

11th SICE Symposium on Computational Intelligence

June 26-27, 2017, Otsu

第11回コンピューターショナル・インテリジェンス研究会

講演論文集

期 日：2017年6月26日(月)，27日(火)

会 場：大津市旧公会堂

**SICE**<sup>®</sup>

主 催：計測自動制御学会 システム・情報部門

企 画：コンピューターショナル・インテリジェンス部会，知能工学部会

協 賛：情報処理学会，システム制御情報学会，電子情報通信学会，電気学会，日本  
神経回路学会，日本知能情報ファジィ学会，日本機械学会，人工知能学会，  
ヒューマンインタフェース学会，進化計算学会，IEEE Computational  
Intelligence Society Japan Chapter (CISJ)，IEEE Systems, Man, and Cybernetics  
Society Japan Chapter

著作権 © 2017

公益社団法人計測自動制御学会(SICE)  
〒101-0052 東京都千代田区神田小川町 1-11-9 金子ビル 4 階

カタログ番号 17 PG 0008

著作権は、計測自動制御学会がもっている  
ので、個人の使用のための複製以外の  
目的で掲載の記事の一部または全文を  
複製する場合には、著作権者に許可を求  
め規定の複製料を支払うこと。

発行日：2017年6月26日

発行者：公益社団法人計測自動制御学会 システム・情報部門  
ニューラルネットワーク部会， 知能工学部会

## 第 11 回コンピューテーショナル・インテリジェンス研究会プログラム

発表 25 分（発表 20 分、質疑応答 5 分）

6 月 26 日（月）

13:00~14:40（100 分 4 件） 司会 畠中利治（大阪大学）

- 1 針筋電図に混入する心電図抽出方法の提案 1  
山田早紀, 満倉靖恵（慶應義塾大学）
- 2 簡易脳波計を用いた脳波による疼痛有無の識別 5  
鍵田潤一郎, 満倉靖恵（慶應義塾大学）
- 3 ニューラルネットワークによる胃 X 線二重造影像における腫瘍領域検出システム 9  
磯川悌次郎, 西村明浩, 峯本俊文, 上浦尚武, 松井伸之（兵庫県立大学）
- 4 多目的最適化問題における評価時間の偏りが半非同期進化計算法に与える影響の分析 13  
原田智広（立命館大学）, 高玉圭樹（電気通信大学）

15:00~16:15（75 分 3 件） 司会 磯川悌次郎（兵庫県立大学）

- 5 近赤外分光法を用いた不快音聴取時における脳機能活性箇所の特化 21  
遠藤真央香, 満倉靖恵（慶應義塾大学）
- 6 乳がん病変検出のための深層学習を用いた計算機支援画像診断システム 24  
鈴木真太郎（東北大学）, 張曉勇（仙台高等専門学校）, 本間経康, 市地慶, 魚住洋佑,  
高根侑美, 柳垣聡, 川住祐介, 石橋忠司, 吉澤誠（東北大学）
- 7 深層学習による乳房 X 線画像上の腫瘍陰影鑑別 32  
鈴木真太郎（東北大学）, 張曉勇（仙台高等専門学校）, 佐々木拓也（京都大学）, 本間経康,  
市地慶, 魚住洋佑, 高根侑美, 柳垣聡, 川住祐介, 石橋忠司, 吉澤誠（東北大学）

16:30~17:45（75 分 3 件） 司会 本間経康（東北大学）

- 8 Design of an EEG-based brain-computer interface using motor imagery for wheel chair control 36  
Philemon Roussel, Yasue Mitsukura（Keio University）
- 9 前頭前野における脳波を用いた耳鳴の検出および苦痛度の推定 39  
松岡将司, 満倉靖恵（慶應義塾大学）
- 10 Improvement of Consumer-Grade Brain-Machine Interface Accuracy through Preliminary Signal Analysis 43  
Goussarov Gleb Andreevitch, Yasue Mitsukura（Keio University）

6 月 27 日（火）

9:30~11:10（100 分 4 件） 司会 内種岳詞（神戸大学）

- 11 セパトトリクスを実現する遺伝子ネットワークの設計問題の解の存在性と解法 47  
森禎弘, 黒江康明（京都工芸繊維大学）
- 12 状態跳躍を導入した Morris-Lecar 型ニューロンモデルにおけるカオス誘起現象の検討 55  
信川創, 吉田拓生（千葉工業大学）, 西村治彦（兵庫県立大学）, 山西輝也（福井工業大学）

13	適合度を用いた Re-labeling Differential Evolution の高効率化 有木大悟, 船木亮平, 村田純一 (九州大学)	60
14	ビジュアルリフティングアプローチによる二足歩行の安定性解析 神克礼, 李想, 田宏志, 井澤大時, 見浪護, 松野隆幸 (岡山大学)	66
11:20~12:20 Invitation to Computational Intelligence Fons Scientiae Project		
13:30~14:45 (75 分 3 件) 司会 田川聖治 (近畿大学)		
15	進化的実験計画法における実験数と感度分析精度の関係 内種岳詞 (神戸大学), 周晨婷, 畠中利治 (大阪大学)	73
16	CNN を用いた環境音スペクトログラムによる授業状況の推定に関する考察 北橋未先, 半田久志 (近畿大学)	78
17	機械学習を用いた高リスク学生の早期発見 近藤伸彦 (首都大学東京), 大久保緑, 畠中利治 (大阪大学)	82
15:00~16:15 (75 分 3 件) 司会 近藤伸彦 (首都大学東京)		
18	多目的機会制約問題に対するチェビシェフの不等式に基づく差分進化 田川聖治, 綿谷剛至 (近畿大学)	86
19	パズルゲームに対する強化学習を用いた汎化性能の向上 大西鴻哉, 飯間等 (京都工芸繊維大学)	93
20	球面上のダイジェーワールドモデルにおけるパターン形成 陰山真矢, 八木厚志 (大阪大学)	101

# 針筋電図に混入する心電図抽出方法の提案

○山田早紀 満倉靖恵 (慶応義塾大学)

## Electrocardiogram signal extraction from signal-channel Electromyogram signal

\*Saki Yamada and Yasue Mitsukura (Keio University)

**Abstract**— In this paper, we propose the method for extraction ECG signal from EMG signal measured by single-channel electrode. In this research, we perform optogenetics in mice. Optogenetics is a new technology which allows us to stimulate the targeted neurons illuminating the targeted area. In this study, we stimulate serotonin neurons in the dorsal raphe nucleus which would be related to anxiety, fear and depression. In order to analyze the effect of the stimulation, we measure bio-signals in our experiments. We especially focus on ECG signals to observe autonomic nervous system. However, there are problems of invasive and instability with ECG measuring. Then, we propose the method for extraction ECG signal from EMG measured by single-channel electrode using Constrained Non-negative Matrix Factorization. Since we are going to use RR intervals of ECG signal in the future, the accuracy of separation is evaluated by the number of R waves detected from extracted ECG signal.

**Key Words:** source separation, ECG, EMG

### 1 はじめに

本稿では、針電極を用いて測定した筋電図から精度よく心電図を抽出することを目的とする。

本研究は情動の研究を背景に持つ。例えば情動疾患に対する投薬に伴う身体反応の観察や、情動の神経メカニズムを解明するために動物実験は必要不可欠である。オプトジェネティクスという新しい手法は、遺伝子改変マウスを用いて、特定の神経細胞を発火させることができる。本研究では、マウスの背側縫線核に存在するセロトニン神経を任意のタイミングで発火させる。セロトニン神経の細胞体は、脳幹の正中部に位置する縫線核部に多く存在し、恐怖や不安、うつなどの情動に強く関与することが指摘されている [1, 2]。しかし、セロトニン神経の発火自体が恐怖や不安の増幅または減少のどちらに関与しているのかということや、神経細胞の存在位置とそれに対応する役割も未だ解明されていないことが多く、長い間議論が行われている [3-5]。オプトジェネティクスは目的の位置に存在する特定の神経細胞を発火させることができるため、この議論に結論をだすことができる手法だと期待されている。一般的に、オプトジェネティクスを行う際は、神経発火の確認や脳の反応を観察するために、深部脳波を同時に測定する。

本研究では深部脳波に加えて筋電図および心電図も取得する。筋電図は、深部脳波への混入が近年指摘されており、ノイズ除去の際のリファレンス信号として取得しておく必要がある。心電図は、心拍変動の解析に用いる。心拍変動の周波数成分を算出すると、交感神経・副交感神経の活動をとらえることができるとする先行研究が多くある [6, 7]。セロトニンは交感神経・副交感神経の調節および情動の調節を行う。そのため、オプトジェネティクスによりセロトニン神経を刺激している最中の心拍変動を解析することは妥当だといえる。

しかし、3つの生体信号計測に加えてオプトジェネティクスも行うため、マウスの侵襲性は非常に高い。それぞれの生体信号計測を行うための手術もしなければならぬ。さらに、本研究ではマウスを頭部固定しているため、首付近の電極は電極位置もずれることなく

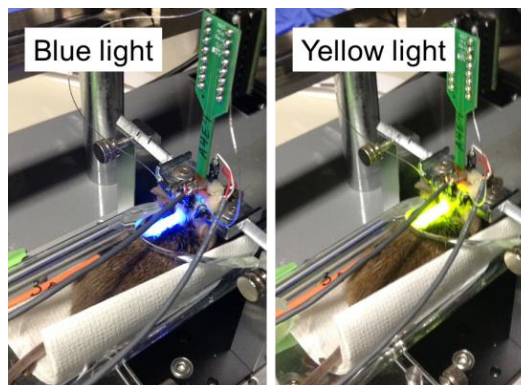


Fig. 1: Optogenetics experiments

計測を行えるが、心電図測定位置である腹側の胸や足付近はマウスが激しく動けば電極がずれてしまう可能性もある。以上より、本稿では一電極から取得した筋電図に信号分離手法を適用し、心電図を精度よく取得する方法を提案する。筋電図は振幅が大きくかつ非定常に混入している信号である。この中にある微小信号でありかつ定常的な信号の心電図の抽出を行う。マウスを用いて、筋電図および心電図を同時に計測し、筋電図から抽出した心電図を用いて算出した心拍変動と正解信号として測定した心電図の心拍変動を比較し、精度の評価を行う。

### 2 実験

本研究ではマウスの頭部を固定した状態でオプトジェネティクスを用いてマウスのセロトニン神経を操作する。その際に深部脳波、筋電図および心電図を計測する。

#### 2.1 オプトジェネティクス

オプトジェネティクスとは特定の神経細胞の発火を光で操作をすることができる手法である。まず、遺伝子改変技術を用いてマウスの特定のニューロンにチャネルロドプシン2 (channelrhodopsin-2; ChR2) を発現させる。ChR2は、青色光照射により開き、黄色光照射によりチャネルが閉じられるまで神経を活性化させる。光はマウスの頭蓋に穴をあけて照射する (Fig.1 参



Fig. 2: ECG mesaurment

照). 本研究では, マウスの背側縫線核のセロトニン神経に ChR2 の改変型である ChR2 (C128S 型) を出現させたマウスを用いる.

## 2.2 細胞外活動記録

細胞外活動記録法とは, 細胞膜におけるイオンの流入に伴って発生する電流を細胞外電位として記録する方法である. 深部脳波とよばれることもある. 計測部位は情動と関連が強いと言われている腹側海馬である. 計測機器には 16 チャンネルのシリコンプローブを用いる.

## 2.3 筋電図

本研究では, マウスの僧帽筋から針電極を用いて筋電図を取得する. サンプルング周波数は 24.4kHz である.

## 2.4 心電図

本研究では, マウスの腹側の胸の位置からワイヤー電極を用いて心電図を取得する. リファレンス電極は足位置に設置する (Fig.2 参照). Fig.2 はマウスの仰向けの状態の図であり, ×印の部分が電極位置を表す. サンプルング周波数は 24.4kHz である.

## 3 解析

本稿では, 取得した筋電図に対して信号分離を行い, 正解信号として取得した心電図との比較を行う.

### 3.1 前処理

はじめに, 得られた筋電図および心電図に対して 1017.3kHz までダウンサンプリングを行う. 次に, ダウンサンプリングを行ったデータに対して 30-250Hz のバンドパスフィルターを適用する. この帯域は実験により経験的に決定した.

### 3.2 信号分離

#### (a) 学習データの準備

学習データは, 正解信号として取得した心電図を用いる.

#### (b) 周波数解析

学習データおよび取得した筋電図を短時間フーリエ変換を用いて周波数領域に変換する. 窓幅はサンプルング周波数と同様に 1017.3kHz とし, シフト量は 80 とした. これは, およそ 80 サンプル前後で心拍の波が出現するためである. 区間  $[0, I-1]$  の  $I$  点で与えられた離散時間信号  $y(i)$  の離散フーリエ変換  $F(k)$  は, 次式

で定義される  $I$  個の複素数列である.

$$F(k) = \sum_{i=0}^{I-1} y(i) e^{-j \frac{2\pi k i}{I}} \quad (k = 0, 1, 2, \dots, I-1) \quad (1)$$

この複素関数の絶対値  $|F(k)|$  は振幅スペクトル, 2乗値はパワースペクトルである.

#### (c) 非負値因子行列因子分解

本稿では, 信号分離のために非負値行列因子分解を用いる. この手法は多変量解析手法の 1 つであり, 振幅スペクトルやパワースペクトルのような非負値データの行列を 2 つの非負値行列  $\mathbf{H}$  および  $\mathbf{W}$  に分解することを目的としている. 行列を計算する際に全ての値に非負制約を設けるため, 非負値因子行列因子分解とよばれる. (b) より取得した筋電図の  $N$  個の非負値 (パワースペクトル) ベクトル  $\mathbf{p}_1, \dots, \mathbf{p}_N \in \mathbb{R}^{\geq 0, M}$  を得る. ただし,  $M$  は非負値ベクトルの次元を表す. このベクトルを観測ベクトルと呼ぶ.

非負値行列因子分解では, 近似的に加法性が成り立つと仮定し, この観測ベクトル  $\mathbf{p}_n (n = 1, \dots, N)$  を,  $K$  個の非負の基底ベクトルの重み付き和によって表わされたものとみなして分解する. すなわち, 基底ベクトル  $\mathbf{h}_1, \dots, \mathbf{h}_K \in \mathbb{R}^{\geq 0, M}$  と時間方向変化量のアクティベーション  $w_{1,n}, \dots, w_{K,n} \in \mathbb{R}^{\geq 0, M}$  に基底数  $K$  だけ分解する.

$$\mathbf{p}_n \approx \sum_{k=1}^K \mathbf{h}_k w_{k,n} \stackrel{\text{def.}}{=} \mathbf{x}_n \quad (2)$$

ここで,  $\mathbf{x}_n$  は再構成ベクトルである. また,  $K \ll \min(M, N)$  であり,  $\mathbf{x}_{k,n} = \mathbf{h}_k w_{k,n}$  とすると,  $\mathbf{x}_n = \sum_k \mathbf{x}_{k,n}$  が成り立つ.

観測ベクトルを並べたものを  $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_N] \in \mathbb{R}^{M \times N}$ , 基底ベクトルを並べたものを  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K] \in \mathbb{R}^{M \times K}$ , アクティベーションベクトルを並べたものを  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_K] \in \mathbb{R}^{M \times N}$ , さらに再構成ベクトルを  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{M \times N}$  とすると, (2) 式は以下の式に書き直すことができる.

$$\mathbf{P} \approx \mathbf{H} \mathbf{W} \stackrel{\text{def.}}{=} \mathbf{X} \quad (3)$$

非負値行列因子分解では, 観測ベクトル  $\mathbf{p}_n$  と再構成ベクトル  $\mathbf{x}_n$  の誤差を最小化して分解を行う. 本稿では, この誤差の評価関数に IS ダイバージェンスを用いる. ほかの評価関数としては, KL ダイバージェンスや EU ダイバージェンスが一般的に利用されているが, パワースペクトルには IS ダイバージェンスが適している [8]. さて, 誤差を最小化して  $\mathbf{H}$  および  $\mathbf{W}$  を求めることは, 非負制約つき非線形最適化問題であり, 解析的に解くことはできない. そこで, 補助関数を用いるがそのアルゴリズムについては本稿では割愛する. 最終的に, 初期値をランダムな非負値に設定し, 以下の乗法更新式を用いて  $\mathbf{H}$  および  $\mathbf{W}$  を求めることができる.

$$h_{m,k} \leftarrow h_{m,k} \left( \frac{\sum_n p_{m,n} w_{k,n} / x_{m,n}^2}{\sum_n w_{k,n} / x_{m,n}} \right)^{1/2} \quad (4)$$

$$w_{k,n} \leftarrow w_{k,n} \left( \frac{\sum_m p_{m,n} h_{m,k} / x_{m,n}^2}{\sum_m h_{m,k} / x_{m,n}} \right)^{1/2} \quad (5)$$

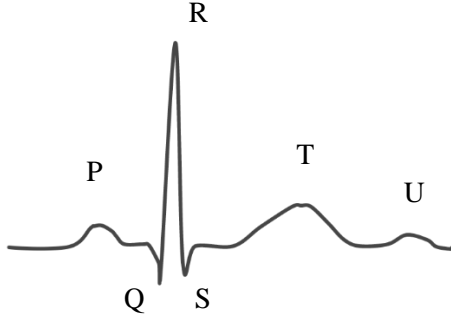


Fig. 3: Electrocardiography

(d) 教師基底の固定および時間制約項

本稿では, (a) で作成した学習データにまずはじめに非負値行列因子分解を行う. 学習データのパワースペクトル  $P_{ECG}$  を基底行列  $H_{ECG}$  およびアクティベーション行列  $W_{ECG}$  に基底数  $K_1$  だけ分解する. 次に, 取得データに対して非負値行列因子分解を行うが, この時に学習データの非負値因子行列因子分解から得られた基底  $H_{ECG}$  を固定して基底数  $K_2$  の分解を行う. 本稿では, 実験より経験的に  $K_1 = 6, K_2 = 12$  と決定した.

また, 抽出をしたい心電図は定常的な信号である一方, 筋電図は非定常に大きい振幅で混入するというデータの特徴を利用する. 前述のとおり,  $H$  および  $W$  は (4), (5) 式を繰り返して求められるが, (5) 式のアクティベーション行列に対して時間的制約を設け,  $H_{ECG}$  に対応するアクティベーション行列は時間方向で大きく変化をしないという制約を与える.

$$J = \sum_{m,n} D_{IS}(P, HW) + \sum_{k,n} (w_{k,n} - w_{k,n-1})^2 \quad (6)$$

(6) 式の第2項は, 時間変化が小さいほど小さい値をとるため, この項をコスト関数に入れることでアクティベーション行列の時間方向に制約を与える.

最後に信号を再構成するとき, 下記の (7) 式のとおり, 固定した基底を用いて再構成をする.

$$P'_{ECG\ m,n} = p_{m,n} \times \frac{\sum_{k=1}^{K_1} h_{m,k} w_{k,n}}{\sum_{k=1}^K h_{m,k} w_{k,n}} \quad (7)$$

### 3.3 R波の特定

心電図は本来, P, Q, R, S, T, U という波を持っている (Fig.3 参照). 本研究では, R波の間隔を周波数解析することで自律神経系の評価を行うという目的があるため, R波のみに着目する. R波は過去10サンプルぶんのRR間隔を用いて窓幅を変化させながら, ピークとなる箇所を算出する. この方法を用いると, はじめにR波を正解値とずれて検出した箇所から, 少しずつ誤差が重なっていく可能性があるが, ピーク値の値を閾値として設定することでずれを修正する探索方法である. しかし, ピーク値を使用するには信号分離の際にピーク信号が抽出できている必要がある.

## 4 結果および考察

### 4.1 信号分離

信号分離結果の一部を Fig.4 に示す. 横軸は時間, 縦軸はマイクロボルトである. 測定された筋電図が黒破

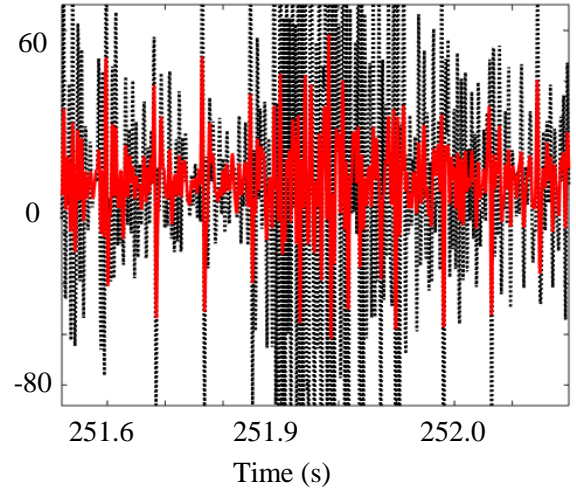


Fig. 4: ECG separation

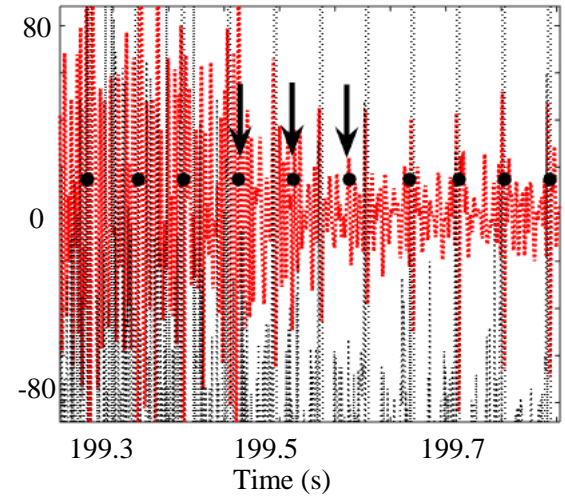


Fig. 5: R wave detection

線, 抽出された信号が赤実線で示されている. 抽出信号が測定信号に比べて信号が減衰している様子がみえる. これは, 分解の基底数が少ないためである. 一般的に, 基底数を増やすとより多くの信号を表現できるようになるため, 信号の振幅の減衰は抑えられるが, 一方で不要である筋電図も表現できるようになってしまうため, 信号の分離がより困難になる. 本稿では, 経験的に基底数を決定しているが, 基底数の決定には更に検討が必要である. また, 教師基底に対応するアクティベーションに時間制約を設けているため, 筋電図が多く混入している部分 (Fig.4 251.9s 付近) からも心電図に似た波形が抽出できている様子が見られる. 時間制約のパラメータも実験から決定した.

