減算合成シンセサイザにおけるSuperSawのユニゾン数とデチューンの推定*

☆ 松本和樹 (早大)，矢田部浩平 (農工大)

1　はじめに

減算合成シンセサイザは現代の音楽に不可欠な楽器である．その音色は，オシレータが発振する音に対しフィルタを適用することで作られる．このとき，オシレータの波形，適用するモジュレーションやフィルタ，エフェクタなどの選択，およびそれらのパラメータの設定が必要となる．

現代の楽曲制作では，アーティスト自らすべての音色を作ることは稀であり，多くの場合，サンプルと呼ばれる音楽制作用の素材を取り入れている．サンプルは楽曲を構成する各要素を個別のデータとして収録した数秒から数十秒程度の音の素材である．サンプルには大きく分けて，ワンショットサンプルとループサンプルの二種類がある．ワンショットサンプルには，ドラムのキックやスネア，シンセサイザのワンショット素材，サウンドエフェクトなどがあり，これらを並べたり，サンプリングシンセサイザで読み込んだりして用いられる．ループサンプルには，パーカッションのループ素材，メロディやコード進行を伴うシンセサイザや生楽器のフレーズなどが存在する．ループサンプルは楽曲にそのまま貼り付けるか，あるいは特定の部分を切り貼りすることで用いられる．このように，サンプルを用いることで，作成の手間のかかる音色を手軽に取り入れることができる．

しかし，サンプルの問題点として，メロディやテンポ，和音の変更に伴う音質劣化が挙げられる．例えばメロディを伴うサンプルの編集では，タイムストレッチやピッチシフトなどによる音質劣化が生じる．さらに，和音の変更では重なった音の分離が要求されるため，高品質な編集は難しくなる．

そこで，既存のサンプルからその音色を再現するシンセサイザのパラメータを推定する技術が望まれる．サンプルからシンセサイザのパラメータを推定できれば，和音の構成やメロディ，テンポの変更，細かい音色の調整などが可能となり，サンプル利用の自由度が向上すると考えられる．類似の研究としては，[1, 2] などが挙げられる．

SuperSaw はポップスや EDM などにおいて重要な音色の一つである．SuperSaw は，少しずつ音高の異なる複数の鋸歯状波を重ねたものであり，シンセサイザのユニゾンの機能を用いて作られる．ユニゾンは，波形が同じで音高が少しずつ異なる複数のオシレータを同時に発振する機能である．ユニゾンを用いることで，単一のオシレータによる音色と比較して派手で厚みのある音色を作ることができる．本稿では，SuperSaw の音作りにおけるユニゾンのパラメータを推定する．特に，ユニゾンの基本的なパラメータであるオシレータの数（ユニゾン数）と，音高をずらす幅（デチューン量）に着目する．提案手法では，ユニゾンされた音のパラメータ推定に有用なコスト関数を導入し，その有効性を実験によって確認した．

2　SuperSaw の定義と性質

2.1　減算合成シンセサイザとオシレータ

減算合成シンセサイザの音色は，オシレータが発振する音にフィルタを適用することで作られる．多くの音色では，音量やフィルタのカットオフ周波数などが時間的に変化する．ただし，本稿では研究の第一歩として，フィルタは適用しないものとし，音量も時間変化しないものとする．

オシレータが発振する音の波形を $y(f,t)$ とおく．ただし，f は基本周波数，t は時刻を表す．y は周期が $1/f$ の周期関数である．ナイキスト周波数を f_{Nyquist} とし，M を $fM < f_{\mathrm{Nyquist}}$ を満たす最大の自然数とすると，y は周波数が kf $(1 \leq k \leq M)$ の M 本の正弦波の和で表される．すなわち，k 番目の倍音の振幅を A_k，位相を φ_k とすれば，y は以下のように書ける．

$$y(f,t) = \sum_{k=1}^{M} A_k \sin(2\pi k f t + \varphi_k) \tag{1}$$

本稿ではオシレータの波形 y として鋸歯状波を扱う．鋸歯状波は，以下の式で表せる．

$$\mathrm{saw}(f,t) = \frac{2}{\pi}\sum_{k=1}^{M} \frac{(-1)^{k-1}}{k}\sin(2\pi f k t) \tag{2}$$

