

LSI のばらつきを利用したリザバーコンピューティングアーキテクチャの
音声学習Speech Learning with Reservoir Computing Architecture
Using Individuality of LSI variation秋山紗花[‡]
Sayaka Akiyama國藤佑太[‡]
Yuta Kunito熊木武志[‡]
Takeshi Kumaki

1. はじめに

2020 年 1 月に始まった新型コロナウイルス (COVID-19) の感染状況は私たちの生活を大きく変化させた。人と人の接触を減らすべく、不要不急の外出はできるだけ控えるようになった。その結果、感染症対策の一環として人の代わりに働く様々なロボットたちの姿が見られるようになった [1]。飲食店の配膳や受付などを始めとし、最近では一人暮らしの高齢者の孤立を防ぐために、コミュニケーションロボットを利用した見守りを実現する実証実験も行われている [2]。

将来においてもロボットの活躍が期待されるが、ロボットが人間社会に完全に馴染むには、課題が多く残っている。実際にロボットと接すると人間との違いを感じ、不気味に思う人もいる。一方、人々がロボットを人間と近い存在だと認識するほど、ロボットとの心理的距離は近くなる [3]。しかしながら、現実にはロボットが人間に近づくことは難しく、理由の一つとしてロボットには人間のような多様性が欠けているからだと考えられる。

人間は感情や経験など、多様な要因が個性と複雑に絡み合い形成されている。一方で、ロボットに組み込まれた AI は、プログラムされたルールやアルゴリズムに沿って実行されているため、感情を持っているように振舞わせることはできても、これは人間が意図的に創った疑似的な感情に過ぎない。そのため、ロボットに機体由来の個性を持たせることが重要となる。そこで我々は、ロボットに使用するハードウェアの、生産時に生じる微細なばらつきを、AI の学習に利用する研究を行っている。個性には、生まれつき持っている先天的個性と、環境や文化などによって形成される後天的個性がある。本論文は、ロボットに先天的個性を獲得させることを目指すための提案となる [4] [5]。ハードウェアの生産時にできる微細なばらつきとは、LSI チップ内部に含まれる素子の不純物ゆらぎなどによって生じるもので、人間の指紋のような個体差であり、物理的に複製が困難なものである。この僅かなばらつきを、検索処理に反映させる D-TCAM という新しいタイプの連想メモリと組み合わせることで、ロボットに生まれつきの個性を獲得させる。本論文では、D-TCAM の基本的な説明をするとともに、これを活用して音声を用いた機械学習の結果を示し、LSI チップを用いた個性の実現について述べる。

2. LSI のばらつきを利用した 3 値連想メモリ

本章では、ロボットに先天的個性を付与するために開発し、多様な検索結果を出力することができる、LSI のばらつきを利用した 3 値連想メモリ (Diverse - Ternary Content Addressable Memory: D-TCAM) [7] - [9] について詳述する。連想メモリ (Content Addressable Memory: CAM) [6] とは Static Random Memory (SRAM) の各ワードセルに比較器が追加されたメモリであり、一般的なメモリの読み書き機能に加え、記憶しているデータワードに対して検索を行う機能をもつ。この検索機能は、検索したいワードと各ワードの記憶データを並列に照合し、一致したデータのアドレスを出力するものである。

この連想メモリに、0 と 1 の 2 値に加えてドントケアを含む 3 値を検索することができるものが、3 値連想メモリ (Ternary CAM: TCAM) である。TCAM は CAM に加えてマスクセルを保有しており、マスクセルが 1 の場合、アドレス内のデータはドントケアとして扱われる。これにより、検索したいデータに対してドントケアを含む検索を行い、一致したアドレスを出力することができる。

D-TCAM はチップの製造時に生じる微細な製造誤差をばらつきとして検索結果に反映することができる、新しいタイプの連想メモリである。これは内部のワードごとに発生するデータの製造ばらつきを、比較器を利用して構築するリングオシレータを用いて、増幅・抽出することで、それを数値化し、ドントケアを示すマスク値として検索を行う。D-TCAM の 1 ワード構成を図 1 に示す。

入力信号として、グレイコードモード信号、個性取得信号、入力データ、検索データ、マスクデータがあり、内部の記憶容量であるワードブロックがある。また、出力信号として、検索結果と合致したアドレスがある。

