

修士論文

fNIRS データの Smith Waterman 法を
利用した類似部分抽出システムの
提案と評価

同志社大学大学院 工学研究科 情報工学専攻
博士前期課程 2010 年度 748 番

西井 琢真

指導教授 三木 光範教授

2012 年 1 月 20 日

Abstract

Functional Near-Infrared Spectroscopy: fNIRS which measure the change of brain blood flow is one of brain functional imaging spectroscopic methods and has been started to utilized in practical cases. Recently fNIRS have a lot of channels and users store the long time data. This lease to make the data huge and it takes a long time and difficulties to analyze the data effectively. To solve this problem, it is useful that the similar parts are chosed automatically from multi channel time series data and these parts are showed to analysis.

In this thesis, the novel algorithm, which can extract similar subsequence of fNIRS data is proposed. In the proposed method, real values of time series data are transformed into string data. Then similar subsequence is extracted using homology search algorithms. In the transformation algorithm is introduced with considering the characteristics of fNIRS data. There are many parallel libraries for homology search is released and extraction of similar subsequence can be speeded up using these libraries. In this thesis, not only the algorithm is proposed but also the fNIRS data analysis system is implemented and described. The effectiveness of the proposed algorithm is discussed through the experiment. At the same time, using the proposed system, actual data of brain function is analyzed and the usefulness of the system was illustrated.

目次

1	序論	1
2	fNIRS とデータの特徴	2
2.1	fNIRS とは	2
2.2	実験機器と計測部位	2
2.3	問題点	4
2.4	関連研究	4
3	相同性検索を用いた 2 つの時系列データの類似部分の抽出方法	6
3.1	提案手法の概要	6
3.2	時系列データの再量子化	7
3.3	fNIRS の出力データに適した前処理手法	9
3.4	相同性検索と SW(Smith Waterman) 法	11
3.5	SW 法の並列化	13
4	提案手法の検証	16
4.1	提案手法と DTW による類似部分の比較	16
4.2	時系列クラスタリングデータセットを用いた数値実験	20
4.3	fNIRS の出力データに適した前処理手法の適用	23
5	解析システムの実装	27
5.1	システムの概要	27
5.2	システムの機能	27
5.3	構築したシステムによる fNIRS データの解析	27
6	結論	33

1 序論

近年、脳血流の変化を測定することにより非侵襲的に脳機能イメージングを行う装置である fNIRS (functional Near-Infrared Spectroscopy) や fMRI(functional Magnetic Resonance Imaging) が注目を集めている¹⁾。これらの装置は、脳機能の解明に役立ち、種々の病理の判定や生体信号によるコンピュータ操作などに利用されている²⁾。fNIRS は、脳の計測部位 (チャンネル) ごとの脳血流の増減を測定することで脳の活性度を把握する装置である。例えば、“テレビを観賞している時”や“歌っている時”に活性化する脳の部位を調べることができる。また、“楽しく音楽を聴いている時”や“楽しく食事している時”の実験データを解析すれば、“楽しい”という感情で共通して活性化する部位を調べることができる。これらの装置の性能は向上しつつあり、出力される時系列データ量が増大している。そのため、解析者が効率的にデータを解析できないという問題が生じている。効率的にデータを処理するための課題はいくつか存在するが、その1つに解析者が実験結果を検討する際、出力された多数の時系列データの中で、どの時系列データに着目するかが解析者に依存してしまうという問題がある。この問題を解決するためには、多チャンネルの時系列データ間における類似部分を探索し、特徴的な部位を探索することが有効である。

そこで本論文では、相同性検索と時系列データの再量子化を用いて2つの時系列データから類似部分を抽出する方法を提案し、多チャンネルの時系列データ間における類似部分の探索を行った。提案手法は、時系列データを再量子化し、相同性検索を用いて、2つの時系列データから類似部分を抽出する。相同性検索は、バイオインフォマティクスの分野で並列アルゴリズム³⁾の開発が盛んに行われており、高速処理が可能である。fNIRS の多チャンネルの時系列データは大規模であり、その計算コストが問題となるが、並列アルゴリズムを用いることで解決できると考えられる。しかしながら、相同性検索は文字列検索用のアルゴリズムであるため、実数値の時系列データにはそのままでは適用できなかった。そこで、本論文では実数値の時系列データを文字列データに置き換える事を考え、そのための再量子化手法を検討した。また、fNIRS の出力データに適した前処理手法を考え、より解析者の感性に近い類似部分の抽出を試みた。これまでに時系列データの再量子化を2本以上の時系列データに対して適用し、類似部分を抽出する文字列検索アルゴリズムを利用した例はなく、提案手法には非常に新規性がある。また、数値実験により提案手法と DTW(Dynamic Time Warping)⁴⁾⁵⁾ による類似部分の比較、再量子化手法の性質の違いの検討を行った。

次に、提案手法を実装し fNIRS 出力データの解析システムを作成した。この解析システムを用いることで、解析者は簡単に多チャンネルの時系列データ間における類似部分を探索し、特徴的な部位を探索することが可能となり、検討時にかかる負担が軽減する。本論文では、解析システムを用いて実際の fNIRS の出力データを解析し、その有効性を示している。

本論文の構成を以下に述べる。まず第2章で fNIRS とその問題点について述べる。第3章では提案手法である相同性検索を用いた2つの時系列データの類似部分の抽出方法について述べる。第4章で数値実験を行い、第5章で fNIRS 出力データ解析システムを実装し、そのシステムを用いた fNIRS 出力データの解析について述べる。最後に第6章で結論を述べる。

2 fNIRS とデータの特徴

2.1 fNIRS とは

fNIRS(functional Near-Infrared Spectroscopy) は、近赤外線分光法を用いた非侵襲の脳機能イメージング装置の一種である。近赤外光を用いて脳血流の相対的変化量を多点で計測する。大脳皮質には脳機能の局在性があること⁶⁾、神経活動による酸素代謝の亢進に伴い脳血管が拡張し脳血流が上昇する現象⁷⁾(神経血管カップリング)が確認されていることにより、fNIRSを用いた脳機能の解析が可能となる。以下にfMRI(functional Magnetic Resonance Imaging)と比較した際のfNIRSの長所と短所を示す⁷⁾。

- 長所

- 非侵襲性

近赤外光を用いた測定であるため生体への有害な影響はなく、幼児に対しても測定が可能である。

- 高時間分解能

0.1秒毎の測定が可能であり、脳機能の時間経過による変化を検討しやすい。

- 実用性

装置が小型で移動可能であり、様々な場所で測定が可能である。また、fMRIは臥位で無動を保たなければならないのに対し、座位や立位など自然な姿勢で測定が可能である。

- 短所

- 低空間分解能

空間分解能は10-30mm程度の大きさである。そのため、脳の部位と機能の対応関係をそれ以上の精度で決定することは困難である。

- データの不確実性

fNIRSから得られるデータは血流の相対的変化量である。変化量の絶対量を把握することはできない。さらに、近赤外光の光路長が不明確であるため⁷⁾、被験者間でのデータを単純に直接比較できない。

2.2 実験機器と計測部位

本論文では脳血流の計測に日立メディコ社製のfNIRSであるETG-7100を用いた。実験機器ETG-7100をFig.2.2に示す。本装置は約120箇所の脳部位を計測する。各計測部位はチャンネル(CH)と呼ばれ、0.1秒ごとに血流の相対的変化量が出力される。本装置では、脳を前頭部、右側頭部、左側頭部、後頭部、頭頂部の5つの部位(プローブ)に分けて扱う。前頭部は24CH、右側頭部は22CH、左側頭部は22CH、後頭部は24CH、頭頂部は22CHが計測される。前頭部、右側頭部、左側頭部のプローブの設置とチャンネル(以下、CH)の関係をFig.2.1, Fig.2.3, Fig.2.4に示す。左側頭部にお

ける出力データの一例を Fig.2.5 に示す. Fig.2.5 において横軸は時間 [100ms], 縦軸は血流の相対的変化量 [mM・mm] を示す.

なおプローブの設置は再現性を保つために国際 10-20 法⁸⁾ の Fpz, T4, T3 を参照点として用いる. 具体的には, 前頭葉のプローブにおける 2CH-3CH の間を Fpz, 右側頭葉のプローブにおける 2CH を T4, 左側頭葉における 2CH を T3 とした.

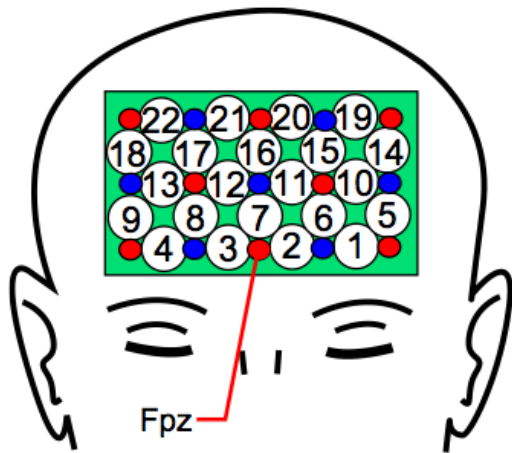


Fig. 2.1 前頭部のプローブ設置位置



Fig. 2.2 ETG-7100

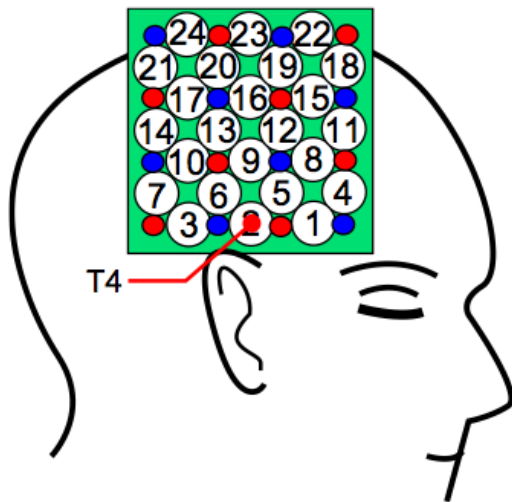


Fig. 2.3 右側頭部のプローブ設置位置

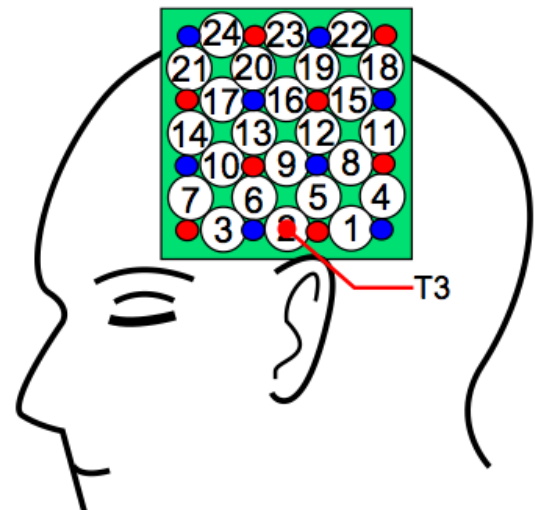


Fig. 2.4 左側頭部のプローブ設置位置

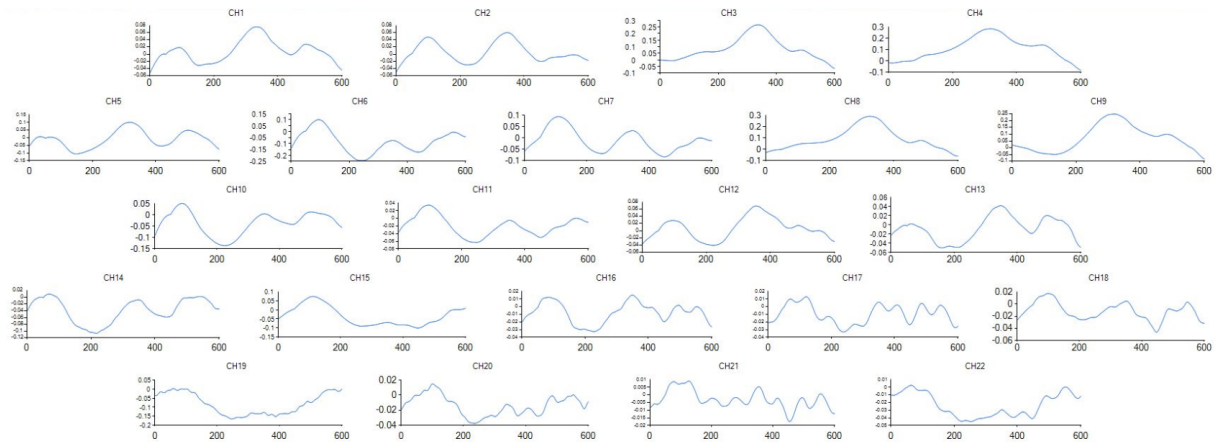


Fig. 2.5 左側頭部における出力データの一例

2.3 問題点

fNIRS 装置の出力する時系列データ量の増大により、解析者が効率的にデータを解析できないという問題がある。現在は、得られた多数の時系列データのうち一部しか有効に活用されておらず、特に実験結果を検討する際、出力された多数の時系列データの中でどの時系列データに着目するかが解析者に依存してしまうという問題がある。現在は、解析者自身が任意で注目した脳部位のデータに対してのみデータ処理を行っており、注目している脳部位以外に重要な要素がある場合にそれを見落としてしまう可能性がある。

また、fNIRS は得られるデータが血流の相対的变化量である。変化量の絶対量は把握することができない。さらに、近赤外光の光路長が不明確であるため⁷⁾、被験者間でのデータを単純に直接比較できない。そのため、複数の被験者間での加算平均や統計的手法を用いたデータの定量的な解析だけでは不十分であり、複数の被験者間によく現れたパターンなどを用いてデータを定性的にも解析する必要がある。定性的に解析をするためには特定のパターンを探索し、それを可視化することが必要である。特定のパターンを探索するためには、多チャンネルの時系列データ間における類似部分を探索し、特徴的な部位を探索することが有効であると考えられる。この探索の詳細を提案手法として第3章に後述する。

