

セルオートマトンを用いた音楽信号記述とデータ圧縮への応用

和田 昌人, 黒岩 丈介, 奈良 重俊

広島大学 総合科学部

Digital sound data description with cellular automata and its application to data compression

Masato Wada, Jousuke Kuroiwa, and Shigetoshi Nara

Faculty of Integrated Arts and Sciences, Hiroshima University,

Higashi-hiroshima 739-8521, Japan

Abstract: Recently, we have proposed a novel method to describe digital sound data by means of rule dynamics in two states and three neighbors cellular automata. In this paper, we report the results of evaluation about the reproducibility and the performance of the sound description by calculating distortion rate, power spectra and return map of signal amplitude. A typical result in our numerical evaluations gives 21.29 dB as averaged distortion rate. Comparison between the two return maps, the one taken from the original signals and the other from the amplitude data generated by the present method shows some difference, which indicates that the present method of rule extraction is not yet optimized in describing sound data. In order to improve the fidelity of describing the original data, we have tried to apply all the possible rule sequences which are generated by combining three rules within applying them eight times to a certain initial bit pattern. In the best case, the trial can give strong possibility of describing digital data with high fidelity and with considerable compression.

1 はじめに

19世紀末期に、数学者 Poincaré によって、天体の3体運動の研究により、カオス現象の存在が発見された [1]. その当時は非線形微分方程式を数値的に解く手段がなかったので、その実像に迫ることはできなかったが、彼はカオス現象を見抜いていたと言われている。それから約70年後の1961年に、気象学者 Lorenz は、簡単な微分方程式による気象モデルにおいて、カオスがみられることを計算機シミュレーションにより発見した。1970年代には、数理生態学者の May が、簡単な反復の過程が大変複雑な振舞いを発生することを発見した [1]. ロバートメイが発見した一次元離散力学系であるロジスティック写像でさえ、わずかな初期状態のずれがのちの振舞いに大きな影響を与える。このように少数自由度におけるカオス現象は単純なルールで記述できるにもかかわらず非常に複雑なふるまいを示し、ほんのわずかな初期値の違いが全く違うふるまいを生ぜしめるので予測し難い現象であり、多くの科学者によって詳細に研究されてきた。そのため、少数自由度におけるカオス現象は、理論的にもその分岐過程やフラクタル次元等の動的構造が明らかになりつつある。

一方、生体（特に脳のニューロン活動を含む）においてもカオス現象が発見されている [2, 3, 4]. しかしながら、生体系といった多自由度系から生じるカオス現象は、その自由度の高さから構造があまりに複雑すぎるゆえに、少数自由度系のカオスほどの理解はされていない。我々は、このよう

な多自由度系におけるカオス現象にも、背後に何らかの単純な法則が存在するかもしれないと考えている。

単純な法則・ルールによってカオス現象といった複雑なダイナミクスを生成するものの別な例として、セルオートマトンがあげられる。セルオートマトンは戦後 Von Neumann によって提唱され、1980 年代に Wolfram によって動的特性が詳細に研究された [5]。特に、1 次元 2 状態 3 近傍、あるいは、2 次元 2 状態 5 近傍総和型などのセルオートマトンが、力学的観点から系統的に研究された。更に、相沢・永井らは、単純なルールのみからでも複雑なダイナミクスが生成されることを、1 次元 2 状態 3 近傍セルオートマトンにおけるルールの時間遷移（以後、“ルールダイナミクス”と呼ぶ）という観点から研究している [6]。また、永井・鬼頭・相沢らは、脳波データをルールダイナミクスの観点から解析している [7]。

以上のように、(i) 単純な法則・ルールからでもカオス現象といった、初期値敏感性に起因する長期的予測困難性を有する複雑なダイナミクスが発生し得ること、(ii) 生体（脳におけるニューロン活動を含む）においてでさえ、カオス現象が見い出されてきていることに強く動機付けられ、我々は生体の情報処理や制御のからくりをカオス現象といった側面より明らかにすることが可能ではないかと考える [8, 9]。つまり、生体といった多自由度系から発生される現象にも、背後に何らかの単純なルールが存在し、そのルールを見い出すことから生体の情報処理や制御のからくりを明らかにできるのではないかと期待している。そこで、本論文では、生体が発生し処理しているものの 1 例として音声・音楽信号をとりあげ、これらのデジタル信号を 1 次元 2 状態 3 近傍セルオートマトンのルールダイナミクスを用いて実現することが研究の目的である。その一例として、1 次元 2 状態 3 近傍セルオートマトンのルールダイナミクスを用いたデジタル化された音声・音楽信号の記述、及び圧縮の可能性を本論文で報告する。