### 4.2 R波検出による評価

正解信号として同時に計測していた心電図からR波を検出し, RR間隔を求める. 抽出した心電図からも同様に算出したRR間隔の比較を行った. データ長は30分であり, 正解信号からは21454個, 抽出信号からは21382個のR波を検出した. Fig.5にR波検出の様子を示す. Fig.4同様に横軸は時間, 縦軸はマイクロ

ボルトである。黒点線は心電図の正解信号，赤点線は抽出した心電図，●は抽出信号を用いて検出した R 波の位置を示す。筋電図の混入が大きく，心電図抽出がうまくできていないところは R 波の検出も失敗している様子が分かる (Fig.5 の矢印部分)。R 波の検出された個数のみではなく，それぞれの位置も評価基準に加える必要がある。R 波の検出が正解値と比較して前後 10 サンプル (およそ 0.01s) 以上位置を誤って検出しているデータは 21382 個のうち，1786 個あった。すなわち，全ての R 波に対して 91.3% の R 波の検出に成功している。

## 5 おわりに

本稿では，針筋電図に混入する心電図を抽出する方法を提案した。他個体や他記録においても同様の学習データやパラメータを用いることができるのかということは今後検討する必要がある。測定によって SN 比が異なるので，学習データにも人為的にノイズをのせる必要があると考えられる。また，非負値行列因子分解において重要なパラメータである基底数や時間制約項のパラメータも現在は実験によって決めているが，計算的に求める方法も試してみる必要がある。今後は，抽出した心電図を用いて RR 間隔の周波数解析を行う。信号分離の評価法として周波数成分を評価指標とすることも検討している。

## 謝辞

本研究は，JSPS 科研費基盤研究 (A)YYK5A05 の助成を受けたものです。

## 参考文献

- 1) Canli Turhan Lesch and Klaus-Peter: "Long story short: the serotonin transporter in emotion regulation and social cognition", *Nature neuroscience*, Vol.10, pp.1103-1109(2007)
- 2) Lira Alena Zhou et al.: "Altered depression-related behaviors and functional changes in the dorsal raphe nucleus of serotonin transporter-deficient mice", *Biological psychiatry*, Vol.54, pp.960-971(2003)
- 3) Marcinkiewicz Catherine A et al.: "Serotonin engages an anxiety and fear-promoting circuit in the extended amygdala", *Nature*, Vol.537, pp.97-101(2016)
- 4) Ohmura Yu et al.: "Optogenetic activation of serotonergic neurons enhances anxiety-like behaviour in mice", *International Journal of Neuropsychopharmacology*, Vol.17, pp.1777-1783(2014)
- 5) Miyazaki Kayoko W et al.: "Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards", *Current Biology*, Vol.24, pp.2033-2040(2014)
- 6) Pagani Massimo et al.: "Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog", *Circulation research*, Vol.59, pp.178-193(1986)
- 7) Kemp Andrew H et al.: "Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis", *Biological psychiatry*, Vol.67, pp.1067-1074(2010)
- 8) 亀岡弘和: 非負値行列因子分解，計測と制御，vol.51, no.9, pp.835-844(2012)



# 簡易脳波計を用いた脳波による疼痛有無の識別

○鍵田 潤一郎 満倉 靖恵 (慶應義塾大学)

## Classification of Pain Using the Simple EEG Device

\* J. Kagita, Y. Mitsukura, (Keio University)

**Abstract**— The final aim in this paper is to classify pain degree by only using the simple EEG (electroencephalogram) device. In clinical care, pain degree is imperative for choice and assessment of the treatment. Pain rating scales such as the VAS (Visual Analogue Scale) and NRS (Numerical Rating Scale) are usually used to attain pain degree, although they are only capable of acquiring subjective values. From that, study for the completion of a system which can measure objective degree of pain has great importance. As the preceding studies have suggested a decrease of grey matter in subjects who have experienced pain, we first tried to achieve the difference in EEG between subjects who have experienced pain and not. Using frequency analysis and SVM (Support Vector Machine), we have classified 100% in LOOCV (Leave-One-Out Cross Validation) accuracy. We tried to achieve the difference in EEG while the subject is in pain and pain-free. Using the same methods as above, we achieved 97% LOOCV accuracy. These results show that classification of pain degree using only the simple EEG device may be feasible.

**Key Words:** Pain, Electroencephalogram, Signal processing, Machine learning

## 1 はじめに

本稿は、簡易脳波計から得られる脳波のみを用いた疼痛有無の識別を目的とする。

疼痛は痛みを意味する医学的用語であり、侵害刺激が脳に伝わることで引き起こされる感覚的要因に加え、非感覚的要因が複雑に関連するため測定および評価が非常に難しいとされている<sup>1)</sup>。ここで非感覚的要因とは、感覚的要因が作用することによって生じる感情的要因および認知的要因を意味する。

医療現場において患者が有する疼痛の度合いを把握することは、治療法の選択および評価に必要である。一般的に患者が有する疼痛の度合いを把握するためには、VAS(Visual Analogue Scale)やNRS(Numerical Rating Scale)といった評価手法が用いられる<sup>2)</sup>。VASとは10cmの直線上で左端0cmの位置を全く痛みのない状態、右端10cmの位置を想像し得る中で最も痛みのある状態とし、患者自身に自分の感じる疼痛の度合いを直線上に表してもらう方法である。NRSとは患者自身に自分の感じる疼痛の度合いを0から10の11段階の点数で表してもらう手法である。これらは場所を選ばず容易に実施することが可能である一方で、患者の主観的な評価に偏る。そのため一人の患者内で疼痛度合いの変遷を計測するには有効であるが、異なる患者間での比較ができないといった問題点がある。従って、疼痛度合いの評価における客観的な指標構築のための研究は重要な研究であると言える。疼痛の測定において考慮すべき点は有用性、信頼性、および妥当性である<sup>3)</sup>。簡易脳波計は装着に要する時間が非常に短く、頭部への締め付けが小さいことから患者への負担も軽いため、有用性に優れている。また簡易脳波計から得られる脳波は再現性が高いため信頼性に優れている。最後に、疼痛の伝達路を考えると妥当性に優れている。疼

痛は侵害刺激が侵害受容器に加わって引き起こされる感覚的要因が前頭前野に伝わることで認知する<sup>4)</sup>。また前頭前野はヒトの感情とも関わっていることが知られている。従って、前頭前野部位の脳波が計測可能な簡易脳波計を用いることで、疼痛がもたらす感覚的要因および非感覚的要因を計測できる可能性が高いため、妥当性に優れていると言える。以上より、簡易脳波計による疼痛度合いの識別は考慮すべき点を全て満たすことができるため、優れた測定法であると言える。

本研究の最終目的は、簡易脳波計から得られる脳波のみを用いた疼痛度合いの識別である。そこで今回の目的は最終目的の前段階として、簡易脳波計から得られる脳波のみを用いた疼痛有無の識別とした。まず、継続的な疼痛の経験年数に比例して、疼痛に関与するとされている脳部位における灰白質の体積が減少しているという症例から<sup>5)</sup>、疼痛を継続的に経験したことがある被験者と経験のない被験者の脳波を比較した。その後、疼痛度合いの識別をするために疼痛がある時の脳波とないときの脳波の特徴量を抽出し、それらの比較検討を行った。

## 2 実験

実験機器にはサンプリング周波数が512 Hzで左前頭前野部位(Fp1)から脳波が取得可能であるMindWave Mobile(簡易脳波計)を用いた。Fp1は国際10-20法によって定められた計測部位であり、その位置をFig. 1に示す。また、同じくFig. 1にMindWave Mobileを装着した様子を示す。上記の簡易脳波計を用いて、各被験者に対し暗く静かな部屋にて安静閉眼状態で1分間の計測を行った。

被験者は全て20代で、男性健常被験者12名、女性生理痛被験者14名、女性片頭痛被験者1名であっ

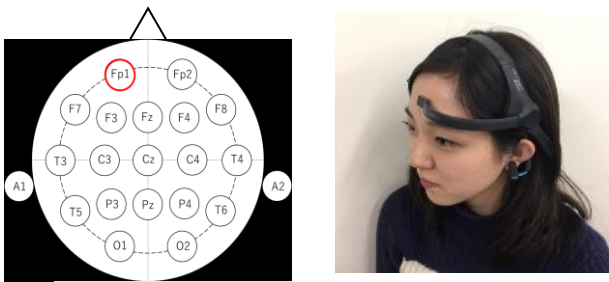


Fig. 1: International 10-20 system and the mounted Mind Wave Mobile

た。女性生理痛被験者および女性片頭痛被験者には、生理痛および片頭痛を感じている際と、感じていない際の両方で脳波計測を行った。

本稿では今後データを扱う際に、女性生理痛被験者および女性片頭痛被験者が疼痛を感じている際に取得したデータをそれぞれ生理痛時データおよび片頭痛時データと呼び、両者をまとめて疼痛時データと呼ぶ。また、女性生理痛被験者および女性片頭痛被験者が疼痛を感じていない際に取得したデータを平常時データと呼び、男性健常被験者から取得したデータを健常被験者データと呼ぶ。

### 3 解析

平常時データと疼痛時データの識別、および健常者データと平常時データの識別で用いた解析手順について述べる。上記 2 つの識別において、前処理と周波数解析は同様に行い、特徴量選択と識別は異なる過程で行った。

#### 3.1 前処理と周波数解析

前処理では、得られたデータに対してトレンド除去を施した後、1 秒ごとのエポックを作った。エポックを作る際に、設けた閾値を超えてより明らかに体動や瞬きから由来すると思われるノイズを含んだエポックは除外した。その後通過帯域が 1-35 Hz のバンドパスフィルタで筋電の影響を除去した後、平均 0 分散 1 となるよう算出することで被験者間の電圧差を正規化した。上記のような前処理を施すことで、脳波以外の大きなノイズを取り除くことが可能であると考えている。

周波数解析には、FFT (Fast Fourier Transform) を用いた後、周波数毎のパワースペクトル密度 (PSD: Power Spectral Density) 値推定を行った。FFT は式 (1)、PSD 値は式 (2) で算出を行った。

$$F(\Omega) = \sum_{-\infty}^{+\infty} f(n)e^{-j\Omega n} \dots \dots \dots (1)$$

$$PSD(\Omega) = |F(\Omega)|^2 \dots \dots \dots (2)$$

$f(n)$  は離散時間信号、 $F(\Omega)$  は  $f(n)$  のフーリエ変換、 $PSD(\Omega)$  は  $F(\Omega)$  の PSD 値を示す。

#### 3.2 健常者データと平常時データの識別

健常者データと平常時データを識別するために、前処理と周波数解析で算出した平常時データと生理痛時データの各周波数におけるパワースペクトル密度値に対して二標本  $t$  検定を行った。特徴量は二標本  $t$  検定で算出した  $t$  値の大きい周波数を順に増加させて用いた。  $t$  値は式 (3) で算出される。

$$t = (\bar{X}_1 - \bar{X}_2) / \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} \dots \dots \dots (3)$$

式 (3) において  $\bar{X}_1$  および  $\bar{X}_2$  は各標本の平均、 $m$  および  $n$  は各標本数、 $S_1$  および  $S_2$  は各標本分散を示す。  $t$  値が大きいということは、二標本間ではその特徴量で有意差が大きいことと等しいので、特徴量として  $t$  値の大きい周波数を順に増加させていく手法は高い識別率を得るために有用な手法であると考えられる<sup>6)</sup>。

その後 RBF カーネルを用いた SVM (support Vector Machine) により識別を行った。SVM とはクラス分類手法の一つである。識別面から最小距離に位置するデータをサポートベクタ、識別面からサポートベクタの距離をマージンと呼び、SVM ではそのマージンが最大となるような識別面を決定する。SVM に用いる特徴量は全て、1 秒のエポックを用いて算出したパワースペクトル密度値を被験者毎に平均し、その後周波数毎に最小値 0 最大値 1 となるようスケールした値を用いた。健常者データと平常時データの識別を行うために、健常者データと平常時データにおける特徴量を使用して識別器を学習した。識別器の評価には LOOCV (Leave-One-Out Cross-Validation) を用いた。LOOCV とは、全標本から 1 つの標本を抜き出してそれ以外を学習に用い、学習後抜きだした標本で評価を行うという過程を全標本が評価データに用いられるまで繰り返す評価法である。

#### 3.3 平常時データと疼痛時データの識別

健常者データと平常時データの識別と同様に、健常者データと平常時データを識別するために二標本  $t$  検定をこれらのデータに対して行った。

識別では平常時データと生理痛時データにおける特徴量を使用して識別器を学習し、評価は LOOCV を用いた。また、上記で学習した識別器で片頭痛時データおよび健常者データを識別し、識別率を算出した。

## 4 結果および考察

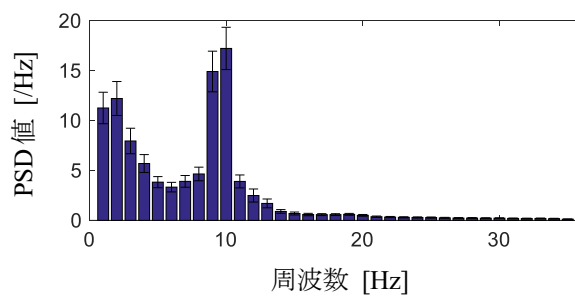
### 4.1 健常者データと平常時データにおける結果

Table 1 に健常者データと平常時データにおける 2 標本  $t$  検定の結果を示す。Table 1 内で赤く示された周波数は有意差が確認できた周波数である（優位水準  $\alpha=0.001$ ）。

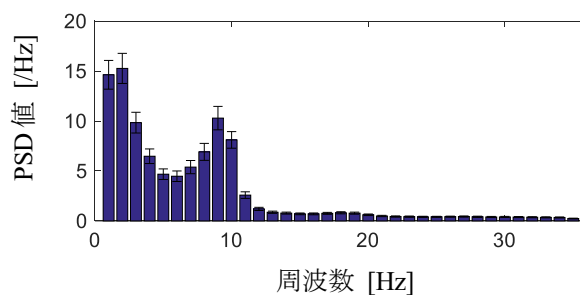
Table 1 より、31 個の周波数で有意差が確認することができた。 $t$  値の絶対値を見ても、最大で 10 Hz の際に 17.9 となっており、有意差が顕著に表れていることがわかる。

次に健常者データと平常時データを用いた識別結果について述べる。 $t$  値の大きい周波数を順に特徴量として用いたところ、1-35 個の特徴量全てのときに LOOCV では 100% の識別率を得た。

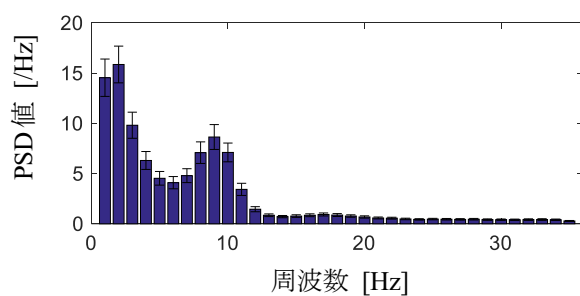
高い識別率の理由としては、推測になってしまうが、疼痛による眼窩前頭皮質の体積減少が考えられる。継続的な疼痛により、疼痛に関連する脳部位の灰白質が減少する事例は発見されており、特に眼窩前頭皮質において体積の減少が報告されている<sup>5)</sup>。疼痛に関連する脳部位における灰白質の減少が、疼痛の原因なのか、若しくは結果なのかは未だ明らかとなっていないが、本研究の健常者データと平常時データにおける有意差は上記のことが理由であると推測できる。ここで Fig. 2 に健常者データ、平常時データおよび生理痛時データにおける 1-35 Hz の PSD 値を示す。Fig. 2 より、平常時データを示す(b)と生理痛時データを示す(c)の概形が類似していることが分



(a)



(b)



(c)

Fig. 2. PSD Value of 1-35 Hz for respective data

(a): Healthy Control data

(b): Normal data

(c): Menstrual data

Table 1:  $t$  ratio between Healthy Controls and pain-free subjects

周波数 (Hz)	$t$ 値	$p$ 値	周波数 (Hz)	$t$ 値	$p$ 値
1	-5.9	4.1E-09	19	-4.5	6.6E-06
2	-5.1	4.4E-07	20	-3.5	4.7E-04
3	-4.4	1.1E-05	21	-4.9	8.9E-07
4	-2.6	9.7E-03	22	-4.6	4.7E-06
5	-4.1	4.4E-05	23	-5.3	1.4E-07
6	-5.8	9.4E-09	24	-5.4	7.5E-08
7	-6.0	2.6E-09	25	-6.3	3.7E-10
8	-7.3	5.0E-13	26	-8.6	1.4E-17
9	8.2	4.3E-16	27	-9.7	1.2E-21
10	18	6.0E-67	28	-8.1	1.2E-15
11	7.7	1.5E-14	29	-8.6	2.1E-17
12	9.1	3.2E-19	30	-9.0	3.9E-19
13	8.6	1.8E-17	31	-9.9	9.2E-23
14	2.3	2.3E-02	32	-9.9	1.1E-22
15	-0.44	6.6E-01	33	-8.5	3.9E-17
16	-3.1	2.0E-03	34	-9.4	1.3E-20
17	-4.9	8.7E-07	35	-8.8	2.9E-18
18	-6.0	2.1E-09			

かる。上記の事象から、疼痛を感じたことのないヒトの脳波は(a)のような概形であるが、継続的な疼痛により疼痛を感じていない際の脳波も(c)のような概形に変化する可能性が考えられる。従って簡易脳波計を用いて疼痛の度合いを識別するためには、疼痛を感じたことのある期間や頻度も考慮する必要があると言える。

### 4.2 平常時データと疼痛時データにおける結果

Table 2 に平常時データと生理痛時データにおける 2 標本  $t$  検定の結果を示す。Table 2 内で赤く示された周波数は有意差が確認できた周波数である（優位水準  $\alpha=0.001$ ）。

Table 2 の結果を見ると、6 個の周波数で有意差が確認できる。これは、疼痛による精神状態の変化およ

Table 2: *t* ratio between pain-free and menstrual subjects

周波数(Hz)	<i>t</i> 値	$\rho$ 値	周波数(Hz)	<i>t</i> 値	$\rho$ 値
1	0.14	8.9E-01	19	0.15	8.8E-01
2	-0.98	3.3E-01	20	-1.9	6.2E-02
3	0.056	9.6E-01	21	-3.0	2.7E-03
4	0.57	5.7E-01	22	-4.3	2.1E-05
5	0.67	5.0E-01	23	-2.1	3.7E-02
6	1.8	6.9E-02	24	-1.2	2.4E-01
7	2.3	1.9E-02	25	-2.6	9.7E-03
8	-0.49	6.3E-01	26	-1.2	2.5E-01
9	3.7	2.6E-04	27	0.062	9.5E-01
10	3.1	2.2E-03	28	-1.6	1.0E-01
11	-5.1	3.1E-07	29	-0.93	3.5E-01
12	-3.4	6.7E-04	30	-0.43	6.7E-01
13	0.38	7.0E-01	31	-0.65	5.1E-01
14	1.8	7.6E-02	32	-1.9	5.7E-02
15	-1.6	1.2E-01	33	-2.7	6.8E-03
16	-3.5	4.5E-04	34	-2.5	1.2E-02
17	-3.9	1.1E-04	35	-2.8	5.8E-03
18	-1.2	2.1E-01			

び疼痛を伝える神経活動から由来する信号が前頭前野部位の脳波に現れていると推測できる。前野部位はヒトの精神状態を現す一方で、疼痛を伝える信号の伝達路となる。従って、本研究で確認された平常時データと生理痛時データにおける有意差は精神状態の変化および疼痛を伝える神経活動によるものと考えられるのは妥当であると言える。

次に、平常時データと生理痛時データの識別結果について述べる。SVMにおいて*t*値の大きい周波数から9個を特徴量としたとき、LOOCVにて97.0%の識別率を得た。上記の識別器で片頭痛時データと健常者データを識別させたとき、片頭痛時データを疼痛があると識別し、91.7%の健常者データを疼痛がないと識別した。

平常時データと生理痛時データの識別率は97.0%と高く、簡易脳波計から得られる脳波のみを用いて平常時と生理痛時を識別することが可能であると言える。また、上記の識別器を用いて片頭痛データも疼痛のあるデータとして正しく識別できたことに関しては、疼痛を感じる場所が異なる場合も、脳内において疼痛を処理する伝達路が同じであるため、前頭前野部位の脳波に現れた傾向が類似していたことが理由であると考えられる。従って、本研究において簡易脳波計から得られる脳波のみを用いて平常時と疼痛時を識別することができたと言える。しかし、本研究で片頭痛被験者は1名と少ないため、結果の精度を向上させるためには様々な疼痛において被験者を増加させる必要がある。また、健常者データの91.7%が疼痛のないデータとして正しく識別されていたこ

とに関して、疼痛有無の識別においては健常者データと平常時データの有意差は考慮しなくてよいと言える。しかし今後疼痛の度合いを識別するためには患者の初期状態にも考慮が必要となるため、健常者データと平常時データの有意差にも考慮しなければならないと考えられる。

## 5 おわりに

本研究の最終目的は、簡易脳波計から得られる脳波のみを用いた疼痛度合いの識別である。最終目的を達成するための前段階として本稿の目的は、簡易脳波計から得られる脳波のみを用いた疼痛有無の識別とした。実験では簡易脳波計を用いて20代被験者27名の脳波を計測した。その後得られた脳波に対して前処理、周波数解析、特徴量選択、および識別を行った。解析の結果、簡易脳波計から得られた脳波のみを用いて、疼痛を感じていないときの被験者と疼痛を感じているときの被験者をLOOCVにて最高で97.0%の識別率で識別可能であることが分かった。また、疼痛を感じたことがない被験者と、慢性疼痛を感じたことがある被験者では有意差があることも確認できた。しかしその有意差は、疼痛有無の識別においては考慮しなくてもよい差であることを確認した。

今後、最終目的を達成するためには、様々な種類の疼痛を感じている際の脳波を解析し、特徴を掴む必要がある。

## 謝辞

本研究は、JSPS 科研費基盤研究(A)YK5A05の助成を受けたものです。

## 参考文献

- 1) 東山篤規, 宮岡徹, 谷口俊治, 佐藤愛子: 触覚と痛み, 株式会社おうふう (2009)
- 2) 平川奈緒美: 痛みの評価スケール, *Anesthesia* 21 Century, Vol.13, pp.2538-2544 (2011)
- 3) Jenny Strong, Anita M. Unruh, Anthony Wright, and G. David Baxter: Pain: A Textbook for Therapists, Churchill Livingstone (2002)
- 4) Anthony K. P. Jones, Nathan T. M. Huneke, Donna M. Lloud, Chris S. Brown, and Alison Watson: Role of Functional Brain Imaging in Understanding Rheumatic Pain, *Curr Rheumatol Rep*, Vol.14, No.6, pp.557-567 (2012)
- 5) Karen D. Davis, and Massieh Moayedi: Central Mechanism of Pain Revealed through Functional and Structural MRI, *Journal of Neuroimmune Pharmacology*, Vol.8, No.3, pp.518-534 (2013)
- 6) 高橋弘武, 古川大弘, 古橋武, 田中一男: 統計的手法による変数選択を用いた脳波による思考認識に関する一考察, 22<sup>nd</sup> Fuzzy System Symposium, pp.518-522 (2006)

# ニューラルネットワークによる胃 X 線二重造影における腫瘍領域検出システム

○磯川悌次郎 西村明浩 峯本俊文 上浦尚武 松井伸之 (兵庫県立大学)

## A Tumor Detection System on Gastrointestinal Walls in Double Contrast X-ray Images based on Autoencoding Neural Network

\*T. Isokawa, A. Nishimura, T. Minemoto, N. Kamiura, and N. Matsui (University of Hyogo)

**Abstract**— Double contrast (DC) X-ray images are useful and cost-effective for the diagnosis of stomach tumors. However, it is difficult to automatically extract tumors from these images, due to the variations in the tumors appearing in the images. This paper presents an automated scheme for detecting tumors in DC X-ray images, based on feedforward neural networks as autoencoders for patches in DC X-ray images. The performances of the scheme are demonstrated by using artificial and actual DC X-ray images.