2.4 関連研究

本論文と関連する研究として、fNIRS などの脳機能イメージング装置のソフトウェア、2つの時系列データからの類似部分の探索手法が挙げられる。

脳機能イメージング装置の解析ソフトウェアとしては、日立メディコ社製の ETG-7100 における Wave Analysis ソフトウェア⁹⁾ やスペクトラテック社製の Spectratech OEG-16¹⁰⁾ の解析ソフトウェアなど装置に付属するソフトウェアが挙げられる。これらのソフトウェアを用いることで生データのグラフ化や簡単な処理は可能である。また、Matlab を用いた解析ソフトウェア「NIRS-SPM」¹¹⁾ や Source Signal Imaging 社製の脳波解析プログラム「EMSE」¹²⁾ を用いることでより高度な波形処理や脳の 3D モデリングが可能である。しかしながら、このような解析ソフトウェアの中でユーザが解析すべきデータのどこに着目すべきかを提示するためのソフトウェアはない。

また、2つの時系列データの類似部分を求める手法としては、時系列探索法が有名である。時系列探索法は、音声や映像の時系列データを扱う分野においてよく用いられる、ある音や映像の信号が長大な時系列データ内のどこに存在するかを探索する手法である。時系列探索法には、一致検索を目的とした時系列アクティブ探索法 (Time-Series Active Search:TAS)¹³⁾ や、時間伸縮を許容する検索を目的としたDTW(Dynamic Time Warping)⁴⁾⁵⁾ がある。これらの手法を効率的に繰り返す手法であるRIFTAS(Reference Interval-Free Time-Series Active Search)¹⁴⁾、RIFCDP(Reference Interval-Free Continuous DP)¹⁵⁾ やRDDs-n法¹⁶⁾ を用いて、2つの時系列データの類似部分を求めることができる。しかし、これらの手法はウィンドウサイズを決め、部分時系列データを作成し、探索を繰り返す必要があるため、時系列データ数の増加に伴い処理時間が爆発的に増加するという問題がある。計算回数の軽減が試みられている¹⁷⁾ が、計算量は膨大でありfNIRS出力データへの適用には向いていないと考えられる。

3 相同性検索を用いた2つの時系列データの類似部分の抽出方法

3.1 提案手法の概要

本論文では時系列データの再量子化と相同性検索の組み合わせによる、2つの時系列データからの類似部分の抽出手法を提案する^{18) 19)}。Fig.3.1に提案手法の流れを示す。まず、実数値の時系列データを文字列データに変換する。本論文ではこれを再量子化と呼ぶ。例えば、Fig.3.1の上部の時系列データは“BAABCCB”に、下部の時系列データは“ABCCBAA”に変換される。再量子化の手法としては、SAX(Symbolic Aggregation approXimation)や等間隔領域分割がある²⁰⁾。時系列データを文字列に変換することによって、相同性検索を適用することが可能となる。この手法により“BAABCCB”と“ABCCBAA”から類似部分として“ABCCB”を取り出すことができる。このように時系列データの再量子化と相同性検索による部分文字列の抽出によって時系列データの類似部分の抽出が可能になる。本論文では、提案手法を用いて抽出した部分を類似部分として定義した。fNIRSの出力データに対して提案手法を適用した例をFig.3.2(a), Fig.3.3(b)に示す。

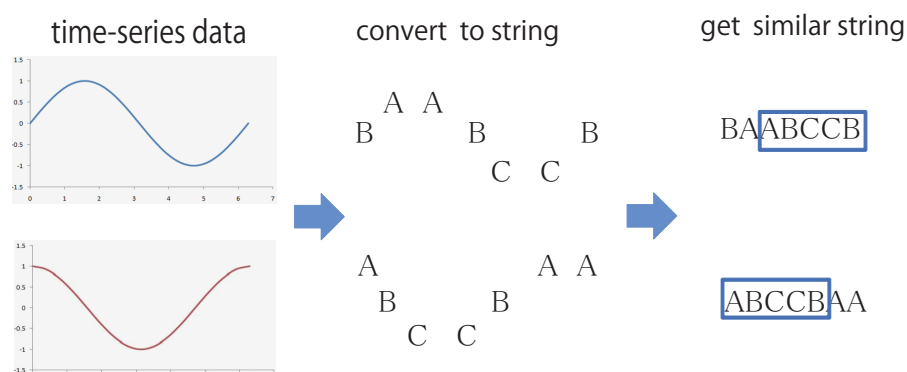


Fig. 3.1 提案手法の流れ

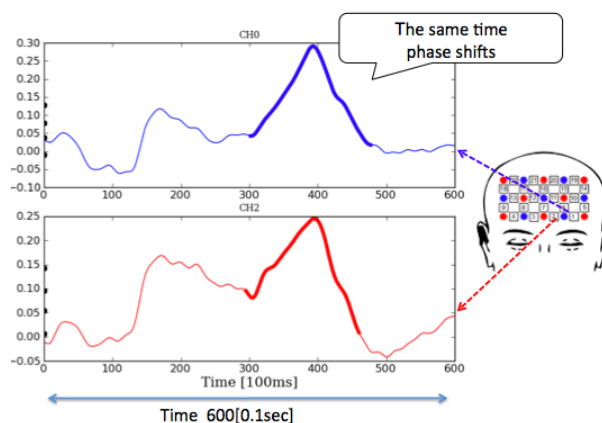


Fig. 3.2 fNIRSの出力データに対して提案手法を適用した例 (類似部分の位相が同じ)

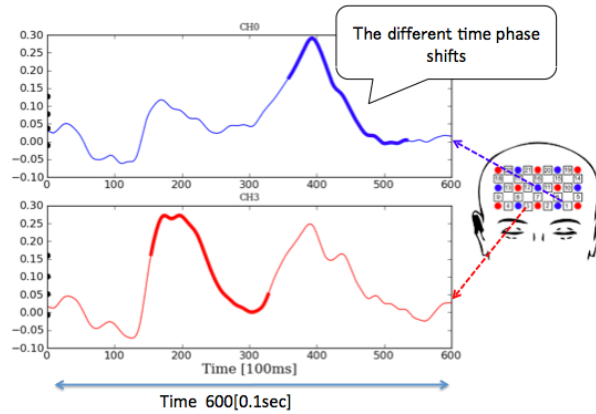


Fig. 3.3 fNIRS の出力データに対して提案手法を適用した例 (類似部分の位相が異なる)

3.2 時系列データの再量子化

時系列データを再量子化するためには、時系列データの数値と文字の対応関係が必要である。対応関係を定めるためには分割線が必要である。例えば、時系列データを3つの文字に再量子化することを考える。分割線を $\{-0.6, 0.6\}$ に設定すると、時系列データ $T = \{0.5, 1.5, -0.8\}$ は“BAC”に再量子化される。再量子化の手法であるSAXと等間隔領域分割は、分割線の定め方が異なっている。再量子化では、1つの波形を何種類の文字で表現するか定める必要がある。ある時系列データの分割数は任意に定めることができる。なお、両手法における外れ値の影響については4章で後述する。

3.2.1 SAX(Symbolic Aggregation approXimation)

SAXは、Eamonn Keoghらによって提案された時系列データの表現手法である²¹⁾。この手法は時系列データが正規分布することを仮定し、データを文字列に変換する。SAXは波形が正規分布することを仮定しているため、時系列データを文字列に変換する前に時系列データの標準化が必要となる。標準化とは、平均が0、分散が1となるようにデータを変換することである。例えば、時系列データ

$$X(t) = \{x(1), \dots, x(M)\}$$

を標準化すると

$$X(t) = \left\{ \frac{x(1) - \mu}{\sigma}, \dots, \frac{x(M) - \mu}{\sigma} \right\} (\mu: X \text{ の平均値}, \sigma: \text{標準偏差})$$

となる。SAXによる文字列変換の流れをFig.3.4に示す。step1によって標準化がなされ、step2によってデータが文字列に変換される。これにより標準化された時系列データを等領域に分割する分割線を定めることができる。これらの分割線は標準正規分布表に基づいている。分割線を $B = (\beta_1, \dots, \beta_n)$ 、分割記号数を α としてTable.3.1に示す。Fig.3.4において、下の分割線より下にあるデータは“c”に変換され、上の分割線と下の分割線の間にあるデータは“b”に、上の分割線より上にあるデータは“a”に変換される。結果的に時系列データは“aaabcc”に変換される。

Table 3.1 分割記号数が2から7のときの正規分布を等領域に分割する分割線

α	2	3	4	5	6	7	8
β_1	0.00	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15
β_2		0.43	0.00	-0.25	-0.43	-0.57	-0.67
β_3			0.67	0.25	0.00	-0.18	-0.32
β_4				0.84	0.43	0.18	0.00
β_5					0.97	0.57	0.32
β_6						1.07	0.67
β_7							1.15

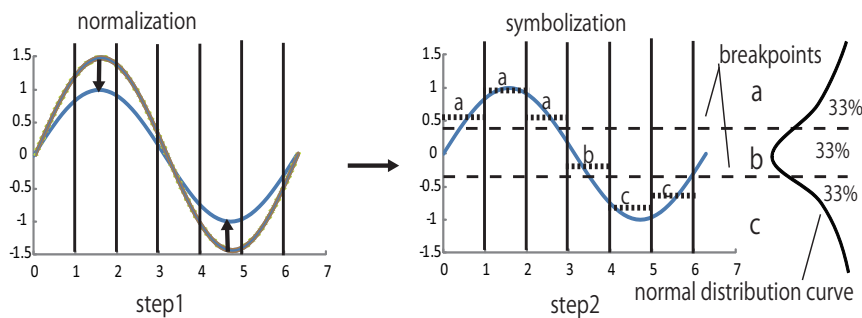


Fig. 3.4 SAXによる文字列変換の流れ

3.2.2 等間隔領域分割

等間隔領域分割は、時系列データ中の最大値と最小値の間を等分割する線を定める手法である。等間隔領域分割による文字列変換の流れを Fig.3.5 に示す。例えば、波形を3つの文字に変換する場合、波形の最小値と最大値を等分割する2つの分割線を定める。最大値 = 1, 最小値 = -1 であれば、領域幅は 0.67 となり、2つの分割線は (0.33, -0.33) となる。

例えば、時系列データ

$$X(t) = \{x(1), x(2), x(3), \dots, x(M)\}$$

の領域幅は (3.1), 分割線は (3.2) のように決定できる。X は時系列データ, w は領域の間隔, num は分割文字数を表す。

$$w = | \max(X) - \min(X) | / \text{num} \quad (3.1)$$

$$\text{breakpoints}(n) = \max(X) - w * \text{num} \quad (1 \leq n \leq \text{num} - 1) \quad (3.2)$$

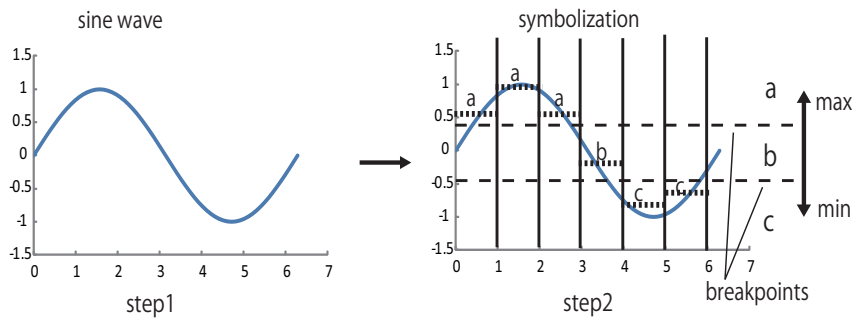


Fig. 3.5 等間隔領域分割による文字列変換の流れ

3.3 fNIRS の出力データに適した前処理手法

3.3.1 概要

ここでは、fNIRS の出力データに適した前処理手法を考える。Fig.3.6 に fNIRS から得られたある時系列データを 4 つ示す。Fig.3.6 は各図とも 3 つの山（3 つの極大値と 2 つの極小値）を持つ多峰性の時系列データである。これは被験者の計測部位が実験中に大きく 3 回活性化したことを意味している。fNIRS の出力データでは、同一被験者内で計測部位により血流の増減度合いが異なる反応を示す場合や、複数の被験者間の同一計測部位で山の大きさの度合いが多少異なる反応を示す場合がある。被験者の生理的变化により、あるタスクに反応を示すが前半タスクと後半タスクでは山の増減の大きさが異なる場合があり、反応を示している場合はこれを類似するものとして定義する。よって大きさの度合いが異なる山の頂点を引き伸ばし、山を最大値に向けて伸長、最小値に向けて伸長（以下、山の最大化、山の最小化）することで適切な類似部分を得ることができると考えられる。

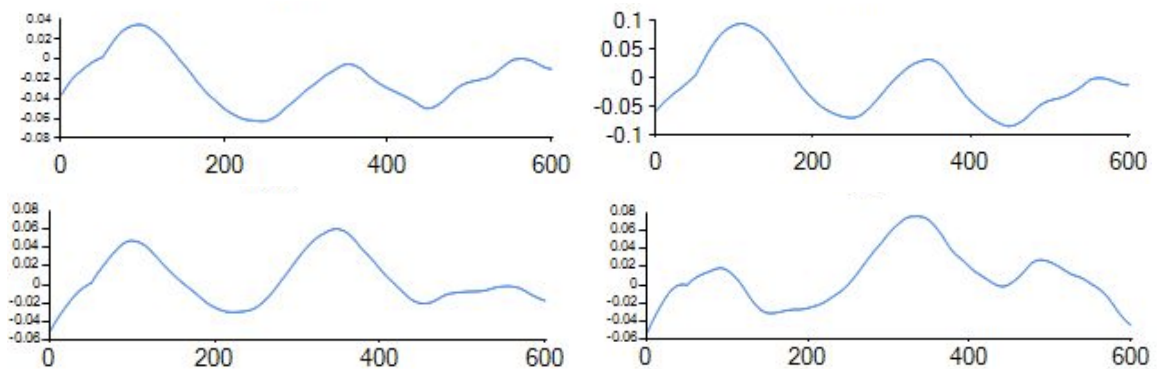


Fig. 3.6 大きく 3 つの山がある fNIRS の時系列データ

3.3.2 方法

山の最大化、最小化のアルゴリズムの流れを以下に示す。Fig.3.7 に山の最大化、最小化の流れを示す。Fig.3.7 は 3 つの極大値を最大値に向けて伸長、2 つの極小値を最小値に向けて伸長している。これにより山の最大化、最小化が可能となる。

