + F_3W_3$$
 (7)

### 2.10 3か所の特徴量を組み合わせた特徴量

これまでの手法は波形データの先頭から任意のフレーム長のデータを用いたが, 1音の「始め, 中ごろ, 末尾」の3か所から特徴量を取得し, 混合した特徴量を用いることで波形の音色や振幅の時間的な変化を含むことができ, 高い相関が得られると考えた. 各特徴量単体で3か所の算出を行い, それぞれの特徴量を求め, 式(6), 式(7)を用いて重み付けし, 算出する.

## 3. 特徴量の追加検証

本節では, 先行研究[5]から音色の時間変化の無い音源と聴覚実験から得た主観評価値を用いて, 各特徴量単体, 2.8節, 2.9節で述べた特徴量の相関を求める.

各フレーム長と次元数について, MFCC はフレーム長 44100 の 13 次元, FFT はフレーム長 1024 の 1024 次元, DFT はフレーム長 1000 の 1000 次元, 倍音成分を用いた特徴量はフレーム長 100 の 50 次元とする.

### 3.1 聴覚実験

実験試料は表 3.1 の仕様で生成された 1 組 2 音の 50 組, 計 100 音であり, これらの音は音色の時間変化の無い定常的な波形である. 被験者は 2 つの音を聴き, 2 つの音がどの程度似ているのかを 1 (全く同じ) ~ 7 (全く違う) の 7 段階で評価する. 聴覚実験は表 3.2 に示す環境で行い, 被験者は 5 名で行う.

表 3.1 聴覚実験で使用したシンセサイザの仕様

オシレータ	サイン波, ノコギリ波, 矩形波, 三角波, ホワイトノイズ
基本周波数	440 [Hz]
フィルタ	ローパス, ハイパスフィルタ
カットオフ周波数	100 ~ 20000 [Hz]
レゾナンス	0.01 ~ 10.0
サンプリング周波数	44100 [Hz]

表 3.2 聴覚実験の環境

実施場所	熊本高専熊本キャンパス無響室
ヘッドホン	SENNHEISER HD580 Precision
オーディオインターフェース	Steinberg UR22

### 3.2 特徴量の算出

2 音それぞれから特徴量を算出し, 算出した特徴量のユークリッド距離を 2 音の音色差とする. ユークリッド距離が小さいほど 2 つの音は似ているといえる.

### 3.3 相関係数の算出

聴覚実験で得た主観評価値と各特徴量の 2 音のユークリッド距離を比べて相関係数を算出する. 相関係数は -1 から 1 の範囲で表され, 値が大きいほど相関が強く, 人間の聴覚特性に適した特徴量といえる.

### 3.4 検証結果

#### 3.4.1 特徴量単体の結果

各特徴量単体の結果を表 3.3 に示す. 各特徴量に対する被験者 5 名の相関係数の平均値を示す.

表 3.3 各特徴量単体と主観評価値との間の相関係数

	相関係数の平均値
MFCC	0.53
FFT	0.62
DFT	0.63
倍音成分を用いた特徴量	0.50

表 3.3 の結果より, すべての特徴量で相関係数の平均値が 0.5 以上となり, 比較的高い相関が認められた. その中でも FFT, DFT 特徴量は相関係数の平均値が 0.6 以上と最も高い相関係数を得た. これは, 実験試料の音源が音色の変化の無い定常的な波形であったため, 要約せず FFT の値をそのまま使うことでより相関が高くなったと考える.

#### 3.4.2 ZCR を組み合わせた特徴量の結果

2.8 節で述べた ZCR 特徴量を組み合わせた特徴量の結果を表 3.4 に示す. 最も相関の高いパターンの重みと相関係数のペアを示し, 記載する重みは 4 つの各特徴量の重みである. 比較のため最右列に各特徴量単体の相関係数の平均値を再掲している.

表 3.4 各特徴量の相関係数最大時の重みと相関係数の関係

	最大時の重み	最大時の相関係数	単体の相関係数
MFCC - ZCR	0.6	0.59	0.53
FFT - ZCR	0.8	0.63	0.62
DFT - ZCR	0.7	0.66	0.63
倍音 - ZCR	0.6	0.57	0.50

表 3.4 の結果より、ZCR 特徴量を組み合わせた場合と特徴量単体の相関係数にほとんど差は無く、ZCR 特徴量を組み合わせることでの効果は少ないといえる。

### 3.4.3 3 つの特徴量を組み合わせた特徴量の結果

2.9 節で述べた 3 つの特徴量を組み合わせた特徴量の検証結果を表 3.5 に示す。最も相関の高い時の重みと相関係数のペアを示し、比較のため最右列に各特徴量単体の相関係数の平均値を再掲する。

