

相対伝達関数を用いたステレオ収録音の近似

Approximation of stereo sound based on relative transfer function

1W143016-6 大木 大夢

OGI Hiromu

指導教員 及川 靖広 教授

Prof. OIKAWA Yasuhiro

概要： サンプル方式は、楽器音などの収録音源を保持再生するシンセサイザの音源方式の一種であり、多くのシンセサイザで用いられている方式である。しかし、サンプル方式は保持しなければならないデータ量が多いため、ハードウェアコストがかかってしまうというデメリットを持つ。モノラル音源からステレオ音源を復元できれば、ステレオ音の記録に必要なデータ量を削減できる。そこで、本研究ではチャンネル間の相対的な伝達関数を推定することで、単一チャンネルのフィルタ処理により元のステレオ音源を近似する方法を提案する。フィルタ設計にあたっては、楽器音の特徴を活用することで、数値安定で効率的なフィルタ作成した。実際にシミュレーションをすることによりこの手法の有用性を確かめると共に、近似誤差の評価を行なった。

キーワード： シンセサイザ、サンプル方式、ピアノ音、IIR フィルタ、real Kautz モデル

Keywords: synthesizer, sample-based synthesis, piano sound, IIR filter, real Kautz model

1. まえがき

シンセサイザの音源方式として、加算合成方式や減算合成方式、変調合成方式などが挙げられる。中でも楽器の収録音を基に再生音を合成するサンプル方式は、他の手法よりも楽器音を忠実に再現できる。しかし、ピアノなど、打鍵の強さやダンパーペダルの有無で複数種類の収録音を保持する必要がある楽器音を再現するには、多くのコストがかかってしまうため、データ量を削減する手法が求められる。データ量が大きくなる一つの要因として、サンプル音源にステレオ収録音を用いることが挙げられる。

そこで、データ量削減の一つの方法として、ステレオ収録音の片側を、フィルタ処理によりもう一方の収録音に近似することが考えられる。片側の収録音に対するもう一方の収録音への相対的な伝達関数(以下、相対伝達関数)を持つフィルタを用いることで、収録音を近似でき、データ量が半分に削減されることが期待できる。

ただし、次数の高いフィルタを用いると、ハードウェアコストがかかりすぎるため、用いるフィルタとしては、低い次数で近似能力の高い無限インパルス応答フィルタ(IIR フィルタ)が有効である。IIR フィルタは設計次第では不安定なフィルタになりやすいが、real Kautz モデル [1] を用いれば、数値安定的なフィルタを設計することが容易となる。しかし、このモデルはフィルタの極配置が任意であるため、ランダムに極を配置すると近似能力が限られてしまう。そこで本研究では、real Kautz モデルの極をデータに基づき配置することで、近似効率が良く数値安定な IIR フィルタを設計し、ステレオ収録音を近似する手法を提案する。

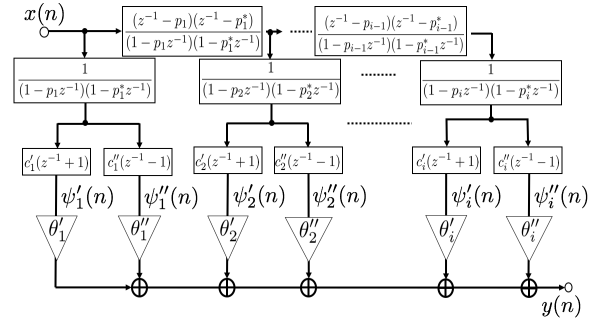


図-1 real Kautz モデル

2. 相対伝達関数の推定

2.1 real Kautz モデル

モデルパラメータ同士の積や商が存在するので、IIR フィルタの設計は容易ではない。その解決策の1つとして、フィルタ設計時に極の位置をあらかじめ固定する手法が提案されている [1]。図-1 に示す real Kautz モデルは、定めた極ごとに互いに正規直交なフィルタを作成し、それらを線形結合することで、所望のインパルス応答を近似するモデルである。各フィルタにおける入力信号に対する応答を $\psi(n) = \{\psi_k'(n), \psi_k''(n)\}_{k=1}^i$ とし、各応答に対する重み付けのパラメータを $\theta = \{\theta_k', \theta_k''\}_{k=1}^i$ とすると、real Kautz モデルの応答は

$$h_{p,\theta}(n) = \sum_{k=1}^i (\theta_k' \psi_k'(n) + \theta_k'' \psi_k''(n)) \quad (1)$$