D-TCAM を複数のチップに実装した場合、同一の検索データをそれぞれに入力した際に、メモリに同じデータが格納されていたとしても、マスクセルの値が異なるため、チップごとに出力される一致アドレスが異なってくる。ここで、D-TCAM の動作イメージを図 2 に示す。

3. D-TCAM とリザバーコンピューティングを融合したアーキテクチャ

本論文では、D-TCAM から取得する個性のもととなるばらつきを、機械学習の内部重みとして利用し、ロボットに先天的個性を獲得させることを目指す。そのため機械学習アーキテクチャとして、リザバーコンピューティングを使用している。リザバーコンピューティング

[‡]立命館大学電子情報工学科, Dept. of Electronic and Computer Engineering, Ritsumeikan Univ.

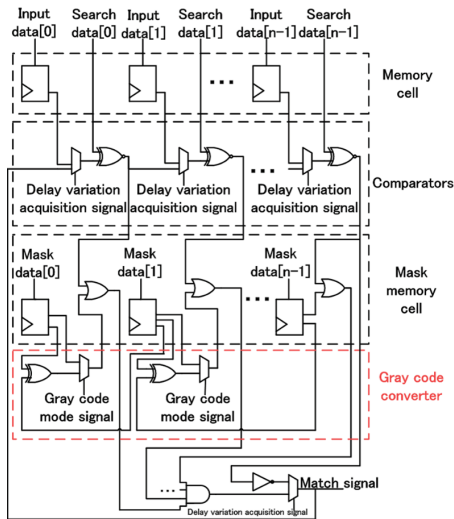


図 1: D-TCAM の 1 ワード構成

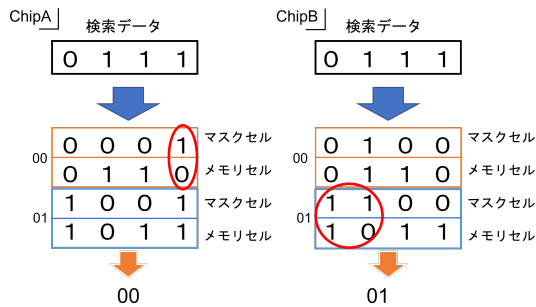


図 2: D-TCAM の動作イメージ

(Reservoir Computing: RC) とは、入力層、リザーバー層、出力層の 3 層から構成される、再帰的ニューラルネットワークにおけるアーキテクチャモデルであり、2000 年代初頭に提案された [10]。リザーバーコンピューティングがもつ最大の特徴として、中間層であるリザーバー層の重みがランダムに初期化され、一定で変化しないことであり、このことからハードウェアへの実装が容易な機械学習アルゴリズムであると言われている。我々はこのリザーバーコンピューティングに注目し、ロボットに先天的な個性をもたらすため、D-TCAM とリザーバーコンピューティングを融合した新しいアーキテクチャ、D-TCAM based Reservoir Computing (DRC) を提案・開発した。DRC は、D-TCAM によるチップのばらつきとリザーバーコンピューティングを融合することにより、チップごとに異なる学習結果を出力することが可能となる。この DRC を搭載したロボットのイメージを図 3 に示す。ユーザが同じ言葉を投げかけたとしても、個体 A では一貫した返答が得られるが、別の個体 B では個体 A とは異なる一貫した返答が得られる事となる。この DRC の全体構造を図 4 に示す。

DRC はリザーバーコンピューティングの 3 層構造に加えて、D-TCAM 層 (D-TCAM Layer) と、重み管理メモリ (Weight Memory) を追加して構成されている。