表 3.5 各特徴量の相関係数最大時の重みと相関係数の関係

	単体	差分	ZCR	相関係数	単体の相関係数
MFCC	0.6	0.0	0.4	0.59	0.53
FFT	0.0	0.7	0.3	0.65	0.62
DFT	0.7	0.0	0.3	0.66	0.63
倍音成分	0.0	0.5	0.5	0.59	0.50

表 3.5 の結果を見ると、3.4.2 節と同様に各特徴量単体との相関係数にほとんど差は無く、3 つの特徴量を組み合わせることでの効果は少ないといえる。また、FFT および倍音成分を用いた特徴量に関して、3 つの特徴量を組み合わせた場合特徴量単体を含めず差分特徴量と ZCR 特徴量を組み合わせることによって全体の相関が高くなる結果となった。

ここで、FFT と FFT の差分特徴量それぞれ単体での相関係数の結果を表 3.6、2.8 節の ZCR を組み合わせた特徴量におけるそれぞれの相関係数を表 3.7、FFT と ZCR を組み合わせた特徴量と FFT 差分と ZCR を組み合わせた特徴量の重みと相関係数の関係を表すグラフを図 3.1 に示す。

表 3.6、3.7 の結果より、お互い似た相関係数が算出された。また、図 3.1 より両者のグラフの推移がほとんど変わらない結果が得られた。これより、お互いが似た特徴量であるため、多重共線性の可能性が考えられる。よって、同時使用せず、どちらか一方を使用するのが妥当と考える。

表 3.6 FFT, FFT 差分特徴量単体での相関係数

	相関係数
FFT	0.62
FFT 差分	0.63

表 3.7 FFT, FFT 差分特徴量と ZCR を組み合わせた特徴量の相関係数最大時の重みと相関係数

	最大時の重み	最大時の相関係数
FFT	0.8	0.63
FFT 差分	0.7	0.65

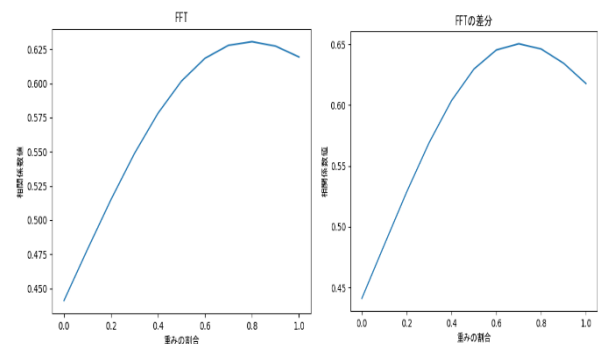


図 3.1 左図：FFT と ZCR の重みごとの相関係数の推移、右図：FFT 差分と ZCR の重みごとの相関係数の推移

## 4. 音色の時間変化を考慮した特徴量の検証

前節の検証は音色が時間変化しない定常的な音を用いて検証したものであった。本節では、音色の時間変化を考慮した実装を行うため、実機のシンセサイザ[6]のプリセット音源を用いて聴覚実験を行う。特徴量について、各特徴量単体、ZCR を組み合わせた特徴量に加えて、フレーム長とフレーム数に対する相関係数の関係について調査する。各特徴量のフレーム長と次元数については前節と同様である。

### 4.1 聴覚実験

本聴覚実験は 3.1 節で述べた聴覚実験の実験試料を音色の時間変化の無い音源から音色の時間変化のある音源に変更し、同様の手順で行う。実験試料は表 4.1 に示すシンセサイザ[6]に初期搭載されたプリセット 1 ~ 100 の音を録音し、1 秒間に切り取った音源である。これらの音源をランダムに 1 組 2 音の 50 組のセットにし、被験者は 2 つの音を聴き、2 つの音がどの程度似ているかを 1 (全く同じ) ~ 7 (全く違う) の 7 段階で評価する。音の聴き直しは可能とし、聴覚実験は表 4.2 に示す環境で、被験者は 6 名で行う。

表 4.1 聴覚実験で使用したシンセサイザの仕様

オペレータ	KORG opsix-ALTERED FM SYNTHSIZER
基本周波数	440 [Hz]
サンプリング周波数	44100 [Hz]

表 4.2 聴覚実験環境

実施場所	熊本高専熊本キャンパス無響室
ヘッドホン	SONY MDR-CD900ST
オーディオインタフェース	Native Instruments KOMPLETE AUDIO 2

