

平成 23 年度 博士論文

図形と音声の変換手法とその応用に関する研究

A study on transforming methods between shape and audio
and their applications

岩淵 勇樹

金沢大学大学院 自然科学研究科 電子情報科学専攻
集積回路工学研究室

Abstract

A lot of studies on tone generation methods are proposed, however, tone generation requires numbers of parameters, which is difficult for users to directly understand the relation between the parameters and the generated tone. In most systems, these parameters are manipulated in one-dimensional interface, which is not effective for two-dimensional interface, such as touch pad. In this paper, the method to transform from audio to shapes, from shapes to audio are proposed and discussed, followed by the tone manipulation method using a kind of closed curves called the analytic signal. In order to transform from audio to shapes, the shape can be generated by calculating the analytic signal of original signal that contain Hilbert transformed signal in imaginary component. On the other hand, in order to transform from shapes to audio, the audio signal can be generated by approximated analytic signal by parameter transformation. The waveform of the tone is related to the corresponding analytic signal, and the manipulation for the analytic signal can generate the modified tone. This method is adequate for tone modification from the existing instrument's tone. It is also expected to extend the application of the analytic signal to visual fusion performance.

第1章 序論

1.1 はじめに

電子楽器の発明以降，それまでの物理的な振動現象を用いてきた楽器の音色は格段に生成の自由度が高くなり，今までにない新たな音色を作ることが可能となった．それに伴ってテクノ音楽等の新たな音楽分野が生まれ，それが再び新たな電子楽器の発展を促している．

電子楽器に向けた，楽器を演奏するための入力インタフェースに関する研究は多くあるが，それと並んで，電子楽器に適した音色の入力・生成方法，特にマウス操作やタッチパッド操作など GUI に適した音色入力・生成方法を探ることは，電子音楽全体の発展のためにも重要な課題といえる．

本論文では，既存の音声信号をベースとし，それに対してマウスやタッチパッドなどの平面操作デバイスによって波形を操作することで，新たな音色の音声信号を作成する手法の提案と実装を行う．具体的には，既存の音声信号から生成された解析信号を複素平面上の閉曲線図形とみなし，これに対して平面操作デバイスによる図形操作によって新たな周期音声信号を生成・変更する手法の提案と実装を行う．

第2章 回転によって音色が不変である信号

2.1 解析信号

$U(\omega)$ を以下のように定義する .

$$U(\omega) = \begin{cases} 0 & (\omega < 0) \\ 1 & (\omega = 0) \\ 2 & (\omega > 0) \end{cases} \quad (2.1)$$

音声信号 $s(t)$ に対して, そのフーリエ変換を $S(\omega)$ とおく . このとき, $U(\omega)$ をフィルタとして用いると, 式 (2.2) のように, $S(\omega)$ の正周波数成分の振幅が 2 倍, 負周波数成分の振幅が 0 倍となり, それを $\tilde{S}(\omega)$ とおく .

$$\tilde{S}(\omega) = U(\omega) \cdot S(\omega) = \begin{cases} 0 & (\omega < 0) \\ S(\omega) & (\omega = 0) \\ 2S(\omega) & (\omega > 0) \end{cases} \quad (2.2)$$

それを逆フーリエ変換した複素信号を一般に「解析信号」と呼び, $\tilde{s}(t)$ と表す .

$s(t)$ が区分的に滑らかで常に有限値をもつ周期信号であれば, $\tilde{s}(t)$ は複素平面 $x + iy$ 上の閉曲線を描くことが導かれる . (ただし i は虚数単位) . 図 2.2 に解析信号の例を示す .

2.2 解析信号と音色の関係

周期的な解析信号を複素平面上の閉曲線図形として見ると, 音声信号の振幅が時間とともに変化するのにあわせて, それに対応する解析信号の閉曲線上を対応点が移動し, 音声信号の 1 周期が閉曲線上の 1 周に対応する . すなわちこの閉曲線の形状は, 音声信号の周波数情報を除いた表現であると言える .