**Key Words:** Double Contrast method, tumor, autoencoder, feedforward neural network

### 1 はじめに

胃がんは日本において罹患率・死亡率ともに上位3位に入るがんであり<sup>1)</sup>, その早期発見は非常に重要である。胃 X 線二重造影法は死亡率減少が科学的に証明されている検査であり<sup>2)</sup>, 胃がん検診で用いられる検査手法の一つとなっている。

一般的な集団検診においては, 被験者一人当たり10枚程度の X 線画像が放射線技師により撮影され, それらを元に医師が読影・診断を行う。そのため, がんなどの病変の発見は X 線画像の読影を行う医師の能力に大きく依存している。医療機器の発展に伴い従来よりも高速かつ大量に情報が得られるようになった反面, それらを処理する医師の負担が非常に大きくなっているのが現状である。また, 読影技術の習得には多大な時間を要するため, 読影を行う専門医の不足も問題となってきた<sup>3)</sup>。そこで, 医師の読影の負担軽減や医療過誤の防止などを目的としたコンピュータによる診断支援システム (Computer Aided Diagnosis System: CAD) の開発が進められている<sup>4)</sup>。

しかし, 胃 X 線二重造影の CAD 開発は, 実用化の段階まで進んでいる肺や乳房対象の CAD に比べると遅れているのが現状である。その理由として, 胃は柔軟な組織であり, 撮影条件の違いにより大きく変形してしまうことが挙げられる。これまでに病変検出システムとして, 胃壁のひだ集中パターンを評価したシステム<sup>5)</sup>, 輝度勾配ベクトルの放射分布度に基づく研究<sup>6)</sup>, および胃粘膜表面模様のフラクタル次元に着目した研究<sup>7)</sup>などが存在する。

これまでに報告された研究は, 手動による特徴量設計に基づいた画像パターン検出を採用したものがほとんどである。これらの研究は腫瘍の特徴に着目しており, 見逃しを防ぐためにはさまざまな種類の腫瘍に対応する必要がある。しかし, 腫瘍の形状の多様性や胃の柔軟な特性を考えると, あらゆる腫瘍に対応することは極めて困難である。より多様な腫瘍形状を認識する手法としてニューラルネットワークを用いた手法<sup>8)</sup>も検討されている。この手法では, 胃 X 線画像から算出した特徴量をニューラルネットワークに入力し, 出力を正常・異常の2クラスとしている。この手法によ

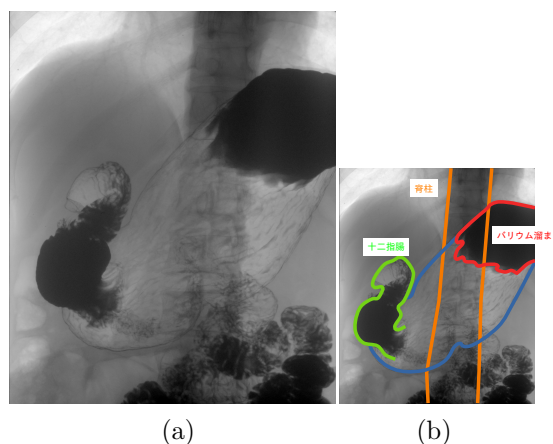


Fig. 1: An image by double contrast gastrography. (a) Whole of an image and (b) subregions in the image

る問題点の一つとして, 胃 X 線画像において腫瘍を含む画像, すなわち異常のクラスに属する画像が少数しかないことや, 学習した異常パターンしか検出できないことが挙げられる。本論文では腫瘍領域よりも多い正常領域に着目し, ニューラルネットワークの一種であるオートエンコーダに正常画像の圧縮・復元を学習させる。このオートエンコーダにより入力画像の再構成誤差を算出し, 正常・異常の判断する。これにより従来法では判断できなかった腫瘍についても検知することが期待できるとともに, システムが異常と判断した領域についてのみ医師が診断することにより, 医師の負担軽減を図ることができる。本論文では胃 X 線画像を模した人工パターンと胃 X 線画像で実験を行い, 提案法の性能を明らかにする。

### 2 胃 X 線二重造影法

硫酸バリウムと発泡剤による炭酸ガスの X 線透過率の差により胃の壁面の凹凸を観測する手法が胃 X 線二重造影法である。胃 X 線二重造影法より撮影された胃 X 線二重造影の例を Fig. 1 に示す。図中では黒い領域として写っているバリウム溜まり, バリウムが流出した十二指腸や脊柱が胃の領域と重なって撮影される

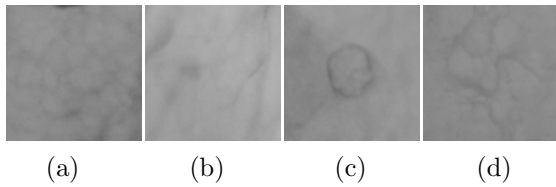


Fig. 2: Patches obtained from a double contrast X-ray image. (a) and (b) Patches without tumors, (c) a patch containing a benign tumor, and (d) a patch containing a malignancy tumor

とその領域の診断は困難である。体位変換を行いながら複数枚撮影することで、胃全体の診断が可能になり、同時に胃壁面に造影剤を流すことが可能となる。

胃では部位により粘膜が異なる。噴門（口側）に近い領域では、平行に走る体部ひだが多く、幽門（肛門側）に近い領域では、網目状の模様として観測される幽門腺粘膜が存在する。体部ひだがある限界を結んだ線を（肉眼的な）F境界線と呼ぶ。病変は隆起性病変と陥凹性病変に分類できる。隆起性の場合バリウムが押しつけられ、陥凹性の場合はその部分にバリウムが溜まるため、濃淡として観測できる。また、胃形の異常やひだの集中を起こすことも多い。正常な胃と腫瘍の例を Fig. 2 に示す。

### 3 胃腫瘍検出システム

#### 3.1 検出法

提案手法は胃 X 線画像の正常部分の特徴を学習することにより、胃 X 線画像が正常かどうかを判断する方法である。特徴を学習する部分にオートエンコーダを用いる。オートエンコーダに正常画像のみを学習することにより、正常画像が入力された場合の残差（入力と出力の誤差）は小さくなり、正常でない画像が入力された場合の残差が大きくなることを期待している。

学習は正常な領域でオートエンコーダを訓練することで行う。正常胃領域から一定サイズの部分領域を切り取る。この部分領域をパッチと呼ぶ。パッチの切り取りではパッチの始点がある幅（ストライド）ごとに移動しながら切り取る。パッチを切り取った後、パッチ単位の前処理を行う。このようにして得られたパッチの各ピクセル値をオートエンコーダに入力し、オートエンコーダを学習する。模式図で示すと Fig. 3 のようになる。

学習段階でテスト誤差が最小になるときのオートエンコーダを用いて異常を検出する。まず、診断したい部分の胃領域パッチをオートエンコーダに入力する。オートエンコーダの入力  $x$  と出力  $y$  の差  $|x - y|$  を算出する。この差分値をしきい値  $T_1$  と比較し、 $T_1$  を超えるピクセル数をカウント  $C$  とする。カウント  $C$  としきい値  $T_2$  を比較し、 $C > T_2$  の場合に異常とする。模式図で示すと Fig. 4 のようになる。

CAD としての性能の評価にしきい値  $T_2$  を変化させたときの ROC 曲線（受信者動作特性曲線）を用いる。ROC 曲線は縦軸に真陽性率、横軸に偽陽性率をプロットする。さらに、ROC 曲線の右下部分の面積である AUC（曲線下面積）を算出する。

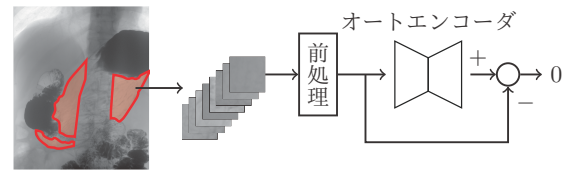


Fig. 3: Block diagram for learning using an autoencoder

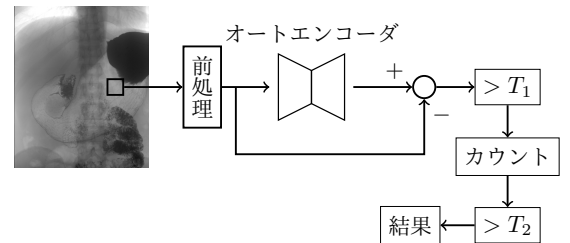


Fig. 4: Block diagram for detecting tumors from patches of a DC X-ray image

#### 3.2 パターン生成法

実験に用いる実際の胃 X 線画像は枚数に限りがあるため、多数の画像を学習した場合の効果や様々な腫瘍形状について検証できるように人工的に胃 X 線画像パッチを生成することを試みる。生成された画像は実際の胃 X 線画像のパッチの代わりとして学習および検出を行う。正常画像の生成法を Fig. 5 に示す。ここで線描画の処理はランダムな間隔で線を描画する処理である。また、「平滑化・変形」の処理は Elastic Distortion<sup>9)</sup> を適用した後に、Gaussian Filter を適用する処理である。Elastic Distortion は近接ピクセルとの位置関係のある程度維持しながら画素ごとにランダムな変位を与える処理である。腫瘍パターンは、ランダムな位置に円を描画し、Elastic Distortion を適用することで生成する。正常パターンから腫瘍パターンを減算し、これを異常パターンとする。生成された人工パターンの例を Fig. 6 に示す。

### 4 実験

#### 4.1 人工パターンを用いた実験

生成した人工パターンを用い、腫瘍検出の実験を行う。各実験に共通の実験条件は以下の通りである。各画像のサイズは  $32 \times 32$  とする。オートエンコーダの中間層のユニット数  $N_{hid}$  は 32 とし、活性化関数は  $f(x) = 1/(1 + e^{-x})$  で与えられる標準シグモイド関数を用いる。前処理として、画像の平均と標準偏差を一定に揃える正規化を行う。しきい値  $T_1$  は 40 に設定する。検出画像枚数は 500 枚とし、半数の 250 枚を正常パターン、残りの 250 枚を異常パターンとする。実験は合計 5 回行い、その中で最良の結果を選ぶ。

学習画像枚数とテスト画像枚数を変化させ、性能の変化を評価する。生成した画像のうち、75% を学習画像にし、残りをテスト画像とする。実験結果を Fig. 7 に示す。生成する画像を  $10^2$  枚、 $10^3$  枚、 $10^4$  枚とすると AUC はそれぞれ 0.658, 0.706, 0.710 である。枚数を増やすことによって性能が向上している。

既存の手法で対応できない場合として、円形の腫瘍の一部分のみにバリウムの溜まった円弧上の腫瘍を考

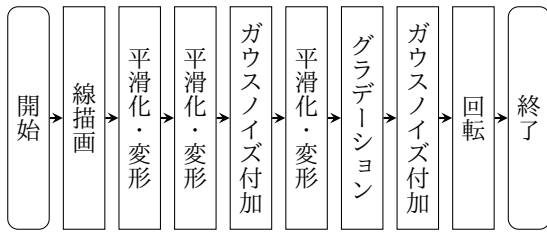


Fig. 5: A scheme for generating artificial patches for gastrointestinal wall

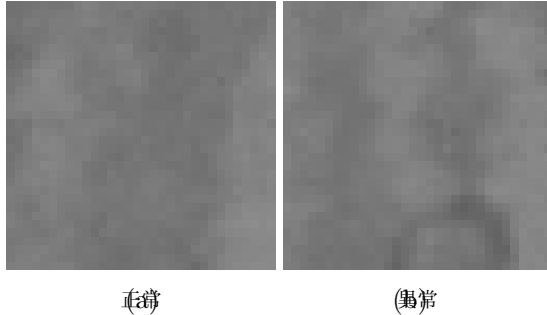


Fig. 6: Examples of artificial generated patches. (a) A patch without tumors and (b) a patch with tumors

える。腫瘍パターンの生成法において、円を描画する部分を円弧を描画するように置き換え、円弧腫瘍パターンの生成方法とする。この人工パターンにおいて円状の腫瘍を検知するシステムである DoRS に基づく手法<sup>6)</sup>の結果と比較する。DoRS の  $r_{\min}$  および  $r_{\max}$  は対象の腫瘍サイズに合わせ 3, 12 に設定する。学習画像枚数は 7500 枚、テスト画像枚数は 2500 枚とする。実験結果を Fig. 8 に示す。本検出法および DoRS に基づく手法の ROC 曲線の AUC はそれぞれ 0.684, 0.580 であり、本手法のほうが性能が高い。

#### 4.2 胃 X 線画像を用いた実験

コンピュータ支援画像診断学会の胃 X 線二重造影データベース<sup>10)</sup>を対象とし、本検出法を適用する。このデータベースには正常 (N 型) 画像が 11 枚、良性隆起 (BP 型) を含む画像が 8 枚、良性陥凹 (BD 型) を含む画像が 8 枚、悪性隆起 (MP 型) を含む画像が 5 枚、悪性陥凹 (MD 型) を含む画像が 44 枚含まれる。縦横の画素数は  $2510 \times 2000$  の画像 (66 枚) と  $1670 \times 2010$  の画像 (10 枚) が存在する。画素サイズはそれぞれ 0.15 mm/pixel, 0.1 mm/pixel である。すべての画像の画素サイズを 0.6 mm/pixel に揃える。濃度分解能は 10 bit である。胃輪郭線および病変領域についてはデータベースに付属の医師によるスケッチを参考に作成する。学習では正常画像 (N 型) を対象とする。検出では正常 (N 型) 画像以外の画像を対象とする。学習・検出ともに胃領域から脊柱領域を除外する。この脊柱領域については実験者が作成したものをを用いる。中間層のユニット数  $N_{\text{hid}}$  を 64 とし、その他の条件は人工パターンを用いた実験と同じとする。ROC 曲線を図 9 に示す。ROC 曲線の AUC は 0.597 となる。結果の例を Fig. 10 に示す。

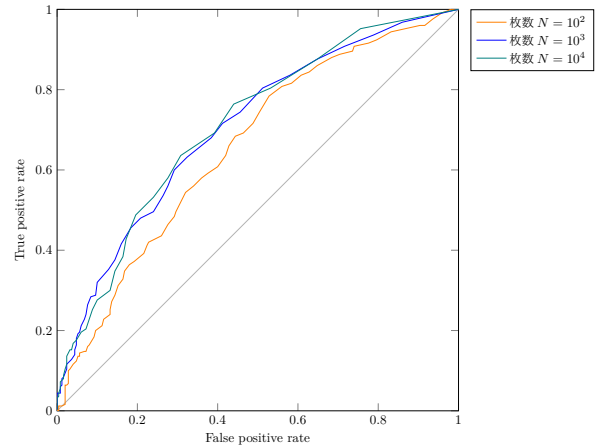


Fig. 7: ROC curves with various numbers of images in learning autoencoders

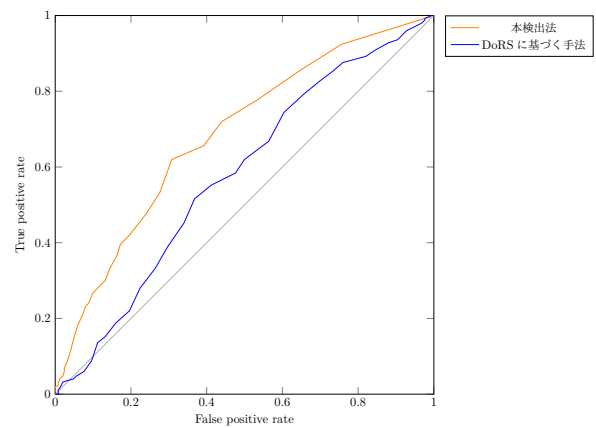


Fig. 8: ROC curves for detecting circular-type tumors, compared with the scheme in <sup>6)</sup>

## 5 まとめ

本論文では、オートエンコーダに基づく胃 X 線二重造影画像からの腫瘍検出法を提案した。このシステムの特性を評価するため、胃 X 線二重造影画像のパッチを模した人工パターンの生成法を提案した。この人工パターンにより腫瘍検出法の性能を評価した。その結果、学習枚数を増やすことによって性能が向上することを確認した。さらに、従来法では検出できない腫瘍について検出できることが示された。実際の胃 X 線画像での結果は、人工パターンでの結果と比較すると性能が低下した。実際の胃 X 線画像では体部ひだや顆粒状の模様など様々なパターンが存在する。これらのパターンの学習に必要な枚数に達しておらず、その結果として性能が低かったと考えられる。

今後の課題として、より実際の胃 X 線画像に近い人工パターンの生成、腫瘍検出法では、ほかの従来法との比較、パッチの大きさの検討、異常検知に適したニューラルネットワークの開発などが挙げられる。

## 謝辞

本研究の一部は日本学術振興会の科学研究補助金 (基盤研究 (C)16K00248) の支援により行われた。

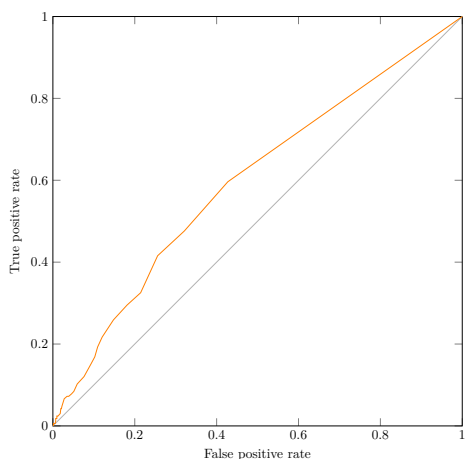


Fig. 9: ROC curves with actual DC X-ray images

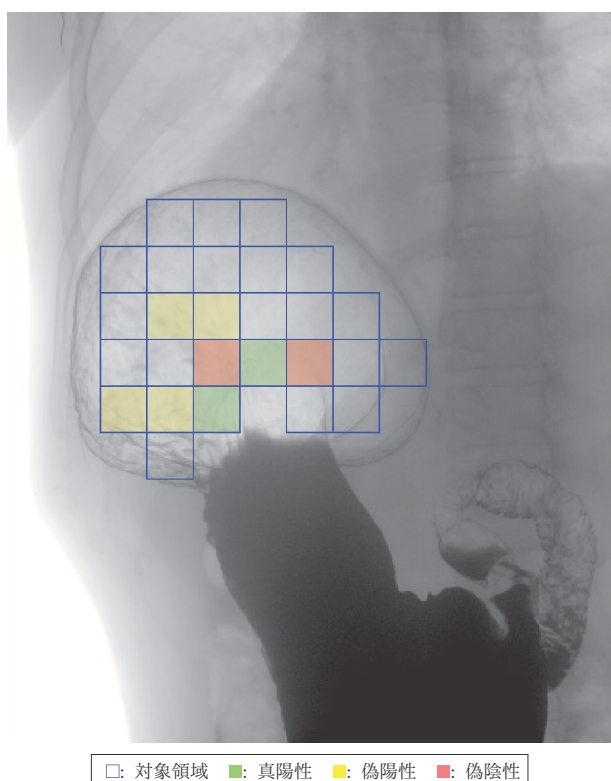


Fig. 10: A detection result for tumors

誌 D-II, **73**(4), 661/669 (1990)

- 6) T. Minemoto, S. Odama, A. Saitoh, T. Isokawa, N. Kamiura, H. Nishimura, S. Ono, and N. Matsui: Detection of tumors on stomach wall in X-ray images, Proceedings of IEEE World Congress on Computational Intelligence (WCCI2010), 1159/1163 (2010)
- 7) 目加田, 長谷川, 春日, 縄野, 宮川: 胃 X 線二重造影像におけるフラクタル次元, *Medical Imaging Technology*, **15**(5), 609/614 (1997)
- 8) 松本, 三浦, 鳥脇: 少数標本時における階層型ニューラルネットの汎化能力の実験的評価-胃 X 線二重造影像の場合-, *Medical Imaging Technology*, **14**(2), 174/181 (1996)
- 9) P. Y. Simard, D. Steinkraus, and J. C. Platt: Best practices for convolutional neural networks applied to visual document analysis, Proceedings of the Seventh International Conference on Document Analysis and Recognition, 958/963 (2003)
- 10) コンピュータ支援画像診断学会 (編): 胃 X 線二重造影像データベース, コンピュータ支援画像診断学会 (1997)

## 参考文献

- 1) 国立研究開発法人国立がん研究センターがん対策情報センター: 最新がん統計 (更新日 2016 年 8 月 2 日), [http://ganjoho.jp/reg\\_stat/statistics/stat/summary.html](http://ganjoho.jp/reg_stat/statistics/stat/summary.html).
- 2) 国立がん研究センター, がん予防・検診研究センター: 有効性評価に基づく胃がん検診ガイドライン 2014 年度版 (2015)
- 3) 細井, 入口, 小田, 茨城県総合健診協会 (編): 症例に学ぶ 胃がん X 線検診読影講座, 医療科学社 (2014)
- 4) 鳥脇: X 線像のコンピュータ支援診断: 研究動向と課題 (次世代医用画像技術論文特集), 電子情報通信学会論文誌 D-II, **83**(1), 3/26 (2000)
- 5) 長谷川, 筒井, 鳥脇: 胃 X 線 2 重造影像におけるひだ集中を伴うがん病変部の自動抽出: 電子情報通信学会論文



# 多目的最適化問題における評価時間の偏りが 半非同期進化法に与える影響の分析

○原田智広 (立命館大学) 高玉圭樹 (電気通信大学)

## Analysis of Semi-Asynchronous Evolutionary Algorithm for Multi-Objective Optimization Problem with Evaluation Time Bias

\*T. Harada (Ritsumeikan University) and K. Takadama (The University of Electro-Communications)

**Abstract**— This paper analyzes the influence of the evaluation time bias in multi-objective optimization problems to a semi-asynchronous evolutionary algorithm. The semi-asynchronous asynchronous evolutionary algorithm we proposed is a novel parallelization approach of evolutionary algorithm that waits some solutions to generate new solutions, unlike the conventional synchronous approach waits all solutions for the next generation. This paper conducts the experiment on the typical multi-objective optimization benchmarks with evaluation time bias depending on the objective function value. We compare the synchronous NSGA-II, the asynchronous NSGA-II, and the semi-asynchronous NSGA-II on the simulated parallel computing environment. The experimental result revealed that the semi-asynchronous NSGA-II that waits small number of solutions outperforms the synchronous and the asynchronous NSGA-II even if the evaluation time is biased. Additionally, it was revealed that the search area of the semi-asynchronous NSGA-II is biased toward the area where the evaluation time is short if the number of waited solutions is small.