## 4.2 検証結果

### 4.2.1 特徴量単体の結果

各特徴量単体の結果を表 4.3 に示す。なお各被験者の相関係数の平均値を相関係数の列に示し、音色の時間変化の無い音声試料で検証した場合の各特徴量単体の相関係数の平均値（表 3.3）を表の最右列に再掲する。

表 4.3 各特徴量単体と主観評価値との間の相関係数

	相関係数	音色変化の無い場合
MFCC	0.44	0.53
FFT	0.15	0.62
DFT	0.16	0.63
倍音成分	0.13	0.50

表 4.3 より、すべての特徴量で音色の変化が無い場合と比較し、相関が低くなった。また、MFCC 以外相関係数が 0.1 台と音色の変化が無い場合と比較し大きく減少し、ほとんど相関の無い結果となった。

### 4.2.2 ZCR を組み合わせた特徴量の結果

ZCR を組み合わせた特徴量の検証結果を表 4.4 に示す。比較のため、最右列に本実験の各特徴量単体の相関係数の平均値を記載する。

表 4.4 各特徴量と ZCR の最大時の重みと相関係数

	最大時の重み	最大時の相関係数	単体の相関係数
MFCC-ZCR	0.6	0.45	0.44
FFT-ZCR	0.8	0.23	0.15
DFT-ZCR	0.7	0.24	0.16
倍音-ZCR	0.6	0.22	0.13

表 4.4 の結果より、3.4.2 節の音色の変化が無い場合と同様に ZCR 特徴量を組み合わせることでの効果は少ないといえる。

## 4.3 フレーム長とフレーム数に対する相関係数の関係

4.2.1 節より特徴量単体で最も相関の高かった MFCC のフレーム長とフレーム数に対する相関係数の関係を調査する。フレーム長、フレーム数が以下の値の時に相関係数を求める。フレーム同士の被りはありとする。

- ・フレーム長：0~44100 の範囲で 2500 刻みの 18 パターン
- ・フレーム数：1~10 の 10 パターン

算出した結果を図 4.1 に示す。横軸にフレーム長、縦軸に相関係数、右ラベルの 10 色がフレーム数を示している。

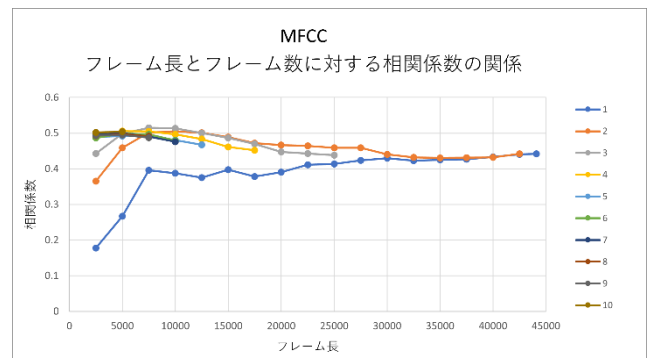


図 4.1 MFCC のフレーム長とフレーム数に対する相関係数の関係

図 4.1 より、フレーム数が増加すると相関が高くなる傾向があり、フレーム長 7500 のフレーム数 3（グレー色）が最も高い相関係数を得た。

この結果から 3 か所の特徴量を組み合わせた特徴量の検証を行う。

### 4.3.1 3 か所の特徴量を組み合わせた特徴量の結果

3 か所の特徴量を組み合わせた特徴量の検証結果を表 4.5 に示す。重みごとに相関係数を求め、最も相関係数の高い時の重みと相関係数を示す。最右列に本実験の各特徴量単体の相関係数の平均値を再掲する。

表 4.5 3 か所の各特徴量の最大時の重みと相関係数

	始め	中頃	末尾	相関係数	単体の相関係数
MFCC	0.4	0.3	0.3	0.64	0.44
FFT	0.1	0.0	0.9	0.30	0.15
DFT	0.2	0.0	0.8	0.23	0.16
倍音	1.0	0.0	0.0	0.30	0.13

表 4.5 の結果より、特徴量単体と比較し相関が高くなる傾向が見られ、MFCC は相関係数が 0.6 以上と強い相関が認められた。しかし、FFT、DFT、倍音成分を用いた特徴量は単体と比較し相関が高くなったものの引き続きほとんど相関のない結果となった。