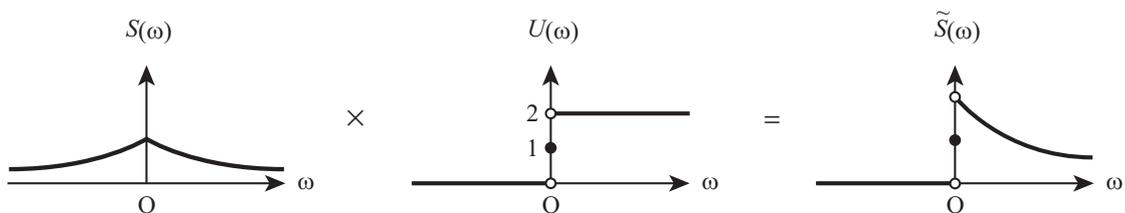


図 2.1: 解析信号フィルタを掛け合わせた信号の概念図

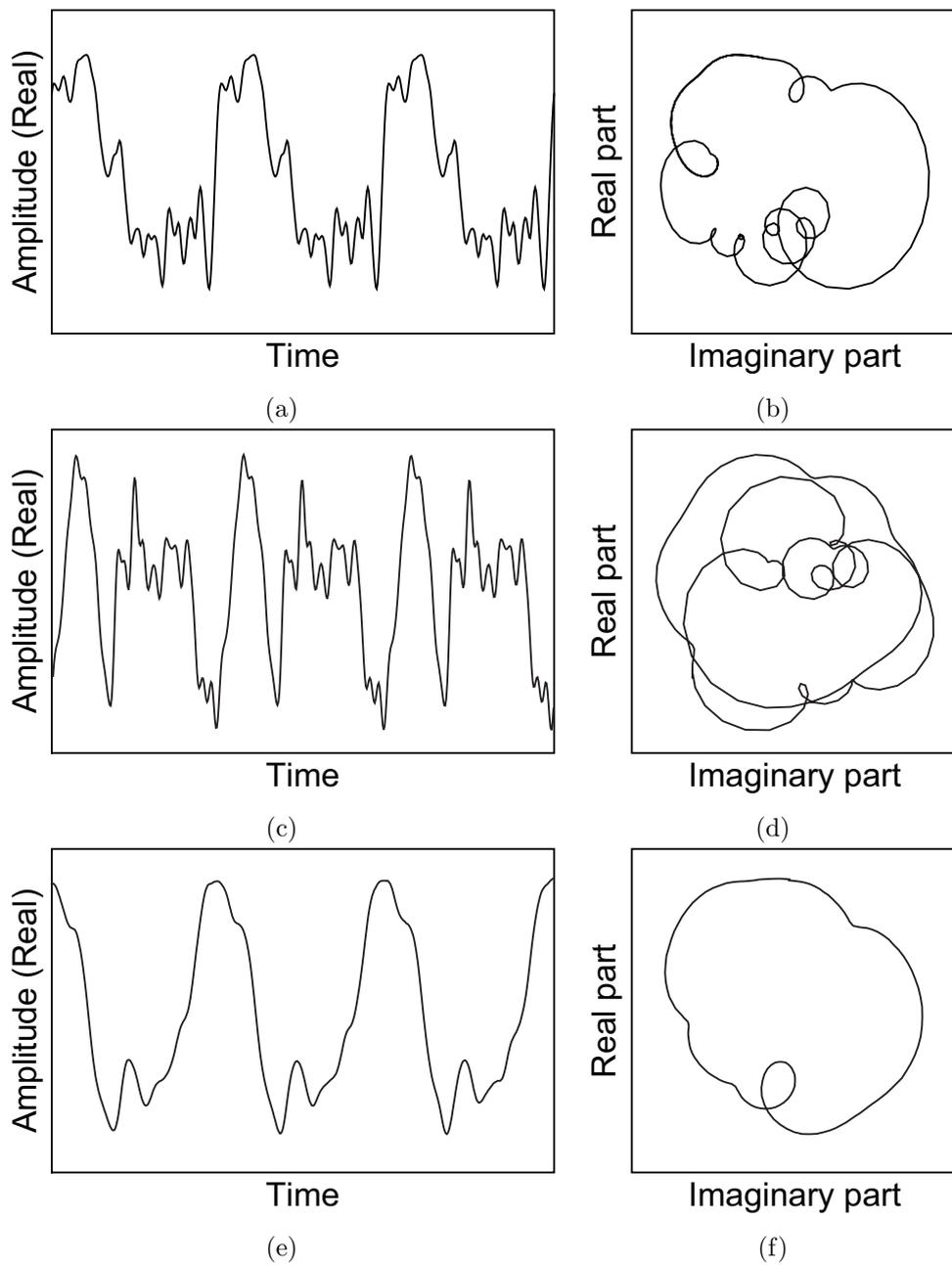


図 2.2: バイオリンの音色 $s(t)$ (a) と対応する解析信号 $\tilde{s}(t)$ (b) , ハーモニカの音色 $s(t)$ (c) と対応する解析信号 $\tilde{s}(t)$ (d) , ピアノの音色 $s(t)$ (e) と対応する解析信号 $\tilde{s}(t)$ (f)

閉曲線の大きさが音声信号の振幅に対応し、閉曲線の平行移動は、時間変化しない定数成分に対応するため音色は変化しない。また位相回転 (定数 $e^{i\theta}$ 倍) は閉曲線の回転に対応するが、一般に音色は基音の位相に依存しない。そのため、相似な閉曲線はすべて同じ音色に対応することになる。

以上のことから、閉曲線図形としての解析信号は音色と密接な関わりをもっているといえる。

2.3 ヒルベルト変換

$F(\omega)$ を以下のように定める。

$$F(\omega) = \begin{cases} i & (\omega < 0) \\ 0 & (\omega = 0) \\ -i & (\omega > 0) \end{cases} \quad (2.3)$$

このとき、

$$H(\omega) = F(\omega) \cdot S(\omega) \quad (2.4)$$

とすると、 $H(\omega)$ の逆フーリエ変換 $h(t)$ は $s(t)$ のヒルベルト変換であるという。 $\tilde{s}(t)$ の実部 $\Re[\tilde{s}(t)]$ は $s(t)$ のままであるという性質をもち、 $h(t)$ は解析信号 $\tilde{s}(t)$ の虚部 $\Im[\tilde{s}(t)]$ に等しくなる。

$$\Re[\tilde{s}(t)] = s(t) \quad (2.5)$$