**Key Words:** Semi-asynchronous evolutionary algorithm, multi-objective optimization, evaluation time

### 1 はじめに

進化的アルゴリズム (Evolutionary Algorithm: EA) は, 事前知識によらず高い最適化性能を得られることから, 実世界の最適化問題に広く適用されている<sup>17, 2)</sup>. 実世界の最適化問題では, 解の評価に物理シミュレーションを用いる場合のように, 解の評価時間に莫大な計算時間を要する問題が多く存在する. このような場合, EA による最適化の高速化のために複数の計算ノードを用いた並列化が有効である. 従来研究において多くの並列 EA が提案されており<sup>5, 23, 4, 10)</sup>, その中でもマスター・スレーブ型の並列化は代表的なアプローチの一つである. マスター・スレーブ型の並列 EA では, 1 台のマスターノードが EA の主要な処理 (初期化, 親選択, 遺伝的操作, 世代交代など) を実行し, 複数台のスレーブノードで新規に生成された解を並列に評価する.

並列 EA は, その世代交代の方法によって同期 EA (*Synchronous EA*: SEA) と非同期 EA (*Asynchronous EA*: AEA) の 2 種類に大別される. 同期 EA は, 一般的な世代に基づく EA と同様にすべての新規解の評価値を算出したのち, 選択・淘汰を実行し, 次世代の解候補を生成する方法である. これに対し, 非同期 EA は, 新規解の一つの評価が完了するごとに選択・淘汰を実行し, 次の解候補一つを継続的に生成する方法である.

同期 EA は, スレーブノードでのすべての解評価が完了してから次世代の新規会を生成するため, 解の評価時間に差がある場合に最も評価時間の長い解を待機する必要がある. そのため, 早期に解評価を完了したスレーブノードの待機時間が増加し, 計算効率が低下する問題がある. これに対し, 非同期 EA は他のスレーブノードの解評価を待たずに継続的に新規会を生成するため, スレーブノードの待機時間なしに効率的に計算資源を活用できる利点がある. しかし, 非同期 EA

は解評価が完了するごとに新規解を生成するため, 解の評価時間が短い局所解に陥りやすいという問題がある<sup>21, 22)</sup>. これに対しては, 同期 EA は評価時間差のある解すべてを待機してから新規解を生成するため, 非同期 EA より評価時間の差によって局所解に陥る可能性は低い.

このように, 同期 EA と非同期 EA は計算効率と探索性能の観点でトレードオフの関係にある. この関係に着目し, 我々は従来研究で計算効率と探索性能を両立する新しい並列化のアプローチとして, スレーブノードの解評価を数個分待機してから次世代の解候補を生成する半非同期 EA (*Semi-Asynchronous EA*: SAEA) を提案した<sup>11)</sup>. 半非同期 EA では, 非同期性を制御するパラメータを導入し, 次世代の新規解生成までに待機するスレーブノード数を決定した. 従来研究<sup>11)</sup>では, 多目的 EA の代表的な手法である NSGA-II (Non-dominated Sorting Genetic Algorithm-II)<sup>6)</sup> に半非同期 EA を適用し, 多目的最適化問題のベンチマークを用いて性能を評価した. その結果, 評価時間の差が小さい場合には同期 NSGA-II が最も有効であるが, 評価時間の差が大きい場合には少ない数のスレーブノードを待機する半非同期 NSGA-II が同期 NSGA-II と (完全) 非同期 NSGA-II を上回る性能を示すことを明らかにした.

従来研究<sup>11)</sup>では, すべての解の評価時間は同様の正規分布に従う条件で実験が行われている. しかし, 実際にはすべての解の評価時間が同様の分布特性を持つとは限らず, 解ごとに評価時間の特性が異なると考えられる. 特に, 文献<sup>22)</sup>では非同期 EA は評価時間の短い局所解に陥りやすい特性があることが示されていることから, 半非同期 EA においても同様の特性を示す可能性がある. そこで本研究では, 解の評価時間の偏りが半非同期 EA に与える影響を明らかにすることを目的とする. 具体的には, 従来研究<sup>11)</sup>と同様に半非同期

NSGA-II を扱い、解の目的関数値に応じて評価時間の分布が異なるベンチマーク問題を用いる実験を行い、その結果を分析する。

本論文の構成は以下の通りである。2 章では、非同期 EA に関する関連研究を概観する。3 章は、我々が従来研究で提案した半非同期 EA について示し、具体的な実現例としてシミュレーション実験でも使用する半非同期 NSGA-II を示す。4 章では、本研究で実施するシミュレーション実験の実験設定を説明する。その後、5 章で実験結果を示し、6 章で考察を述べる。最後に 7 章で本論文をまとめるとともに今後の課題を示す。

## 2 関連研究

### 2.1 非同期進化的アルゴリズム

従来研究では、単目的、ならびに多目的最適化を対象とした非同期 EA が提案されている。本研究では、代表的な並列 EA のアプローチであるマスター・スレーブ型の並列化を対象とする。マスター・スレーブ型の並列 EA では、1 台のマスターノードが EA の主要な処理（初期化、親選択、遺伝的操作、世代交代など）を実行し、複数台のスレーブノードでそれぞれ解の評価を並列に実行する。

まず従来の世代交代に基づく EA の非同期 EA への拡張として非同期 steady-state GP (ASSGP)<sup>14)</sup>、非同期粒子群最適化 (Asynchronous Particle Swarm Optimization: APSO)<sup>3, 12)</sup>、非同期差分進化 (Asynchronous Differential Evolution: ADE)<sup>15, 26)</sup> がある。また非同期多目的 EA としては、APSO を多目的に拡張した MAPSO<sup>13)</sup> や、多目的差分進化の DEMO (Differential Evolution for Multi-objective Optimization)<sup>18)</sup> を非同期に拡張した AMS-DEMO<sup>9)</sup> が提案されている。近年では、多目的 Bat Algorithm (IMOA)<sup>20)</sup> のような比較的新しいアルゴリズムを非同期に拡張した ANIMOBA<sup>19)</sup> の提案や、S-metric に基づく選択手法を用いる SMS-EMOA<sup>1)</sup> の同期的アプローチと非同期的アプローチの比較に関する研究<sup>24)</sup> がなされている。

Scott らは、評価時間の異なる最適化問題における非同期 EA の挙動を分析している<sup>22, 21)</sup>。文献<sup>22)</sup> では、同期型  $(\mu + \lambda)$ EA と非同期型  $(\mu + 1)$ EA を解の評価値と評価時間の関係性（正相関、負相関、相関なし）を変化させた最適化問題で比較し、非同期 EA がすべての関係性で同期 EA を上回る性能を示すことを明らかにした。一方、文献<sup>21)</sup> では、非同期 EA は評価時間の短い準最適解に収束する可能性が高まることが示されている。

### 2.2 並列計算環境における計算時間モデル

Zvoianu らは、並列計算環境における同期 EA と非同期 EA の計算時間モデルを提案している<sup>25)</sup>。提案モデルでは、1 台のマスターノードと  $\lambda$  台のスレーブノードからなるマスター・スレーブ型の並列計算環境を用い、マスターノードは EA の主要な処理を  $t_s$  単位時間で実行し、スレーブノードは解評価を  $t_p (>> t_s)$  単位時間で実行するとしている。

提案モデルにおける同期 EA と非同期 EA の時間経過の模式図を Fig. 1 に示す。Fig. 1 では、スレーブノード数  $\lambda = 4$  としている。同期 EA では、初期化時に  $\lambda = 4$  回の解生成処理がマスターノードで実行され、各解はスレーブノードへ送られる。その後、スレーブノードで

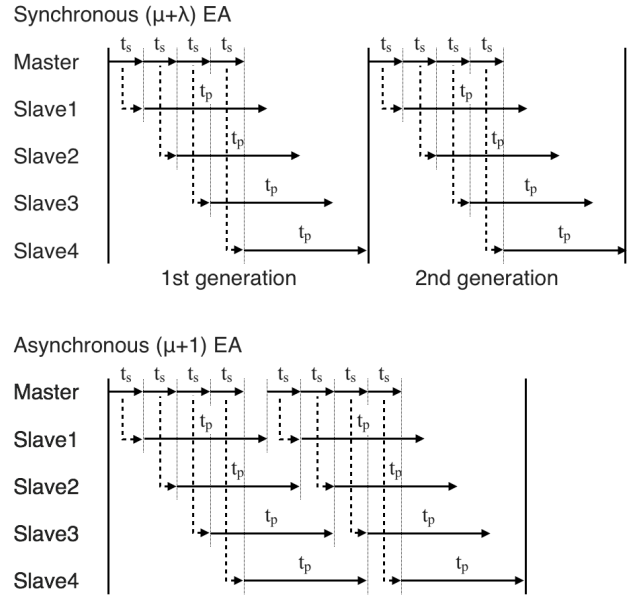


Fig. 1: The computational time model of a synchronous  $(\mu + \lambda)$  EA (top) and an asynchronous  $(\mu + 1)$  EA (bottom) discussed in<sup>25)</sup>

の評価がすべて完了した段階で次の解生成がマスターノードで実行される。この処理が繰り返されることで、同期 EA では  $N$  世代の進化に  $N \times (\lambda \times t_s + t_p)$  単位時間を要する。これに対し、非同期 EA では、初期化時は同期 EA と同様であるが、2 世代目以降は 1 台のスレーブノードでの評価が完了するごとにマスターノードで新規解生成が実行され、生成された解がスレーブノードへ送られる。このように継続的に解生成が実行されるため、非同期 EA では  $N$  世代 ( $N \times \lambda$  回の解評価) の計算が同期 EA よりも短い  $(\lambda \times t_s + t_p) + (N - 1) \times (t_s + t_p)$  単位時間で完了する。

文献<sup>25)</sup> では、このモデルに基づいた同期 EA と非同期 EA の比較実験の結果、評価時間  $t_p$  の分散が大きい場合に非同期 EA が同期 EA よりも大幅に計算効率が高くなり、同期 EA より短時間で最適解を獲得可能であることが示された。一方、分散が小さい場合には非同期 EA の計算効率が同期 EA と比較して高くないため、探索性能が劣る非同期 EA が最適解獲得までにより多くの計算時間を要することも示されている。

## 3 半非同期進化的アルゴリズム

### 3.1 概要

従来研究では、様々な種類の非同期単目的/多目的 EA が提案されており、評価時間差のある最適化問題においてその有効性が示されている。しかし、非同期 EA は各解の評価が完了するごとに新規解を生成することから、一部の解情報しか利用することができず、探索性能が同期 EA と比較して劣る。一方、同期 EA は評価時間差が大きい場合にスレーブノードの待機時間が大きく、計算効率が非同期 EA と比較して大きく劣るといふ問題がある。このように、同期 EA と非同期 EA は探索性能と計算効率の観点でトレードオフの関係にある。

これに対し、我々は従来研究<sup>11)</sup> で、スレーブノードの評価を数個分待機する半非同期 EA (Semi-

### Semi-asynchronous EA ( $\alpha=0.5$ , waiting two evaluations)

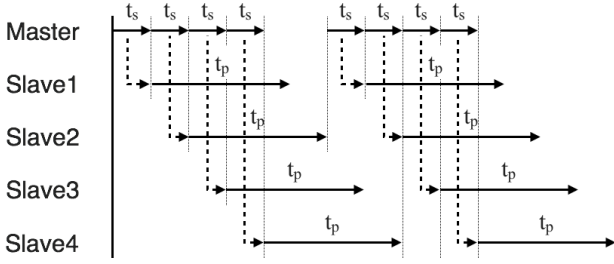


Fig. 2: Semi-asynchronous ( $\mu + \alpha\lambda$ ) EA where  $\lambda = 4$  slave nodes and the asynchrony parameter  $\alpha = 0.5$ , i.e.,  $n = \alpha\lambda = 2$  evaluations out of four slave nodes are waited to execute the master process

### Semi-asynchronous EA ( $\alpha=3/4$ , waiting three evaluations)

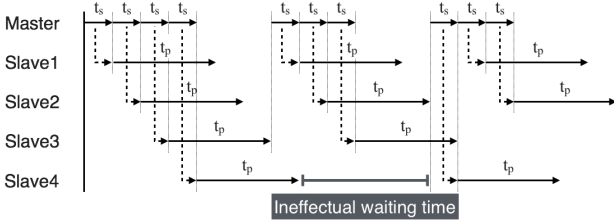


Fig. 3: An example of ineffectual asynchrony of semi-asynchronous ( $\mu + \alpha\lambda$ ) EA where  $\lambda = 4$  slave nodes and the asynchrony parameter  $\alpha = 3/4$ , i.e.,  $n = \alpha\lambda = 3$  evaluations out of four slave nodes are waited to execute the master process

*Asynchronous EA*: SAEA) を提案した。半非同期 EA では、非同期性 (*asynchrony*) を決定するパラメータ  $\alpha$  を定義し、 $n = \lceil \alpha \times \lambda \rceil$  ( $1/\lambda \leq \alpha \leq 1$ ) 個体の評価を待機する。ここで、 $\lambda$  はスレーブノードの台数を表し、 $\lceil x \rceil$  は天井関数を表し、 $x$  以上の最小の整数値を返す。例えば、Fig. 2 に示すようにスレーブノード数  $\lambda = 4$ 、非同期性パラメータ  $\alpha = 0.5$  のとき、半非同期 EA では  $n = \lceil \alpha \times \lambda \rceil = \lceil 0.5 \times 4 \rceil = 2$  台を待機し、新規解を生成する。半非同期 EA において、 $\alpha = 1/\lambda$  の場合は 1 個体を待機するため (完全) 非同期 EA と同等となり、 $\alpha = 1$  の場合は  $\lambda$  個体を待機するため同期 EA と同等となる。以降、特に断りのない限り  $\lceil \alpha\lambda \rceil$  を単純に  $\alpha\lambda$  と表記する。

### 3.2 有効な非同期性パラメータ

半非同期 EA では、非同期性パラメータ  $\alpha$  としてすべての値を取ることはできない。これは、不適切な  $\alpha$  を設定することで、同期のタイミングにずれが生じ、無駄な待機時間が増大してしまうためである。具体的には、非同期性パラメータ  $\alpha$  は以下の条件を満たす必要がある：

$$\lambda \equiv 0 \pmod{\lceil \alpha\lambda \rceil}, \quad (1)$$

ここで、 $a \equiv b \pmod{n}$  は整数の合同を表し、整数  $a$  と整数  $b$  の差  $a - b$  が正の整数  $n$  の倍数であることを示す。もし、非同期性パラメータ  $\alpha$  がこの条件を満たさず、 $\lambda \equiv k \pmod{\lceil \alpha\lambda \rceil}$  を満たす整数  $k (\neq 0)$  が存在するとき、 $k$  台のスレーブノードが他のノードと同期が合わず、無駄な待機時間が生じる。

例えば、Fig. 3 では、スレーブノードの台数  $\lambda =$

4、非同期性パラメータ  $\alpha = 3/4$  とする時、 $\lambda \equiv 1 \pmod{\lceil \alpha\lambda \rceil}$  となることから 1 台のスレーブノードの同期がずれる。具体的には、最初の同期時点では Slave1 から Slave3 の解評価が完了し、新規解が 3 個生成される。その後、Slave4 の解評価が完了するが、他のスレーブノードは評価が始まった直後であるため、それらの解評価が完了するまで Slave4 は待機する必要がある。その後、Slave1 と Slave2 の評価が完了することで Slave4 と合わせた 3 台の同期が完了し、新規解が 3 個生成されるが、この際も次に評価が完了する Slave3 が他のスレーブノードを待機する必要性が生じる。このように、無効な非同期性パラメータは計算効率の著しい低下を招く。以後、本論文では、式 (1) を満たす有効な非同期性パラメータのみを考える。

### 3.3 半非同期 EA の例：半非同期 NSGA-II

ここでは、半非同期 EA の具体例として、文献<sup>11)</sup> で用いられた半非同期 NSGA-II を示す。NSGA-II<sup>6)</sup> は探索性能の高い多目的 EA の代表的な一つであり、非支配ソーティングと混雑度距離に基づく選択によりパレート最適解を求める手法である。マスター・スレーブ型の並列同期 NSGA-II の大まかな流れは下記の通りである：

1. 世代数  $t = 0$
2. 初期化
  - (a) 初期母集団  $P_0$  を生成
  - (b) すべての解をスレーブノードへ送る
  - (c) すべての解の評価を待機
3. 終了条件を満たすまで以下を繰り返す
  - (a) 母集団  $P_t$  と同数の子集団  $Q_t$  をトーナメント選択と遺伝的操作によって生成
  - (b) すべての新規解をスレーブノードへ送る
  - (c) すべての新規解の評価を待機
  - (d)  $R_t = P_t \cup Q_t$ .
  - (e) 次世代母集団  $P_{t+1}$  を非支配ソーティングと混雑度距離に基づいて  $R_t$  から選択
  - (f)  $t = t + 1$ .

これらの処理において、スレーブノードはマスターノードから送られる解の評価を実行し、その評価値をマスターノードに返し、マスターノードは NSGA-II の主な処理を実行する。

NSGA-II は  $(\mu + \lambda)$  型 EA の  $\mu = \lambda = |P|$  とした世代交代方法であり、容易に半非同期法に適用可能である。具体的には、半非同期 NSGA-II は下記の流れで実行される。なお、以下において、同期と半非同期の差異は太字で表される：

1. 世代数  $t = 0$
2. 初期化
  - (a) 初期母集団  $P_0$  を生成
  - (b) すべての解をスレーブノードへ送る

(c) すべての解の評価を待機

3. 終了条件を満たすまで以下を繰り返す

- (a) 待機中のスレーブノード数と同数の子集団  $Q_t$  をトーナメント選択と遺伝的操作によって生成
- (b) すべての新規解を待機中のスレーブノードへ送る
- (c)  $\alpha\lambda$  個の新規解の評価を待機.
- (d)  $R_t = P_t \cup Q_t$ .
- (e) 次世代母集団  $P_{t+1}$  を非支配ソーティングと混雑度距離に基づいて  $R_t$  から選択
- (f)  $t = t + 1$ .

同期 NSGA-II と半非同期 NSGA-II の主な違いは 3.(a) と 3.(c) である. 半非同期 NSGA-II では, 3.(c) において全個体の評価を待機するのではなく,  $\alpha\lambda$  個体の評価を待機し, 3.(a) において同数の子個体を現在の母集団から生成する. 次世代の母集団は  $\lambda$  個体の親集団と  $\alpha\lambda$  個体の小集団を合わせた  $(\lambda + \alpha\lambda)$  個体の中から非支配ソーティングと混雑度距離に基づいて選択される.

## 4 実験設定

### 4.1 実験概要

本実験では, 多目的最適化問題において解の目的関数値に応じて評価時間に偏りがある場合の半非同期 EA への影響を分析する. 半非同期 EA としては, 前章の例に挙げた半非同期 NSGA-II を用いる. 評価時間の偏りとしては, ある解  $\mathbf{x}$  の  $i$  番目の目的関数値を  $f_i(\mathbf{x})$  とするとき, 解の評価時間  $t(\mathbf{x})$  を下記のように定義する:

- 正相関

$$t(\mathbf{x}) = \mathcal{N}(2dt_p g_i(f_i(\mathbf{x})) + (1-d)t_p, c_v \times t_p) \quad (2)$$

- 負相関

$$t(\mathbf{x}) = \mathcal{N}(-2dt_p g_i(f_i(\mathbf{x})) + (1+d)t_p, c_v \times t_p) \quad (3)$$

ただし,

$$g_i(f) = \begin{cases} 0 & f < f_i^{\min} \\ 1 & f > f_i^{\max} \\ \frac{f - f_i^{\min}}{f_i^{\max} - f_i^{\min}} & \text{otherwise} \end{cases} \quad (4)$$

ここで,  $\mathcal{N}(\mu, \sigma)$  は平均  $\mu$ , 標準偏差  $\sigma$  の正規分布を表し,  $t_p$  は基準となる評価時間を表す.  $d(> 0)$  は目的関数値による評価時間の偏り度合いを決定するパラメータであり,  $c_v(> 0)$  は評価時間の分散度合いを決定するパラメータである. また,  $f_i^{\max}$ ,  $f_i^{\min}$  は真のパレート最適解集合における  $i$  番目の目的関数値  $f_i$  の最大値, 最小値をそれぞれ表す.

正相関の関係を持つ場合,  $i$  番目の目的関数値が最大値以上の場合に評価時間平均が最大の  $(1+d)t_p$ , 最小値以下の場合に評価時間平均が最小の  $(1-d)t_p$  となる. 負相関の関係を持つ場合,  $i$  番目の目的関数値が最大値

以上の場合に評価時間平均が最小の  $(1-d)t_p$ , 最小値以下の場合に評価時間平均が最大の  $(1+d)t_p$  となる. 本実験では, いずれのベンチマーク問題でも第 1 目的関数値に基づいて評価時間を決定するものとする.