本論文の構成は以下の通りである。第 2 章では、1 次元 2 状態 3 近傍セルオートマトンについての説明を与える。第 3 章では音声・音楽信号のデジタル化について説明を与える。第 4 章では 1 次元 2 状態 3 近傍セルオートマトンのルールダイナミクスを用いたデジタル化された音声・音楽信号の記述、再現音の評価結果、及び圧縮率の結果を与える。第 5 章は結論であり、本研究の成果と今後の課題について述べる。

2 1 次元 2 状態 3 近傍セルオートマトンとカオス

1 次元 2 状態 3 近傍セルオートマトンの離散時刻 $t (= 0, 1, 2, \dots)$ における各セルの状態は 0 か 1 の 2 値を取り、次の時刻 $t+1$ のセルの状態は自分自身とその両隣りの 2 つのセル計 3 セルの状態によって決定される（図 1 参照）。図 1 においては、1 次元のセルが横一列に配置されている。そして、2 状態ということをお白か黒（0 か 1）であらわしている。

つまり、位置 i 、時刻 t でのセルの状態を $a_i^t (= 0 \text{ or } 1)$ とすると、次の時刻 $t+1$ でのセルの状態は、

$$a_i^{t+1} = f(a_{i-1}^t, a_i^t, a_{i+1}^t) \quad (1)$$

であらわされる。 $f(\cdot)$ は、遷移関数であり、

$$\begin{aligned} f(0, 0, 0) &= f_0, f(0, 0, 1) = f_1, f(0, 1, 0) = f_2, f(0, 1, 1) = f_3, \\ f(1, 0, 0) &= f_4, f(1, 0, 1) = f_5, f(1, 1, 0) = f_6, f(1, 1, 1) = f_7 \end{aligned}$$

とする。但し、 $f_i (i = 0, \dots, 7)$ は、0 か 1 をとるものとする。つまり可能なルールは全部で $2^8 = 256$ 個ある。各 f_i の値を 0 か 1 かに決定することは、256 個のルール中で実際に用いるルール集合を決定することに対応する。このルール集合を記述するために、Wolfram はルール番号を導入した [5]。ルール番号は次のように定義される。

$$f_0 + 2f_1 + 2^2f_2 + \dots + 2^7f_7 \quad (2)$$

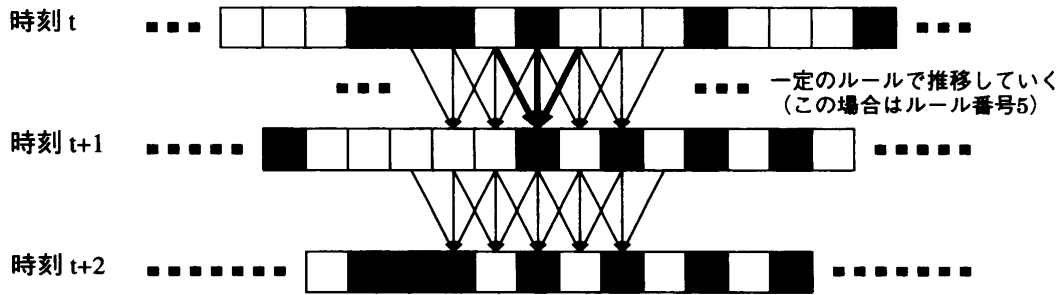


図 1: セルオートマトンの時間発展

例えば、ルール番号を 5 と決めると、図 2 のように用いるルール集合が一意に決定するのである。

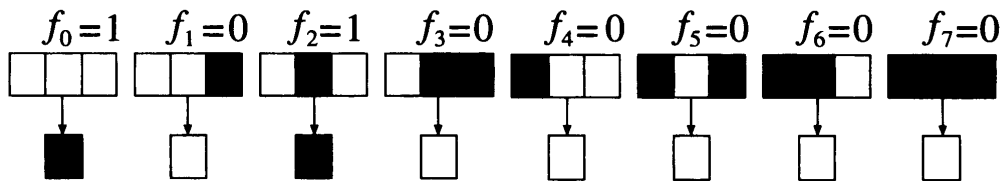


図 2: 遷移関数の一例 (ルール番号 5)

図 3 は、ある初期パターンから始まったセルの時間遷移を縦にならべたものであり、Wolfram の分類に従うと [5], 図 3 (a) はクラス 1, 図 3 (b) はクラス 2, 図 3 (c) はクラス 3 の状態遷移の例である。ここで各 4 クラスは以下のような意味で分類されている。

- クラス 1: このクラスのセルオートマトンは時間がたつとすべてのセルの状態が 0 か 1 の状態におちつく。(図 3 (a))
- クラス 2: クラス 2 では、時間がたつと周期的なパターンへと落ちつく。(図 3 (b))
- クラス 3: クラス 3 は、セルオートマトンのダイナミクスが周期性を見せずカオス的である。(図 3 (c))
- クラス 4: このクラスの振舞いは非常に複雑で、ランダムな初期値をあたえても、ある場合はすべてが、0 におちつき、別の初期値からは、周期的なパターンが表われてたり、ずっとパターンがかわりつづけることもある。