$$\Im[\tilde{s}(t)] = h(t) \quad (2.6)$$

なお、 $s(t)$ は $h(t)$ の逆ヒルベルト変換であるといい、 $F(\omega)$ の代わりに $-F(\omega)$ を用いた時の $s(t)$ と $h(t)$ の関係に等しい。式 (2.4) は畳み込みの関係を用いて以下のように表すことができる。

$$h(t) = \frac{1}{\pi t} * s(t) \quad (2.7)$$

また、 $\tilde{s}(t)$ は負の周波数成分を持たないという性質があり、すなわち解析信号である。なお、 $s_x(t)$ は $s_y(t)$ の逆ヒルベルト変換であるといい、 $F(\omega)$ の代わりに $-F(\omega)$ を用いた時の $s_x(t)$ と $s_y(t)$ の関係に等しい。

2.4 回転によって音色が不変であるということ

$\tilde{s}(t)$ が解析信号であるとき、以下の等式が成り立つ (\mathcal{F} はフーリエ変換を表す)。

$$|\mathcal{F}[\Re[e^{i\theta}\tilde{s}(t)]]| = |S_x(\omega)| \quad (2.8)$$

ところで、人間が周期的な音声信号を知覚する際、聞こえに影響するのはほとんど振幅スペクトルの形状の違いであり、位相が与える影響はごく僅かである。そのため、式 (2.8) より、周期関数 $s_x(t)$ および $\Re[e^{i\theta}\tilde{s}(t)]$ は同等の音色をもつとみなす。すなわち、複素平面上の閉曲線が与えられたとき、その実部を音声信号に対応させるとすると、閉曲線が解析信号であった場合はその回転図形も同等な音色をもつことになる。

また、音声信号の定常成分は音色に全く影響を与えないため、定常成分のみ異なる音声信号もそれぞれ同等な音色をもつと見なせる。なお、本論文では定常成分を含まない信号のみを考える。