### 4.2 並列計算時間モデル

本実験では, 1 台のマスターノードと  $\lambda = 100$  台のスレーブノードを擬似的な並列計算モデル上で検証を行う. ここで, 擬似並列計算モデルは文献<sup>25)</sup>で提案されているモデルを用い, マスターノードでの計算時間は  $t_s = 1$ , スレーブノードでの計算時間は式 (2), (3) に基づいて決定され,  $t_p = 1000$ ,  $c_v = 0.02$  とする. 目的関数値による偏りを表すパラメータ  $d = \{0.2, 0.4\}$  を使用する. マスターノードでは NSGA-II の主な処理である初期母集団生成, 親選択, 交叉, 突然変異, 世代交代などを実行し, スレーブノードでは解の評価のみを実行する.

### 4.3 ベンチマーク問題

本実験では, 性能評価のために, 多目的最適化問題のベンチマーク問題である ZDT<sup>27)</sup> (ZDT1-4, ZDT6), および WFG (WFG1-9) を用いる. ZDT は 2 目的最適化問題であり, 設計変数は ZDT1, 2, 4 では 30 次元とし, ZDT3, 6 では 10 次元とする. WFG では, 目的数は 2 目的とし, 設計変数は 6 次元 ( $k = 4, l = 2$ ) とする.

### 4.4 パラメータ設定

同期, および半非同期 NSGA-II では, 以下の遺伝的操作, およびパラメータ設定を用いる.

- 母集団サイズ = 100 (=  $\lambda$ )
- 最大評価回数 = 50000 (最大世代数 = 500)
- 交叉法: Simulated Binary Crossover (SBX)<sup>7)</sup> ( $P_c = 0.9, \eta_c = 20.0$ )
- 突然変異法: Polynomial Mutation (PM)<sup>8)</sup> ( $P_m = 1/D, \eta_m = 20.0$ )

本実験では, 式 (1) の条件に基づいて半非同期 NSGA-II に対して有効なすべての非同期性パラメータ  $\alpha$  を検証する. 具体的には, スレーブノード数が 100 のため,  $\alpha = \{0.01, 0.02, 0.04, 0.05, 0.1, 0.2, 0.25, 0.5, 1.0\}$  が有効な非同期性パラメータとなる. ここで,  $\alpha = 0.01$  は完全な非同期 NSGA-II と同等 (1 個体のみ待機) であり,  $\alpha = 1$  は同期 NSGA-II と同等 ( $\lambda$  個体待機) である.

本実験の実装は, 進化計算法のフレームワークである jMetal (ver. 5.0)<sup>16)</sup> を用いた. また, 各ベンチマーク問題, 各非同期性パラメータ  $\alpha$ , 相関関係の違いの組み合わせごとに 25 試行ずつ実施する.

## 5 実験結果

本実験では, 評価尺度として Hypervolume (HV)<sup>28)</sup> を用いる. HV は, 非支配解集合が支配する領域の大きさを測る指標である. 各ベンチマーク問題において, 真のパレート最適解集合 (Pareto Optimal Solutions: POS) の HV の一定割合の HV を持つ非支配解集合を獲得するまでの擬似並列計算モデル上での計算時間を比較する. 具体的には, すべての ZDT 問題と WFG2,

4では真のPOSのHVの95%, WFG7, 9では90%, WFG3では85%, WFG5, 6では80%をそれぞれ用いる。なお, WFG1とWFG8については, 真のPOSのHVの80%の非支配解集合を獲得できなかったため実験結果からは除いている。半非同期NSGA-IIが同期NSGA-IIと比較して計算時間を削減できている割合 $\Delta_{time}^{x\%}$ を下記の式に基づいて算出する:

$$\Delta_{time}^{x\%} = 100 \times \left( 1 - \frac{T^{x\%}(\alpha)}{T^{x\%}(1)} \right), \quad (5)$$

$T^{x\%}(\alpha)$ は, 非同期性パラメータ $\alpha$ のときに, 真のPOSの $x\%$ のHVを持つ非支配解集合を獲得するまでの擬似並列計算モデル上での計算時間を表す。 $\Delta_{time}^{x\%} > 0$ の場合は半非同期NSGA-IIが計算時間を削減できていることを意味し, 逆に $\Delta_{time}^{x\%} < 0$ の場合は計算時間が増大していることを意味する。

Table 1に正相関の場合の $\Delta_{time}^{x\%}$ の結果, Table 2に負相関の場合の $\Delta_{time}^{x\%}$ の結果を示す。各表において, 負の結果をグレーで表記する。また, Wilcoxonの順位和検定を用いた検定を行い, 有意差5%で半非同期NSGA-IIが同期NSGA-IIと比較して有意に計算時間を削減できている結果を“ $\Delta$ ”, 有意に増加している結果を“ $\blacktriangledown$ ”で表記している。

Table 1, 2の結果から, 正相関, 負相関いずれの場合も評価時間の偏り $d$ が大きいほど半非同期NSGA-IIが同期NSGA-IIと比較して大幅に最適解獲得までの計算時間を削減できていることがわかる。非同期性パラメータ $\alpha$ に着目すると,  $\alpha \geq 0.1$  (10個体以上待機)する場合に同期NSGA-IIと同等の計算時間か, 問題によっては計算時間が増加することが確認できる。一方,  $\alpha < 0.05$  (5個体以下待機, あるいは完全非同期)の場合は, 同期NSGA-IIと同等以上の計算時間を達成できていることが確認できる。また, 多くの例題で半非同期NSGA-IIが完全非同期NSGA-IIより短い計算時間で最適解を獲得できていることから, 非同期性パラメータ $\alpha$ を適切に調節することでより高い探索性能と計算効率を両立できることがわかる。この傾向は目的関数値によらず評価時間の分布が一様な文献<sup>11)</sup>の実験結果と同様である。このことから, 半非同期NSGA-IIは目的関数値に応じた評価時間の偏りがある場合でも, 非同期性パラメータを低く適切に設定することで同期NSGA-IIを大きく上回る計算効率を実現できることが明らかになった。

## 6 考察

ここで, 評価時間の偏りによる探索範囲の違いを分析するために, WFG2とWFG4における解探索の推移を正相関, 負相関の場合で比較する。具体的には, 100評価ごと(1世代ごと)の非支配解集合のうち, 第1目的関数に基づいて降順に並べ替えた場合の50%番目の解の目的関数空間上での推移をFig. 4とFig. 5に5試行分示す。ここでは非同期性パラメータ $\alpha = \{0.01, 0.02, 0.5, 1.0\}$ の結果のみ示している。Fig. 4, 5において, 横軸, 縦軸はそれぞれ第1目的関数値, 第2目的関数値を表し, 黒線は真のパレート最適面を表す。青点は正相関の結果, 赤点は負相関の結果を表す。正相関の場合, 第1目的関数値が大きいほど評価時間が長いため, 図の右側ほど評価時間が長く, 左側ほど評価時間が短い。逆

に, 負相関の場合, 第1目的関数値が小さいほど評価時間が長いため, 図の左側ほど評価時間が長く, 右側ほど評価時間が短い。

Fig. 4, 5とも, 非同期性パラメータ $\alpha$ が大きい場合は50%番目の解の分布が類似していることから, 評価時間の偏りによる探索領域の偏りが少ないことがわかる。これに対し, 非同期性パラメータ $\alpha$ が小さい場合は評価時間の短い方向, つまり正相関の青点はパレート最適面の中心より左側, 負相関はパレート最適面の中心より右側に50%番目の解が分布していることが確認できる。これは, 非同期な世代交代により, 評価時間の短い領域へ探索のバイアスがかかっているためである。このことから, 評価時間に偏りがある場合, 半非同期NSGA-IIでは評価時間の短い領域への探索バイアスが生じることが確認された。

## 7 おわりに

本研究では, 並列EAの新しい世代交代法である半非同期EAに対し, 評価時間の偏りが最適化に与える影響を分析した。具体的には, 多目的EAの代表的な手法であるNSGA-IIに半非同期法を適用した半非同期NSGA-IIを多目的最適化ベンチマーク問題に適用する実験を行った。実験では, 目的関数値に応じて評価時間に偏りのある問題を設定し, 同期NSGA-IIと比較した。実験の結果, 半非同期EAは評価時間に偏りがある場合でも同期EAと比較して短い計算時間で最適解が獲得可能であり, 特に比較的小さな非同期性パラメータ( $\alpha \leq 0.05$ )が有効であることが明らかになった。また, 半非同期NSGA-IIは評価時間の短い探索領域への探索バイアスが生じることが確認された。

今後は(1)非同期性パラメータ $\alpha$ の動的な調整法の提案や, (2)NSGA-II以外の単目的/多目的EAへの半非同期法の適用と有効性の検証, (3)実並列計算環境での検証などに取り組む予定である。

## 参考文献

- 1) Nicola Beume, Boris Naujoks, and Michael Emmerich. Sms-emoa: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, Vol. 181, No. 3, pp. 1653 – 1669, 2007.
- 2) Christina Bonnington. Teen’s iOS App Uses Complex Algorithms to Summarize the Web. <http://www.wired.com/2011/12/summly-app-summarization/>, Dec 2011.
- 3) A. Carlisle and G. Dozier. An off-the-shelf PSO. In *PSO Workshop*. Indianapolis, IN, April 2001.
- 4) Jui-Fang Chang, Shu-Chuan Chu, John F. Roddick, and Jeng-Shyang Pan. A parallel particle swarm optimization algorithm with communication strategies. *Journal of Information Science and Engineering*, pp. 809–818, 2005.
- 5) A Chipperfield and P Fleming. Parallel genetic algorithms. *Parallel and distributed computing handbook*, pp. 1118–1143, 1996.
- 6) K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on*, Vol. 6, No. 2, pp. 182–197, apr 2002.
- 7) Kalyanmoy Deb and Ram B. Agrawal. Simulated Binary Crossover for Continuous Search Space. *Complex Systems*, Vol. 9, pp. 115–148, 1995.
- 8) Kalyanmoy Deb and Mayank Goyal. A combined genetic adaptive search (geneas) for engineering design. *Computer Science and Informatics*, Vol. 26, pp. 30–45, 1996.

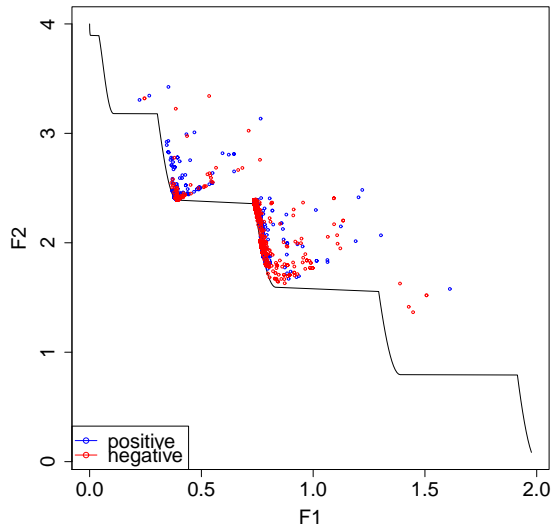
Table 1: The result of  $\Delta_{time}^{x\%}$  with the positive correlation

d	ZDT1 ( $\Delta_{time}^{95\%}$ )								ZDT2 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	11.7% $\Delta$	<b>13.5%</b> $\Delta$	12.2% $\Delta$	9.6% $\Delta$	8.7% $\Delta$	5.2% $\Delta$	2.2%	-5.7% $\blacktriangledown$	2.2%	<b>2.7%</b>	0.3%	1.6%	-3.7%	-2.0%	-5.5% $\blacktriangledown$	-7.4% $\blacktriangledown$
0.4	25.2% $\Delta$	<b>26.8%</b> $\Delta$	25.2% $\Delta$	25.9% $\Delta$	23.5% $\Delta$	20.0% $\Delta$	17.6% $\Delta$	3.9% $\Delta$	<b>28.1%</b> $\Delta$	27.5% $\Delta$	24.0%	23.3% $\Delta$	20.8% $\Delta$	22.9% $\Delta$	18.6% $\Delta$	5.6%
d	ZDT3 ( $\Delta_{time}^{95\%}$ )								ZDT4 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	<b>13.3%</b> $\Delta$	12.0% $\Delta$	9.6% $\Delta$	10.6% $\Delta$	7.2% $\Delta$	4.7% $\Delta$	3.2%	-2.9%	<b>11.7%</b> $\Delta$	5.0%	6.2%	7.1%	3.2%	0.5%	2.5%	-2.1%
0.4	21.6% $\Delta$	<b>23.4%</b> $\Delta$	18.0% $\Delta$	-6.3% $\Delta$	16.6% $\Delta$	12.1% $\Delta$	12.3% $\Delta$	2.7%	<b>23.5%</b> $\Delta$	18.5% $\Delta$	20.8% $\Delta$	20.8% $\Delta$	16.7% $\Delta$	12.2% $\Delta$	12.6% $\Delta$	6.3%
d	ZDT6 ( $\Delta_{time}^{95\%}$ )								WFG2 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	<b>10.4%</b> $\Delta$	9.6% $\Delta$	7.9% $\Delta$	6.4% $\Delta$	4.7% $\Delta$	2.1% $\Delta$	-1.9% $\blacktriangledown$	-5.1% $\blacktriangledown$	8.3%	<b>15.8%</b> $\Delta$	13.3% $\Delta$	1.6%	7.4%	3.4%	-5.1%	3.9%
0.4	<b>22.1%</b> $\Delta$	20.9% $\Delta$	21.4% $\Delta$	20.2% $\Delta$	18.9% $\Delta$	15.5% $\Delta$	12.8% $\Delta$	1.5%	17.7% $\Delta$	<b>22.1%</b> $\Delta$	14.0% $\Delta$	19.4% $\Delta$	15.0% $\Delta$	10.9% $\Delta$	15.1% $\Delta$	-5.1%
d	WFG3 ( $\Delta_{time}^{85\%}$ )								WFG4 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	-3.5%	-0.6%	<b>3.3%</b>	0.9%	-1.2%	-7.9% $\blacktriangledown$	-4.2%	-9.6% $\blacktriangledown$	<b>8.3%</b>	6.3%	7.9%	<b>8.3%</b>	1.3%	-1.6%	-2.1%	-6.3%
0.4	18.1% $\Delta$	<b>21.0%</b> $\Delta$	13.6% $\Delta$	17.1% $\Delta$	10.8% $\Delta$	11.9% $\Delta$	11.4% $\Delta$	-3.8%	19.0% $\Delta$	<b>21.9%</b> $\Delta$	19.2% $\Delta$	16.0% $\Delta$	17.5% $\Delta$	15.0% $\Delta$	12.2% $\Delta$	1.3%
d	WFG5 ( $\Delta_{time}^{80\%}$ )								WFG6 ( $\Delta_{time}^{80\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	-4.6%	-3.4%	-4.6% $\blacktriangledown$	-3.9%	-5.9%	-13.7% $\blacktriangledown$	-10.0% $\blacktriangledown$	-9.4% $\blacktriangledown$	23.8%	<b>28.8%</b>	19.6%	20.3%	21.4%	22.0%	21.9%	27.1%
0.4	5.6% $\Delta$	<b>8.7%</b> $\Delta$	5.3% $\Delta$	4.5% $\Delta$	3.4%	-3.4%	-2.4%	-4.1%	37.1%	34.9% $\Delta$	38.7% $\Delta$	32.5% $\Delta$	<b>43.4%</b> $\Delta$	42.2%	31.3%	7.5%
d	WFG7 ( $\Delta_{time}^{90\%}$ )								WFG9 ( $\Delta_{time}^{90\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	-2.6%	-4.7%	-2.7%	-5.5%	-9.5% $\blacktriangledown$	-7.1% $\blacktriangledown$	-12.2% $\blacktriangledown$	-1.2%	<b>5.2%</b>	<b>5.2%</b>	4.1%	2.6%	-0.5%	-3.6%	3.3%	0.0%
0.4	4.8% $\Delta$	3.9% $\Delta$	4.8% $\Delta$	<b>6.2%</b> $\Delta$	4.4% $\Delta$	-1.6%	-0.6%	-6.4% $\blacktriangledown$	8.2%	<b>10.0%</b> $\Delta$	1.9%	3.0%	7.3%	4.2%	9.0% $\Delta$	-11.0%

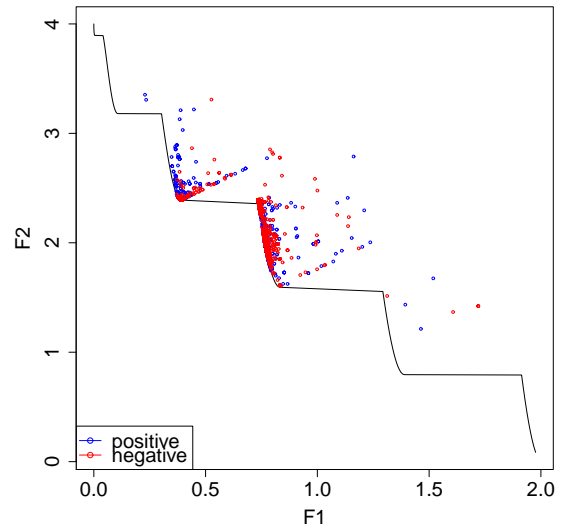
Table 2: The result of  $\Delta_{time}^{x\%}$  with the negative correlation

d	ZDT1 ( $\Delta_{time}^{95\%}$ )								ZDT2 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	2.6%	<b>3.0%</b> $\Delta$	2.5%	0.7%	-0.4%	-3.1% $\blacktriangledown$	-6.6% $\blacktriangledown$	-5.0% $\blacktriangledown$	<b>6.0%</b> $\Delta$	4.3% $\Delta$	0.6%	1.9%	-0.7%	-2.6%	-8.8% $\blacktriangledown$	-9.5% $\blacktriangledown$
0.4	12.9% $\Delta$	<b>14.6%</b> $\Delta$	10.2% $\Delta$	9.7% $\Delta$	8.4% $\Delta$	3.1% $\Delta$	2.9%	-0.9%	14.5% $\Delta$	13.7% $\Delta$	19.8% $\Delta$	14.9% $\Delta$	<b>21.0%</b> $\Delta$	14.2% $\Delta$	8.0% $\Delta$	3.6% $\Delta$
d	ZDT3 ( $\Delta_{time}^{95\%}$ )								ZDT4 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	<b>4.0%</b>	2.7%	0.3%	2.1%	2.5%	-3.7%	-4.1% $\blacktriangledown$	-9.4% $\blacktriangledown$	-2.9%	<b>2.8%</b>	0.9%	2.0%	-3.1%	-4.2%	-6.2%	-4.3%
0.4	<b>20.8%</b> $\Delta$	18.0% $\Delta$	17.6% $\Delta$	19.0% $\Delta$	15.3% $\Delta$	14.3% $\Delta$	9.9%	8.0%	2.4%	6.5%	2.7%	<b>9.5%</b> $\Delta$	8.2% $\Delta$	-0.9%	-1.6%	1.6%
d	ZDT6 ( $\Delta_{time}^{95\%}$ )								WFG2 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	<b>8.2%</b> $\Delta$	7.0% $\Delta$	6.9% $\Delta$	6.3% $\Delta$	2.8% $\Delta$	0.5%	-1.5% $\blacktriangledown$	-6.3% $\blacktriangledown$	-1.9%	-1.5%	<b>1.8%</b>	-5.1%	-1.2%	-7.6%	-10.2%	-2.3%
0.4	<b>20.2%</b> $\Delta$	18.8% $\Delta$	18.7% $\Delta$	17.8% $\Delta$	15.8% $\Delta$	12.8% $\Delta$	10.7% $\Delta$	0.6%	7.0% $\Delta$	<b>8.0%</b>	3.8%	0.9%	3.5%	-12.3% $\blacktriangledown$	-4.4%	-14.2% $\blacktriangledown$
d	WFG3 ( $\Delta_{time}^{85\%}$ )								WFG4 ( $\Delta_{time}^{95\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	4.4%	5.1%	<b>11.8%</b> $\Delta$	3.6%	2.5%	1.6%	1.0%	3.6%	20.5% $\Delta$	19.3% $\Delta$	19.0% $\Delta$	<b>22.4%</b> $\Delta$	17.3% $\Delta$	16.3% $\Delta$	9.1% $\Delta$	4.3%
0.4	12.7% $\Delta$	11.6% $\Delta$	<b>13.2%</b> $\Delta$	10.9% $\Delta$	11.4% $\Delta$	2.2%	2.2%	-2.0%	29.4% $\Delta$	<b>31.5%</b> $\Delta$	26.1% $\Delta$	26.7% $\Delta$	23.7% $\Delta$	23.0% $\Delta$	19.7% $\Delta$	4.9%
d	WFG5 ( $\Delta_{time}^{80\%}$ )								WFG6 ( $\Delta_{time}^{80\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	8.9% $\Delta$	9.3% $\Delta$	<b>9.8%</b> $\Delta$	6.5% $\Delta$	8.1% $\Delta$	2.6%	0.0%	-7.0% $\blacktriangledown$	<b>21.2%</b> $\Delta$	14.0%	17.5% $\Delta$	-5.7% $\Delta$	13.1%	6.6%	-43.8%	-41.9%
0.4	<b>27.7%</b> $\Delta$	24.5% $\Delta$	24.7% $\Delta$	25.7% $\Delta$	24.6% $\Delta$	21.6% $\Delta$	19.0% $\Delta$	3.0%	29.8% $\Delta$	<b>37.8%</b> $\Delta$	29.9% $\Delta$	-6.4% $\Delta$	33.1% $\Delta$	-61.5%	34.6% $\Delta$	5.4%
d	WFG7 ( $\Delta_{time}^{90\%}$ )								WFG9 ( $\Delta_{time}^{90\%}$ )							
	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5	0.01	0.02	0.04	0.05	0.1	0.2	0.25	0.5
0.2	11.1% $\Delta$	9.3% $\Delta$	<b>12.1%</b> $\Delta$	10.1% $\Delta$	7.4% $\Delta$	2.0%	1.1%	2.4%	19.5% $\Delta$	17.6% $\Delta$	<b>19.9%</b> $\Delta$	15.9% $\Delta$	15.0% $\Delta$	10.8%	14.6%	3.4%
0.4	20.6% $\Delta$	<b>22.4%</b> $\Delta$	22.0% $\Delta$	21.8% $\Delta$	22.0% $\Delta$	19.0% $\Delta$	15.2% $\Delta$	-1.4%	<b>32.2%</b> $\Delta$	29.8% $\Delta$	30.8% $\Delta$	26.5% $\Delta$	27.0% $\Delta$	24.1% $\Delta$	24.2% $\Delta$	0.3%