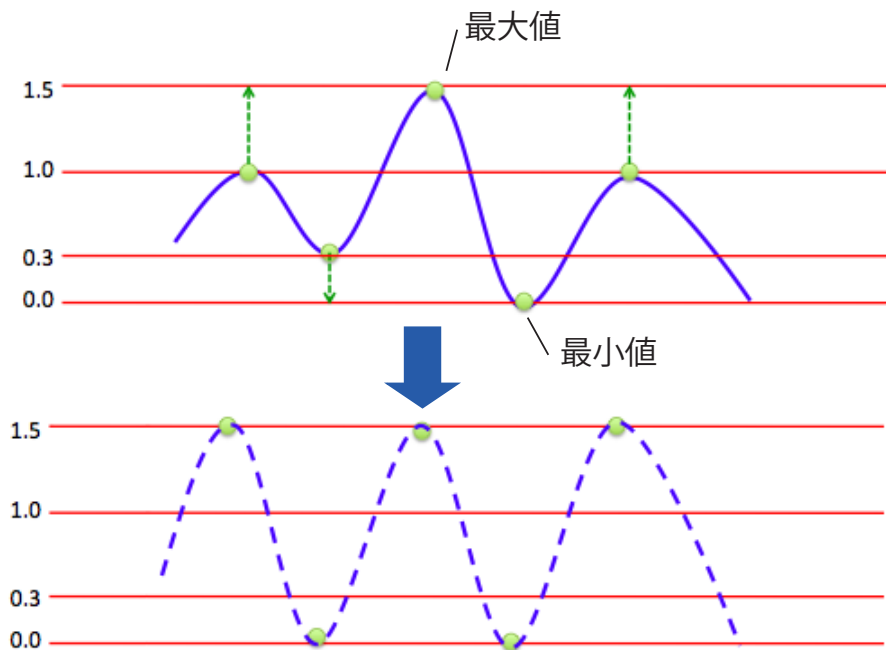


Fig. 3.7 山の最大化, 最小化の流れ

Step1: 時系列データを移動平均法を用いて平滑化する

Step2: 時系列データを微分（前方差分）する

Step3: 差の正負から極値を判断する

Step4: 離散値による誤差を隣の極値との距離差, 大小の差を考慮して修正する.

Step5: 局所解を決定する

Step6: 時系列データを最大值, 最小値に向けて伸長する.



Fig. 3.8 SW法で抽出された類似部分文字列の例

x\y	-	C	B	C
-				
B				
B				
C				

Fig. 3.9 文字列テーブルの例

3.4 相同性検索とSW(Smith Waterman)法

相同性検索は、バイオインフォマティクスの分野で広く用いられている文字列検索アルゴリズムであり、DNA や塩基配列の類似度測定や類似部分の抽出が可能である。例えば、ハツカネズミの未知の遺伝子を発見した際に、ヒトがその配列と類似した遺伝子を持つかどうかを調べる場合などに用いられる。

相同性検索には、精度を重視するSW(Smith Waterman)法^{22) 23)}、速度を重視するFASTA(FAST-All)²⁴⁾、BLAST(Basic Local Alignment Search Tool)²⁵⁾がある。本論文では、より精度の高い類似部分を抽出したいためSW法を選択した。SW法はアルゴリズムの並列性が高く、GPUを用いた高速実行のための様々な実装が試みられている^{26) 27) 28) 3)}。

SW法は動的計画法の1種であり、全ての部分文字列の比較を行うことで類似度を最適化する。

類似度は文字列テーブルのスコアによって評価される。Fig.3.8にSW法で抽出された類似部分文字列の例を示す。Fig.3.9に文字列テーブルの例を示す。文字列テーブルでは文字列Xのそれぞれの文字が行に、文字列Yのそれぞれの文字が列に割り当てられる。長さが m と n の文字列から類似部分を抽出する場合、アルゴリズムのオーダーは $O(mn)$ である。

3.4.1 スコアパラメータ

SW法には $match$, $mismatch$, gap の3つのパラメータがある。 $match$ は文字列の一致に、 $mismatch$ は文字列の不一致に、 gap はスペース発生に関わるパラメータである。これらのパラメータが変化すれば抽出される文字列も変化する。 $mismatch$ が $match$ より低ければ、類似部分の長さは短くなるがその分、一致度の高い文字列が抽出される。また、類似部分の比較を行う際、スペースが入ることによって類似度が高くなる2つの文字列も存在する。 gap は0に近いほど、スペースの多く入った類似部分が抽出される。 gap によるスペースの発生は、時系列データにおける時間的な伸縮を意味することになる。どのパラメータが最適であるかは、元データやどのような類似文字列を抽出するかによって異なる。ここではパラメータを $match = 1$, $mismatch = -1$, $gap = -1$ として説明する。

3.4.2 SW 法のアプローチ

SW 法のアプローチの流れを以下の Step1-Step4 に示す。Fig.3.10(a)-Fig.3.10(d) は、SW 法の探索過程を、式 (3.3)、式 (3.4) はスコアの計算式を表している。SW(x, y) はそれぞれの文字列の x, y 番目のセルにおけるスコアを表す。それぞれの文字列の x, y 番目の文字が一致すれば式 (3.3) が、不一致であれば式 (3.4) が適用される。文字列 X="BBC", Y="CBC" とする。具体的に、文字列 "BBC" と "CBC" の類似部分を SW 法により求めることを考える。

Step1: 文字列テーブルを作成し、それぞれの文字列を列と行に割り当て 0 で初期化する。(Fig.3.10(a))

Step2: それぞれのセルにおけるスコアを文字の一致や不一致及び式 (3.3), (3.4) に基づき計算する。(Fig.3.10(b))

Step3: テーブルの終了までスコアを計算する。(Fig.3.10(c))

Step4: 最も高いスコアを持つセルからスコアが 0 のセルまで経路をたどることにより、文字列を取り出す。(トレースバック)(Fig.3.10(d))

$$SW(y, x) = \max \begin{cases} SW(y-1, x-1) + match \\ SW(y-1, x) + gap \\ SW(y, x-1) + gap \\ 0 \end{cases} \quad (3.3)$$

$$SW(y, x) = \max \begin{cases} SW(y-1, x-1) + mismatch \\ SW(y-1, x) + gap \\ SW(y, x-1) + gap \\ 0 \end{cases} \quad (3.4)$$

例えば、Fig.3.10(a) において最初に計算されるセル (1, 1) のスコアは、セルの上の文字が "C", 左の文字が "B" と不一致なので、式 (3.4) が適用され、式 (3.5) となる。

$$SW(1, 1) = \max\{-1, -1, -1, 0\} = 0 \quad (3.5)$$

また、セル (1, 2) においてスコアは、セルの上の文字が "B", 左の文字が "B" と一致なので、式 (3.3) が適用され、式 (3.6) となる。

$$SW(1, 2) = \max\{1, -1, -1, 0\} = 1 \quad (3.6)$$

途中でスコアがマイナスになった場合は、そのセルのスコアを 0 にする。Fig.3.10(c) の状態から類似部分を得るために、"最大のスコア" のセルからスコアが 0 のセルまでトレースバックを行う。そのため、スコアの計算時にそれぞれのセルに対してどのセルから辿ってきたか目印をつける必要がある。仮に、左セル、左上セル、上セルでスコアが重なっていれば、任意で優先順位を定める必要がある。

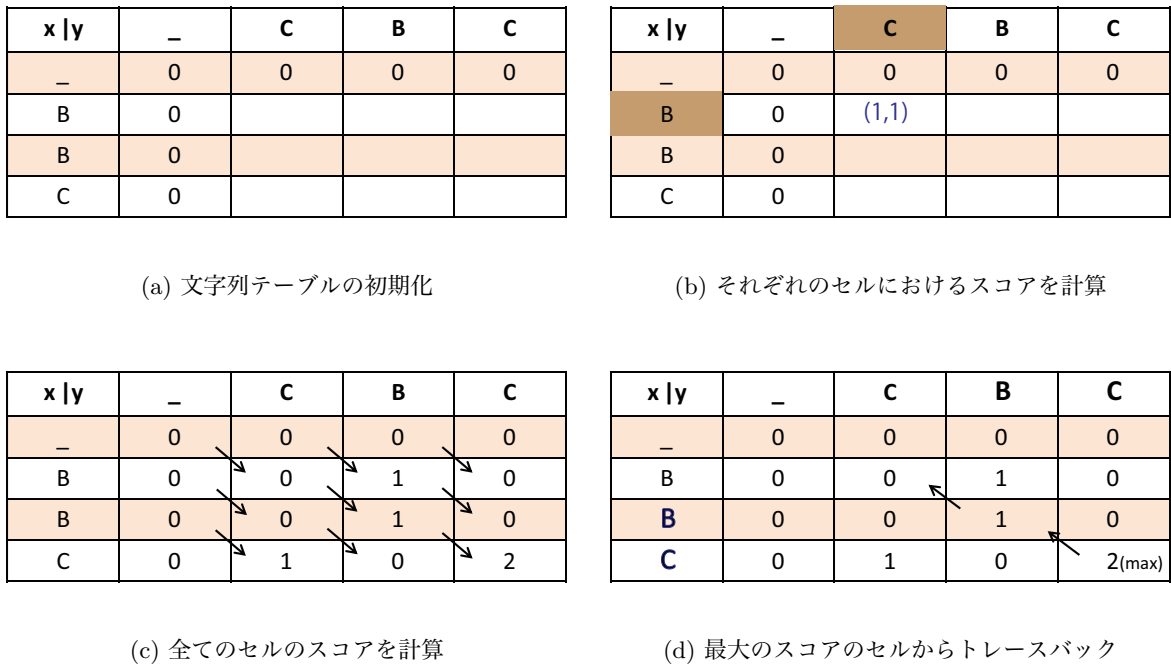


Fig. 3.10 文字列“BBC”と“CBC”の類似部分をSW法により探索した過程

る。左上セル，左セル，上セルの順に矢印をつけることとする。Fig.3.10(d)において，最大値が(3, 3)に当たるのでトレースバックは(3, 3)，(2, 2)，(1, 1)という経路をたどる。0に辿りつけばトレースバックは終了し，辿ってきたセルの上と左の文字から類似部分を抽出する。(3, 3)の上の文字は“C”，左の文字は“C”であり，(2, 2)の上の文字は“B”，左の文字は“B”である。これにより“BBC”から“BC”が“CBC”から“BC”の部分文字列が抽出される。

3.5 SW法の並列化

3.5.1 概要

提案手法において特に処理に時間がかかるのはSW法の部分であると予想される。そのため，インターフェースを実現するための処理時間を計測し，複数のスレッドを利用して高速化することを検討する。マルチスレッドのSW法を実行するためには，スレッド数と部分ブロックのサイズをパラメータとして指定する必要がある。パラメータの設定によっては高速化がうまくいかない場合も考えられるため，事前に最適なパラメータを把握する必要があった。

3.5.2 並列環境におけるパラメータの設定

2つのパラメータと文字列の長さを256, 512, 1024, 2048, 4096と変化させSW法の処理時間を計測した。Fig.3.12に1024文字の場合を示す。パラメータによって処理速度が変化することがわかる。それぞれの文字列サイズにおける最適パラメータをTable.3.2に示す。しかし，この最適パラメータはマシン環境により変化することが考えられる。Table3.2の最適パラメータを用いて文字列の長さを256, 512, 1024, 2048, 4096と変化させ，SW法の処理時間を計測した結果をFig.3.11に示す。使

Table 3.2 それぞれの文字列サイズにおける最適パラメータ

Table size	The number of threads	Sub-block size
256	4	32
512	4	32
1024	4	128
2048	8	32
4096	8	128

Table 3.3 動作環境

OS	Ubuntu10.10 64bit
CPU	Intel Core i7-2600 4cores 3.40GHz
RAM	8GB

用したマシンの環境を Table.3.3 に示す。fNIRS のサンプリングレートは 0.1 秒であるため、1 秒あたり 10 文字となる計算である。なお、文字列の内容によって処理時間が変化することはない。

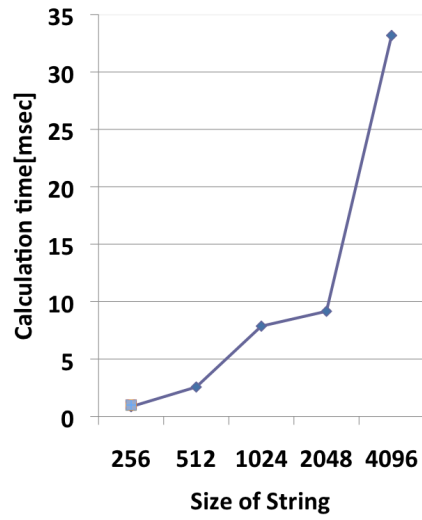


Fig. 3.11 各文字列における最適な処理時間

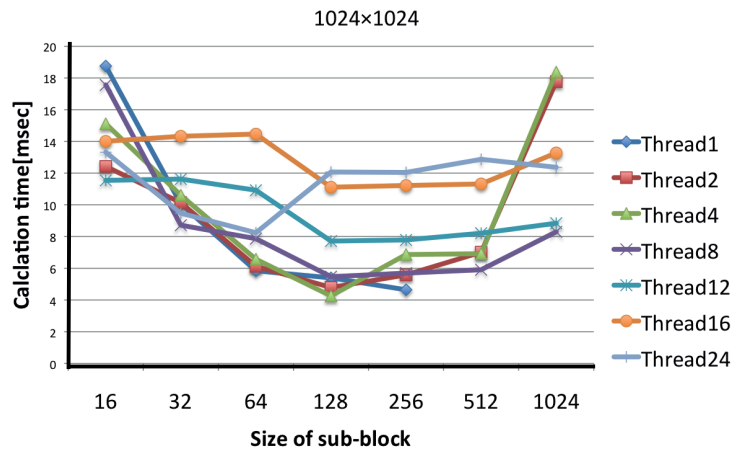


Fig. 3.12 文字列サイズ 1024 × 1024 における各スレッドと処理時間

4 提案手法の検証

4.1 提案手法と DTW による類似部分の比較

4.1.1 目的

提案手法により抽出された類似部分が DTW により抽出された類似部分とどの程度一致するかを比較する。DTW により抽出された類似部分とは、ウィンドウをずらして総当り的に DTW 距離を比較して距離が最小となるものと定義する。詳細については後述する。仮に、提案手法により抽出された類似部分が DTW により抽出された類似部分に一致すると言えれば、提案手法のほうが高速であるため、提案手法の有効性が確認できると考えた。