以上を総合すると、本論文における音色は、定常成分を除く振幅スペクトルの形状によって特徴付けられる。

第3章 音声信号から図形への変換

3.1 音声信号と曲線

音声信号を $s_x(t)$ とし, そのフーリエ変換を $S_x(\omega)$ とする. このとき, $S_x(\omega)$ にあるフィルタ $F(\omega)$ を掛けて

$$S_y(\omega) = F(\omega) \cdot S_x(\omega) \quad (3.1)$$

となる $S_y(\omega)$ を与えると, その逆フーリエ変換 $s_y(t)$ と原信号 $s_x(t)$ を用いて

$$\tilde{s}(t) = s_x(t) + is_y(t) \quad (3.2)$$

という複素関数が定まる (i は虚数単位). $F(\omega)$ が定数の場合などを除けば, $\tilde{s}(t)$ は複素平面 $s_x + is_y$ 上の曲線となり, 特に $s_x(t)$ が周期関数かつ $s_x(t)$ および $s_y(t)$ が区分的に滑らかで常に有限値の場合は閉曲線となる.

本論文では, フィルタ $F(\omega)$ を式 (2.3) のヒルベルト変換フィルタに限定して議論する.

3.2 周期音声信号から閉曲線への変換

周期信号を閉曲線に変換するには 2.1 で述べたように, 信号を解析信号に変換して複素平面上に投影することによって描かれる.

第4章 図形から音声信号への変換

4.1 解析信号への近似

閉曲線を解析信号に漸近させる方法について検討する。

シミュレーションの初期曲線には、約 15 ピクセル角のシルエット画像を 8-連結の境界追跡によって求めた輪郭線を用いる。初期曲線は $m+1$ 個の複素平面上の点列 $\tilde{s}_0(t) = (\tilde{s}_0(0), \tilde{s}_0(1), \dots, \tilde{s}_0(m))$ によって構成され、 $\tilde{s}_0(m) = \tilde{s}_0(0)$ とする。フーリエ変換時にはこの重複する末尾の点を無視して FFT を行う。

初期曲線には図 4.1 に示す曲線 1 と曲線 2 を用いる。それぞれのスペクトルは (c)(d) のようになる。

4.1.1 方法 A: 解析信号対の漸近による方法

解析信号 $\tilde{s}(t)$ では、実部解析信号 $\xi(t) = s_x(t) + i\mathcal{H}[s_x(t)]$ と虚部解析信号 $\eta(t) = \mathcal{H}^{-1}[s_y(t)] + is_y(t)$ は当然等しくなる。なお、 \mathcal{H} 、 \mathcal{H}^{-1} はそれぞれヒルベルト変換、逆ヒルベルト変換を表す。

以下の手順で $\xi_{tmp}(t) - \eta_{tmp}(t)$ の値を最小化させ、入力曲線を解析信号に漸近させることを試みる。

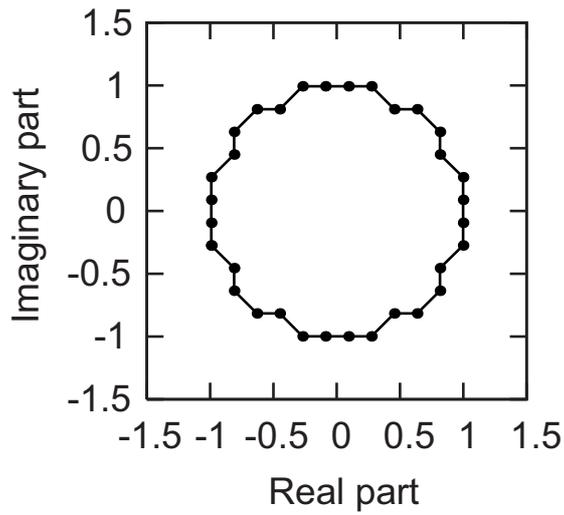
1. $n = 0$ とおく
2. $\tilde{s}_n(k)$ と $\tilde{s}_n(k+1)$ の間の位置に $\frac{\tilde{s}_n(k) + \tilde{s}_n(k+1)}{2}$ を挿入し、 $\tilde{s}_{tmp}(t)$ とおく
3. $\tilde{s}_{tmp}(t)$ について、 $\xi_{tmp}(t) - \eta_{tmp}(t)$ の 2 乗和が最小となる k を求め、そのときの $\tilde{s}_{tmp}(t)$ を $\tilde{s}_{n+1}(t)$ とする
4. n を $n+1$ に置き換え、2. 以降を繰り返す