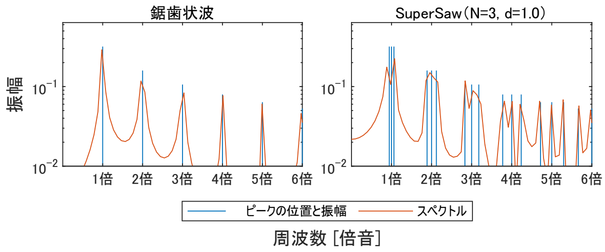
2.2　SuperSaw とユニゾン

SuperSaw は Roland 社のシンセサイザである JP-8000 [3] に搭載されて以来，現代でも数多くの楽曲に用いられている．SuperSaw は音高の少しずつ異なる複数の鋸歯状波を重ねた音であり，シンセサイザの基本的な機能であるユニゾンを用いて作られる．

シンセサイザの音作りにおけるユニゾンは，波形が同じで音高が少しずつ異なる複数のオシレータを同時に発振する機能である．各オシレータの周波数は区間

---

*Estimation of the number of unison and degree of detune of SuperSaw in subtractive synthesizers. By Kazuki Matsumoto (Waseda University) and Kohei Yatabe (TUAT)



図–1　鋸歯状波と SuperSaw の振幅スペクトル

$[2^{-d/12}f, 2^{d/12}f]$ を対数的に等間隔に分けるように定められる．ただし，f は基本周波数であり，$d > 0$ は音高をずらす幅（デチューン量）である．このとき，n 番目のオシレータの周波数 $f_{n,N,d}$ は,

$$f_{n,N,d} = 2^{(d/12)(-1+2(n-1)/(N-1))} f \tag{3}$$

で与えられる．ただし，N はユニゾン数であり，$N_{\max}$ をユニゾン数の上限としたとき $2 \le N \le N_{\max}$ を満たす．$N_{\max}$ はシンセサイザの仕様により決まり，本稿では典型的な値である 16 を採用する．

SuperSaw は鋸歯状波の和で与えられ,

$$\begin{aligned} &\mathrm{SuperSaw}_{N,d,\{\phi_n\}_{n=1}^{N}}(f,t) = \\ &\qquad \sum_{n=1}^{N} \mathrm{saw}\left(f_{n,N,d},\ t - \tau_{n,N,d}\right) \end{aligned} \tag{4}$$

と書ける．ただし，$\tau_{n,N,d}$ は n 番目のオシレータの時間シフト量であり，位相シフト量 $0 \le \phi_n < 2\pi$ $(n = 1, \cdots, N)$ を用いて以下の式で表される．

$$\tau_{n,N,d} = \phi_n / f_{n,N,d} \tag{5}$$

2.2.1　SuperSaw の振幅スペクトルの特徴

図–1 に鋸歯状波及び SuperSaw の振幅スペクトルを示す．鋸歯状波の各倍音成分の振幅は式 (2) の係数 A_k と対応する．SuperSaw の振幅スペクトルでは，各倍音に N 個のピークが存在し，それらの間隔は倍音次数 k に比例する．有限長の信号から得られるスペクトルでは，位置が近いピーク同士の干渉が生じる場合がある．図 1 の SuperSaw のスペクトルでも，倍音次数の低い 1〜3 倍音でピークが干渉し，本来あるはずの 3 つのピークが確認できない状態となっている．

3　提案手法

本稿では SuperSaw のモノラル信号から，ユニゾン数 N とデチューン量 d を推定する手法を提案する．ただし，信号の基本周波数は既知とする．これは，シンセサイザの音は機械的に生成されるため，その基本周波数は 12 平均律上の音高と正確に一致する場合が多く，推定が比較的容易と考えられるためである．