3 音声・音楽信号の計算機での取扱い

3.1 音声・音楽信号のサンプリング

音声・音楽信号は通常、アナログデータなのでそのまま計算機で扱うことはできない。したがって、計算機で扱えるようなデータにする必要がある (これを A/D 変換という)。まず、連続時間信号を離散時間信号に変換する必要がある。これは標本化 (サンプリング) という。この際にどのような間隔 (これをサンプリング周期という、またその逆数をサンプリング周波数という) で標本

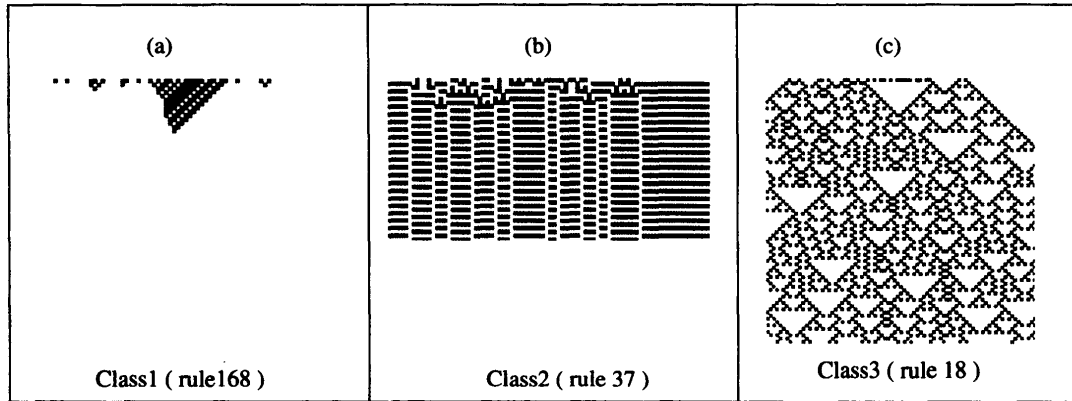


図 3: 各クラスに属するルールによる時間発展

化すればよいのか、という問題がでてくる。もとの連続時間信号を忠実に再現するためには、できるだけ大きいサンプリング周波数でサンプリングすればよいのだが、計算機の演算処理時間や信号のデータ量を考えるとサンプリング周波数を大きくしすぎても、データ量を増すことになるので好ましくはない。しかし、小さくしても信号を再現できなくなるので、適切なサンプリング周波数をもとめる必要がある。この問題を解決するために以下のようなサンプリング定理が知られている [10]。サンプリング定理で述べられていることは、以下の通りである。

「所定の条件をみたすクラスに属する任意の時間に関する関数について、フーリエ変換を施し、フーリエスペクトルが得られたとする。その際スペクトルの帯域幅が周波数でいって $\pm f_s$ の範囲内に制限されていることが成立している場合、サンプリング周波数 $2f_s$ 以上でサンプリングした離散データから、もとの時間に関する関数を完全に再現できる。」

したがって、任意の信号を $2f_s$ でサンプリングして表現しようとするときは、 f_s 以上の周波数をフィルタによって除去しておかないと再現したときには原信号にはなかった歪が発生するのである。一般に人間の可聴限界は、約 20kHz なので [11]、音楽 CD などは、44.1kHz でサンプリングされている。また、人の発話のように、上限が、10kHz などとわかっている場合は、20kHz くらいのサンプリングでも十分に、再現できるのである。図 4 は、連続時間信号から離散時間信号へ変換したときの模式図である。

3.2 音声・音楽信号の量子化

次に量子化である。量子化とは、振幅値に対して、離散化をおこなうことである。図 5(a) はその実際の入出力特性の一例である。ここでは、線形量子化特性の例を挙げている。この階段階数（これを量子化階調と呼ぶ）が多いほど、量子化前のデータとくらべて誤差が少なく再現できるが、あまり多くてもデータ量が膨大になる上処理に時間がかかるようになる。現在の音楽データの主流は 16 ビットであり、65536 階調になる。このようにして、量子化されたデータは離散値で 16 個の 1, 0 の列で記述されることができる。ここで、1 の時を黒、0 の時を白とすれば、セルオートマトンに対応つけることができる。たとえば、34944 の場合、1001001000000 となる。

したがって、音声・音楽データは有限セル列のつくるビットパターンの遷移とみなすことができる。つまり、セルオートマトンのルールダイナミクスを用いて、デジタル音信号を記述することが出来るのではないかと考える。