この方法での $n = 100$ におけるシミュレーション結果を図 4.2, 4.3 に示す。

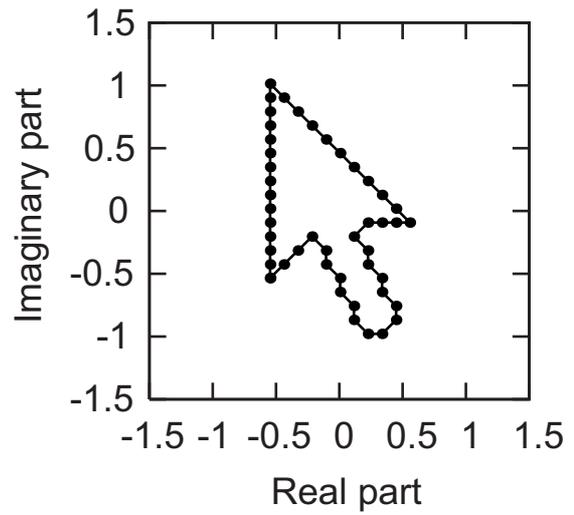
4.1.2 方法 B: 負周波数成分除去による方法

既に述べた通り、解析信号はその周波数成分のうち負周波数成分は全て 0 である。また、定数成分も含まないを考える。そのため、上記の方法のうち 3. における評価値を $\tilde{s}_{tmp}(t)$ の 0 以下の周波数成分の 2 乗和に置き換える。

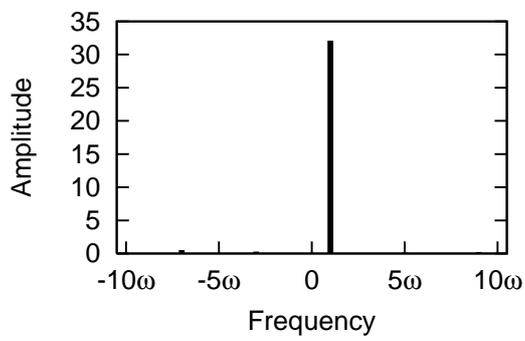
この方法での $n = 100$ におけるシミュレーション結果を図 4.4, 4.5 に示す。