3.1　コスト関数

SuperSaw のパラメータである N および d を推定するためのコスト関数は，パラメータの真値で最小値をとるように定義する．以下ではまず，単音のコスト関数を定義し，その後和音へと拡張する．

単音の SuperSaw に対するコスト関数 $\mathfrak{C}_f(N,d)$ は,

$$\mathfrak{C}_f(N,d) = -\frac{1}{M_{\mathrm{est}} N} \sum_{n=1}^{N} \sum_{k=1}^{M_{\mathrm{est}}} |(\mathscr{F}x)(k f_{n,N,d})| \tag{6}$$

と定義する．ただし，添え字の f は信号の基本周波数を表す．また，$(\mathscr{F}x)(f)$ は元信号 x のフーリエスペクトルにおける周波数 f での値を意味する．M_{est} は周波数のもっとも高いオシレータにおいてナイキスト周波数以下となる最大の倍音の次数であり，d の探索範囲の上限 $d_{\max}$ を用いて以下のように書ける．

$$M_{\mathrm{est}} = \left\lfloor \frac{f_{\mathrm{Nyquist}}}{f_{N,N,d_{\max}}} \right\rfloor \tag{7}$$

コスト関数値は，各 N および d においてピークが存在するはずの全ての周波数で振幅スペクトルを計算し，それらの平均値の符号を反転した値である．そのため，振幅スペクトルを計算したすべての点がピークに該当するとき，コスト関数は最小値をとる．

和音の SuperSaw に関しては，和音を構成する各音の基本周波数について式 (6) に定義するコスト関数を計算し，その平均値を用いる．すなわち，和音の SuperSaw に対するコスト関数 $\mathfrak{C}_F$ は,

$$\mathfrak{C}_F(N,d) = \sum_{f \in F} \mathfrak{C}_f(N,d)/|F| \tag{8}$$

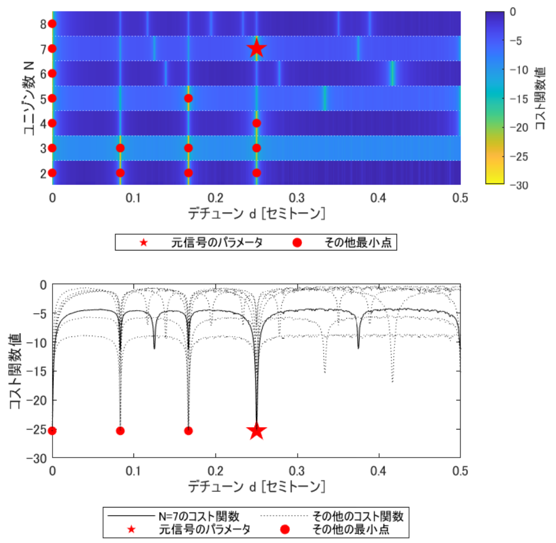
と書ける．ただし，F は和音を構成する音の基本周波数の集合であり，$|F|$ は F の要素数，すなわち和音を構成する音の数である．

例として，図–2 に $N = 7$，$d = 0.25$ の SuperSaw に対するコスト関数を示す．フーリエスペクトルの分解能が十分ならば，元信号のパラメータに対応する点（星印）はコスト関数の最小点となる．その他にも複数の最小点（丸印）が存在し，元信号のパラメータに対応する最小点は，$d \ne 0$ を満たす最小点の中で N と d がいずれも最大の点となる．分解能が十分でない場合には，ピーク同士の干渉の影響により最小点での値にばらつきが生じる可能性がある．

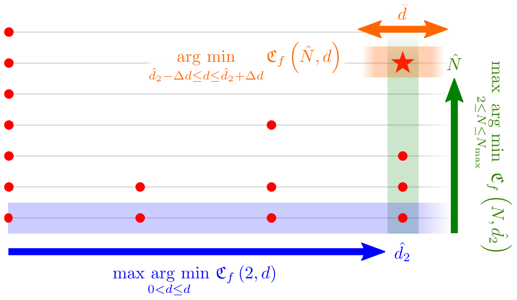
提案するコスト関数は，多数の点でのスペクトルの値を必要とする．実装時に非等間隔 FFT [4] を用いることで計算を効率化できる．