4.1.2 データセット

データセットは、実際の fNIRS を用いた脳機能実験から得られた時系列データを用いた。実験内容は、“照明や空調の異なる環境下で計算作業を行う”というものである。ブロックデザイン法⁷⁾を用いた実験であり、レスト 10 秒、タスク 30 秒、レスト 20 秒となっている。タスクは、“紙面に書かれた簡単な数式の答えを声に出して読み上げる”，レストは、“1 から 9 までの数を声に出して数える”ことである。計測部位は前頭部の 19 箇所部位 (CH1-CH19) である。サンプリングレートは 0.1 秒であり 60 秒間の計測を行ったため、時系列データの長さは 600 である。使用機器は日立メディコ製の ETG-7100 である。

4.1.3 DTW(Dynamic Time Warping) による類似部分の抽出

DTW⁴⁾⁵⁾ は時間伸縮を許容した時系列データの距離測定手法である。例えば、ある言葉をゆっくり話した音声データ、速く話した音声データは、時間軸の伸縮比率の異なるデータだが、DTW は 2 つの時系列データの時間軸の伸縮を許容して距離を算出するため、2 つの音声を類似として判断できる。一回あたりの距離算出のアルゴリズムのオーダーは $O(n^2)$ である。

DTW 距離は式 (4.1) として定義される。

$$\begin{cases} DTW(P, Q) = f(n_p, n_q) \\ f(i, j) = |p_i - q_j| + \min \begin{cases} f(i, j-1) \\ f(i-1, j) \\ f(i-1, j-1) \end{cases} \\ f(0, 0) = 0, f(i, 0) = f(0, j) = \infty \quad (i = 1, \dots, n_p, j = 1, \dots, n_q) \end{cases} \quad (4.1)$$

DTW により類似部分を抽出するためには、ウィンドウサイズを任意で定め、ウィンドウを徐々にずらして DTW 距離を算出していき、距離が最も小さくなった部分を類似部分として定める必要がある。そのため、DTW により類似部分を抽出するためのアルゴリズムのオーダーは $O(n^3)$ である。なお、提案手法で用いられる SW 法のアルゴリズムのオーダーは $O(n^2)$ である。

4.1.4 パラメータ

Table.4.1 に実験に用いたパラメータを示した。まず、提案手法を用いて ${}_{19}C_2(171)$ 通りの組み合わせの時系列データから、類似部分の抽出を行った。DTW で用いるウィンドウサイズは、提案手法で

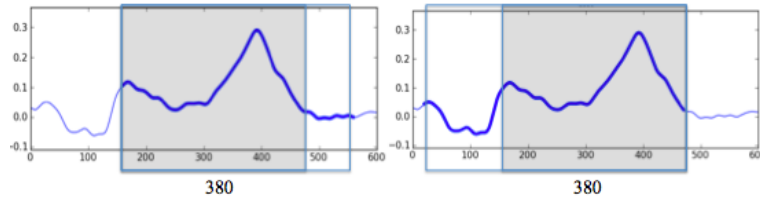


Fig. 4.1 重複率の定義

抽出された類似部分のサイズとした。例えば，CH1 と CH2 の時系列データで得られた類似部分のサイズが L であれば，DTW のウィンドウサイズも L とした。

Table 4.1 実験パラメータ

match	1
mismatch	-1
gap	-1
分割記号数	5
再量子化手法	SAX
DTW のウィンドウサイズ	L

4.1.5 評価方法

DTW を基準とし，提案手法によって抽出された類似部分がどの程度重複するかを検証した。重複とは，提案手法における類似部分が DTW における類似部分を基準として，どれだけ一致するかを示す指標である。重複率を以下の式のように定義した。

$$\text{重複率} = \text{提案手法及び DTW による類似部分の重複する部分の数} / \text{ウィンドウサイズ}$$

例えば，Fig.4.1 において左の時系列データと右の時系列データの類似部分の重なる部分の数は 380 であり，ウィンドウサイズは 460(左右ともに同じ)であるため，重複率は $380/460 = 0.74$ として求められる。

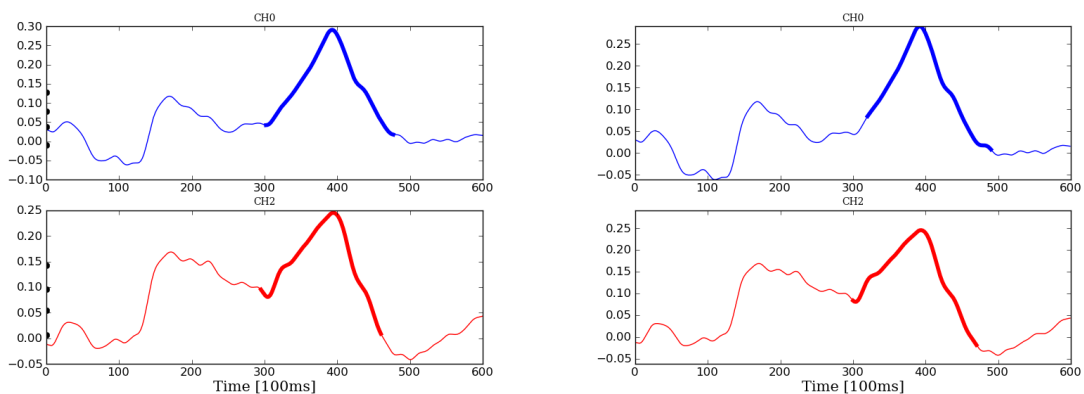
次に DTW による類似部分を基準に，“提案手法による類似部分”と“ランダムによる類似部分”の重複率を算出し比較した。“ランダムによる類似部分”とは，提案手法の類似部分のサイズをもとに類似部分の位置をランダムに決定したものである。また，171 通りの全ての組み合わせにおいて，DTW による類似部分を基準とした場合に，提案手法における類似部分と“ランダムによる類似部分”の重複率のどちらが高いかについてそれぞれのデータを比較した。重複率の定義から，ウィンドウサイズが長いほど重複率が高くなる傾向があるが，比較の際には問題はないと考えた。類似部分のサイズ別にまとめたものを Table.4.2 に示す。

Table 4.2 類似部分のサイズ別の重複率の比較

類似部分の長さ	データ数	1.DTW と提案手法 の重複率 (平均)	2.DTW とランダム の重複率 (平均)	1のほうが 高い確率
300 以下	19	0.40	0.31	0.66
400 以下	37	0.56	0.48	0.69
500 以下	53	0.63	0.57	0.66
600 以下	171	0.87	0.85	0.51

4.1.6 結果と考察

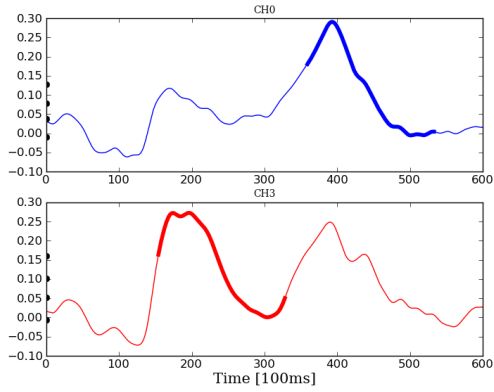
時系列データは 19 本あり、その中から 2 本ずつ選び類似部分の抽出を試みた。Fig.4.2-Fig.4.4 に、いくつかの結果を示す。各図には、4 本の時系列データがあり、類似部分がそれぞれ太線で表されている。また、左の 2 本が提案手法による類似部分を、右の 2 本が DTW による類似部分を表している。なお、SAX のグラフは元の時系列データを標準化したものである。Fig.4.2 を見ると提案手法と DTW による類似部分がうまく一致する場合もあるが、Fig.4.3 のように位相がずれた部分を提案手法が類似部分として抽出する場合もあることがわかった。また Table.4.2 より、データの長さが 500 以下の時は、提案手法による類似部分のほうがランダムによる類似部分よりも重複率が高くなることがわかった。データの長さが 600 以下の時は、どちらの方法を用いても重複率が同じになった。これは類似部分のサイズが 500 以上になると、Fig.4.4 のように類似部分がデータの大半を占めるため、どこを選んでも同じような結果が出るためだと考えられる。以上から提案手法による類似部分は、ランダムによる類似部分よりも DTW による類似部分に近いことが、一致しているとは言えないことがわかった。



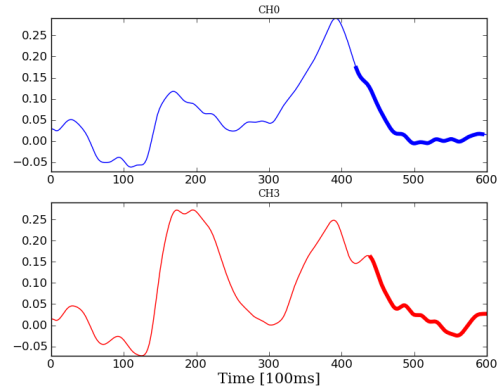
(a) 提案手法による類似部分

(b) DTW による類似部分

Fig. 4.2 上側 (青線) の時系列データが CH1, 下側 (赤線) の時系列データが CH2

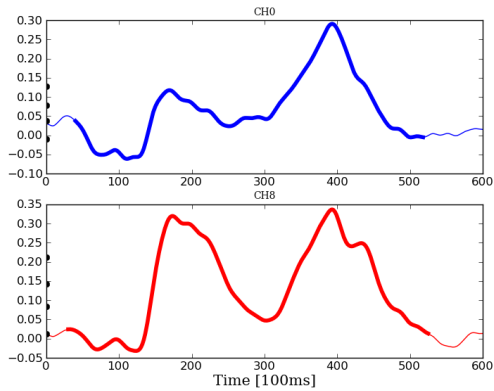


(a) 提案手法による類似部分

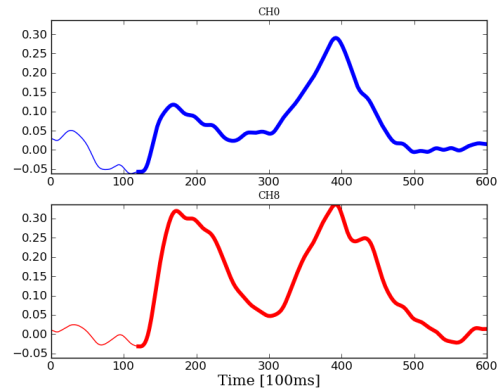


(b) DTW による類似部分

Fig. 4.3 上側 (青線) の時系列データが CH1, 下側 (赤線) の時系列データが CH3



(a) 提案手法による類似部分



(b) DTW による類似部分

Fig. 4.4 上側 (青線) の時系列データが CH1, 下側 (赤線) の時系列データが CH8

4.2 時系列クラスタリングデータセットを用いた数値実験

4.2.1 目的

ここでは、SAX と等間隔領域分割で抽出される類似部分がある時系列データに対して提案手法を適用したときどのように異なるかを比較する。また、提案手法を fNIRS データに対して適用する際にどのような条件が必要となるか検討する。

4.2.2 データセット

使用したデータセットは時系列クラスタリング用データセット²⁹⁾である。このデータセットにはあるタイプの実数値の時系列データが複数のクラス別に含まれている。同じクラス内に属する時系列データは互いに類似しているといえるため、提案手法を用いて同じクラス内に属する2つの時系列データの全体同士を類似部分として抽出できるかどうかを検証した。なお、SAX のグラフは元の時系列データを標準化したものであるため、等間隔領域分割のグラフとは多少異なって見える。分割文字数は予備実験より5文字とした。SW 法のパラメータは $match = 1$, $mismatch = -1$, $gap = -1$ とした。

4.2.3 評価方法

SAX と等間隔領域分割では異なる部分が類似部分として抽出される。今回は、類似する2つの時系列データを用いたため、データ全体を類似部分として抽出することが望ましいと考えた。2つの手法のどちらを用いるべきかを判断するため、抽出された類似部分の長さ¹と DTW 距離を用いて評価を行った。

類似部分は長くて誤差が大きな場合と短くて誤差が少ない場合が考えられた。そのため、DTW 距離を抽出された類似部分の長さで割ることにより、長さ1あたりの DTW 距離を誤差として定義した。この誤差が小さいほど、抽出された類似部分は似ていると考えられ、SAX と等間隔領域分割でどちらが DTW 距離に近い類似部分を抽出しているのか判断できる。Table.4.3 にそれぞれの図におけるデータの長さ、抽出された類似部分の長さ¹、類似部分の長さの割合²、2つの時系列データの類似部分間の DTW 距離、長さ1あたりの DTW 距離³の結果を示す。

4.2.4 結果

4つのデータに対して SAX と等間隔領域分割を適用しそれぞれの特徴を調べた。Fig.4.5-Fig.4.8 には2本の時系列データがあり、類似部分が黒色の太線でそれ以外の部分が灰色の線で表されている。簡単のため、変換で得られた文字列は標記しない。等間隔領域分割と SAX の両方で類似部分が抽出された時系列データ、片方のみで抽出された時系列データ、抽出されなかった時系列データの4パターンを Fig.4.5-Fig.4.8 に示す。ここでは Table.4.3 における類似部分の長さの割合が0.8以上だったものを類似部分が抽出されたとしている。