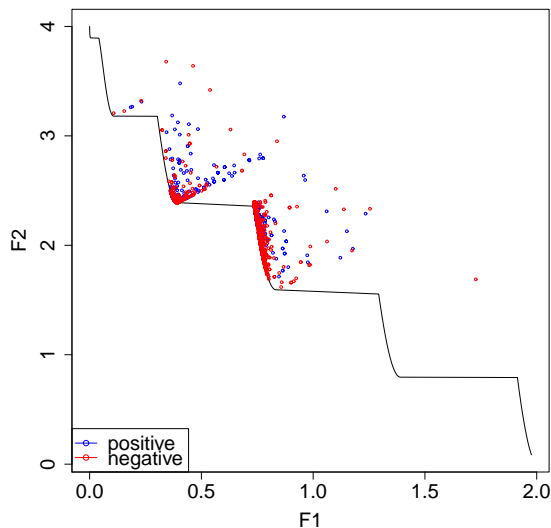
- 9) M Depolli, R Trobec, and B Filipic. Asynchronous master-slave parallelization of differential evolution for multi-objective optimization. *EVOLUTIONARY COMPUTATION*, Vol. 21, No. 2, pp. 261 – 291, 2013.
- 10) Juan J. Durillo, Qingfu Zhang, Antonio J. Nebro, and Enrique Alba. Distribution of Computational Effort in Parallel MOEA/D. In Carlos A. Coello Coello, editor, *Learning and Intelligent Optimization*, Vol. 6683 of *Lecture Notes in Computer Science*, pp. 488–502. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- 11) Tomohiro Harada and Keiki Takadama. Performance Comparison of Parallel Asynchronous Multi-Objective Evolutionary Algorithm with Different Asynchrony. In *Proceedings of the 2017 IEEE Congress on Evolutionary Computation (CEC 2017)*, 2017.
- 12) Byung-Il Koh, Alan D George, Raphael T Haftka, and Benjamin J Fregly. Parallel asynchronous particle swarm optimization. *International journal for numerical methods in engineering*, Vol. 67, No. 4, pp. 578–595, 07 2006.
- 13) Andrew Lewis, Sanaz Mostaghim, and Ian Scriven. Asynchronous multi-objective optimisation in unreliable distributed environments. In Andrew Lewis, Sanaz Mostaghim, and Marcus Randall, editors, *Biologically-Inspired Optimisation Methods*, Vol. 210 of *Studies in Computational Intelligence*, pp. 51–78. Springer Berlin Heidelberg, 2009.
- 14) Sidney R. Maxwell III. Experiments with a coroutine model for genetic programming. In *Proceedings of the 1994 IEEE World Congress on Computational Intelligence*, Vol. 1, pp. 413–417a, Orlando, Florida, USA, 27-29 June 1994. IEEE Press.
- 15) A. Milani and V. Santucci. Asynchronous differential evolution. In *Proceedings of the 2010 IEEE Congress on Evolutionary Computation (CEC 2010)*, pp. 1–7, July 2010.
- 16) Antonio J. Nebro, Juan J. Durillo, and Matthieu Vergne. Redesigning the jmetal multi-objective optimization framework. In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation, GECCO Companion '15*, pp. 1093–1100, New York, NY, USA, 2015. ACM.
- 17) Shigeru Obayashi, Shinkyu Jeong, Koji Shimoyama, Kazuhisa Chiba, and Hiroyuki Morino. Multi-Objective Design Exploration and its Applications. *International Journal of Aeronautical and Space Sciences*, Vol. 4, No. 4, Dec 2010.
- 18) Tea Robič and Bogdan Filipič. *DEMO: Differential Evolution for Multiobjective Optimization*, pp. 520–533. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- 19) Sergio Santander-Jiménez and Miguel A. Vega-Rodríguez. Asynchronous non-generational model to parallelize metaheuristics: A bioinformatics case



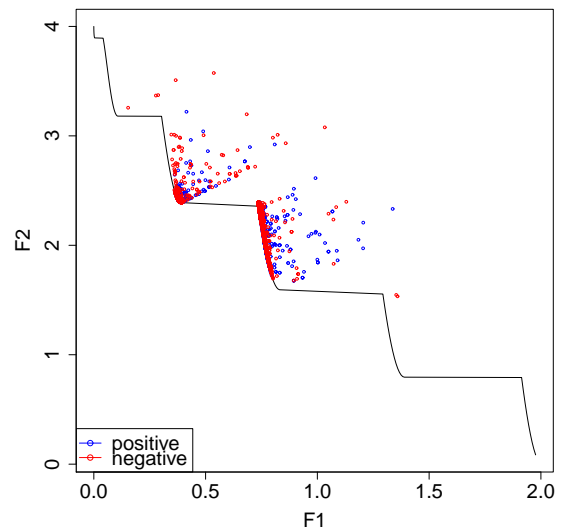
(a)  $\alpha = 0.01$



(b)  $\alpha = 0.02$



(c)  $\alpha = 0.5$



(d)  $\alpha = 1.0$

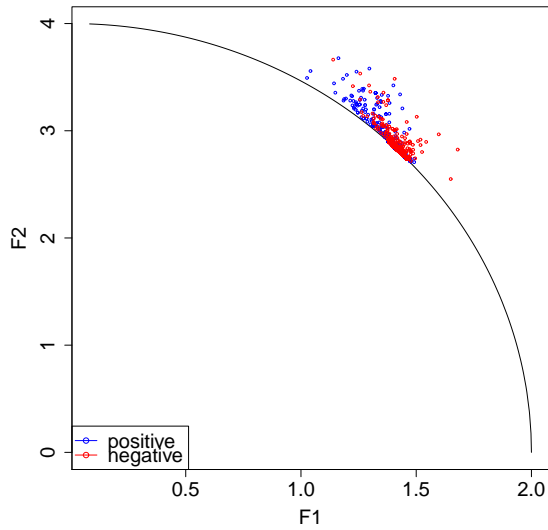
Fig. 4: Distribution of solutions at 50% of the population in each generation (WFG2)

study. *IEEE Transactions on Parallel and Distributed Systems*, Vol. PP, No. 99, pp. 1–1, 2016.

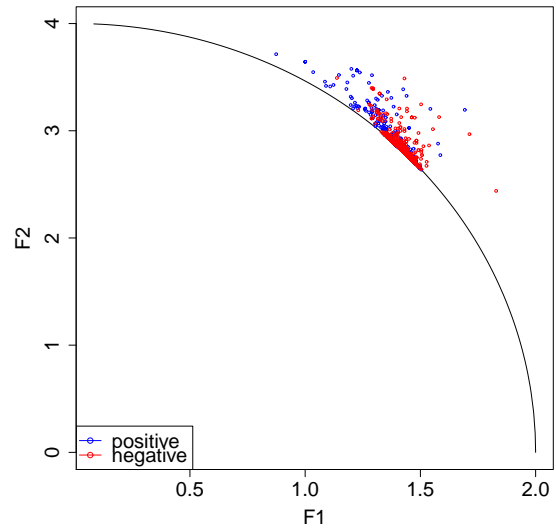
- 20) Sergio Santander-Jiménez and Miguel A. Vega-Rodríguez. Performance evaluation of dominance-based and indicator-based multiobjective approaches for phylogenetic inference. *Inf. Sci.*, Vol. 330, No. C, pp. 293–314, February 2016.
- 21) Eric O. Scott and Kenneth A. De Jong. Evaluation-time bias in asynchronous evolutionary algorithms. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, GECCO Companion '15*, pp. 1209–1212, New York, NY, USA, 2015. ACM.
- 22) Eric O. Scott and Kenneth A. De Jong. Understanding simple asynchronous evolutionary algorithms. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII, FOGA '15*, pp. 85–98, New York, NY, USA, 2015. ACM.
- 23) D.K. Tasoulis, N.G. Pavlidis, V.P. Plagianakos, and M.N. Vrahatis. Parallel differential evolution. In *Evo-*

*lutionary Computation, 2004. CEC2004. Congress on*, Vol. 2, pp. 2023 – 2029 Vol.2, June 2004.

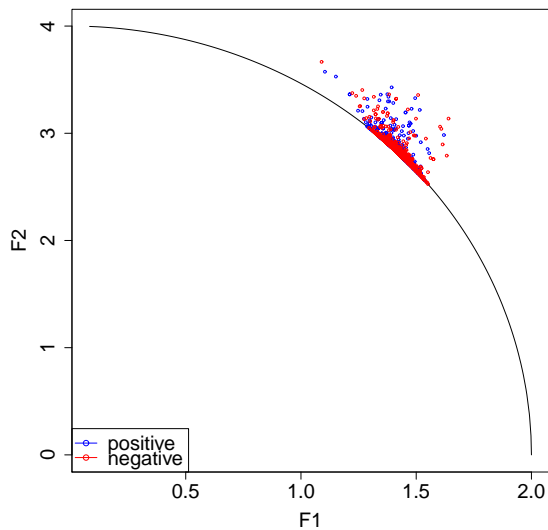
- 24) Simon Wessing, Günter Rudolph, and Dino A. Menges. *Comparing Asynchronous and Synchronous Parallelization of the SMS-EMOA*, pp. 558–567. Springer International Publishing, Cham, 2016.
- 25) Alexandru-Ciprian Zvoianu, Edwin Lughofer, Werner Koppelsttter, Günther Weidenholzer, Wolfgang Amrhein, and Erich Peter Klement. Performance comparison of generational and steady-state asynchronous multi-objective evolutionary algorithms for computationally-intensive problems. *Knowledge-Based Systems*, Vol. 87, pp. 47 – 60, 2015. Computational Intelligence Applications for Data Science.
- 26) Evgeniya Zhabitskaya and Mikhail Zhabitsky. Asynchronous differential evolution with restart. In Ivan Dimov, Istvan Farag, and Lubin Vulkov, editors, *Numerical Analysis and Its Applications*, Vol. 8236 of *Lecture Notes in Computer Science*, pp. 555–561. Springer Berlin Heidelberg, 2013.



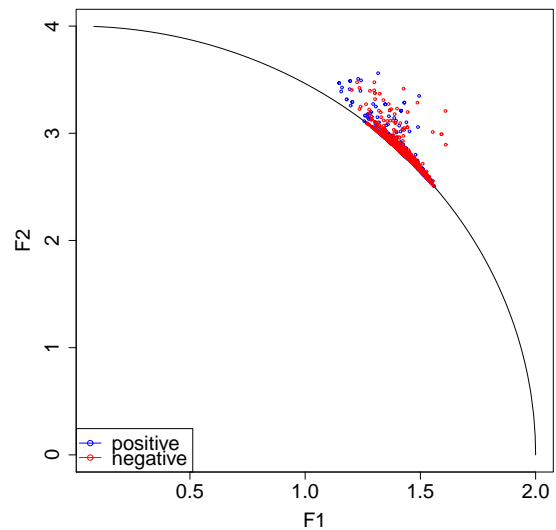
(a)  $\alpha = 0.01$



(b)  $\alpha = 0.02$



(c)  $\alpha = 0.5$



(d)  $\alpha = 1.0$

Fig. 5: Distribution of solutions at 50% of the population in each generation (WFG4)

- 27) E. Zitzler, K. Deb, and L. Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, Vol. 8, No. 2, pp. 173–195, 2000.
- 28) Eckart Zitzler and Lothar Thiele. *Multiobjective optimization using evolutionary algorithms — A comparative case study*, pp. 292–301. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.



# 近赤外分光法を用いた 不快音聴取時における脳機能活性箇所の特

○遠藤真央香 満倉靖恵 (慶應義塾大学)

## Identification of Functional Activities in the Brain when Hearing Unpleasant Sounds using Near Infra-Red Spectroscopy

\* M. Endo and Y. Mitsukura (Keio University)

**Abstract**— The purpose of this study is identification of functional activities in the brain when hearing unpleasant sounds using Near Infra-Red Spectroscopy (NIRS). Hearing unpleasant sounds has deleterious emotional effects and a previous study found particular electroencephalogram signals when hearing unpleasant sounds. That is, hearing them have a relationship between hearing sounds and functional activities in the brain.

We measured using NIRS. The data are analyzed by 1<sup>st</sup> order auto-regression model and two samples t-test. As the result, we confirmed 16 significant activities and more significant activities on the right side of the brain than the left side. From this result, we could identify the activating area by hearing unpleasant sounds.

**Key Words:** Functional activities in the brain, Unpleasant sounds, Near infra-red spectroscopy

### 1 Introduction

Sound is vibration of air which always multifariously exist in our life, and they elicit human emotions. The sounds are transmitted from any organ near the ear to auditory region in the brain, and transformed into an electrical signal. The electrical signal is transmitted to the frontal region in the brain. The frontal region in the brain is the region where takes a decision on human unpleasant feeling based on processing at the frontal region. That is to say, the frontal region assumes high-order functions such as thinking and creativity. Above all, the unpleasant sounds multifariously exist as “noise,” and hearing unpleasant sounds has deleterious emotional effects<sup>1)</sup>. Moreover, according to the previous study, it is clear that there are particular electroencephalogram (EEG) signals when hearing unpleasant sounds<sup>2)</sup>. Namely, hearing unpleasant sounds has a connection with functional activities in the brain. However, the degree and influence on activation in the brain by hearing these sounds haven’t been clear yet.

To clarify that the degree and influence on function in the brain, observation of varying oxygenation in the brain cortex is effective for identification of functional activities in the brain when hearing unpleasant sounds. It is well-known that the amount of variation of oxygen during brain functional activities can be measured by non-invasive brain functional measurement device named Near Infra-Red Spectroscopy (NIRS)<sup>3)</sup>. The device captures information about the amount of oxygenated hemoglobin (Oxy-Hb) by difference between amount of incident light on light emitter and amount of detection light on photo-detector. Oxygen in the human body is carried by hemoglobin. If the oxygenated hemoglobin releases oxygen, it will be deoxygenated hemoglobin. Therefore, the amount of oxygen consumed will increase if the point of activated brain region is expanded or the function is

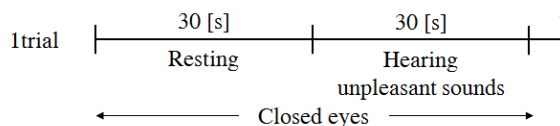


Fig. 1: The experimental flow

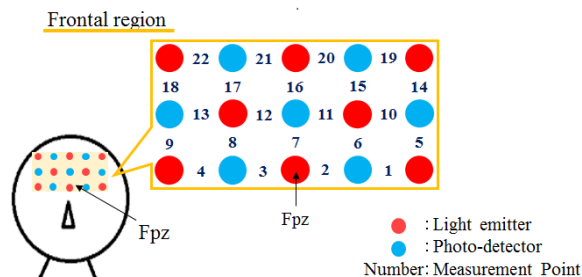


Fig. 2: Measurement points

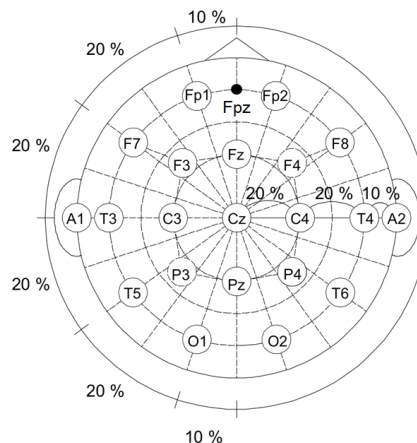


Fig. 3: The international 10-20 system

reacted. NIRS has some advantage of easy to prepare for measurement, low impact on subject, and higher time resolution than other non-invasive brain measurement devices such as fMRI(functional magnetic resonance imaging) and PET(positron emission tomography).

Therefore, the purpose in this study is identification of functional activities in the brain when hearing unpleasant sounds using NIRS.

## 2 Experiments

Four students (three male and one female aged from 22 to 25 years old) participated in this study. All subjects were healthy. The experimental flow consists of consecutive resting for 30s and hearing unpleasant sounds for 30s (see Fig. 1). During the experiment, the eyes were closed.

The sounds for experiment are five kinds of daily life noise mandated by the environmental agency, the sounds of vacuum cleaner, a fan gunning, a motorcycle engine, a crowd, and a laundry machine. Subjects listen to these sounds randomly.

The measurement device is NIRS. It can assess changes in the concentration of Oxy-Hb of the human cortex. NIRS had less binding force to the head and is not uncomfortable, so it could measure brain functional activities in a state which is close to real life conditions. Data were acquired using a 3×5 optode set, that included 22 channels (seven photo-detectors and eight light emitters) (see Fig. 2). The probe set was placed on a specialized bathing cap for fixation of the electrodes and attached on the scalp. The middle optode of the bottom row was set to Fpz according to the international 10-20 system (see Fig. 3) for assuring minimum variability between sensor position across multiple subjects as well as coverage of frontal cortex regions.

In this study, data about the amount of Oxy-Hb were used for analysis.

## 3 Analysis Methods

### 3.1 Preprocessing

Acquired data were separated into two kinds of sections (resting and hearing unpleasant sounds), and baseline correction was applied on each section. Thereby, the two types of section were acquired each subject and each measurement point on Oxy-Hb.

### 3.2 1<sup>st</sup> order auto-regressive modeling

The auto-regressive (AR) model is one of the time-series modeling methods which estimates future value based on past data. Furthermore, since data are stochastically modeled, all of the data can be utilized. We employed first-order AR model for preprocessing data. The higher the order of AR modeling was, the lower its general versatility was. It was because high order AR model will be very complicated. This analysis method was employed for each subject and each measurement point. The first-order AR model is shown as follows:

$$x(t) = \alpha x(t-1) + \epsilon(t)$$

where  $x(t)$ ,  $x(t-1)$ ,  $\alpha$ , and  $\epsilon(t)$  were values at the times  $t$  and  $t-1$ , AR coefficient, and error value, respectively. For fitting observations with AR models, maximum likelihood estimation and Yule-Walker's method were used. The AR coefficients express weight of the past measurement data. The AR coefficient is a value representing the importance of measurement data.

### 3.3 Two samples t-test

The two samples t-test is a method to compare the change between two individual samples based on t-distribution. In this study, we employed the statistical test (level of significance is 5%) using AR coefficients acquired during resting and hearing unpleasant sounds for each measurement point. Then, we obtained the t-value which expresses the degree of difference between two states for each measurement point. The t-value  $T$  was calculated as follows:

$$T = \frac{\bar{a}_h - \bar{a}_r}{\sqrt{\frac{1}{n}(s_h^2 + s_r^2)}}$$

where,  $\bar{a}_h$ ,  $\bar{a}_r$ , and  $n$  were the average AR coefficient when hearing unpleasant sounds, average AR coefficient during resting state, and number of subjects, respectively. Furthermore,  $s_h$  and  $s_r$  were the standard deviation when hearing sounds and the standard deviation during resting state. The t-value would be large if the difference between two states was large, variance was small, or the number of subjects was large.

Subsequently, we made a color map using the t-value. In this paper, the t-value translated to the degree of brain functional activity. When the difference is large, the t-value is large and turns red on the color map. Reddening in the color map expresses that oxygen concentration in blood of the observed region of the brain is large. Therefore, it means that the neurons are active.

## 4 Results and Discussions

The result of the two-sample t-test is shown in Fig. 4 and table 1. The color map is colored based on the t-value, and the number characters in it represent the measurement points. The side of numbers 5 and 14 is right side of the brain. Black number characters represent points where a significant difference was confirmed. On the other hand, white number characters represent points where a significant difference was not confirmed. Significant difference was confirmed mean that t-value exceeded the critical value (2.07). Furthermore, the bigger the t-value, the more different the functional activities in the brain are between two states. Fig. 4 and table 1 shows that there are significant differences on 16 of the measurement points of the frontal region on the brain. The areas showed clear statistically significant difference in activation when hearing unpleasant sounds.

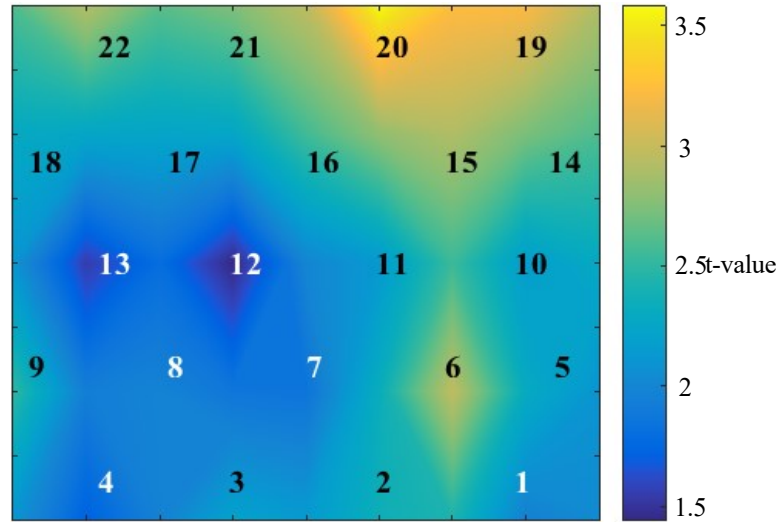


Fig. 4: The color map of t-value

Table 1: Result of two samples t-test (“\*” means that there is a significant difference, and “n. s.” means that there is not a significant difference.)

Measurement point		1	2	
Result		n.s.	*	
3	4	5	6	7
*	n.s.	*	*	n.s.
8	9	10	11	12
n.s.	*	*	*	n.s.
13	14	15	16	17
n.s.	*	*	*	*
18	19	20	21	22
*	*	*	*	*

That is, the degree and affected areas when hearing unpleasant sounds can be identified. This result will help to decide measurement points in future experiments. Next we intend to measure the electroencephalogram around the area which showed significant activation in this study. The electroencephalogram is able to measure action potentials in the brain, and we can get multimodal influence of unpleasant sounds on the brain.

## 5 Conclusion

The purpose of this study was identification of functional

activities in the brain when hearing unpleasant sounds using NIRS. We measured functional activities in the brain when resting and hearing unpleasant sounds using NIRS, and analyzed acquired data. As the result, the activating area was identified by using color map.

However, the number of subjects is statistically few. Future works include increasing the number of subjects and measuring electroencephalogram and NIRS at the same time. In doing so, we would multimodally identify the influence of functional activities in the brain when hearing unpleasant sounds.

## Acknowledge

This study was supported by JSPS KAKENHI (Scientific Research (A)) Grant Number YYK5A05.