この検証結果より、最も強い相関を得た3か所のMFCCを組み合わせた特徴量を採用し、次節で音色の最適化について述べる。

## 5. 最適化によるパラメータの探索

4節の検証で最も強い相関を得た3か所のMFCCを組み合わせた特徴量と貪欲法、乗算方式シンセサイザを用いて音色の最適化を行う。

### 5.1 貪欲法

貪欲法は最適化アルゴリズムの一種で、暫定的に最適とされる地点の近傍を探索し、その中で得られた最適な地点に場所移動することを収束まで繰り返す比較的単純な手法である。貪欲法を用いることで近似の結果が単調減少になり、突然変異を起こさないという利点がある。また、初期値と最適値との間の中間的な音色を容易に取得できる。そこで本研究で用いるシンセサイザシステムの振幅、周波数の各値について貪欲法を使い音色の最適化を試みる。

### 5.2 乗算方式シンセサイザ

乗算方式とは元となる音にもう1つの音を用いて周波数変調をかけることで音を作る方式である。少数のパラメータで複雑な周波数成分を持った音色を作り出すことができるが、音色変化を予測するのが難しく、音作りの難易度は高いという特徴がある。ある時点でのキャリアの振幅を $A_c$ 、位相角を $\theta_c$ 、モジュレータの振幅を $A_m$ 、位相角を $\theta_m$ と置くと、乗算方式シンセサイザの瞬時値 $Y$ は式(8)で算出される。

$$Y = A_c \sin(\theta_c + A_m \sin \theta_m) \quad (8)$$

本研究では先行研究[5]で作成された乗算方式シンセサイザを用いて最適化を行う。乗算方式シンセサイザの仕様を表5.1、音色最適化の処理フローを図5.1に示す。

表 5.1 乗算方式シンセサイザの仕様

使用言語	Processing 3.5.4
音源方式	乗算方式
オペレータの初期値群	振幅, 基本周波数, フェードイン, フェードアウトがランダムなオペレータを4つ用意し, ランダムに配線
プリセットの初期値群	任意の数のプリセットをランダムで生成
サンプリング周波数	44100 [Hz]
使用したライブラリ	javax.sound.sampled, ddf.minim
オペレータの数	4
各オペレータの振幅	0.0 ~ 1.0
各オペレータの基本周波数にかかるとの倍率	0.0 ~ 8.0

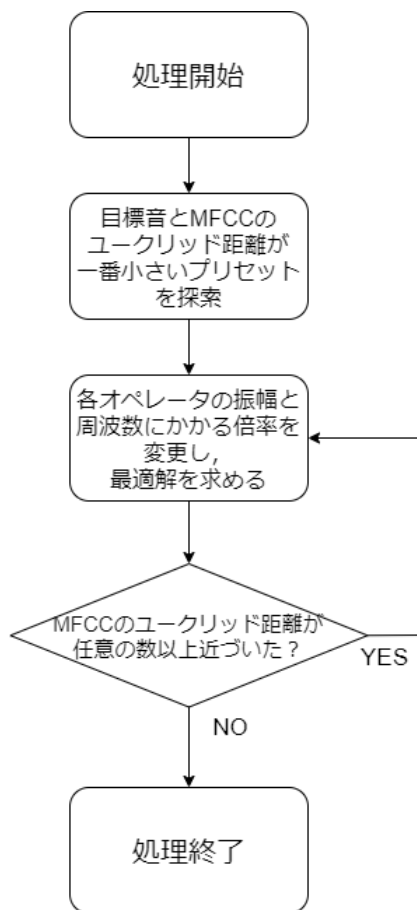


図 5.1 音色最適化の処理フロー

### 5.3 最適化結果の検証

実装したシンセサイザシステムについて最適化を行った結果について述べる。あらかじめ MFCC のユークリッド距

離が近いプリセットを探索し、そのプリセットのパラメータを基に最適化を行う。表 5.2 に検証の条件を記し、クレッシェンドのかかったオルガンに似た音（目標音：プリセット 16）を目標音として最適化を行う。

表 5.2 本研究における最適化の検証の条件

MFCC の次元数	13
ランダムに生成するプリセット初期値群の数	50000
最適化に用いる音	バイオリン風, バル風, ランダム生成された音
サンプリング周波数	44100 [Hz]
振幅と基本周波数にかかる倍率の変更幅	0.01
最適化に際して変更する値	振幅, 基本周波数にかかる倍率
最適化終了の判定	最適化後に 2 音間の MFCC のユークリッド距離が任意の数以上縮まらなかった際