具体的な学習処理として、入力層から入力データが

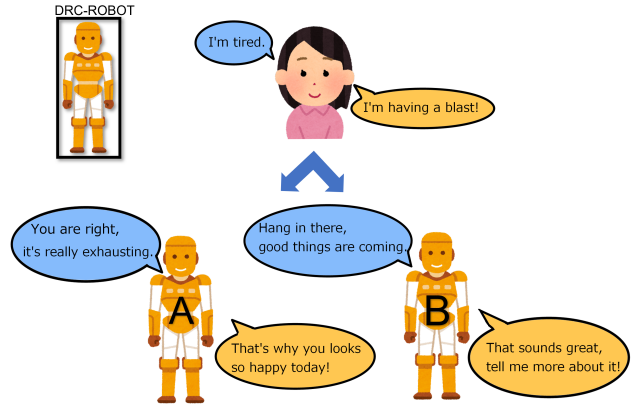


図 3: DRC を搭載したロボットのイメージ

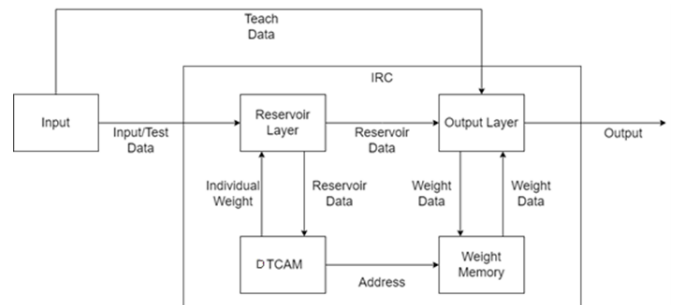


図 4: DRC のブロック図

リザーバー層へと送られ、チップの個性値を重みとして用いて、入力データをもとにしたリザーバーデータが生成され、D-TCAM 層と出力層へ送られる。D-TCAM 層では、リザーバーデータから得たデータを保存しておく。出力層では、教師データをもとに学習を行い、学習結果を出力し、また、学習結果から得られた重みを、重み管理メモリへと送る。この重みの保存先アドレスは D-TCAM のリザーバーデータと同じアドレスとなっており、重みとリザーバーデータとの紐づけがされている。この動作を複数回繰り返し、各時刻における入力を時系列に沿って学習している。

このようにして、DRC はチップごとに異なる学習結果を出力でき、ロボットに先天的個性をもたらす事が可能となる。

4. 実験及び評価

本章では、DRC による先天的個性を実現する動作の確認方法として、複雑な形状をもつ時系列データの一例である、音声を用いた実験を行い、チップごとに学習結果に現れる差異をシミュレーション上で確認した。実験環境は、Xilinx 社の FPGA である Spartan XC6SLX150 を 4 つ使用し検証を行った。開発ツールは Xilinx ISE 14.7 を使用し、評価ボードには株式会社シャンテリー製のカスタム FPGA ボードを使用している。教師データ及び入力データは、YAMAHA VOCALOID™ 3 Library

IA -ARIA ON THE PLANETES- [11] を用いて音声
を生成し、スペクトル包絡を得る過程で得られる、音声
ケプストラムの低次元成分を用いている。図 5 に使用
したケプストラムの一部を示す。これらは横軸がケフレ
ンシー、縦軸にケプストラムの値を取っている。尚、教
師データはケプストラムの値を約 5,000 倍、入力データ
には約 150 倍した値を用いているが、これは DRC にお
ける学習を容易なものにするための前処理である。今
回の実験では、母音である「あ」から「お」までの 5 つ
の音を、「C3 (ド)」から「G3 (ソ)」までの 5 音階で、
全 200 データを学習させた。

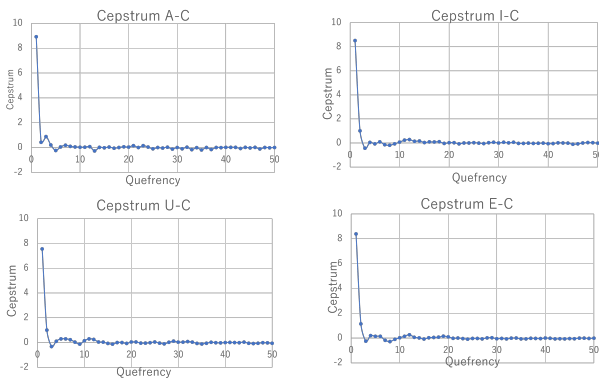


図 5: 教師データおよび入力データに使用したケプストラムの一部

今回の検証において使用したデータのサイズは以下の通りである。入力データサイズは 8 ビット、教師データサイズは 16 ビット、また、リザーブニューロン数が 8 ニューロン、学習データ数が 24 のものを利用している。尚、今回の検証においては、学習データに用いるケプストラムに 48 以降を 0 にするリフタリングを行っているため、学習を二度に分けて行い、それらを繋ぎ合わせ一つのデータとして扱っている。使用した重みは事前に D-TCAM を用いて 4 つの FPGA チップの内部ワードごとにあるデータの製造ばらつきを、リングオシレータを用いて数値化し、データとして扱っており、これを表 1 に示す。