Fig.4.5 から、上部と下部の時系列データのうち位相がずれた部分がうまく抽出されていることが分かる。このことから提案手法は位相のずれたデータに対して適用できることが分かる。Fig.4.6 か

¹2つの時系列データの類似部分間の長さの平均

²抽出された類似部分の長さ/時系列データ全体の長さ

³2つの時系列データの類似部分間の DTW 距離/抽出された類似部分の長さ

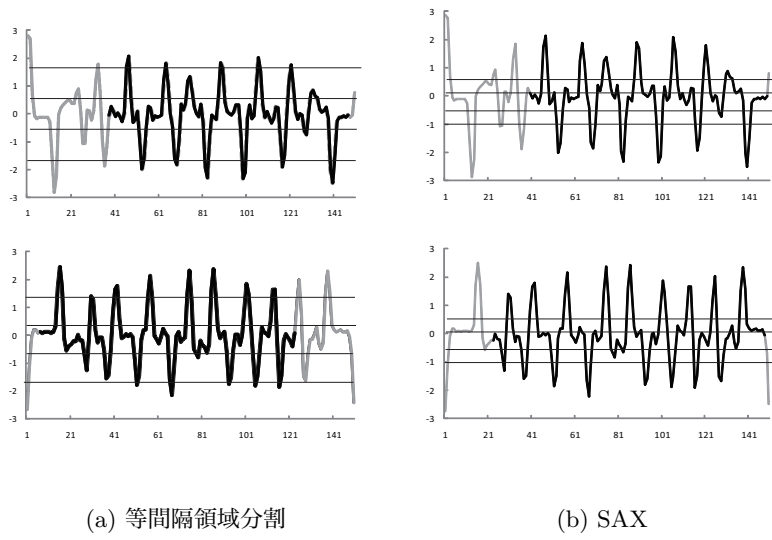


Fig. 4.5 等間隔領域分割と SAX の両方で類似部分が抽出された例

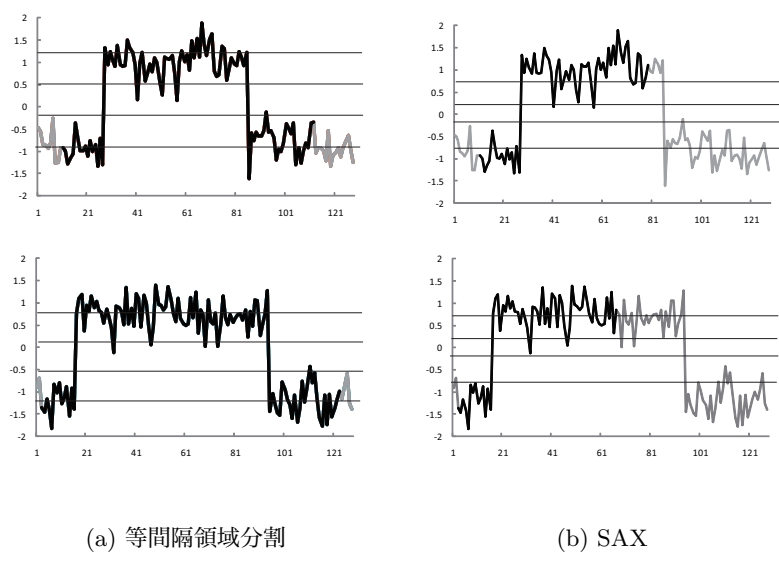


Fig. 4.6 等間隔領域分割では類似部分が抽出されたが SAX では抽出されなかった例

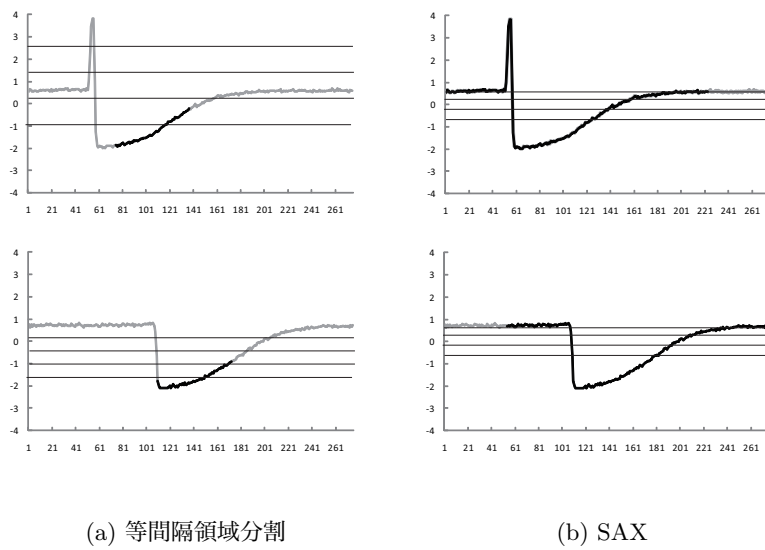


Fig. 4.7 SAX では類似部分が抽出されたが等間隔領域分割では抽出されなかった例

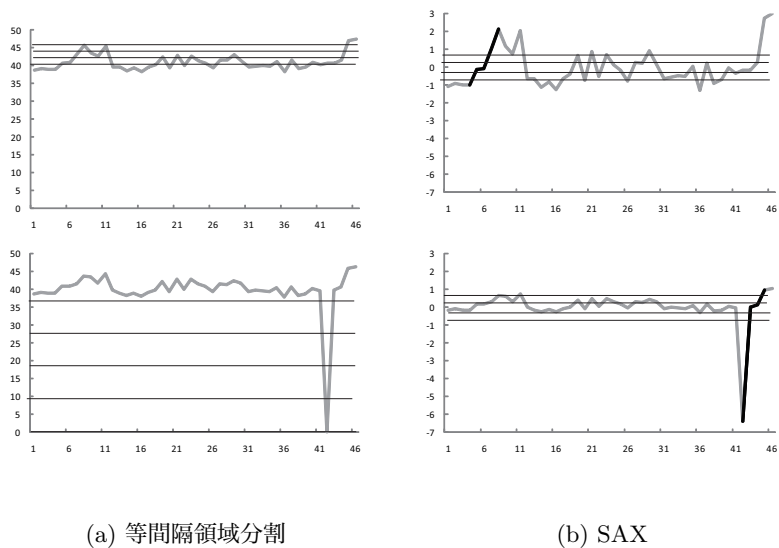


Fig. 4.8 等間隔領域分割と SAX の両方で類似部分が抽出されなかった例

Table 4.3 それぞれの図における類似部分の長さの割合, 長さ 1 あたりの DTW 距離

図の名称	データ長	抽出された 類似部分の長さ	類似部分の 長さの割合	2つの時系列データの 類似部分間の DTW 距離	長さ 1 あたり の DTW 距離
Fig.4.5(a)	150	114	0.76	41.6	0.36
Fig.4.5(b)	150	118	0.79	53.3	0.45
Fig.4.6(a)	130	112	0.86	32.2	0.29
Fig.4.6(b)	130	67	0.52	16.3	0.24
Fig.4.7(a)	270	63	0.23	10.6	0.17
Fig.4.7(b)	270	222	0.82	23.2	0.10
Fig.4.8(a)	46	1	0.02	1.1	1.09
Fig.4.8(b)	46	4	0.09	42.6	10.65

ら, SAX では類似部分が半分程度しか抽出されないことが分かる. データが頻繁に上下に振動すると適切な分割線を定めにくく, 提案手法を適用しづらいことが考えられる.

Fig.4.7 から, 外れ値の影響で等間隔領域分割は類似部分を抽出できてないことが分かる. Fig.4.8 では, 外れ値の影響が大きく, どちらの手法でも適切な抽出が行えなかったことが分かる. また, Table.4.3 より長さ 1 あたりの DTW 距離は SAX でも等間隔領域分割の間で明確な差は見られなかった.

4.2.5 考察

以上の結果から SAX と等間隔領域分割が適用できる時系列データには向き・不向きがあると言える. 時系列データに外れ値がない場合は等間隔領域分割の方が分割の幅を広くとることができ, より適切な分割ができると推測される. 逆に, 外れ値がある場合は SAX を用いたほうが良いと思われたが, Fig.4.8 より実際は SAX を用いても必ずしも外れ値に対応できるわけではないことが分かった.

SAX と等間隔領域分割の差異は明確には判明しなかった. しかし, fNIRS に適用するための条件として時系列データの外れ値の除去や移動平均による平滑化の前処理が必要であることがわかった.

4.3 fNIRS の出力データに適した前処理手法の適用

4.3.1 概要

fNIRS が計測する血流の増減データは絶対的変化量ではなく相対的変化量である. そのため, 相対的なパターンを定性的に解析する必要がある. また, 被験者の生理的変化により, あるタスクに反応を示すが山の頂点の増減の大きさが異なる場合があり, 反応を示している場合はこれを類似するものとして定義した. そのために有効となる前処理手法を 3.3 節にて提案した. ここでは実際の fNIRS データに対して 3.3 節の前処理手法を適用し, この手法が有効であることを示す.

4.2 節から再量子化手法の等間隔領域分割と SAX では, 時系列データの再量子化が適切に行われないうちがあることが分かった. そこで 3.3 節の前処理手法を用いることで, 山のパターンが同じデータを類似部分として抽出可能となることを示す.

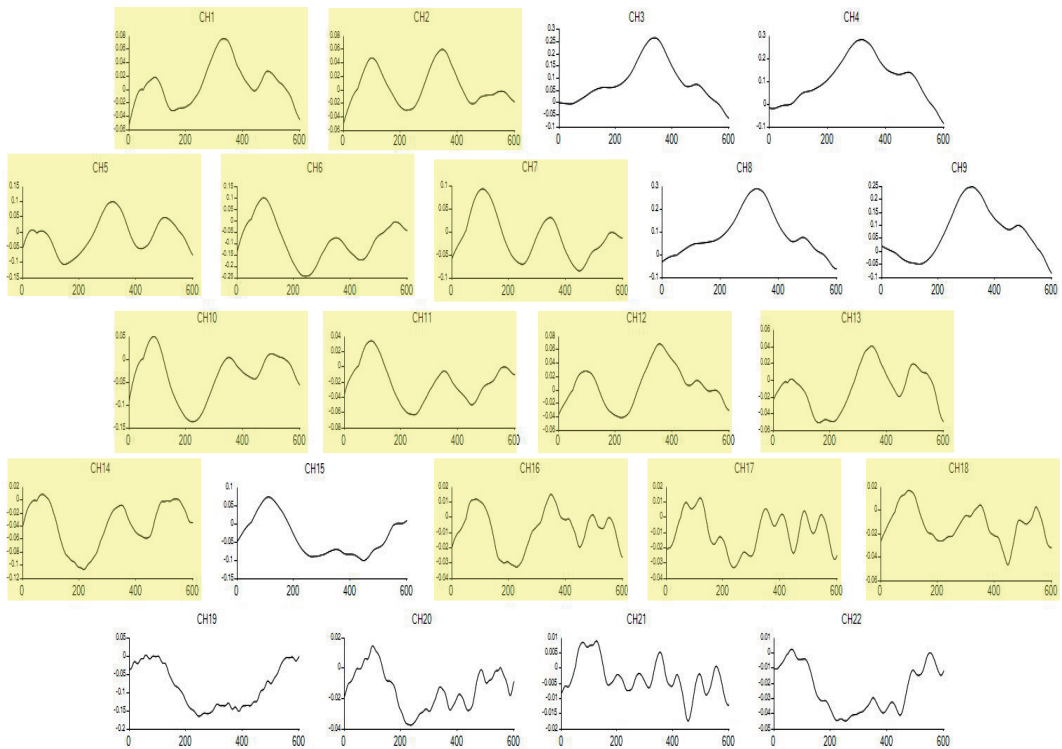


Fig. 4.9 山が3つあるfNIRSの出力データ

Fig.4.9にfNIRS出力データの一例を示す。黄色で囲われたCHは3つの山を持つ多峰性のデータである。Fig.4.10, Fig.4.11にCH2を基準として、他のCHとの類似部分を提案手法により抽出した結果を示す。時系列データの再量子化の際、Fig.4.10は等間隔領域分割を、Fig.4.11はSAXを用いた。Fig.4.12に前処理手法を用いた結果を示す。各図において赤色のCH(CH2)が基準CH、青色の範囲が類似部分である。また、Fig.4.12において青色線が伸長後のfNIRSデータ、黄色の丸点が極値を表す。

Fig.4.9とFig.4.10, Fig.4.11, Fig.4.12を比較すると、等間隔領域分割、SAXでは抽出できなかった3つの山があるパターンをFig.4.12では抽出できていることがわかる。これにより、fNIRSにおけるパターンを調べる上では山の最大化、最小化する手法が有効であるといえる。