## References

- 1) K. D. Kryter and F. Poza: Effects of Noise on Some Autonomic System Activities, *The Journal of the Acoustical of America*, **(67)** 6, 2036/2044 (1980).
- 2) S. Nishifuji and I. Miyahara: Amplitude and Phase Response of Brain Alpha Wave to Unpleasant and Pleasant Sounds, *The Society of Instrument and Control Engineers*, **(44)** 10, 776/778 (2008)
- 3) R. Takizawa and M. Fukuda: Functional Neuroimaging as a Laboratory Test and The First Advanced Medical Technology in Psychiatry, *Medix*, **(53)**, 30/35 (2010)

# 乳がん病変検出のための 深層学習を用いた計算機支援画像診断システム

○鈴木真太郎 (東北大学) 張曉勇 (仙台高等専門学校) 本間経康 市地慶  
魚住洋佑 高根侑美 柳垣聡 川住祐介 石橋忠司 吉澤誠 (東北大学)

## Deep Learning based Computer-Aided Diagnosis System for Mammographic Lesion Detection

\*S. Suzuki (Tohoku University), X. Zhang (Sendai National College of Technology),  
N. Homma, K. Ichiji, Y. Uozumi, Y. Takane, S. Yanagaki, Y. Kawasumi,  
T. Ishibashi, and M. Yoshizawa (Tohoku University)

**Abstract**— In recent years, deep convolutional neural network (DCNN) has widely been applied for image recognition, and shown a remarkable performance in various natural image-related applications. However, for medical image-related application, due to the limitation of training data and the difference between the natural and medical images, training the DCNN for medical image recognition is still a research topic. In this paper, we propose a DCNN-based method for lesion detection in mammograms. The proposed method consists of the following two steps. Given a mammogram, several lesion candidates with various locations and sizes are firstly detected and extracted from the mammogram based on their intensity characteristics. Secondly, a transfer learning-based method is applied for training an existing DCNN to classify the lesion candidates into lesions or normal tissues. The candidate detection and the DCNN-based classifier are tested on a public mammogram database. Compared with several previous studies, our proposed method researched a higher true positive rate and a lower false positive of lesion detection.

**Key Words:** Deep convolutional neural network (DCNN), Computer-aided diagnosis (CAD), Mammogram

## 1 はじめに

近年、乳がん罹患率の急激な上昇<sup>1)</sup>と早期発見の重要性が周知されてきたことを背景に、乳房X線撮影(マンモグラフィ)を用いた画像検診受診者数が増加傾向にある。しかしこれに伴い読影医の負担も増加し、疲労に起因した病変の見落とし等の発生が懸念されている。この対策として期待されているのが、コンピュータを用いて乳房X線画像上の病変を自動検出し、第2の意見として医師に提示することで医師の診断を補助する、計算機支援診断(computer-aided diagnosis: CAD)システムの開発である<sup>2)</sup>。

従来のCADシステム<sup>3, 4, 5, 6)</sup>は、医師の読影論理に倣った画像解剖学的な病変の特徴を手動で設計し、抽出された特徴量を基に病変か正常かを識別するものが多い。しかし、医師の読影論理は専門知識のほか、経験に基づいて定性的に表現される場合が多く、病変特徴量を定量的に表現することは極めて難しいため、CADシステムの病変検出性能向上は容易ではなかった。特に、病変のみを正確に表す特徴量の定義は難しく、病変を見逃さないように検出すると、同時に似たような特徴をもつ正常組織も多く誤検出してしまいう問題があった。

そこで本研究では、この特徴抽出の困難さを解決するため、近年注目を集めている深層学習の一種であるdeep convolutional neural network (DCNN)を用いた乳房X線画像上の病変検出システムを提案する。

DCNNは、特徴量および識別規則を学習的に自動獲得することができ、自然画像認識の分野においては従来法を上回る高い性能を示すことが知られている<sup>7)</sup>。しかし、DCNNの学習に必要な大量の医用画像を収集することは極めて困難である。また、DCNNの長所の一つである位置や大きさの違いにロバストな認識性能は、

逆に病変の位置や大きさなど診断上有用な情報を失うことに繋がりがねない問題もある。

このため本研究では、はじめに乳房X線画像上では病変が周囲に比べて高輝度、すなわち高コントラストを有することなど、従来法で蓄積された特徴の知見や手法を基にしたシンプルな前処理により、病変サイズや位置情報を保持したまま、大まかな候補選定を行う。このような従来法の知見に基づく特徴は、多くの病変に共通するため真陽性(病変を正しく検出する)数は多い一方、従来法の問題で述べたように病変だけを正確に表現することはできないため、偽陽性(正常組織が病変と誤検出される)候補も多く選定される。そこで、これらをDCNNにより正しく識別する2段階検出法を提案し、最終的な偽陽性数を低減する。またDCNNの学習にあたっては、事前に大量の自然画像を用いて学習済みのDCNNに対し、医用画像を用いた転移学習<sup>8)</sup>を適用することで大量の医用画像の収集困難性を解決しつつ、病変の識別性能向上を試みた。

## 2 提案手法

### 2.1 病変候補抽出

前処理としての病変候補抽出の目的は、DCNNでは違いが吸収されてしまう病変の位置や大きさを抽出・保持したまま、大まかな病変候補の選定を行うことである。したがって、多くの病変に共通するシンプルな特徴に基づいて出来る限り多くの病変を検出することを優先し、偽陽性(誤検出)を許容するものとする。

提案する病変候補抽出法は、Fig.1に示すように、検出対象領域の抽出、病変候補領域の抽出、病変候補ROI画像の作成の3つのステップからなる。

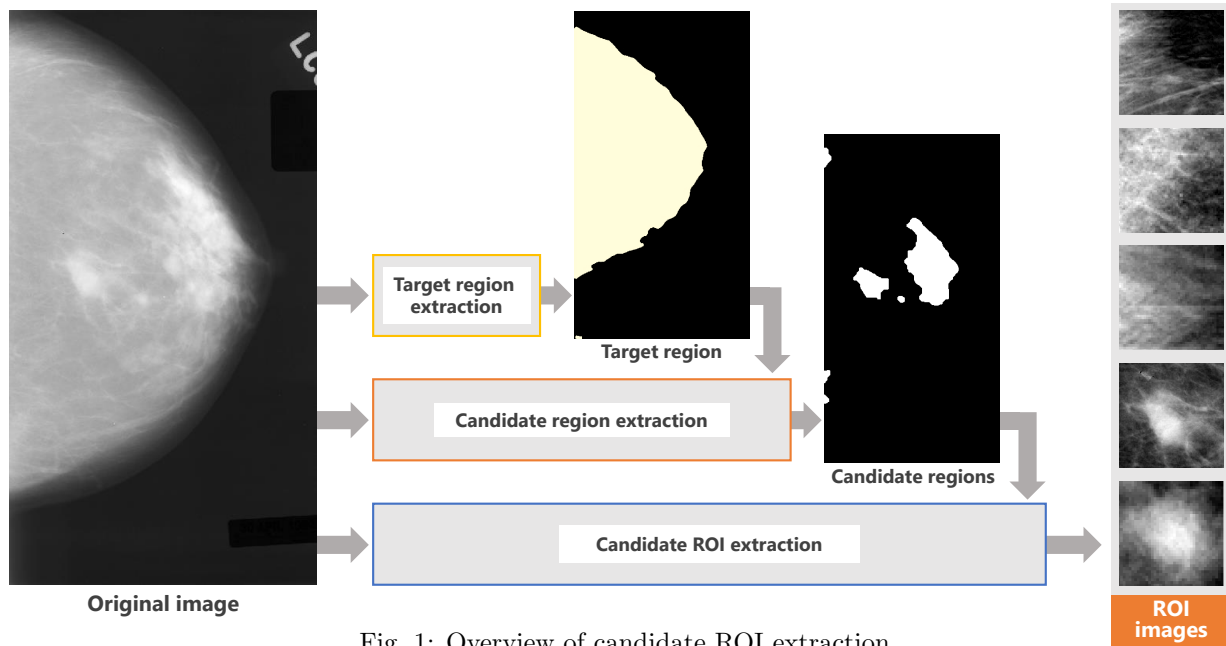


Fig. 1: Overview of candidate ROI extraction

### 2.1.1 検出対象領域の抽出

乳房 X 線画像では、Fig. 1 左端の原画像のように、しばしば乳房領域外に撮影時の諸情報を記録したラベル等が存在する。これらは以降の検出対象領域抽出の妨げとなるため、事前に乳房領域のみを抽出しておく必要がある。本研究では、原画像を大津の閾値法<sup>9)</sup>を用いて二値化し、最も面積の大きい領域を乳房領域として抽出することで、Fig. 2(a) に示すような乳房領域のみの画像を得る。

乳房は解剖学的に乳腺と脂肪の主に 2 つの組織からなり、乳房 X 線画像の輝度値は主にこの 2 つの重ね合わせによって構成される。すなわち輝度値は X 線透過方向における両者の厚さで決定するといえるが、撮影は乳房を均一な厚みに押しつぶした状態で行われるため、全域に渡って厚さが一定に保たれていると仮定すると、輝度値は両者の割合で決まり乳腺の割合が大きいほど高輝度になると言える。ここで、輝度に基づいて以下のように画像解剖学的な乳腺密度分布（割合）ごとに 4 つの領域に分割する。

- class 1:** 乳房と背景領域の境界 (marginal area)
- class 2:** 脂肪組織 (fat tissue)
- class 3:** 脂肪組織と乳腺組織の混合組織 (fat and fibroglandular tissue)
- class 4:** 高密度乳腺組織 (dense tissue)

すなわち、class 1 が最も低輝度で、class 2, 3 の順に高輝度となり、class 4 が最も輝度が高い。

Fig. 2(a) に示す画像のヒストグラムを Fig. 3 に示す。ヒストグラムは 3 ~ 4 つ程度の極大値を持ち、各々の両側はなだらかに減少する正規分布的な形状となっており、各々の分布が各 class に相当するものと推測される。そこで各 class から求めた確率密度関数が正規分布であると仮定し、その確率密度関数をガウス混合モデル (Gaussian mixture model: GMM) を用いて 4 つのガウス分布の混合により近似することで、画像の乳房領域を上記 4 つの class に分割する<sup>10)</sup>。Fig. 3 上の 4 本の実線および破線は、それぞれ 4 つの class に対応

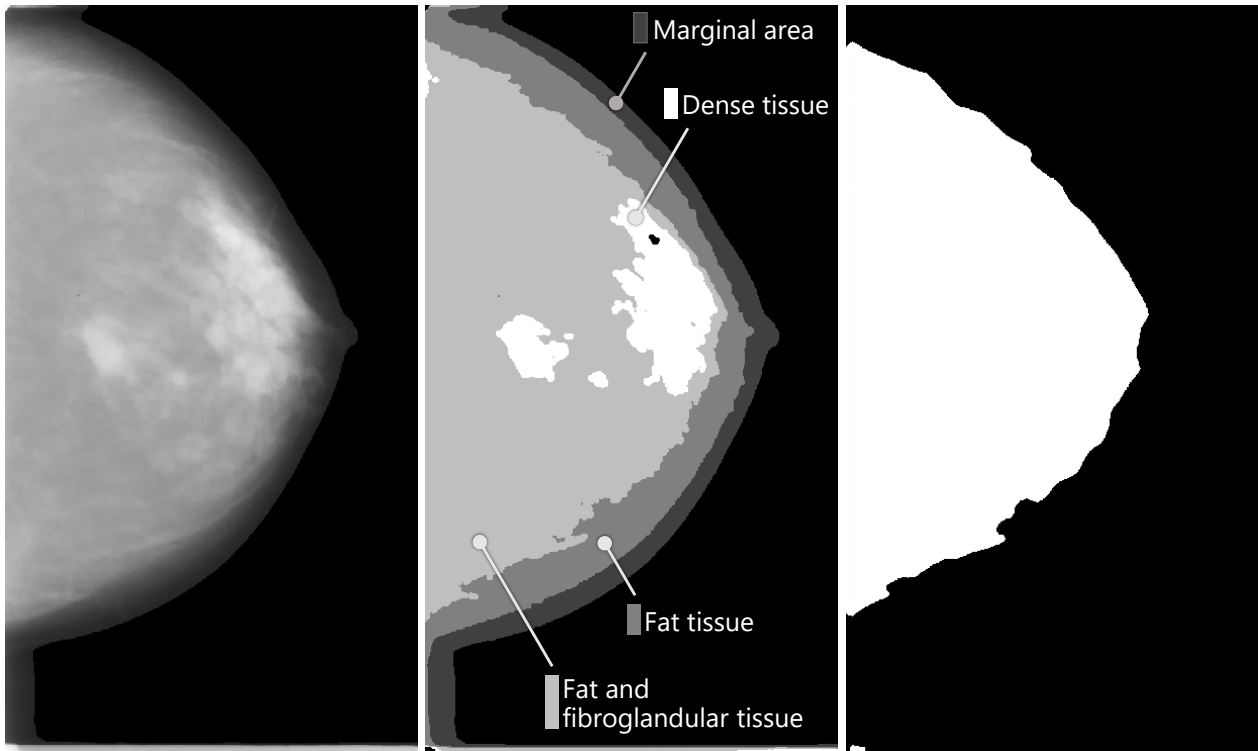
するガウス分布と、それらの混合により近似された確率密度分布を表す。画像の各輝度値について、4 つのガウス分布のうち確率密度が最も大きいものに対応する class に属するものとして、画像上の乳房領域を分割した結果を Fig. 2(b) に示す。上記の画像解剖学的な乳腺分布ごとに、4 つの領域に分割されていることが分かる。ここで、乳がん病変は一般に乳腺組織に発生することから、病変検出の対象とする領域を Fig. 2(c) に示すように乳腺が含まれる class 3 以上の高輝度領域に限定する。これにより、誤検出の低減と計算コストの低減が期待される。

### 2.1.2 病変候補領域の抽出

ここでは、Fig. 1 左端の原画像より Fig. 4(d) のような病変候補領域を示す二値画像を生成する。

はじめに、Fig. 1 左端の原画像より、脂肪組織などの非病変の背景組織を、形態学的フィルタの一種であるトップ・ハット変換<sup>11)</sup>を用いて除去した画像 Fig. 4(a) を作成する。トップ・ハット変換は、原画像に対してオープニング画像を差し引く処理であり、オープニングに用いた構造要素よりも小さな要素を原画像上から抽出する。乳房 X 線画像上において、一般的に病変は局所的に高い輝度値を持つのに対し、脂肪組織等の背景組織は低周波成分が支配的である。このため、構造要素を適切に設計することにより、トップ・ハット変換によって背景組織の輝度成分のみを減じることができると期待される。このような手法は乳房 X 線画像上の多くの病変に対して有効と考えられるが、構造要素の設計には検出対象の想定が必要なため、以下では代表的な乳がん所見の一つである腫瘤陰影の検出を例に説明する。

乳房 X 線画像上において腫瘤は類円形の高輝度領域として観察される。また、腫瘤陰影は一般に直径 5 cm 未満であることから、本実験では構造要素に直径 5 cm の円を用いた。Fig. 1 左端の原画像と背景除去後の画像 Fig. 4(a) を比べると、背景組織の低周波成分が除去され、局所的に高コントラストな領域がより明瞭に強調



(a) Label removed image (b) Segmented breast (c) Target region

Fig. 2: GMM-based breast segmentation and target region detection

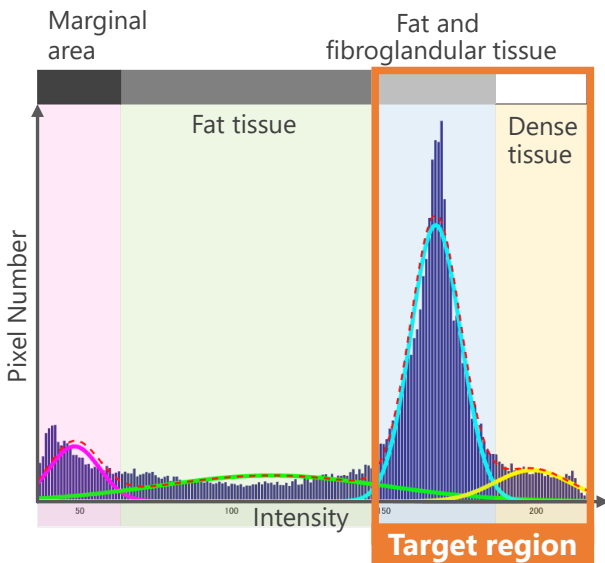


Fig. 3: Histogram approximation by GMM

されていることが確認できる。

次に、画像 Fig. 4 (a) に対し 2.1.1 節で得た検出対象領域 Fig. 2 (c) との論理積を取ることで、Fig. 4 (b) のように病変検出の対象となる領域のみを抽出する。

続いて、一般に病変は局所的に高い輝度値を持つことに着目し、Fig. 4 (b) で強調された領域ををもちに、閾値処理により二値化された Fig. 4 (c) のような病変候補領域を抽出する。閾値決定の手法については後述する。

次に、高輝度領域の二値化画像 Fig. 4 (c) に対し、オープニングとクロージング処理により高輝度領域の整形を行い、ここでの病変候補領域として Fig. 4 (d) のよう

な二値化画像を得る。オープニングおよびクロージング処理は、形態学的な平滑処理であり<sup>11)</sup>、対象の大きさ(面積)をあまり変化させずに平滑化を行うことができる。その効果を Fig. 5 を例に説明する。

Fig. 5 (a) は、閾値処理により二値化された高輝度領域の模式図であり、画像中央の高輝度領域は病変領域を模したものである。乳房 X 線画像上の病変領域は、乳腺組織等との重なりにより複雑な輝度分布を持つことが多く、二値化の閾値によっては Fig. 5 (a) のように、しばしば複数の高輝度領域に分割された状態で抽出される。このため、高輝度領域内の小さな穴状の欠損や領域間の隙間を埋める働きのあるクロージング処理を行うことにより、Fig. 5 (b) に示すように複数の分裂した高輝度領域をひとつの病変領域に整形する。また、Fig. 5 (a) 上の小さな輝点や線状の高輝度領域は、乳房 X 線画像上のノイズや正常乳腺による高輝度領域を模したものである。これらの明らかな非病変領域を除去するため、微小な高輝度領域を除去する働きをもつオープニング処理を行う。Fig. 5 (a) に対し、オープニング処理を行ったものを Fig. 5 (c) に示す。画像中央の病変領域はほぼ元のまま、ノイズや乳腺陰影を模した微小な高輝度領域が除去されたことが確認できる。このオープニング処理を、クロージング処理後の二値化画像 Fig. 5 (b) に対して行った結果を Fig. 5 (d) に示す。クロージング処理後に残った微小な非病変領域をオープニング処理により除去することで、整形された病変候補領域のみが抽出される。

最後に、二値化のための閾値決定の流れを Fig. 6 に示す。はじめに、閾値を最高輝度値(8bit 階調の場合は 255)に設定して二値化処理を行い、抽出された候補領域の数を求める。ここでの処理は、前処理の目的に

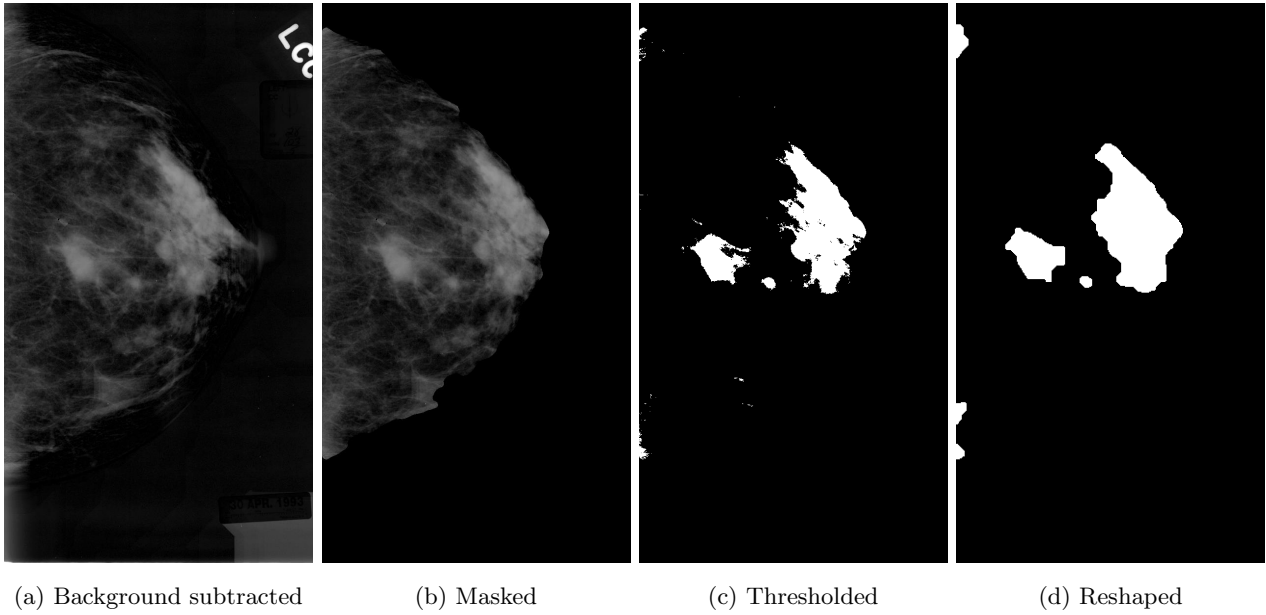


Fig. 4: Workflow of candidate region extraction

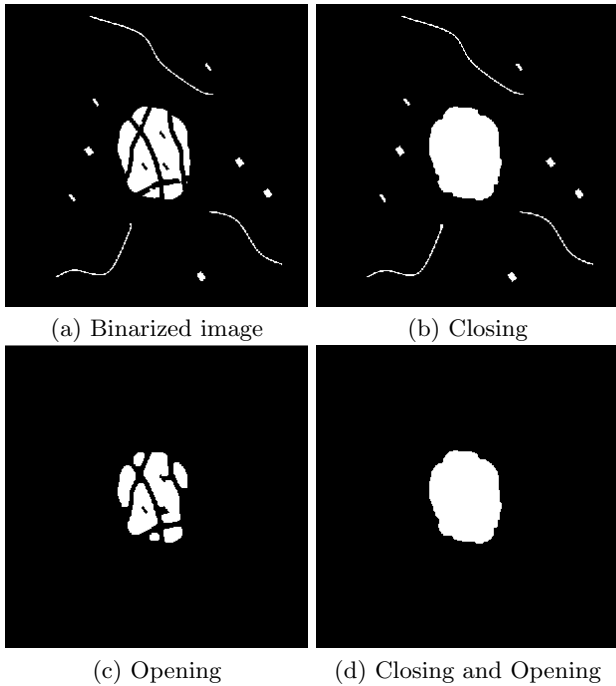


Fig. 5: Region reshaping

沿って病変領域の確実な抽出を優先し、非病変領域の誤抽出はある程度許容するが、極端に多くの誤抽出領域は計算量の観点からも好ましくない。また、一般に乳房 X 線画像上の腫瘍病変は 1,2 個程度であり、3 箇所を超えることはまれである。そこで、誤抽出も考慮に入れつつ 1 枚あたりの病変候補領域数の上限目安を 5 とし、抽出された候補領域の数が 5 よりも小さければ、閾値を 1 減じて同様に候補領域を抽出しその数を求める処理を、候補領域の数が 5 以上になるか、あるいは閾値が 0 に達するまで続ける。抽出された候補領域の数が 5 以上に達した場合、閾値をその値に決定し、病変候補領域の抽出を終了する。一方、候補領域数が 5 以上になることなく閾値が 0 に達した場合、候補領域数が最大となる閾値を用いて病変候補領域の抽出を