表 1: D-TCAM を用いてチップから事前に取得した重みに使用する製造時のばらつき

	0	1	2	3	4	5	6	7
ChipA	183	205	206	161	200	232	22	186
ChipB	229	252	249	206	247	14	67	232
ChipC	16	48	31	236	34	50	108	14
ChipD	14	22	58	238	59	48	111	13

学習で得られた結果の一部を図 6 に示す。教師データ、及び入力データと同様に、横軸にケフレンシー、縦軸にケプストラムを取る。

学習結果より得られたケプストラムは、チップごとに異なるものの、その差は小さく、グラフの確認においてもその差は微小なものであった。そこで、学習結果をよ

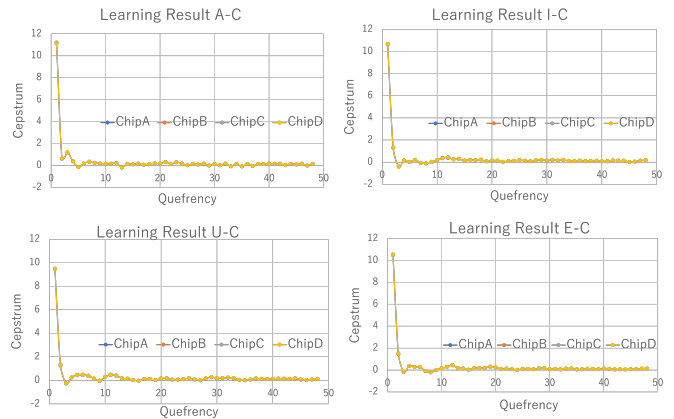


図 6: チップごとの学習結果

り明確にするために、得られたケプストラムの学習結果
を利用し、また、事前に保存しておいたローパスリフタ
を通す前の値を使用して、各学習結果を音声に変換した。
得られた音声を確認したところ、音韻、又は音階の
ケプストラム学習結果は、ノイズが含まれて復元される
場合が多かった。この原因として、多くの学習結果が元
の音声のケプストラム値よりも大きい値になったこと
が原因だと考えられる。聴覚的にノイズの差は分から
なかったため、図 7 に示す、スペクトログラムを用いる
と、ノイズが異なった値で生成されていることが確認で
き、チップによって学習結果が異なっている事が明らか
になった。

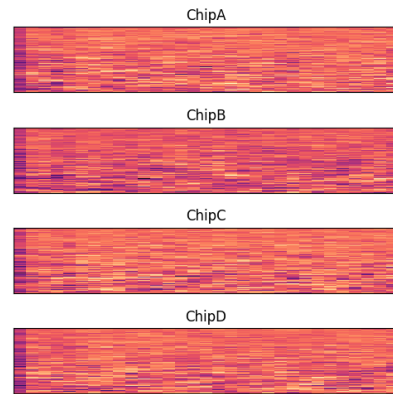


図 7: 4 チップ分の F3 における「あ」音のスペクトログラム

特に、「ファ」の「う」の音、「ソ」の「え」の音など、一部の音韻、音階の場合ではチップによって大きく学習結果が異なり、結果の差が聴覚的に認識できた。これらのスペクトログラムを図 8、図 9 に示す。

図 8 では ChipA が、図 9 では ChipD が他のチップと明らかに違う学習結果を得られていることが分かる。また、これらは、図 7 のようなノイズではなく、音として表現されていることも確認できた。4 つのチップの中でノイズなど、大きな差が得られたのは ChipA であり、