しかし、Fig.4.12のCH17, CH21を見ると極値が適切に設定されず、そのために類似部分の抽出も適切にされないことがわかった。

以上から、fNIRSにおけるパターンを調べる上では山の最大化、最小化する手法が有効であるが、極値を適切に定める必要があるといえる。

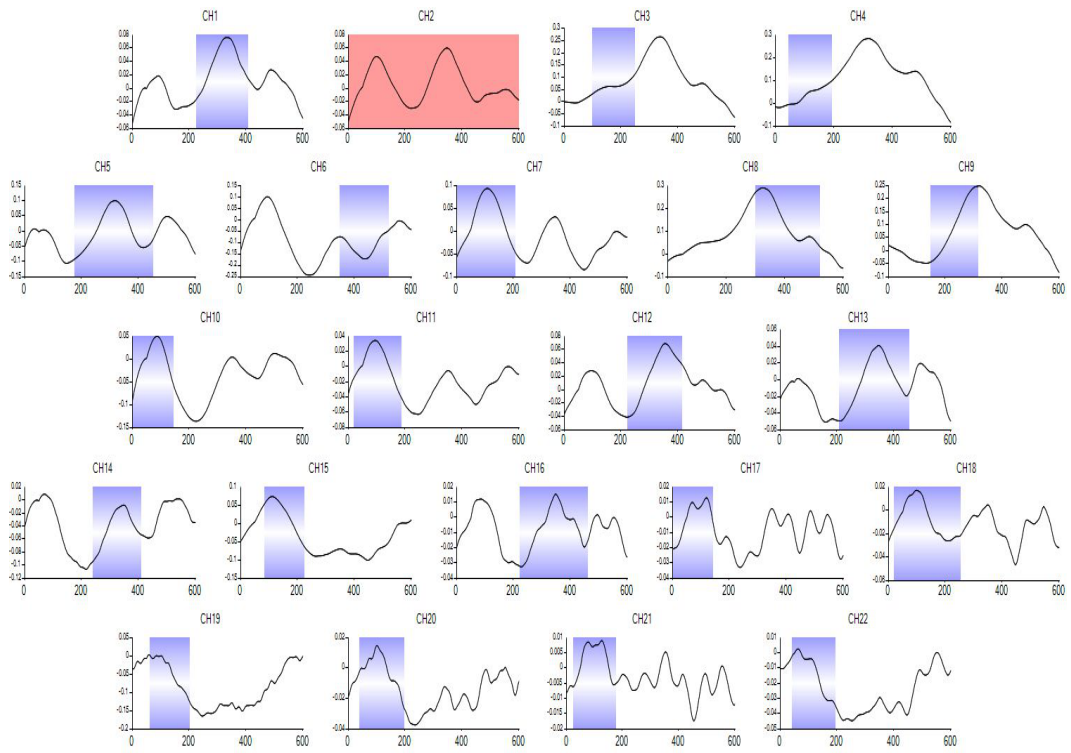


Fig. 4.10 等間隔領域分割を適用

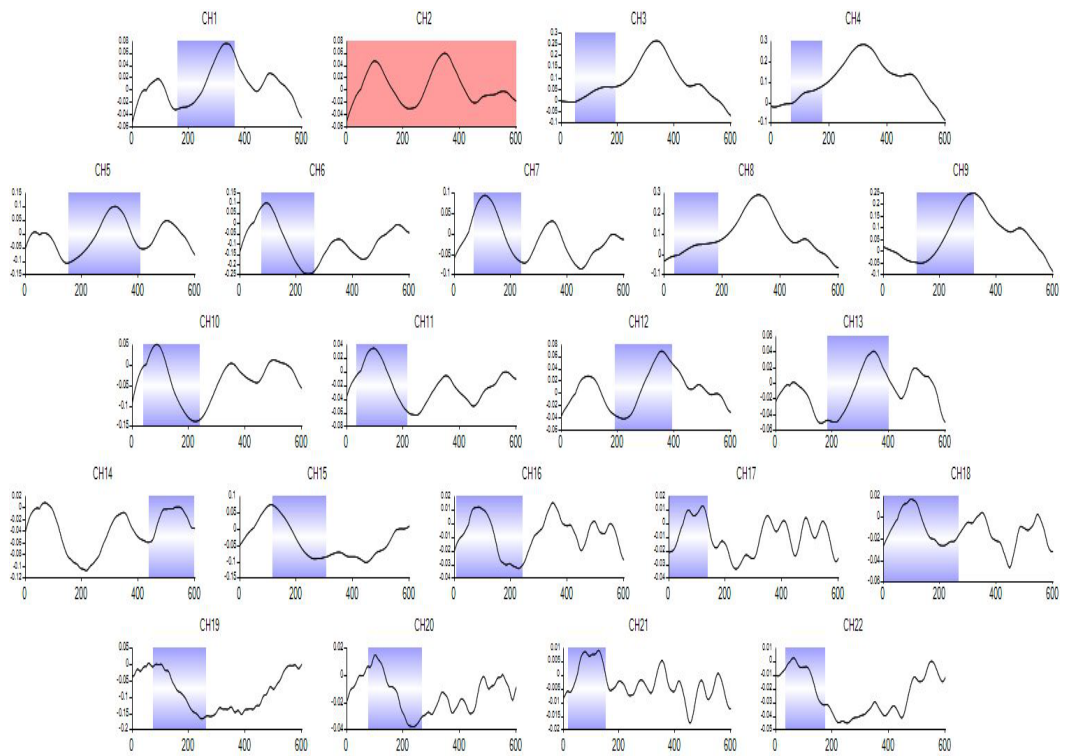


Fig. 4.11 SAX を適用

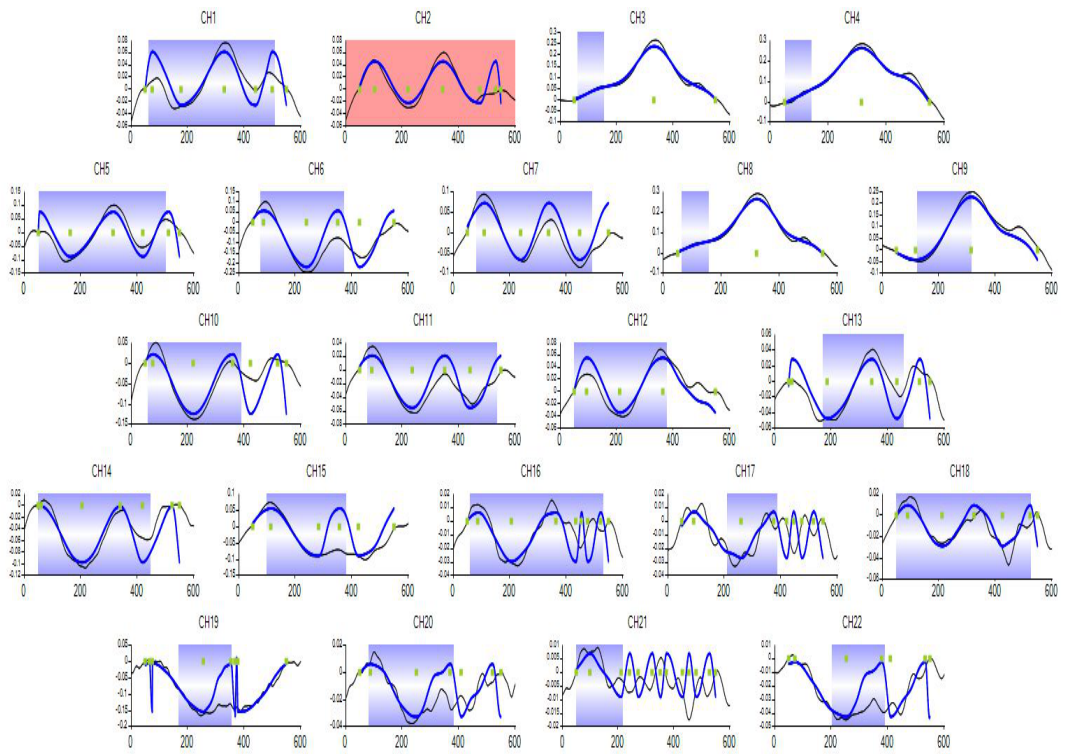


Fig. 4.12 山の伸長を適用

5 解析システムの実装

5.1 システムの概要

このシステムは、データ解析において従来は注目されなかった脳部位データのどれに着目すべきかを提示することで解析者に新たな知見を与えることを目的とする。現在、fNIRS の解析者は自身が任意で注目した脳部位のデータに対してのみデータ処理を行っており、その方法では注目している脳部位以外に重要な要素がある場合にそれを見落としてしまう可能性がある。これを解決するためには、ある特徴的な波形に類似した波形が他の脳部位のどこにあるのかを特定することが有効であると考えた。これにより、解析者はデータ量が多いために今までは気がつかなかったが、着目すべき可能性がある部位を低負担で見つけることが可能となる。

5.2 システムの機能

システムの機能は大きく分けて2つある。複数プローブ内における探索と同一プローブ内における探索である。ユーザはまず複数プローブを探索し、新たな発見があれば同一プローブでの探索に切り替え詳細に解析を行う。システムの機能は以下の4つである。

【同一のプローブ内での探索】

機能1: fNIRS のファイルを読み込み表示する

機能2: CH を1つ選択し、その CH の時系列データに類似するデータを表示する

機能3: ある CH の時系列データの範囲を指定し、それと類似するものを他の CH の時系列データ群から表示する

【複数のプローブ内での探索】

機能4: 複数のプローブにおいて、ある CH データに類似するデータを探索し表示する

5.3 構築したシステムによる fNIRS データの解析

5.3.1 概要

構築したシステムを用いて実験から得られた fNIRS データを解析した。実際の fNIRS データに対して、どのようにシステムを適用することができるのかを示す。ここでは一例として、ストループ効果の実験データを用いた。ストループ効果³⁰⁾³¹⁾³²⁾とは、文字色と語の意味が不一致な色文字に対して反応するとき、被験者に認知的葛藤が発生し反応時間が増大するという現象である。Fig.5.5 にストループ効果における一致課題、不一致課題の例を示す。

5.3.2 データセット

fNIRS データの実験内容は、”認知的葛藤場面におけるストループテストの影響の検討”である。タスクは、“文字色と文字意味が不一致のとき色を答える”，レストは、“「あいうえお」を繰り返し発生する”ことである。ブロックデザイン法を用いて、レスト10秒、タスク30秒、レスト20秒と設定した。計測部位は前頭部、右側頭部、左側頭部、後頭部、頭頂部の5箇所である。使用機器は日立メディコ製のETG-7100である。サンプリングレートは0.1秒であり60秒間の計測を行ったため、

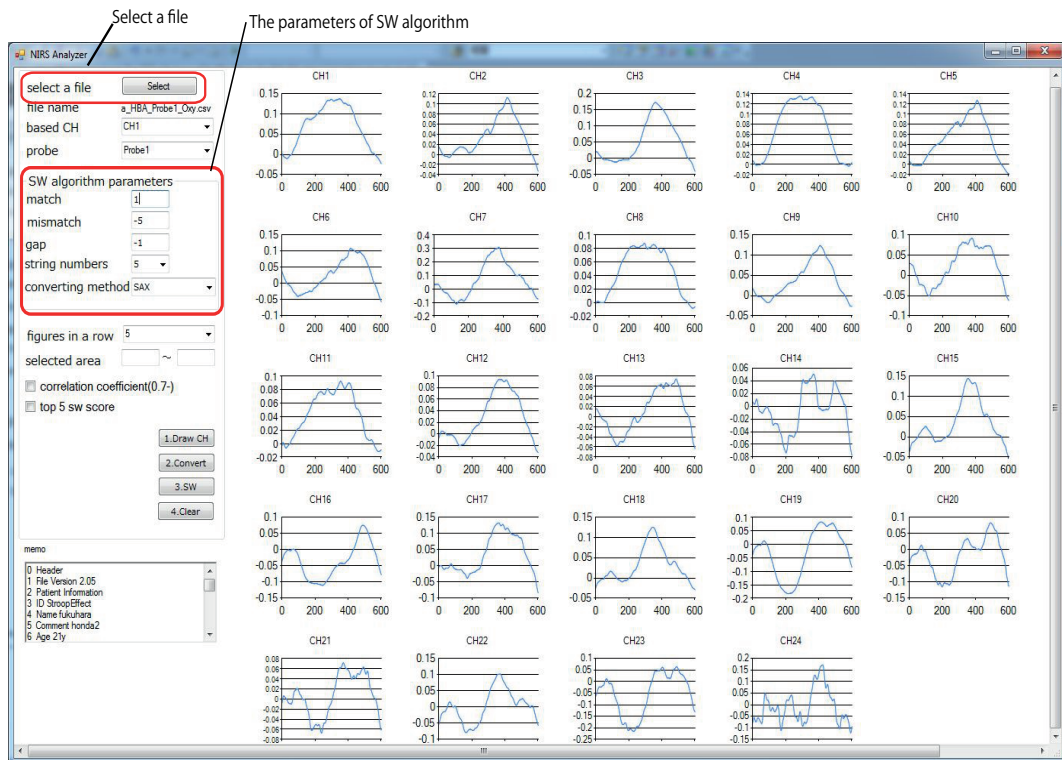


Fig. 5.1 fNIRS の出力データを読み込み，各 CH データを表示

時系列データの長さは 600 である。この場合，文字列の長さが 600 になることから，Table.3.2 より，スレッド数を 4，サブブロックサイズを 32 に設定した。

5.3.3 解析の目的

従来の先行研究ではストループ効果における認知的葛藤は脳の左側頭部のみが活性化されると言われていた。しかし，福原等によって左側頭部の他に右側頭部も活性化することが示唆されている³³⁾。そこで，解析者が左側頭部を見た場合に，システムがそれ以外に見るべき部位を提示できるのかどうかを検討した。仮に，右側頭部に左側頭部と類似する波形が多くあると言えれば，システムが有効であるといえる考えた。

5.3.4 結果と考察

ストループ効果では左側頭部（特に下側）の活性化が示唆されていたため，左側頭部の CH1 を基準として他の CH との類似部分を抽出した。Fig.5.6 にストループ効果の実験データにおける類似部分抽出の結果を示す。Fig.5.7 は左側頭部の拡大図，Fig.5.8 は右側頭部の拡大図，Fig.5.9 は頭頂部の拡大図である。Fig.5.6-Fig.5.9 により，右側頭部に左側頭部の CH1 に類似する CH が多くあることがわかる。左側頭部の CH1 は，ストループ効果によって活性化が示唆されている部位である。よって右側頭部もストループ効果によって活性化しているのではないかと考えることができる。このことからシステムが着目すべき部位の提示に役立つことがわかる。