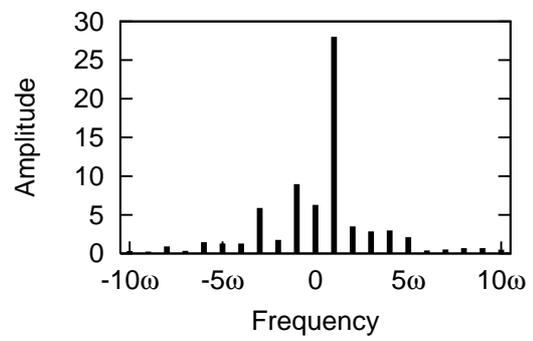
(a) 曲線 1



(b) 曲線 2



(c) 曲線 1 のスペクトル



(d) 曲線 2 のスペクトル

図 4.1: 入力図形とそのスペクトル

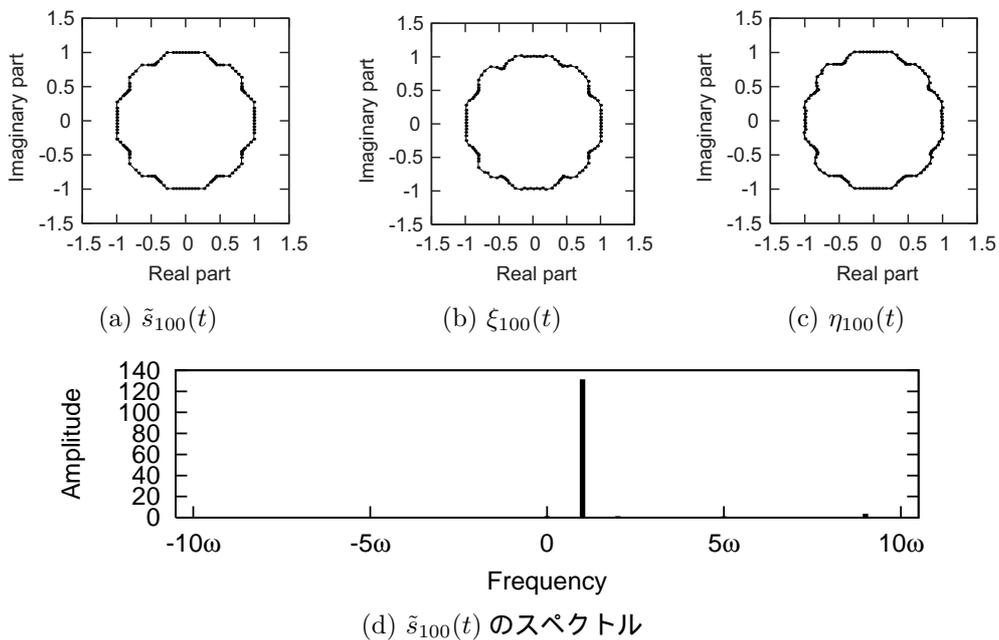


図 4.2: 方法 A のシミュレーション結果 (曲線 1)

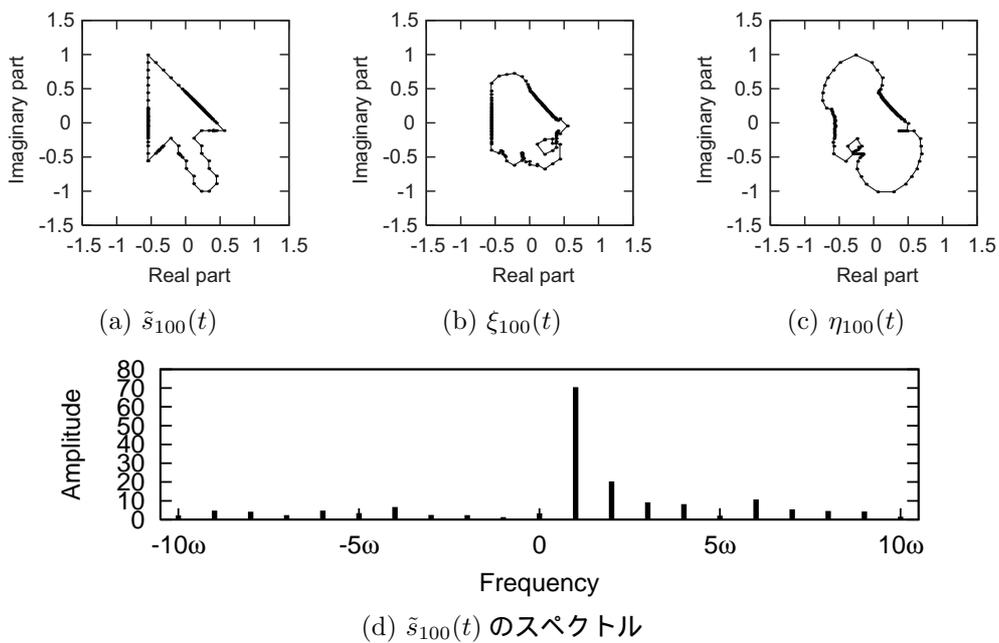


図 4.3: 方法 A のシミュレーション結果 (曲線 2)