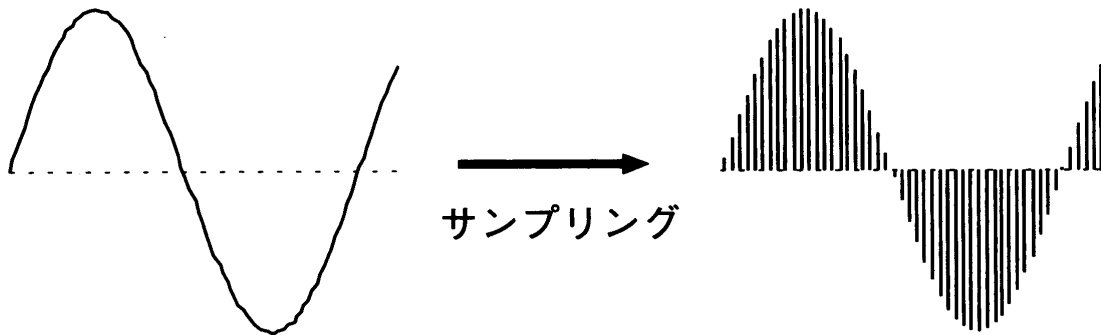


図 4: サンプリングの過程

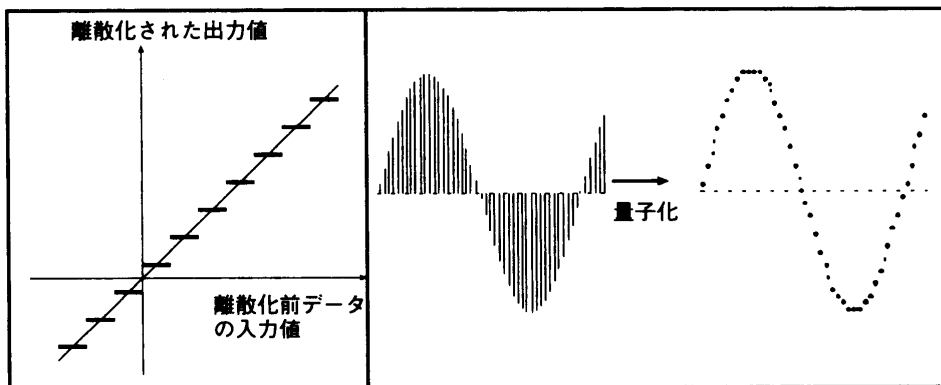


図 5: (a) 量子化器の入出力特性の例 (b) 量子化されたデータの例

4 セルオートマトンによる音信号の記述

4.1 ルールによるの音の記述方法

第 3 章では、アナログデータからデジタルデータへの変換を述べた。ここでは、デジタル化された音声・音楽信号をどうやってセルオートマトンで扱うかについて説明する。まず、デジタル化されたデータは、各ステップのデジタル化された 0, 1 の音振幅振幅バイナリデータをセルオートマトンの各セルの状態値に対応させる。つまり、図 6 のようになる。図 6 では、左上の 35000 が時刻 t での実際の原音振幅値であり、振幅値の正負を考慮して、2 進表現で表したものが各セルの上に記してある $\{0, 1\}$ の列である。つまり、 $\{0, 1\}$ の列が振幅値バイナリデータである。図中では、0 を白色、1 を黒色で示してある。ここで次の時刻 $t+1$ の音振幅値バイナリデータを適当なセルオートマトンのルール番号で与えることが可能ならば、音信号はルール番号の時間発展により記述することが可能となる。結局、音信号をセルオートマトンで記述するかは如何にしてルールダイナミクスを決定するかが重要となる。節からこのルールダイナミクスの決定方法を考察する。

第 1 章でも述べたように 5 近傍ではなく 3 近傍によるダイナミクスを用いる一番の理由は、我々は多自由度系における複雑な現象の背後にも何らかの単純なルールが存在すると考えるからである。実際 5 近傍の場合はルール数は 2^{32} の約 43 億個となる。これは 3 近傍の 2^8 の 256 個に比べ、ルールがはるかに複雑になる。また実際にルールを総当たりで探す際には、5 近傍では、膨大な量の計算を必要とする。記述のことを考えてもルール 1 つに 32bit 必要となり、原音が 1 時刻あたり