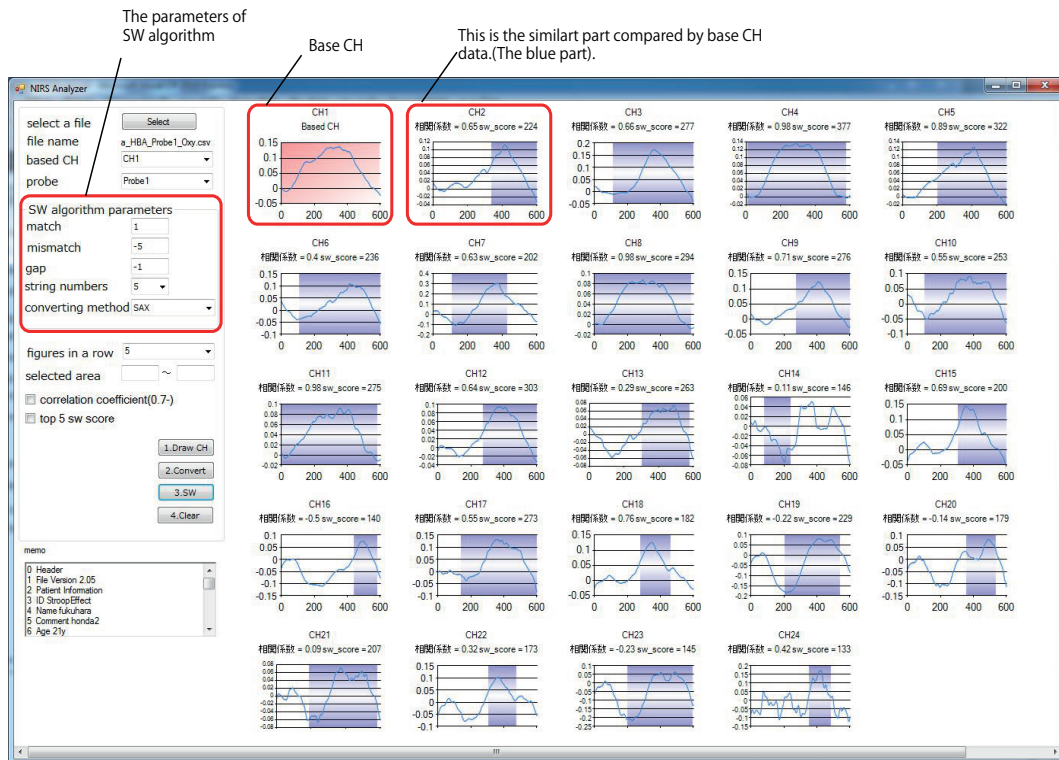


Fig. 5.2 CH データを1つ指定して類似するデータを探索

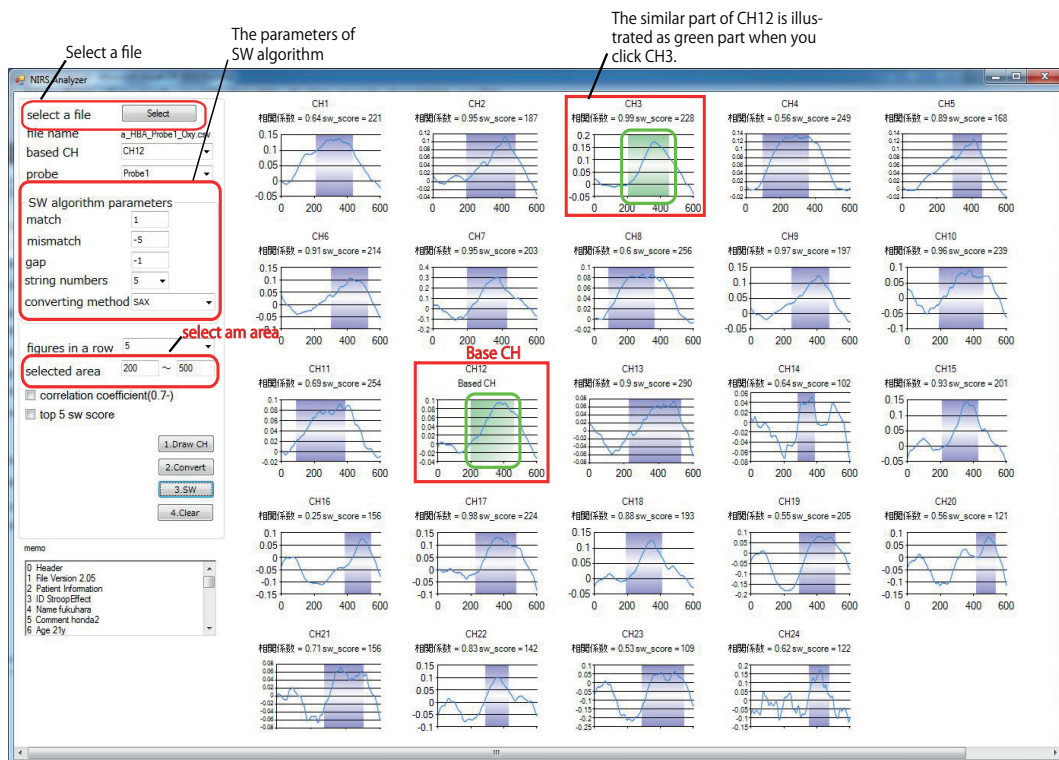


Fig. 5.3 CH データを部分的に指定して類似するデータを探索

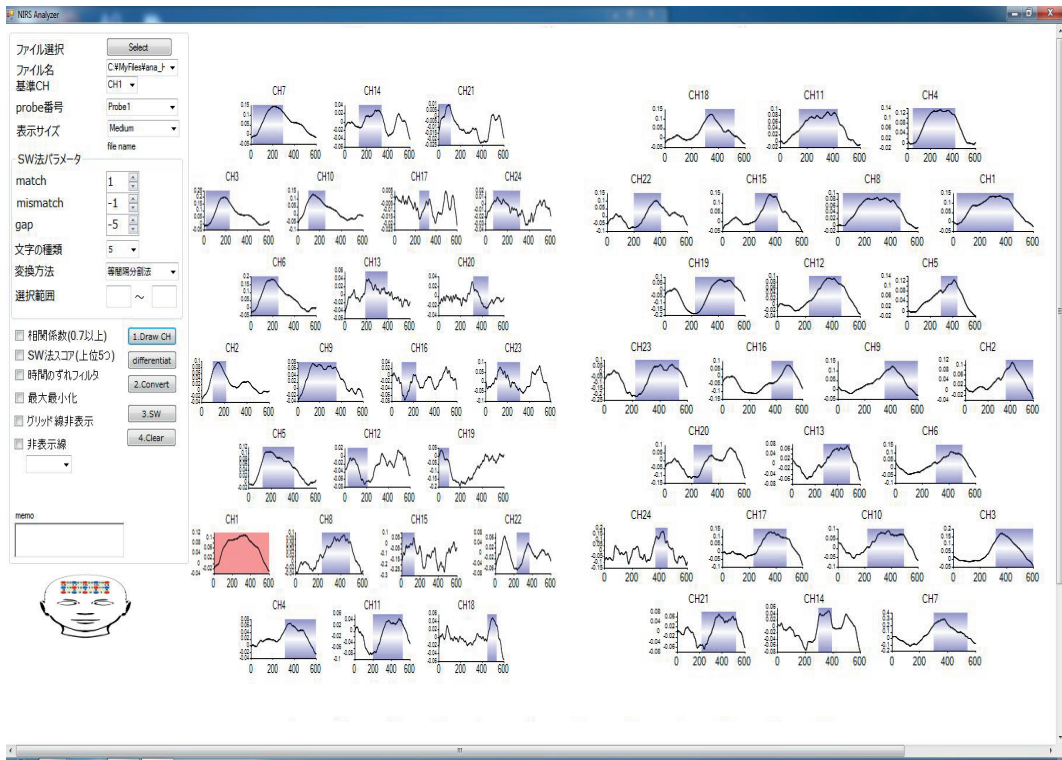


Fig. 5.4 複数のプローブにおいて、あるCHデータに類似するデータを探索し表示



Fig. 5.5 ストロープ効果実験における一致課題，不一致課題の例

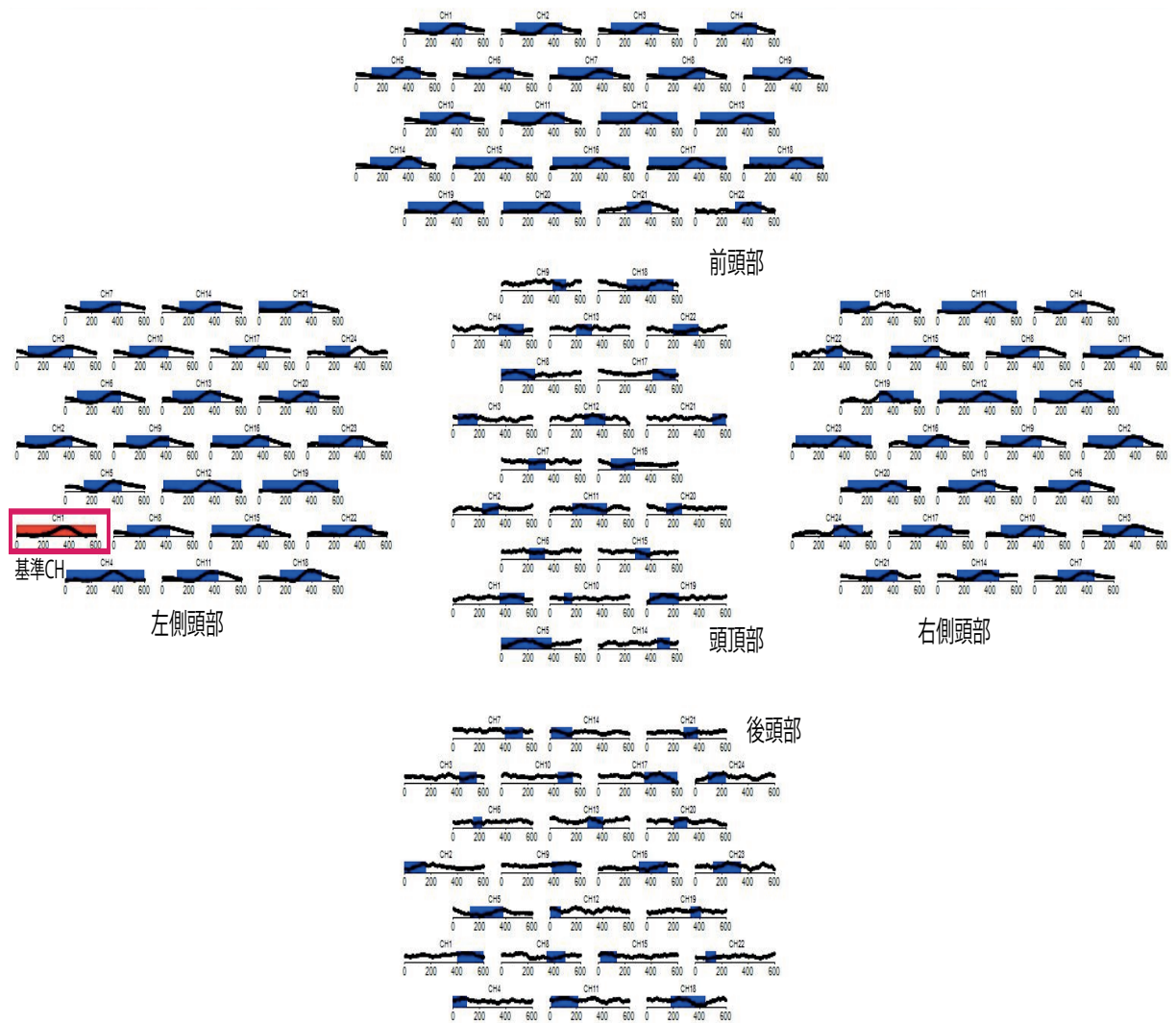


Fig. 5.6 ストロープ効果実験における探索

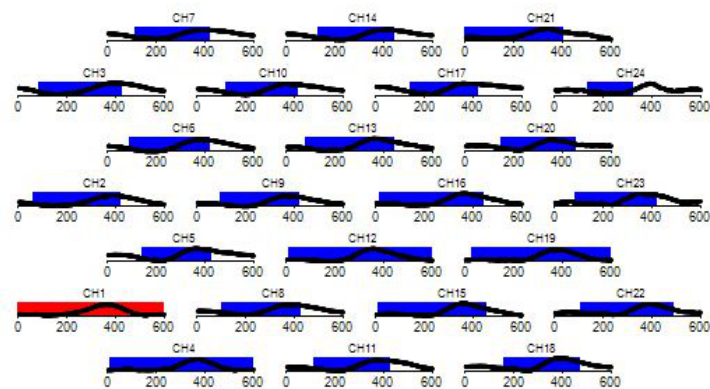


Fig. 5.7 ストロープ効果実験における探索結果（左側頭部の拡大図）

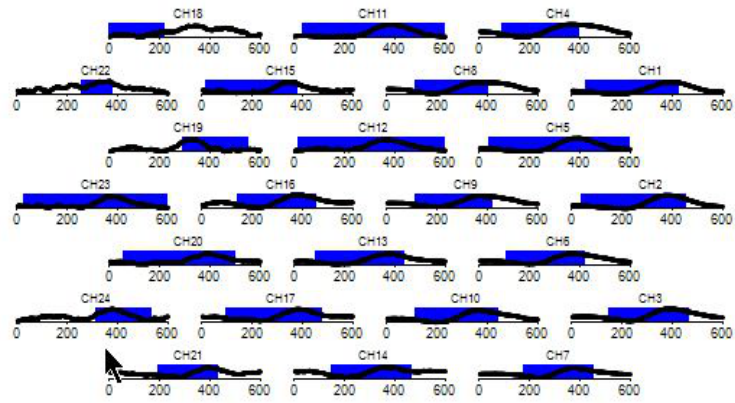


Fig. 5.8 ストループ効果実験における探索結果（右側頭部の拡大図）

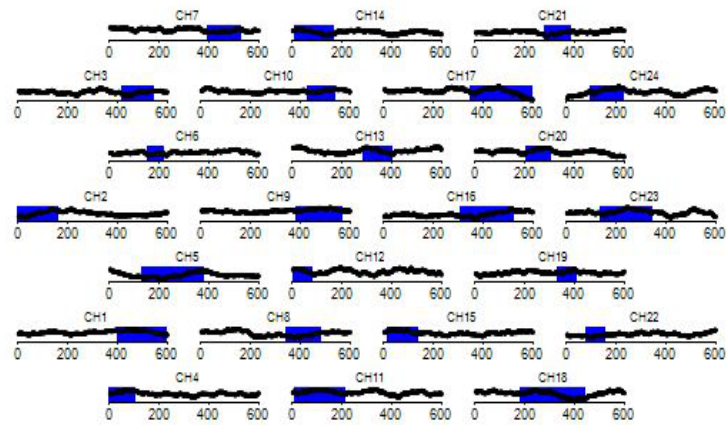


Fig. 5.9 ストループ効果実験における探索結果（頭頂部の拡大図）

6 結論

fNIRS 装置には出力する時系列データ量の増大により、解析者が効率的にデータを解析できないという問題がある。特に実験結果を検討する際、出力された多数の時系列データの中でどの時系列データに着目するかが解析者に依存しているという問題がある。現在は、解析者自身が任意で注目した脳部位のデータに対してのみデータ処理を行っており、注目している脳部位以外に重要な要素がある場合にそれを見落としてしまう可能性があった。これを解決するためには、複数の被験者間によく現れたパターンなどデータを定性的に解析する必要があった。