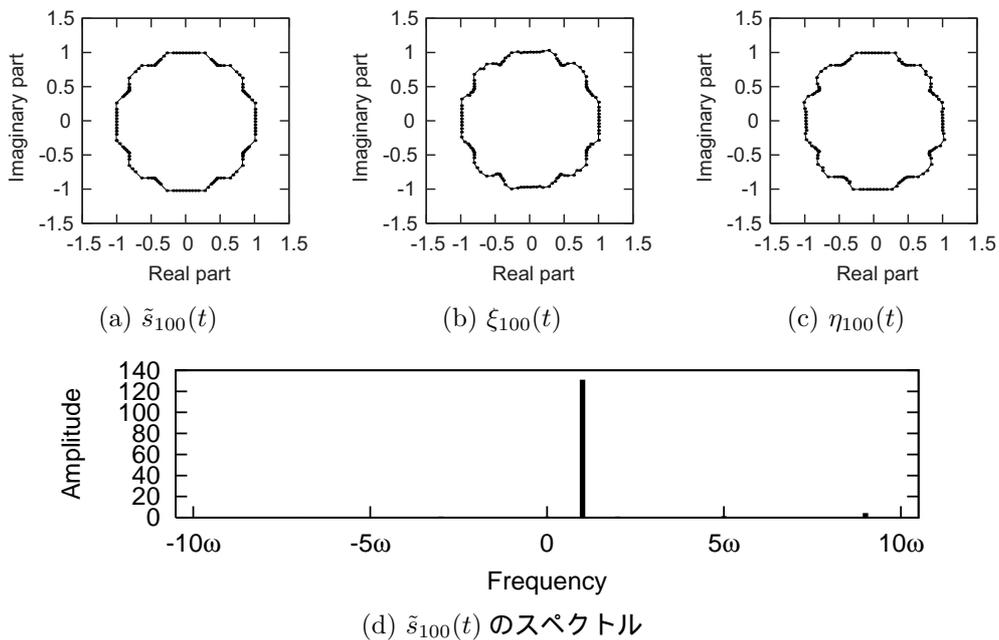


図 4.4: 方法 B のシミュレーション結果 (曲線 1)

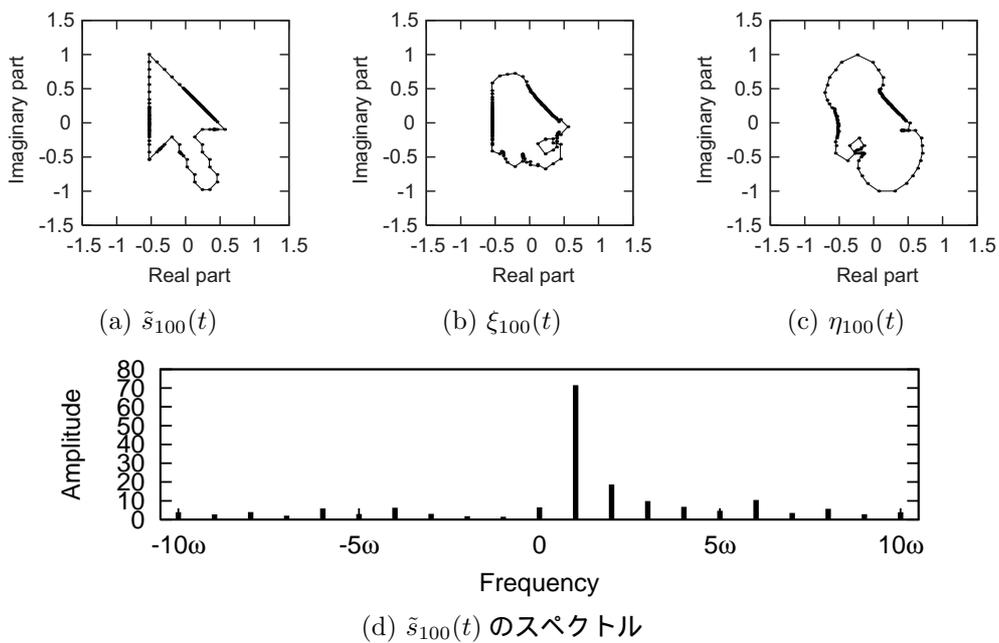


図 4.5: 方法 B のシミュレーション結果 (曲線 2)

第5章 解析信号シンセサイザ“ CloSynth ” の開発

5.1 CloSynth

開発した解析信号を用いる音色操作型のシンセサイザ“ CloSynth ”¹を図 5.1 に示す。図 5.1 の例では、実行画面中段の左の閉曲線には 64 個の制御点があり、制御点をマウス等によるドラッグによって音色の加工が可能である。なお操作する対象である閉曲線図形としては、あらかじめ楽器音を解析信号化したものがいくつか用意されており、プルダウンメニューから読み込みができる。すなわち、既存の楽器等の音色を元に、それに図形操作によって音色に加工を加える形での音色生成が可能である。