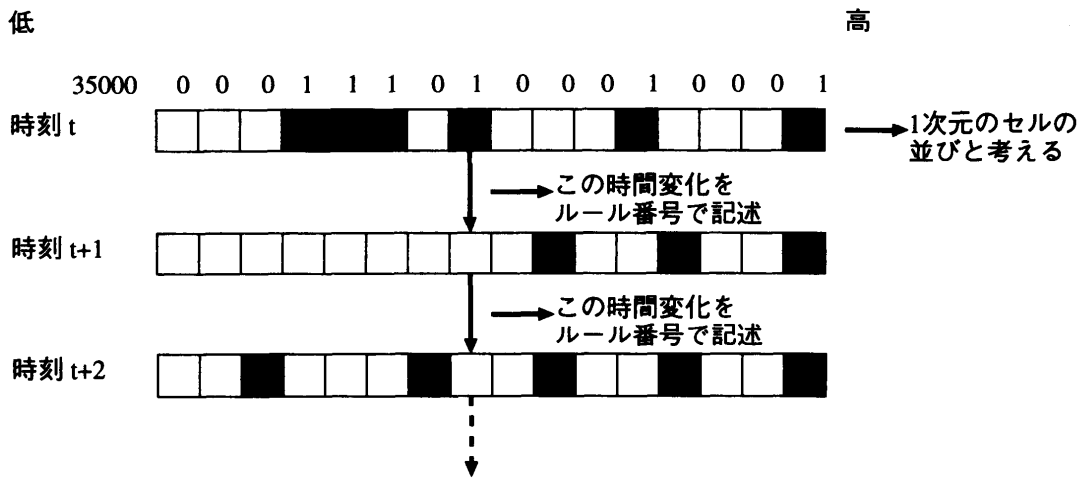


図 6: デジタル音データとセルオートマトンの対応関係

16bit で記述されているので 2 倍のデータ量となる。また装置の実現性においても 3 近傍では回路配線が交差することなく設計できるのに対し、5 近傍でははるかに複雑な回路設計を要求する。このような理由からも 3 近傍セルオートマトンを採用している。

4.2 具体的なルール抽出法

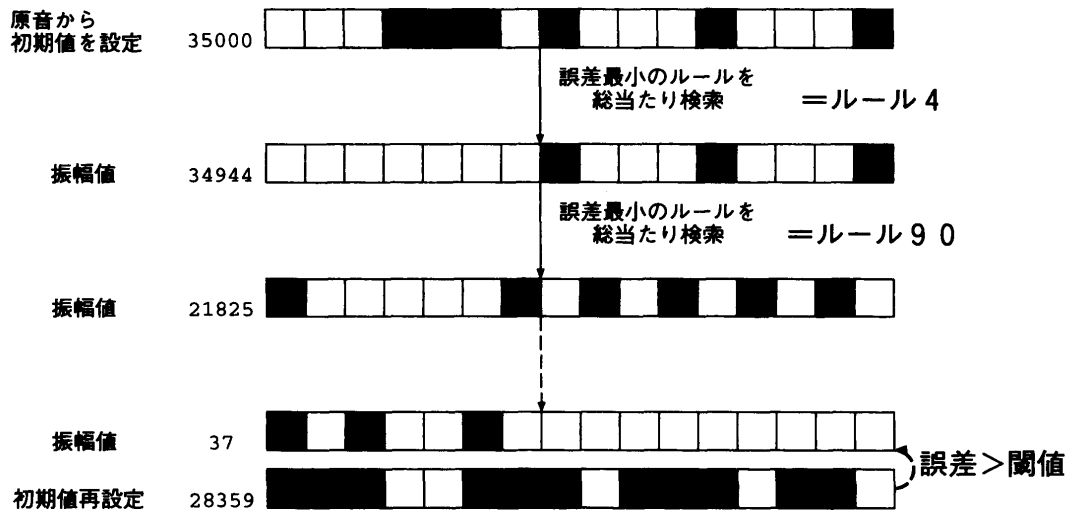


図 7: ルール抽出法 (1)

本節では、いかにして各ステップで適用するルールを抽出するかということを考える。ここでは、

いくつかの試みの中の 2 例を与える。

ルール抽出法 1 まず、1 番目の方法として、図 7 に示す方法である [9]。この方法は以下のようにある。まず、原音から初期値を設定する。そして、この初期値に対するビットパターンに対し、256 個の全ルールを 1 度適用し、ある閾値以内で原音との誤差が最小となるルールをその時刻のルールとして採用する。但し、この際の誤差は、得られたビットパターンを 10 進表現に直し、原音の振幅値との差を用いて評価される。図 7 の場合、ルール 4 が最もよいルールとして得られたとする。そして次の時刻のパターンは、ルール 4 によって得られるセルのパターンで、今と同様の手順を繰り返す。こうやって順次得られたルールによって、記述していくのである。ただし、あるステップにおいて、最もよいとして得られたルールによる記述と原音との誤差が、閾値をこえたら、そのステップは、原音から初期値を与えなおす。ここで、閾値は、振幅値が 1024 以上のところでは、原音の 5% とし、それ以下のところでは、100% とした。