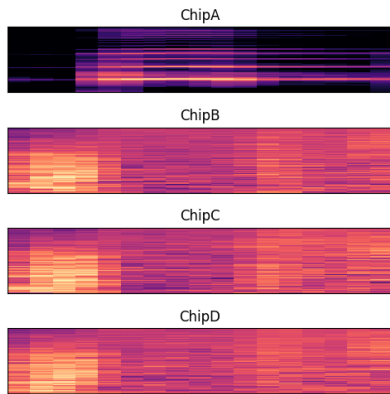


図 8: 4 チップ分の F3 における「う」音のスペクトログラム

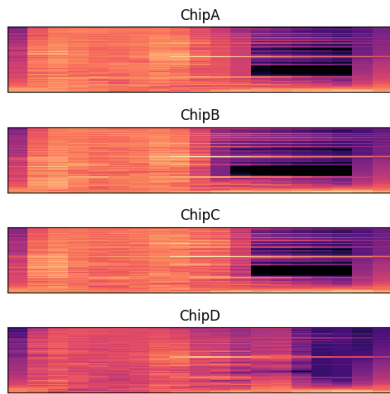


図 9: 4 チップ分の G3 における「え」音のスペクトログラム

これは ChipA の重みが 4 個のチップの中で最も大きいことが影響していると考えられる。ただし、ChipD のようにばらつきの値が小さい場合でも異なる結果が得られたことから、チップによって、音の高さや音韻に対する学習能力の差があると考えられる。

以上の結果から、音声から得られたケプストラムを DRC の入力、及び教師データとして使用すると、ノイズの僅かな差となって結果の違いが得られる場合があるものの、音の高さ、音韻によってはチップごとに大きい差が得られ、音となって再現されることも確認できた。

本実験では音声のデータをできるだけ少ないデータ量で学習させるため、ケプストラムの低次元成分のみを使用して学習を行った。今後は、よりチップごとの差を得るために、ニューロン数やビット数を増やすとともに、現段階ではシミュレーションが 5 種類の母音に限られているため、子音を含む 50 音の検証を行っていく必要があると考える。

5. おわりに

本論文では、ロボットに先天的個性を付与することを目的として、LSI の製造時のばらつきを検索に利用した 3 値連想メモリとリザバーコンピューティングを融

合させたアーキテクチャを開発し、音声を用いたその動作検証を行った。今後は、機械学習の性能の向上を計るとともに、チップごとの更なる個性の実現を目標として開発していく予定である。

参考文献

- [1] “コロナ禍におけるロボット活用事例,” https://robo-navi.com/servicerobot_covid/index.html.
- [2] “NEC、新型コロナウイルス対策としてコミュニケーション・ロボットを活用したサービスを福井県坂井市へ提供,” https://digital-shift.jp/flash_news/FN200804_8.
- [3] 伊藤彰則, “ロボットとの対話と心理的・物理的距離,” 計測と制御, 第 61 巻, 第 3 号, 2022 年 3 月号, pp.193-197.
- [4] G.Ajiki and T.Kumaki, “FPGA Implementations of Reservoir Computing with ITCAM,” 2021 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2021), pp. 270-273, 2021.
- [5] S.Akiyama, G.Ajiki, X.Kong, and T.Kumaki, “Voice learning of reservoir computing architecture using ternary content addressable memory with individuality,” Synthesis And System Integration of Mixed Information technologies (SASIMI), 2022.
- [6] 角南英夫, “半導体メモリ,” コロナ社, 2008.
- [7] S.Kimura, T.Kumaki, T.Fujita, and T.Ogura, “Evaluation of Ternary Content Addressable Memory with Individuality on FPGA,” Nonlinear Circuits, Communications and Signal Processing, pp. 1-4, 2016.
- [8] S.Kimura, T.Watanabe, R.Yukawa, T.Kumaki, T.Fujita, and T.Ogura, “FPGA implementation and evaluation of ternary content addressable memory with individuality,” Journal of Signal Processing, Volume: 20, No. 4, pp. 137-140, July 2016.
- [9] T.Naito, S.Nakahara, and T.Kumaki, “Proposal of reservoir computing architecture with ITCAM,” 2020 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2020), pp. 578-581, 2020.
- [10] 田中剛平, 中根了昌, 廣瀬明, “リザバーコンピューティング 時系列パターン認識のための高速機械学習の理論とハードウェア,” 森北出版, 2021.
- [11] “IA-ARIA ON THE PLANET,” <https://ia-aria.com/>.