3.2　推定の流れ

ユニゾン数 N およびデチューン量 d の推定は式 (9) から式 (11) に示す三段階で行われる．初めに，N を



図–2　コスト関数の例



図–3　推定の流れ

2 に固定したコスト関数 $\mathfrak{C}_f(2,d)$ が最小値を取る d を探索し，その中で最大のものを $\hat{d}_2$ とする．次に，d を $\hat{d}_2$ に固定したコスト関数 $\mathfrak{C}_f(N,\hat{d}_2)$ が最小値を取る N を探索し，その中で最大の N をユニゾン数の推定値 $\hat{N}$ とする．最後に，N を $\hat{N}$ に固定したコスト関数 $\mathfrak{C}_f(d,\hat{N})$ を用い，$\hat{d}_2$ を中心とする両幅 Δd の区間 $[\hat{d}_2-\Delta d,\ \hat{d}_2+\Delta d]$ 上の最小点をより詳細に探索し，最終的なデチューン量の推定値 $\hat{d}$ を得る．

$$\hat{d}_2 \leftarrow \max \mathop{\arg\min}_{0<d\le d_{\max}} \mathfrak{C}_f(2,d) \tag{9}$$

$$\hat{N} \leftarrow \max \mathop{\arg\min}_{2\le N\le N_{\max}} \mathfrak{C}_f\left(N,\hat{d}_2\right) \tag{10}$$

$$\hat{d} \leftarrow \mathop{\arg\min}_{\hat{d}_2-\Delta d\le d\le \hat{d}_2+\Delta d} \mathfrak{C}_f\left(\hat{N},d\right) \tag{11}$$

上記の式は推定の概要を示すためのものであり，実用上はコスト関数のサンプリングに起因する誤差やピーク同士の干渉などを考慮し，これらの式を改良する必要がある．詳細については後述する．

図–3 に推定の流れを示す．提案手法はコスト関数の最小点を d 方向，N 方向の順に辿り，最後に $\hat{d}$ をより詳細に推定することで元信号のパラメータを効率的に推定している．

3.3　$\hat{d}_2$ の探索

$\hat{d}_2$ の探索では，コスト関数を等間隔にサンプリングした配列 $\tilde{\mathfrak{C}}_f[N,i]$ を用いる．d に関するサンプリング点数を L とすると，配列の各要素 $\tilde{\mathfrak{C}}_f[N,i]$ は，

$$\tilde{\mathfrak{C}}_f[N,i] = \mathfrak{C}_f(N,\tilde{d}[i]) \tag{12}$$

と書ける．ただし，$\tilde{d}[i]$ は d のサンプリング点であり，探索範囲が $[d_{\mathrm{L}},d_{\mathrm{R}}]$ となるとき以下の式で書ける．

$$\tilde{d}[i] = d_{\mathrm{L}} + \frac{(i-1)(d_{\mathrm{R}}-d_{\mathrm{L}})}{L-1} \tag{13}$$

サンプリングしたコスト関数を用いる場合には，サンプリング点が最小点に一致せず，厳密な最小値が得られない場合が多い．また，ピーク同士の干渉が影響し，本来同一の最小値をとる点の値にばらつきが生じる場合もある．そのため，ヒューリスティックに最小点を求める．本稿では，$\hat{d}_2$ のインデックス $\hat{i}_2$ は，

$$\hat{i}_2 \leftarrow \max\left\{ i\in V \,\middle|\, \tilde{\mathfrak{C}}_f[2,i] < \mathfrak{C}_{\mathrm{thresh}} \right\} \tag{14}$$

とする．ただし，V はコスト関数の谷のインデックスの集合である．また，閾値 $\mathfrak{C}_{\mathrm{thresh}}$ は，

$$\mathfrak{C}_{\mathrm{thresh}} = r\cdot \min \tilde{\mathfrak{C}}_f[2,i] \tag{15}$$