ルール抽出法 2 手法 2 のおおまかな流れは、図 8 に示す。手法 1 と違う点は、手法 1 は 256 個のルールを 1 度だけ適用し次のビットパターンを求めるのに対し、手法 2 は使用するルールを 3 つのみに限定するかわりルールを適用する回数を最大 8 回までとする点である。そしてこの場合は原音の振幅値との誤差が閾値以内でかつ最低適用回数となるようなルール列を採用する。その他の点においては手法 1 と同様である。

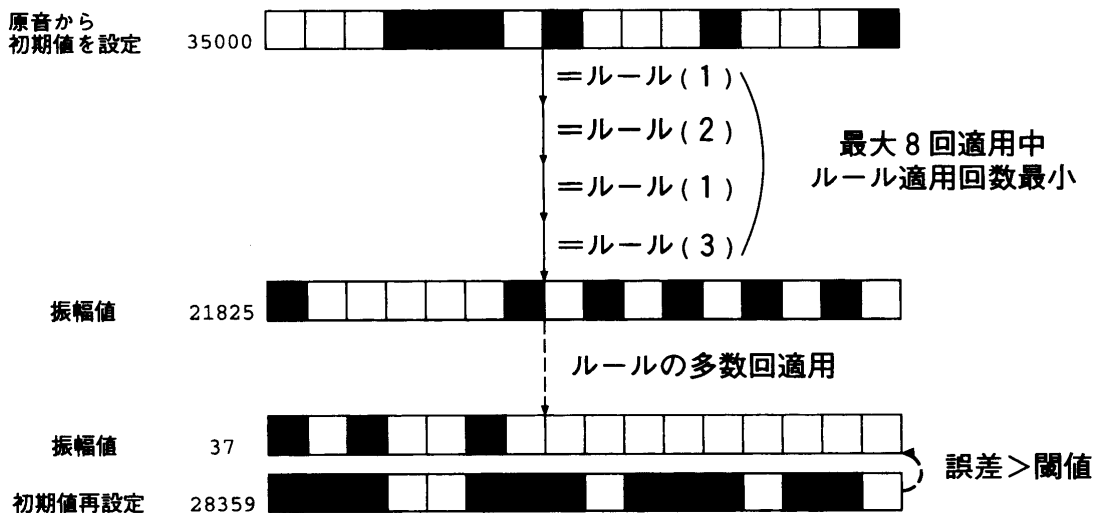


図 8: ルール抽出法 (2)

4.3 実験条件 および 記述能力

今回の実験に際して用いたデータは、音楽 CD のクラシック 4 曲と 日本のポップミュージック 1 曲を 30 サンプル採取した。但し、1 サンプルのデータ長は、1 秒である。音楽 CD は、44100Hz でサンプリング、16 ビットで量子化されている。

再現性の評価の方法として、以下の 3 種類を採用した。

表 1: 復元率. 手法 2 に用いるルールセットは各々 #1 (45, 86, 101), #2 (75, 86, 101), #3 (101, 105, 154).

	(単位: dB) 手法 2			
	手法 1	# 1	# 2	# 3
平均	21.29	17.35	16.51	17.36
最良	28.69	25.85	25.57	25.80
最悪	15.43	10.56	9.20	10.66

- 実音による聴取
- 復元率
- パワースペクトラム

但し, 復元率式は以下のように定義される.

$$\text{復元率} = 10 \log_{10} \frac{\sum_t a(t)^2}{\sum_t (a'(t) - a(t))^2} \quad (3)$$

で定義した. ここで, $a(t)$ は, 原音の時刻 t における振幅値 (10 進表示), $a'(t)$ は再現音の振幅値である.

4.3.1 実音による聴取

区間平均振幅が大きいものは, 手法 1 に比べ手法 2 はわずかながら高周波ノイズが存在するが両手法とも聴取性はよかった. 区間平均振幅が小さいものは, 手法 1, 2 とも音楽の内容はわかるが, 音質は必ずしも良くはなかった. これは, 振幅値の小さいところでは誤差を大きく許しているためであると思われる.

4.3.2 復元率

復元率の結果は表 1 に与える. ここで定義されている復元率は, 再現音が原音と比べ全て 5% の誤差を含んだ場合, 約 25 dB であり, ふつうに聞いている場合には原音との違いはほとんどわからない. また, 数値が高いほどよい再現性が良いといえる.

復元率も再現音の聴取と同じく, 音の振幅値が大きいものは再現率も高く, 小さいものは再現率も低い. 手法 1 の方がよりよい音質で再現されているのは, 各ステップで誤差最小となるものを与えるルールを採用しているからである. 一方, 手法 2 では最小組み合わせ数を採用するようにしているので, 必ずしも誤差が最小とは限らないからである. また, 手法 2 では用いるルールセットによってわずかではあるが復元率に差異が見られる. つまり, 音質向上のためにより良いルールセットを選ぶことが重要となる.