と表される。これにより IIR フィルタの推定問題を、線形パラメータ θ の推定問題に置き換えることができる。

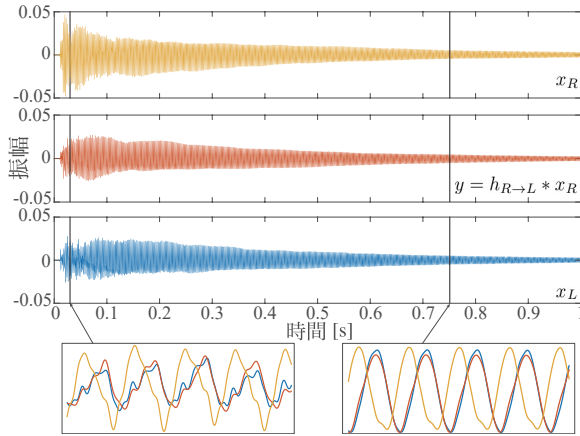


図-2 元音源と近似した音の波形

表-1 シミュレーション条件

対象音	ピアノ音 A4 (442 Hz)
解析時間 [ms]	1000
極の数	16
極半径	0.999
サンプリング周波数 [Hz]	48000

3. 極の配置

real Kautz モデルを用いると線形パラメータの推定が容易となるが、極をあらかじめ定める必要があり、極の配置により近似の能力が限られてしまう。そのため、人の聴覚特性に合わせて極を配置するなど、様々な手法が提案されている [2]。対して、本研究で扱う楽器音は、調波成分の存在する特定の周波数にエネルギーが集中しているため、その周波数に極を配置することで、少ない次数でより良い近似ができることが期待される。そこで、近似目的信号を自己回帰全極モデル (AR モデル) を用いて近似し、そのスペクトルのピークに対応する周波数を、極の周波数とした。

4. シミュレーション

提案手法を用いてステレオ収録音を近似した。左音源を x_L 、右音源に対する左音源の相対伝達関数のインパルス応答の推定値を $h_{R \to L}$ 、右音源を x_R とすると、近似した音 y は以下のように表される。

$$x_L \approx y = h_{R \to L} * x_R \quad (2)$$

なお、 $*$ は畳み込みを表す。

シミュレーションの条件を表-1、 x_L 、 x_R の波形と y の波形を図-2 に示す。図-2 から y が x_L の振幅や位相を近似できていることがわかる。

また、極半径をどのような値にすると、より精度の高い近似ができるかを調べるために、極半径を複数設定し x_L と y の相対誤差を求めた。相対誤差 ε は、

$$\varepsilon = \frac{\|x_L - y\|}{\|x_L\|} \quad (3)$$

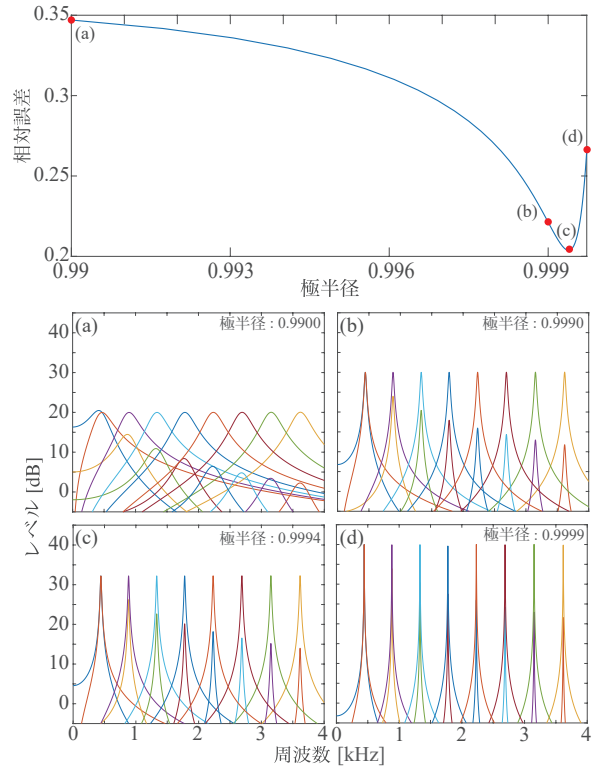


図-3 相対誤差と極ごとに作成されるフィルタの特性

で求めた。ただし、 $\|\cdot\|$ はベクトルのユークリッドノルムを表す。図-3 の上部に相対誤差を示し、下部に図中の (a)~(d) に対応する正規直交なフィルタの振幅特性を示す。極半径が 1 に近づくほど、フィルタの特性のピークは急峻になることがわかり、0.9994 付近で、相対誤差が最少となることが確認できた。これは、調波成分を個別に扱うことで少ない次数でも相対伝達関数を近似できるが、フィルタが局所的になりすぎると近似能力が不足することを示唆している。

5. むすび

本論文ではステレオ収録音の左右チャネルの相対伝達関数を推定し、右チャネル収録音から左チャネル収録音を近似する手法を提案した。シミュレーションにより、収録音を近似することができたが、ある程度の近似誤差があることもわかった。今後は、さらに近似誤差が小さくなるよう、配置した極の極半径を最適化する手法を検討する所存である。

参考文献

- [1] G. Vairetti, T.V. Waterschoot, M. Moonen, M. Catrysse and S.H. Jensen, "A Scalable Algorithm for Physically Motivated and Sparse Approximation of Room Impulse Responses With Orthonormal Basis Functions," IEEE/ACM Trans. Audio Speech Lang. Process. Vol. 25, No. 7, pp. 1547–1561, July, 2017.
- [2] M. Karjalainen and T. Paatero, "Equalization of Loud speaker and Room Responses Using Kautz Filters: Direct Least Squares Design," EURASIP J. Adv. Signal Process., Vol. 2007, No. 1, pp. 060949-1–060949-13, 2006.