とする．ただし，r は 0.9 程度の定数である．最終的に，$\hat{d}_2$ は $\hat{i}$ を用いて以下のように書ける．

$$\hat{d}_2 \leftarrow \tilde{d}[\hat{i}_2] \tag{16}$$

3.4　$\hat{N}$ の探索

$\hat{N}$ の探索でもヒューリスティックな探索が必要となる．$\hat{N}$ は以下の式に示すように求める．

$$\hat{N} \leftarrow \max\left\{ 2\le N\le N_{\max} \,\middle|\, \tilde{\mathfrak{C}}_f\left(N,\hat{i}_2\right) < \mathfrak{C}_{\mathrm{thresh}} \right\} \tag{17}$$

3.5　$\hat{d}$ の探索

$\hat{d}$ の探索では，まず区間 $[\hat{d}_2-\Delta d,\ \hat{d}_2+\Delta d]$ においてコスト関数をサンプリングし，その最小点を探索する．最小点のインデックス $\hat{i}$ は以下の式で得られる．

$$\hat{i} \leftarrow \mathop{\arg\min,}_{1\le i\le L} \tilde{\mathfrak{C}}_f[\hat{N},i] \tag{18}$$

その後，二次関数の補間曲線 $\mathfrak{C}_{\mathrm{interp}}$ を用い，より詳細に最小点を求める．$\mathfrak{C}_{\mathrm{interp}}$ は以下の式で定義される．

$$\mathfrak{C}_{\mathrm{interp}}(d) = ad^2+bd+c \tag{19}$$

ただし，係数 $a,\ b,\ c$ は補間曲線がサンプリングされたコスト関数の $\hat{i}-1,\ \hat{i},\hat{i}+1$ に対応する点を通るように定める．デチューン量の推定値 $\hat{d}$ は補間曲線の最小点であり，以下の式で得られる．

$$\hat{d} \leftarrow -b/2a \tag{20}$$

表–1　和音の種類と構成音のルート音に対する周波数比

和音の種類	ルート音に対する周波数比			
	1 度	3 度	5 度	7 度
ルート音のみ	1	-	-	-
パワーコード	1	-	$2^{7/12}$	-
メジャー	1	$2^{4/12}$	$2^{7/12}$	-
マイナー	1	$2^{3/12}$	$2^{7/12}$	-
メジャーセブンス	1	$2^{4/12}$	$2^{7/12}$	$2^{11/12}$
マイナーセブンス	1	$2^{3/12}$	$2^{7/12}$	$2^{10/12}$

4　実験

SuperSaw 信号に対し提案手法を用いたパラメータ推定を行い，その精度を確認する．

4.1　実験方法

元信号のサンプリング周波数は 44100 Hz とし，信号長は 0.3, 0.5 秒とした．和音の構成は，表–1 に示す 6 種を対象とした．ルート音の周波数は 220〜880 Hz を対数等間隔に分ける 10 段階を，オシレータ数 N は 2〜16 の整数を，デチューン量 d は 0.05〜0.50 セミトーンを線形等間隔に分ける 10 段階を対象とした．

推定のパラメータに関しては，$d_{\max}$ を 0.55 セミトーン，分割数 L を 1000，閾値のパラメータ r を 0.9，$\hat{i}_2$ 探索時の探索範囲の幅 Δd を 0.05 セミトーンとした．デチューン量の推定時の $\hat{N}$ の値には N の真値を用いた．提案手法の精度は，オシレータ数 N に誤り率，デチューン量 d に平均絶対値誤差を用い定量化した．