4.3.3 パワースペクトル

図 9 はパワースペクトルの結果である。低音域に関してはほとんど良い再現をしているが、高音域に関しては、少しノイズが見られる。また、この傾向は手法 2 により顕著に見られ、実音聴取結果と一致する。このような高音域にノイズが発生する原因は誤差を許す音記述を採用しているからだと考えられる。

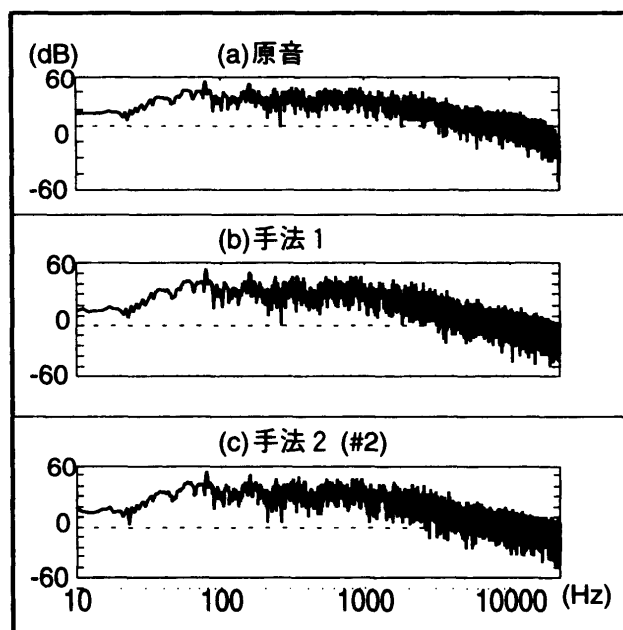


図 9: 音データ 1 秒間のパワースペクトラム。(a) 原音, (b) 手法 1, (c) 手法 2 (ルール 75, 86, 101)。

4.4 音楽データ圧縮の考察

更に圧縮率という観点から本手法の記述能力について考察する。ルールダイナミクスを用いたセルオートマトンによるデジタル記述方法では、初期値とルールを基みして音を記述する。この際、初期値 16 ビット、ルールは、手法 1 においては 8 ビット、手法 2 においてはルール 1 つにつき 2 ビットであるので、原音の音データ量を 1 とすると、表 2 のように評価される。

最初のみ初期データ 16 ビットを与え、以後はルールのみで記述できるような理想的な場合は、手法 1 においては、

$$\text{圧縮率} = \frac{16 + 8n}{16n} \simeq 0.5 \quad (\text{for large } n) \quad (4)$$

となるはずである。しかしながら実際は、ほど遠い値になっている。これは、頻繁に初期値を変更しているからである。これは我々が期待しているほど、ルールのみによる記述が、長くつづいていないことを意味する。

では、どうして長くつづかないのかを、ルールから生成されるビットパターンの分布を用いて考察する。図 10(a) は原音の時刻 t の振幅値 (10 進表示) を x 軸に、時刻 $t+1$ の振幅値を y 軸にとりプロットしたものである。図 10(b) では可能な振幅値に対し (横軸)、256 通り全てのルールを

表 2: 圧縮率 (原音を 1). 各々手法 2 のルールセットは, 表 1 と同様

	手法 1	手法 2		
		# 1	# 2	# 3
平均	0.697	0.683	0.594	0.667
最良	0.621	0.675	0.510	0.639
最悪	0.796	0.693	0.653	0.677

適用した際に生成される全ての値を縦軸にプロットしている. 図 10(c) では可能な振幅値に対し, ルールセット (75,86,101) を最大 4 回繰り返し適用して生成される全ての値を縦軸にプロットしている.

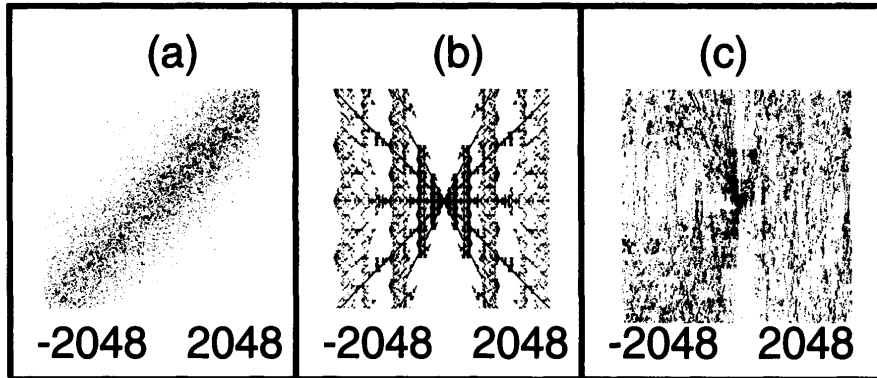


図 10: (a) 横軸に時刻 t の振幅値, 縦軸に時刻 $t+1$ の値をプロット. (b) 256 個のルールによって生成される値の分布, (c) 3 個のルール (75, 86, 101) の 4 回以内の適用 (120 通り) によって生成される値の分布 (b) (c) の横軸は初期値, 縦軸は生成される値をプロット.