そこで本論文では、特定のパターンを探索するために、多チャンネルの時系列データ間における類似部分を探索法を提案した。提案手法は、時系列データの再量子化と相同性検索の組み合わせによる2つの時系列データからの類似部分の抽出法である。相同性検索は文字列検索アルゴリズムであるため、時系列データに対して適用するためには再量子化が必要であった。さらに再量子化の際はfNIRSデータに適切な前処理を行う必要があった。

本論文では提案手法の性質を検討するため、3つの数値実験を行った。

1つ目の数値実験では、提案手法により抽出された類似部分とDTWにより抽出された類似部分との程度一致するかを比較した。提案手法により抽出された類似部分がDTWにより抽出された類似部分に一致すると言えれば、提案手法のほうが高速であるため、提案手法の有効性が確認できると考えたが、一致率が高いとは言えないことが分かった。

2つ目の数値実験では、時系列クラスタリングデータセットを用いて、SAXと等間隔領域分割で抽出される類似部分がどのように異なるかを比較した。その結果、2つの手法のどちらも位相にずれがあるデータに対してはうまく適用できることが分かった。しかし、SAX、等間隔領域分割の両手法において、外れ値があるデータや頻繁に上下に振動するデータは分割線が適切に定められにくく、類似部分がうまく抽出できないことが分かった。また、提案手法をfNIRSデータに対して適用する際にどのような条件が必要となるか検討した。提案手法を適用するためには時系列データの外れ値の除去や平滑化等の前処理が必要であることが分かった。

3つ目の数値実験では、SAXと等間隔領域分割では抽出できなかったパターンの類似部分を、fNIRSに適した前処理手法を用いることで抽出可能となることを示した。しかし、前処理手法を用いる際の極値の設定が問題になることが分かった。

最後に、提案手法のアルゴリズムを実装し解析者のためのシステムを作成した。実際のfNIRSの出力データに対してシステムを適用し、類似部分がどのように表示され、着目すべき部分を提示できるかを示した。

謝辞

本研究を遂行するにあたり、多大なる御指導そして御協力を頂きました、同志社大学生命医科学部の廣安知之教授に心より感謝致します。度重なる研究 MTG において貴重な御意見を頂き、本研究を完成に近づけることができました。研究活動を通して自分自身を成長させることができました。

知的システムデザイン研究室を理想的な環境に育てていただいた同志社大学工学部の三木光範教授に感謝致します。本研究について様々なアドバイスとディスカッションの機会を設けて頂きました吉見真聡助教に心より感謝致します。吉見真聡助教とのディスカッションにより本研究の成果を飛躍的に高めることができました。本研究について様々なアドバイスを頂きました横内久猛教授に心より感謝致します。

また、コーチである横田山都さんには研究活動における相談や研究方針の確認をさせて頂きました。ここに深く感謝申し上げます。データマイニング研究班の一員として、研究ミーティングにて多くの議論をしていただきました宮部洋太さん、宮地正大さん、大堀祐一さんに御礼申し上げます。本論文を校正して頂き、貴重な御意見を頂いた福原理宏さんに深く感謝申し上げます。

最後に、知的システムデザイン研究室および医療情報システム研究室の皆さまには、本研究に関して数多くの議論や助言をして頂きました。また、皆様のおかげで、充実した研究生生活を送ることができました。この場を借りて厚く御礼申し上げます。

参考文献

- 1) James C. Eliassen; Erin L. Boespflug; Martine Lamy; Jane Allendorfer; Wen-Jang Chu; Jerzy P. Szaflarski. Brain-Mapping Techniques for Evaluating Poststroke Recovery and Rehabilitation. *Neuroplasticity: Changing Minds and Changing Function*.
- 2) Xu Cuia; Signe Braya; Daniel M. Bryanta; Gary H. Gloverc; Allan L. Reissa. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*.
- 3) nVIDIA. Tesla bio workingbench. http://www.nvidia.co.jp/object/tesla_bio_workbench_jp.html.
- 4) 石川 雅弘; 吉川 昂伯; 陳漢雄. 類似部分区間検索のためのタイムワーピング距離の下限値計算. 日本データベース学会 letters, Vol. 6, pp. 25–28, 2007.
- 5) 櫻井 保志; Faloutsos Christos; 山室雅司. ダイナミックタイムワーピング距離に基づくストリーム処理. 電子情報通信学会論文誌, Vol. J92-D, pp. 338–350, 2009-03.
- 6) 新井康允. 脳のしくみ. 日本実業出版社, 1997.
- 7) 福田正人. 精神疾患と NIRS-光トポグラフィ検査による脳機能イメージング. 中山書店, 2009.
- 8) 電極の配置 (国際 10-20 法) . <http://eeg.cloudcomputing-navi.com/denkyoku.html>.
- 9) 株式会社日立メデコ. Optical topography etg-7100. <http://www.hitachi-medical.co.jp/product/opt/etg/func.html>.
- 10) Spectratech Inc. Spectratech oeg-16. <http://www.spectratech.co.jp/>.
- 11) BiSPL(Bio Imaging Signal Processing Lab). Nirs-spm. <http://bisp.kaist.ac.kr/NIRS-SPM.html>.
- 12) Source Signal Imaging Inc. Emse 脳波解析プログラム. <http://www.miyuki-net.co.jp/jp/product/emse.htm>.
- 13) KASHINO Kunio; SMITH Gavin A; MURASE Hiroshi. A Quick Search Algorithm for Acoustic Signals Using Histogram Features. *The transactions of the Institute of Electronics, Information and Communication Engineers*.
- 14) NISHIMURA Takuichi; MIZUNO Michinao; OGI Shinobu; SEKIMOTO Nobuhiro. Same Interval Retrieval from Time-Sequence Data Based on Active Search.
- 15) ITOH Yoshiaki; KIYAMA Jiro; KOJIMA Hiroshi; SEKI Susumu. Reference Interval-free Continuous Dynamic Programming for Spotting Speech Waves by Arbitrary Parts of a Reference Sequence Pattern.

- 16) 菅井康祐; 杉山雅英. 任意個数の時系列に含まれる類似部分探索法. 情報処理学会研究報告 SLP, 音声言語情報処理, 2007-12-20.
- 17) 杉山雅英. 距離に基づく時系列 Active 探索法における距離の探索効率の比較. 電子情報通信学会論文誌 情報・システム J91-D(8), 2008-08-01.
- 18) Takuma Nishii; Tomoyuki Hiroyasu; Masato Yoshimi; Mitsunori Miki; Hisatake Yokouchi. Similar subsequence retrieval from two time series data using homology search. *Systems Man and Cybernetics*, pp. 1062–1067, 2010.
- 19) 西井琢真; 廣安知之; 吉見真聡; 三木光範; 横内久猛. Smith Waterman 法を利用した fNIRS データの類似部分抽出アルゴリズムの提案. 同志社大学理工学研究報告, Vol. 52, No. 4, pp. 37–45, 2012.
- 20) 廣安知之; 西井琢真; 吉見真聡; 三木光範; 横内久猛. 相同性検索を用いた 2 つの時系列データからの類似部分抽出手法と DTW による類似部分の評価. 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告, Vol. 2010, No. 24, pp. 1–6, 2010-09-21.
- 21) Abdullah Mueen; Eamonn Keogh; Qiang Zhu; Sydney Cash; Brandon Westover. Exact Discovery of time-series Motifs. *SDM*., pp. 473–484, 2009.
- 22) T F Smith; M S Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, Vol. 147, No. 1, pp. 195–197, March 1981.
- 23) Hideyuki Takagi. Implementation of the smith-waterman algorithm on a reconfigurable supercomputing platform. HPRCTA '07 Proceedings of the 1st international workshop on High-performance reconfigurable computing technology and applications, 2007.
- 24) 岩出智行; 石原達也; 松田秀雄; 橋本昭洋. 配列類似度に基づく大規模配列集合の高速分類アルゴリズム. 情報処理学会研究報告 アルゴリズム研究会報告, Vol. 96, No. 67, pp. 33–40, July 1996.
- 25) The European Bioinformatics Institute. Blast. <http://www.ebi.ac.uk/Tools/blast/>.
- 26) Keisuke Dohi; Khaled Benkrid; Cheng Ling. Highly efficient mapping of the smith-waterman algorithm on cuda-compatible gpus. ASAP ' 10, Vol. 36, pp. 29–36, 2010.
- 27) Yongchao Liu; Douglas L Maskell; Bertil Schmidt. Cudasw++: optimizing smith-waterman sequence database searches for cuda-enabled graphics processing units. *BMC Research Notes*, 2009.
- 28) Svetlin A Manavski; Giorgio Valle. Cuda compatible gpu cards as efficient hardware accelerators for smith-waterman sequence alignment. *BMC Bioinformatics*, 2008.
- 29) Eamonn Keogh; Xiaopeng Xi; Li Wei. Welcome to the ucr time-series classification / clustering page. <http://www.cs.ucr.edu/eamonn/timeseriesdata/>.

- 30) 嶋田博行. ストループ効果-認知心理学からのアプローチ. 培風館, 1994.
- 31) Stroop J.R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, pp. 643–661, 1935.
- 32) A. Ehlis et al. Multi-channel near-infrared spectroscopy detects specific inferior-frontal activation during incongruent stroop trials. *Biological psychology*, pp. 315–331, 2005.
- 33) 福原理宏; 廣安知之; 三木光範; 吉見真聡; 横内久猛. 認知的葛藤場面における近赤外分光法を用いたストループテストによる左右脳半球の検討. 第 50 回 日本生体医工学会大会.

付 図

2.1	前頭部のプローブ設置位置	3
2.2	ETG-7100	3
2.3	右側頭部のプローブ設置位置	3
2.4	左側頭部のプローブ設置位置	3
2.5	左側頭部における出力データの一例	4
3.1	提案手法の流れ	6
3.2	fNIRS の出力データに対して提案手法を適用した例 (類似部分の位相が同じ)	6
3.3	fNIRS の出力データに対して提案手法を適用した例 (類似部分の位相が異なる)	7
3.4	SAX による文字列変換の流れ	8
3.5	等間隔領域分割による文字列変換の流れ	9
3.6	大きく 3 つの山がある fNIRS の時系列データ	9
3.7	山の最大化, 最小化の流れ	10
3.8	SW 法で抽出された類似部分文字列の例	11
3.9	文字列テーブルの例	11
3.10	文字列 “BBC” と “CBC” の類似部分を SW 法により探索した過程	13
3.11	各文字列における最適な処理時間	15
3.12	文字列サイズ 1024 × 1024 における各スレッドと処理時間	15
4.1	重複率の定義	17
4.2	上側 (青線) の時系列データが CH1, 下側 (赤線) の時系列データが CH2	18
4.3	上側 (青線) の時系列データが CH1, 下側 (赤線) の時系列データが CH3	19
4.4	上側 (青線) の時系列データが CH1, 下側 (赤線) の時系列データが CH8	19
4.5	等間隔領域分割と SAX の両方で類似部分が抽出された例	21
4.6	等間隔領域分割では類似部分が抽出されたが SAX では抽出されなかった例	21
4.7	SAX では類似部分が抽出されたが等間隔領域分割では抽出されなかった例	22
4.8	等間隔領域分割と SAX の両方で類似部分が抽出されなかった例	22
4.9	山が 3 つある fNIRS の出力データ	24
4.10	等間隔領域分割を適用	25
4.11	SAX を適用	25
4.12	山の伸長を適用	26
5.1	fNIRS の出力データを読み込み, 各 CH データを表示	28
5.2	CH データを 1 つ指定して類似するデータを探索	29
5.3	CH データを部分的に指定して類似するデータを探索	29
5.4	複数のプローブにおいて, ある CH データに類似するデータを探索し表示	30
5.5	ストループ効果実験における一致課題, 不一致課題の例	30
5.6	ストループ効果実験における探索	31
5.7	ストループ効果実験における探索結果 (左側頭部の拡大図)	31

5.8	ストループ効果実験における探索結果（右側頭部の拡大図）	32
5.9	ストループ効果実験における探索結果（頭頂部の拡大図）	32

付表

3.1	分割記号数が2から7のときの正規分布を等領域に分割する分割線	8
3.2	それぞれの文字列サイズにおける最適パラメータ	14
3.3	動作環境	14
4.1	実験パラメータ	17
4.2	類似部分のサイズ別の重複率の比較	18
4.3	それぞれの図における類似部分の長さの割合，長さ1あたりのDTW距離	23