4.2　実験結果

4.2.1　ユニゾン数の推定

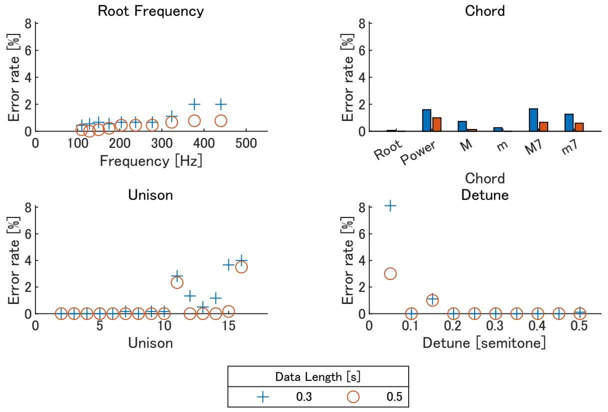
ユニゾン数の推定における誤り率は 0.67 % となった．また，図–4 に示すように，誤差はデータ長が短いとき，基本周波数が高いとき，N が大きいとき，d が小さいときに増加する傾向が見られた．

デチューン量 d が 0.2 セミトーン以上のときの誤り率は平均で 0.0079 % となった．これは 10,000 個のサンプルにつき 1 つの誤りが生じる程度の割合であり，実用に足る水準であるといえる．

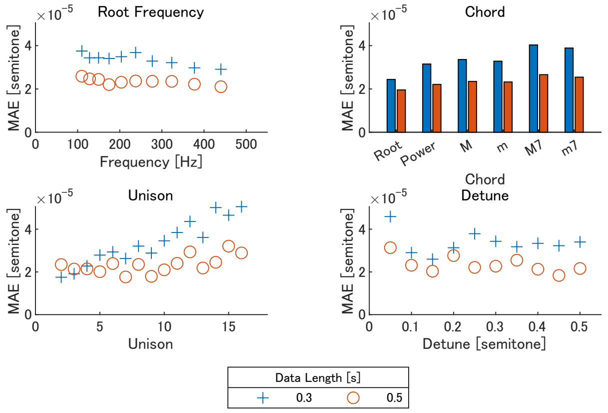
一方，デチューン量 d が 0.05 セミトーンのときの誤り率は顕著に大きく，8.1 % となった．これは，ピークの間隔が狭く干渉の影響が大きいことが要因だと考えられる．実際，推定を誤る場合には最小点での値がばらつき，真のパラメータに対応する点が閾値を超えないことが確認できた．何らかの方法で閾値を自動調整できれば誤り率は低減されると考えられる．

4.2.2　デチューン量の推定

誤差の最大値，平均値はそれぞれ 2.6×10^{-4}，2.8×10^{-5} セミトーンとなった．また，図–5 に示すように，誤差はデータ長が短いとき，基本周波数が低いとき，N が大きいとき，d が小さいとき，和音を構成する音の数が多いときに増加した．



図–4　ユニゾン数の推定結果



図–5　デチューン量の推定結果

本実験で用いた信号では，デチューン量の差が 1.0×10^{-3} セミトーン以下のとき，音色としての差は知覚できなかった．したがって，提案手法の誤差は実用上十分に小さいといえる．

5　おわりに

本稿では SuperSaw のユニゾン数 N およびデチューン量 d を推定する手法を提案した．提案手法では，ユニゾン機能のパラメータの推定に特化したコスト関数を導入した．実験の結果，提案手法は多くの場合で実用に足ることが確認できた．デチューン量が小さいときの閾値の自動調整が今後の課題である．

参考文献

[1] 糸山克寿, 奥乃博. “楽器音に対する仮想音源のパラメータ推定,” 情報処理学会研究報告 (MUS), **2013**(5), 1–6 (2013).
[2] O. Barkan, D. Tsiris, O. Katz, and N. Koenigstein, “Inversynth: Deep estimation of synthesizer parameter configurations from audio signals,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **27**(12), 2385–2396 (2019).
[3] Roland, “Support - JP-8000 - OWNER'S MANUALS,” https://www.roland.com/global/support/by_product/jp-8000/owners_manuals/ (参照 2022-07-06)
[4] A. Dutt and V. Rokhlin, “Fast Fourier transforms for nonequispaced data,” *SIAM J. Sci. Comput.*, **14**(6), 1368–1393 (1993).