実際の音データは $y = x$ の直線近傍にかたよって分布している (図 10(a)). 一方手法 1 によって生成される音データ (図 10(b)) は, 非常にきれいな構造的空白を持ち, 音データとは異なる分布である. 特に, $y = x$ の直線近傍に大きな構造的空白が存在し, このことが音記述を困難にしていると考えられる. 手法 2 によって生成される音データ (図 10(c)) は, 空白領域はわずかに見られるが生成される音データの点がより広範囲に分布している. つまり手法 2 ではひとつの初期値に対して 120 点しかないにもかかわらず, 手法 1 に比べより原音に近い点を効率良く生成しているといえる. このことが手法 1 よりも手法 2 の方が圧縮率があがっている要因であろうと考えられる.

5 まとめ

本論文では, 1 次元 2 状態 3 近傍セルオートマトンのルールダイナミクスを用いてデジタル音信号を記述方法及びデータ圧縮の可能性について研究した. 得られた結果は以下の通りである.

- セルオートマトンのルールダイナミクスを用いて音信号を記述をできることを示した.

- 最適化はまだ必要であるが、音信号の圧縮記述になっていることを示した。
- 音楽信号を記述するにあたって、全ルールを1度適用するよりも数ルールを多数回適用する方が圧縮率向上という意味で効率のいいことを示した。

また今後の課題としては、今回はルールセットを適当に3つ選び調査してみたが、組み合わせによって圧縮率に差異がみられることから、より適切なルールセットを系統的に選び出すことが考えられる。

今回提案した手法は、現在世の中で広まりつつある音楽信号の圧縮方法であるMP3などと比較すると、圧縮率の面からみるとまだまだ十分ではない。MDやMP3などといった圧縮方法は、フーリエスペクトラムを利用した非可逆圧縮となっている。しかし今回我々が本論文で提案した方法は、フーリエスペクトラムを用いず、直接音楽信号そのものをセルオートマトンのルールダイナミクスを用いて圧縮記述する手法である。本論文では誤差を含んだ記述となっているために非可逆記述となっているが、誤差を一切許すことなく記述を行なうことができるならば、セルオートマトンのルールダイナミクスを用いた音記述は、完全可逆な圧縮記述ができる可能性を大いに秘めている。

また、使用されるルールの分布及びそのルールの特徴から、音の複雑さの特徴付けの可能性をさぐることも考えられる。第2章にも述べたように3近傍セルオートマトンには、カオス状態を導くルール（ウォルフラムらの分類によるとクラス3）がある。また今回提案の記述法で使用ルールを限定してする方法に用いたルールセットの大半はクラス3のルールである。この使用ルールが音楽信号の複雑さにどのように関与しているかは、今後の研究の大きな一つのテーマである。

参考文献

- [1] 井庭 崇・福原 義久：“複雑系入門 知のフロンティアへの冒険”，NTT 出版，1998.
- [2] C. A. Sarkada and W. J. Freeman : *Bihav. Brain Sci.*, **10**p.161, 1987.
- [3] K. Aihara and G. Matsumoto : in “Chaos in Biological Systems”, ed. by H. Degn, A. V. Holden and L.F. Olsen (plenum Press. N. Y.) p.121, 1987.
- [4] H. Hayashi and S. Ishizuka : in “Chaos in Biological Systems”, ed. by H. Degn, A. V. Holden and L.F. Olsen (plenum Press. N. Y.) p.157, 1987.
- [5] Wolfram, S.; “Theory and Application of Cellular Automata”, World Scientific Singapore, 1986.
- [6] 相沢 洋二・永井 善則：物性研究 **48** No.4, 1987.
- [7] 相沢 洋二・鬼頭 政・永井 善則：物性研究 **48** No.4, 1987.
- [8] 奈良 重俊・ピーター・デイビス：“カオスと知的情報処理”，ジャストシステム，1994.
- [9] S. Nara, M. Wada, N. Abe & J. Kuroiwa : *Int. J. Bifurcation & Chaos*, **9**(6), pp.1211-1217, 1999.
- [10] 今井 秀樹：“情報理論”，昭見堂，1984.
- [11] 古井 貞熙：“音響・音声工学” 近代科学社，1992.