### 5.3.1 検証結果

表 5.2 の条件のもと、結果の一例としてクレッシェンドのかかったオルガンに似た音の検証結果を図 5.2 に示す。

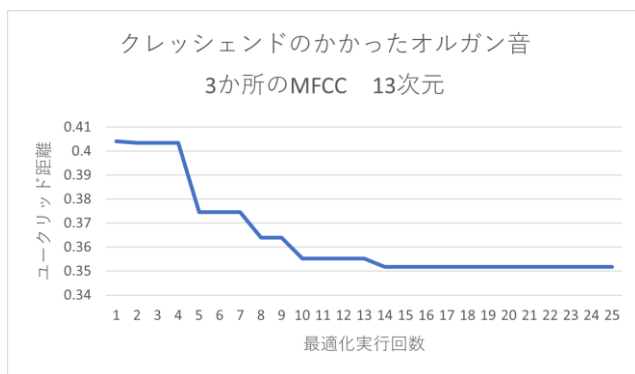


図 5.2 最適化の結果の推移

図 5.2 の結果を見ると、最適化実行回数を重ねるにつれユークリッド距離が小さくなっていることが分かる。聴覚上の所感について、目標音に対して似た音が出力された。

## 6. 考察

特徴量の追加検証について、FFT, DFT 特徴量は音色の変化が無い場合は強い相関（相関係数：0.62, 0.63）を得たが、音色の変化がある場合ではほとんど相関の無い結果（相関係数：0.15, 0.16）となり、相関係数が大きく下がった。これは周波数や振幅、倍音成分などが 1 音の中で大きく変化し、定常的でなくなったためだと考える。また、3 か所の MFCC を組み合わせた特徴量が最も相関の強い特徴量

となった理由としては、MFCC は波形全体を窓の個数分で要約しているため波形の情報や特徴を得やすく、フレーム数を 3 つに増やすことで音色や振幅の時間的な変化を含むことができたためと考える。

## 7. おわりに

本研究では、計 8 つの特徴量について検証を行い、MFCC のフレーム長とフレーム数に対する相関係数の関係を調べた結果、フレーム長 7500（すなわち 170 ミリ秒）のフレーム数 3 の組み合わせが最も相関の強いことを明らかにした。この結果は、聴覚実験を通して人間の聴覚特性に適したフレーム長とフレーム数の関係を示しており、この結果から 3 か所の特徴量を組み合わせた特徴量を検証した結果、3 か所の MFCC を組み合わせた特徴量が最も強い相関を示した。これらの結果を基に 3 か所の MFCC を組み合わせた特徴量と貪欲法、乗算方式シンセサイザを用いて音色の最適化を行った。聴覚上の所感では、目標音と似た音を出力することができた。このように、細かな特徴量の検証を行い、目標音に近い音を出力するシンセサイザシステムを実装できたが、いくつか課題が残っている。例えば、最適化に 10 分程度要する点や実際の楽音について最適化を行えていない点が挙げられる。他にも、本研究では実装した乗算方式シンセサイザの音色最適化の精度について、まだ限られた人間による聴感上の所感で判断した段階であり、具体的な分析を行えていない。評価実験を行うなどし、客観的に実装内容や最適化手法の有効性を確かめたい。

## 参考文献

- [1] Yuyo Lai, Shyh-Kang Jeng, Der-Tzung Liu, and Yo-Chung Liu, AUTOMATED OPTIMIZATION OF PARAMETERS FOR FM SOUND SYNTHESIS WITH GENETIC ALGORITHMS, International Workshop on Computer Music and Audio Technology, pp.33-38, 2006.
- [2] Matthew Yee-King, Martin Roth, SYNTHBOT: AN UNSUPERVISED SOFTWARE SYNTHESIZER PROGRAMMER, International Conference on Mathematics and Computing, 2009.
- [3] Katsutoshi Itoyama & Hiroshi G. Okuno, Parameter Estimation of Virtual Musical Instrument Synthesizers, International Conference on Mathematics and Computing, pp.1426-1431, 2014.
- [4] Hiroko Terasawa, Malcolm Slaney, and Jonathan Berger, PERCEPTUAL DISTANCE IN TIMBRE SPACE, ICAD 05- Eleventh Meeting of the International Conference on Auditory Display Limerick Ireland July6-9, pp.61-68, 2005
- [5] 甲斐大雅, 電子楽器の音色制御パラメータの機械学習に関する研究, 第 130 回音楽情報科学研究会・第 59 回エンタテインメントコンピューティング研究会合同研究発表会, 2021.
- [6] KORG opsix ALTERED FM SYNTHESIZER, <https://www.korg.com/jp/products/synthesizer/opsix>.