また生成された周期音声信号に対して、時間と共に振幅の変化を加えて音として仕上げるための機能として、ADSR (Attack, Decay, Sustain, Release) エンベロープジェネレータを搭載した。生成された音色を用いた簡易的な楽器の機能として、ソフトウェアキーボードのクリック、キーボードの入力、MIDI キーボードからの入力によって簡易的な演奏を可能とした。

¹<http://butchi.jp/documents/d-thesis/>

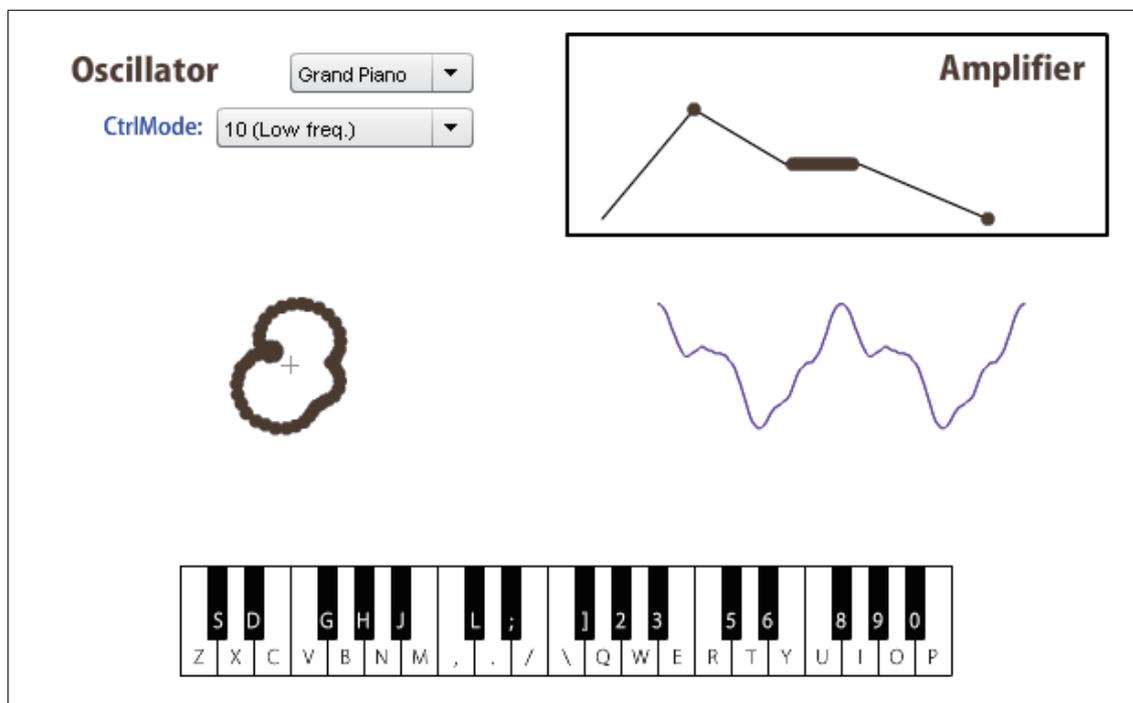


図 5.1: 開発した解析信号シンセサイザの実行画面

第6章 まとめと展望

本研究では音声と図形の変換方法として解析信号を用いる方法を提案した。これにより音声信号を曲線図形として表示することを可能にした。図形から音声への変換としては、任意の閉曲線図形をパラメータ変換することにより解析信号に近似する2つの手法を提案し、シミュレーションを行った。また、解析信号を拡張し、回転しても音色が不変であるような信号の条件を示した。そして、解析信号を用いたシンセサイザ“CloSynth”を提案・実装した。

音を形として見ることは、図形的な処理による音声認識や、音楽を聴く際の視覚エフェクト、VJの演出などへの応用が考えられる。形を音として聞くことは、画像処理による音色生成への応用が考えられる。形を編集して音を編集することは、音色生成・加工のモジュールとして利用することが考えられる。ベクトル画像における曲線表現形式としてベジェ曲線が普及しているように、音声信号の図形的表現形式として解析信号が普及することを私は願っている。