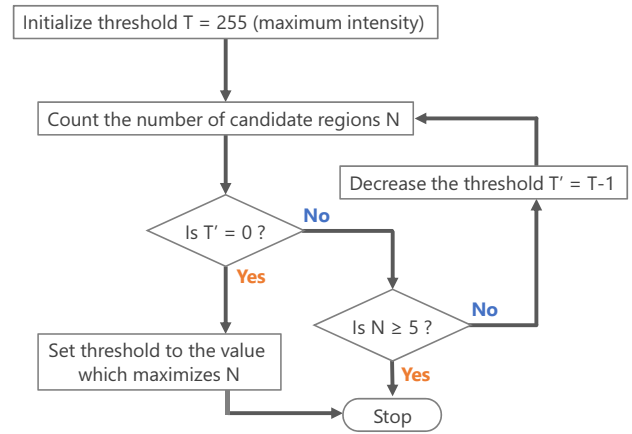


Fig. 6: The flowchart of threshold decision

行う。

### 2.1.3 病変候補 ROI 画像の作成

続いて、抽出された各病変候補領域の位置と大きさに従い、DCNN に入力する病変候補 ROI 画像を、2.1.2 節で述べた Fig.4 (a) のような背景組織除去後の乳房 X 線画像上より切り出す。本研究で用いる DCNN の入力サイズは  $227 \times 227$  画素の正方形画像であり、ROI 画像のサイズをこれに一致させる必要がある。また、病変候補が ROI からみ出してしまうと識別が非常に困難となることから、ROI 画像の一辺の長さ  $L_{ROI}$  は病変候補の水平、垂直方向の長さをそれぞれ  $L_x, L_y$  に対し余裕を持たせて、 $L_{ROI} = 1.5 \times \max(L_x, L_y)$  の正方形とした。ROI の位置は、病変領域の水平、垂直方向それぞれの中央を中心とするが、ROI が原画像上からはみ出す場合には、ROI の大きさは変えず位置を ROI が原画像に収まるまで内側に移動する。DCNN は入力画像上での対象の位置変動に対して非常に高いロバスト性を有することが知られており、候補領域が ROI 内に収まってさえいれば識別性能に影響はないと考えられるためである。各 ROI は切り出し後、前述の DCNN の入力サイズに合わせてリサイズされる。補間法には

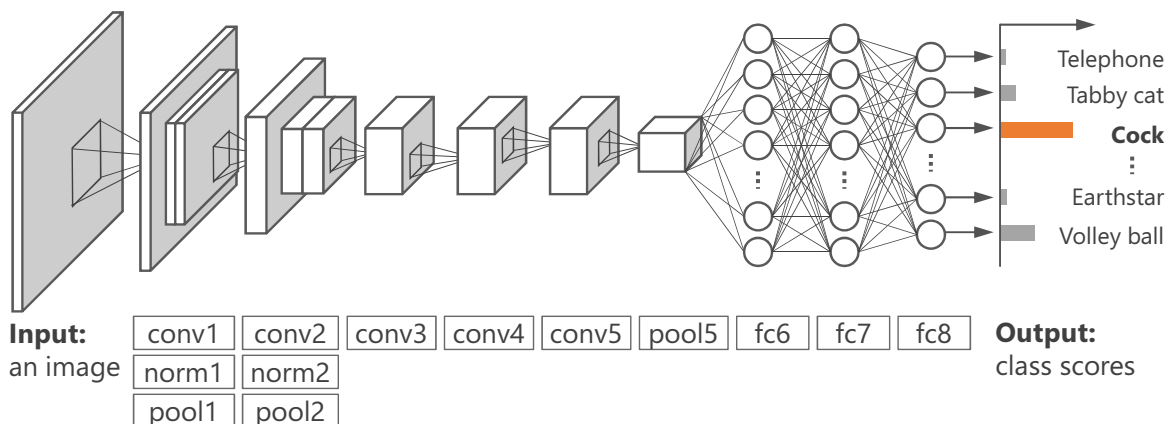


Fig. 7: Architecture of the *AlexNet*<sup>7)</sup> consisting of 5 convolution layers (conv1~conv5), 2 normalization layers (norm1, norm2), 3 pooling layers (pool1, pool2, and pool5), and 3 fully-connected layers (fc6~fc8)

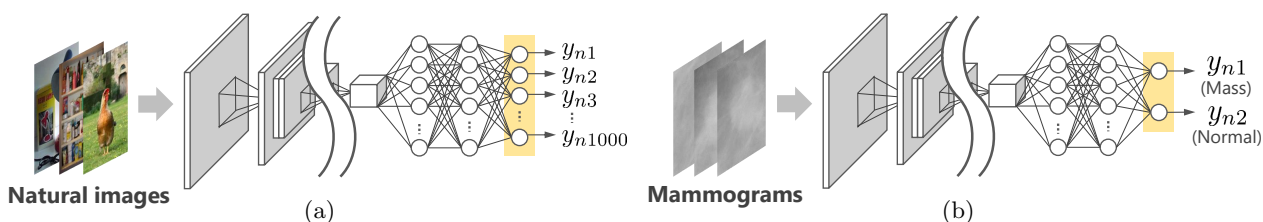


Fig. 8: The fine-tuning transfer learning for mass detection in mammograms. (a) The DCNN is pre-trained by using a large-scale image dataset for 1,000 classes. (b) We modify the last layer into two neurons, corresponding mass and normal classes respectively, and subsequently retrain the DCNN using a small-scale mammogram dataset.

バイキュービック補間を用いた。これにより、Fig. 1 右端に示すように、中心付近に病変候補の高輝度領域を含む ROI 画像を得る。

## 2.2 DCNN による病変候補の識別

### 2.2.1 DCNN の基本構造

本稿で識別器として用いた DCNN は、Krizhevsky et al. による *AlexNet*<sup>7)</sup> と同様の構造を持つものである。Fig. 7 に示すように、この DCNN は 5 層の畳込み層と 3 層の全結合層をもつ。DCNN への入力として 3 チャンネル (R,G,B) の画像輝度値を与え、出力層 (*fc8*) は、入力画像が *ImageNet*<sup>12)</sup> における 1,000 種類のクラスそれぞれに属する確率を出力する。Fig. 7 における *conv1* - *conv5* 層は特徴抽出を担う畳込み層、*pool1*, *pool2*, *pool5* 層は次元と計算量の削減および特徴の位置変動に対するロバスト性の獲得を担うプーリング層、*fc6* - *fc8* 層はクラス識別を行う全結合層である。なお、*AlexNet* について詳細は参考文献<sup>7)</sup> を参照されたい。

### 2.2.2 DCNN の転移学習

転移学習とは、あるタスクで学習した「知識」(対象の識別に有用な特徴や識別規則)を別のタスクに転用する手法である<sup>8)</sup>。本稿では、まず大量に収集可能な自然画像で DCNN を学習し、獲得した知識すなわち特徴量や識別規則を、本来の対象タスクである病変候補の識別に転用し、病変候補の識別を学習させる。具体的な手順は以下のとおりである。まず Fig. 8(a) に示すように、約 120 万枚の自然画像からなる *ImageNet*<sup>12)</sup> データセットを用いて DCNN を学習する。この操作を事前学習と呼ぶ。次に、DCNN の出力層を病変候補識別のタスクにおける分類クラス (病変/正常) に対応

する 2 つのニューロンからなる新たな全結合層で置換する。続いて Fig. 8(b) に示すように、病変候補 ROI 画像を用いて病変候補の識別を学習する。この操作を fine-tuning という。

この手順により、自然画像認識で獲得した基本的な知識 (エッジ等の基本的な特徴量) を病変識別に転用することができ、少量の医用画像でも病変特徴の獲得、識別が可能となると期待される。

## 3 実験結果

### 3.1 実験データセット

DCNN の学習 (fine-tuning) および病変検出テストは、乳房 X 線画像のデータベースである *Digital Database for Screening Mammography* (DDSM)<sup>13)</sup> の画像を用いて行った。学習およびテストに使用した ROI 画像数を Table 1 に示す。DCNN の学習には、乳房 X 線画像より 2.1.3 節の作成した 2,133 枚の ROI 画像を使用した。提案システムの病変検出テストでは、学習用データとは異なる 70 名の乳がん罹患者の乳房 X 線画像より同様に作成した ROI 画像 349 枚を DCNN で識別し、病変検出性能を評価した。なお、乳房 X 線画像には乳房を撮影する向きにより CC (頭尾方向) 画像と MLO (内外斜位方向) 画像の 2 種類があるが、MLO 画像は胸筋領域に起因する大きな高輝度領域を有する事が多い。このため本稿では CC 画像のみを対象に提案法を適用し、実験を行った。

### 3.2 病変検出性能の評価

病変検出性能の評価指標として、CAD システムの性能評価で一般的に用いられる free-response receiver operating characteristic (FROC) 曲線<sup>14)</sup> を用いた。一



Table 1: Experimental data (Number of ROIs)

	Training data	Test data	Total
Mass	509	91	600
Normal	1,624	258	1,882
Total	2,133	349	2,482

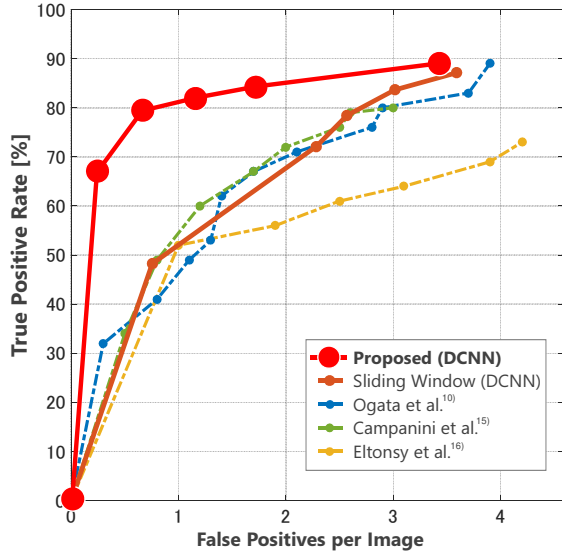


Fig. 9: FROC curves of the proposed system and conventional systems

般に、病変と正常の閾値を変更することにより、病変を正しく病変として検出する割合である真陽性率 (true positive rate: TPR) は変動するが、真陽性率と正常組織を病変と誤検出する偽陽性数 (false positives per image: FPI) は互いにトレードオフの関係にある。Fig. 9 に示すように、FROC 曲線は横軸に乳房 X 線画像 1 枚あたりの偽陽性検出数、縦軸に真陽性率をとる。したがって、FROC 曲線が左上を通るほど、病変検出性能が高いことを表す。Fig. 9 には、比較のため提案システム他、ある大きさの ROI を移動しながら画像全体にわたってスキャンする、sliding window 法による ROI 作成後に DCNN で識別を行う手法、ならびに手動での特徴設計による従来システム 3 手法<sup>10, 15, 16)</sup> の FROC 曲線を併せて示した。提案システムは、真陽性率が 80%、70% のときの偽陽性数がそれぞれ他手法の約 1/4、1/6 となり、実用的な真陽性率を保ったまま偽陽性数を大幅に低減できることが示唆された。また、提案法と sliding window 法の主な違いは 2.1 節で説明した提案法の第 1 段階における病変候補抽出の有無であるから、両者の比較により提案法の病変候補抽出が偽陽性数の低減に有効であることが示された。

### 3.3 考察

提案システムによる有効性が確認できた症例を Fig. 10 に示す。各画像上の緑色および黄色の矩形は、ROI の位置および大きさを表し、DCNN により病変と判定された ROI を黄色で、正常と判定された ROI を緑色で表す。また、医師の指定した真の病変領域を橙色で示している。Fig. 10 (a) の例では、提案システムにより 2 つの ROI が病変として検出されたことが分かる。うち 1 つの ROI が真の病変領域と一致しており、検出す

べき病変を適切に検出できたことを示し、もう一方の ROI は偽陽性検出である。Fig. 10 (b) は比較的大きな腫瘍を有する乳房 X 線画像の例であるが、大きな 2 つの ROI に跨る形で病変が適切に検出されている。この例では偽陽性検出はない。Fig. 10 (c) は、小さな腫瘍をもつ例である。病変の大きさに一致した小さな ROI により、病変が適切に検出されていることが確認できる。また、この例でも偽陽性検出はない。

続いて、提案法の問題点について考察する。Fig. 11 (a) に示した検出例では、病変の検出には成功しているものの、乳房領域のほとんどを覆うような極めて大きな ROI が検出されている。このような ROI が生成された原因として、Fig. 11 (b) に示すように、局所的に分布した高密度の乳腺領域がひとつの大きな高輝度領域として抽出されたことが挙げられる。候補領域の抽出は閾値処理に基づくので、組織の分布によってはこのような大きな領域が抽出されてしまう。しかし、高密度領域内に病変が存在することも多いため、単純に面積の大きな候補領域を除外することは、多くの病変を見逃し真陽性率の低下に繋がることから望ましくない。この問題の解決のためには、病変が検出された ROI 内で sliding window 法により病変の位置をより詳細に検出する方法や、DCNN の中間層の出力を利用するなどして病変検出の位置精度を高める必要がある。

また別の問題として、本システムの第 1 段階の病変候補抽出の過程において、テストデータの全腫瘍のうちおよそ 11% が見落とされている。これにより、Fig. 9 に示すように真陽性率が 90% を超えることができない。一方で DCNN による識別性能は良好であるため、第 1 段階の病変候補抽出においてはより多くの偽陽性検出を許しつつ、病変の見落としをより少なくするよう調整することで、真陽性率の改善が可能であると考えられる。

## 4 おわりに

本稿では、乳房 X 線画像上の病変検出を行う CAD システムとして、多くの偽陽性検出を許容した前処理的なシンプルな手法で病変候補を抽出し、これらを DCNN により正しく識別する 2 段階検出法を提案した。DCNN では違いを吸収してしまう病変の位置や大きさの情報を第 1 段階の候補抽出で保持することにより、極めて困難な特徴量の手動設計を行うことなしに病変検出性能の大幅な向上を実現すると同時に、病変の位置ならびに大きさの提示が可能な CAD システムを実現した。DCNN を応用したこのような CAD システムの構築は、我々の知る限り世界初の成果である。

今後データ数を増やして信頼性を向上させることにより、提案 CAD システムを用いて医師の診断効率および診断精度の向上が実現され、乳がんの早期発見・早期治療による生活の質の向上ならびに医師の負担軽減による医療費削減への効果が期待される。

## 5 謝辞

本研究は JSPS 科研費 15K20852, 26540112 ならびに東北大学スマート・エイジング学際重点研究センターの助成を受けた。

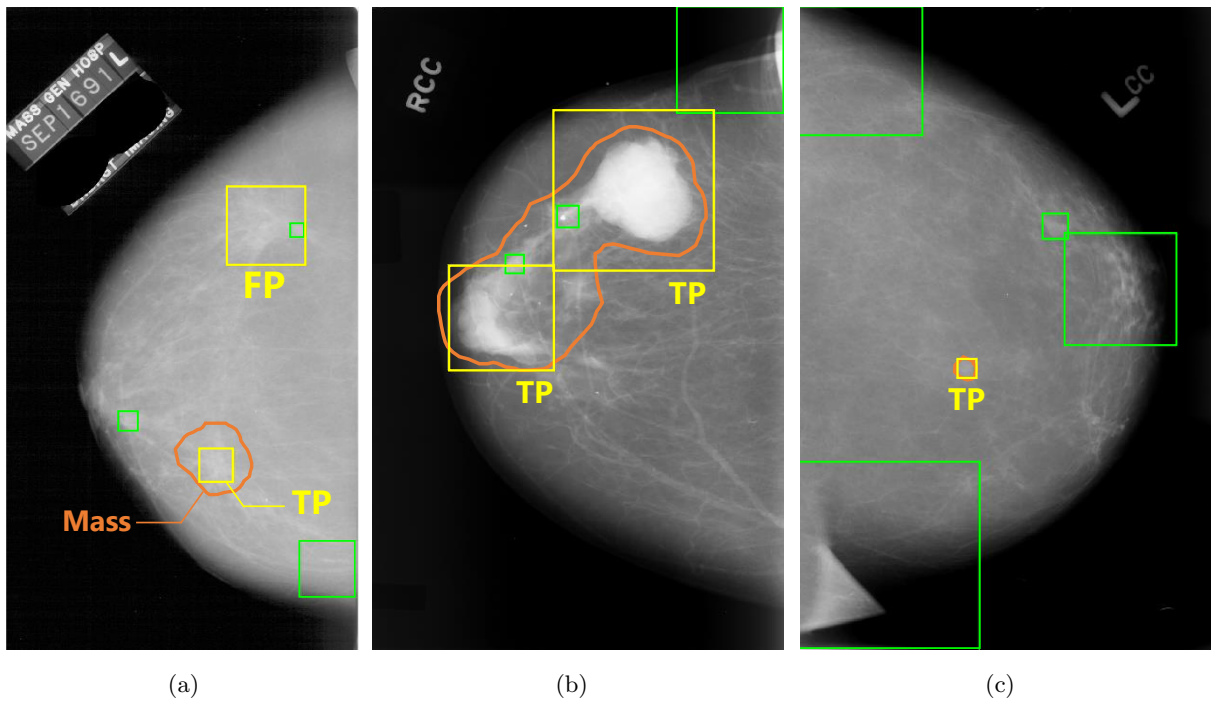


Fig. 10: Examples of detection results

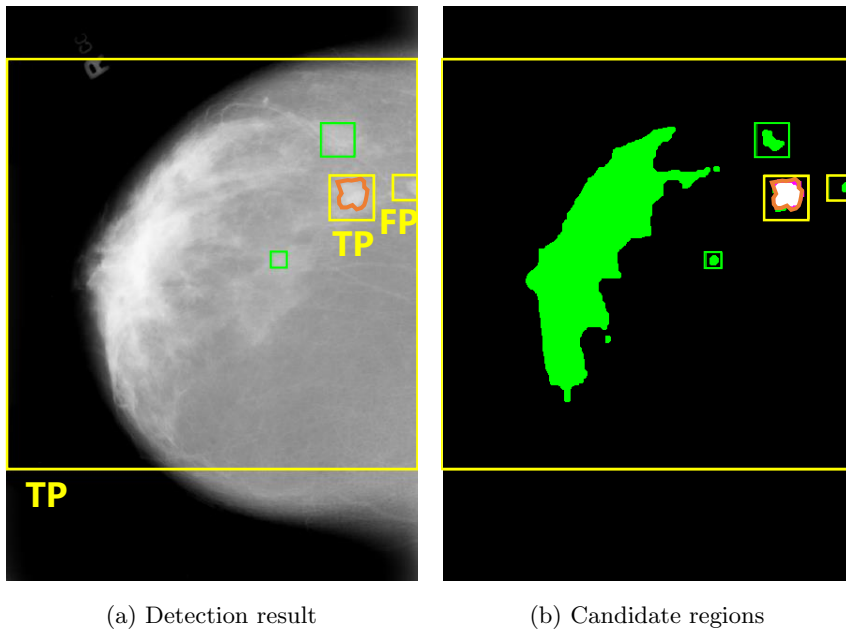


Fig. 11: Detection result with very large ROI

## References

- 1) A. Matsuda, T. Matsuda, A. Shibata, K. Katanoda, T. Sobue, H. Nishimoto, and The Japan Cancer Surveillance Research Group: Cancer Incidence and Incidence Rates in Japan in 2008: A Study of 25 Population-based Cancer Registries for the Monitoring of Cancer Incidence in Japan (MCIJ) Project, *Japanese Journal of Clinical Oncology*, **44-4**, 388/396, (2013).
- 2) J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, and Y. Yang: Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances, *IEEE Trans. Inf. Technol. Biomed*, **13-2**, 236/251, (2009).
- 3) H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick: Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space, *Phys. Med. Biol.*, **40-5**, 857/876, (1995).
- 4) B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt: Classification of mass and normal breast tissue: A convolutional neural network classifier with spatial domain and texture images, *IEEE Trans. Med. Imag.*, **15-5**, 598/610, (1996).
- 5) N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels: Gradient and texture analysis for the classification of mammographic masses, *IEEE Trans. Med. Imag.*, **19-10**, 1032/1043, (2000).
- 6) J. Wei, B. Sahiner, L. Hadjiiski, H. Chan, N. Petrick, M. Helvie, M. Roubidoux, J. Ge, and C. Zhou: Computer aided detection of breast masses on full field digital mammograms, *Med. Phys.*, **32-9**, 2827/2837, (2005).
- 7) A. Krizhevski, I. Sutskever, and G. E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, In *Proc. NIPS*, (2012).
- 8) S. J. Pan and Q. Yang: A Survey on Transfer Learning, *Knowledge and Data Engineering, IEEE Transactions on*, **22-10**, 1345/1359, (2010).
- 9) N. A. Otsu: A threshold Selection Method from Gray-Level Histograms, *IEEE Trans. Sys. Man. Cyber.*, **9**, 62/66, (1979).
- 10) 小形, 本間, 石橋, 張, 大橋, 長谷川, 川住, 阿部, 杉田, 吉澤: 構造情報に基づく乳房 X 線画像上の腫瘤陰影検出法, 計測自動制御学会東北支部 第 285 回研究集會予稿集, **285-2**, (2013).
- 11) 小畑秀文: モルフォロジー, コロナ社, (1996).
- 12) J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei: ImageNet: A Large-Scale Hierarchical Image Database, *CVPR09*, 248/255, (2009).
- 13) M. Heath, K. Bowyer, D. Kopans, R. Moore and W. P. Kegelmeyer: The Digital Database for Screening Mammography, *Proceedings of the Fifth International Workshop on Digital Mammography*, 212/218, (2001).
- 14) Bunch PC, Hamilton JF, Sanderson GK, and Simmons AH: A Free Response Approach to the Measurement and Characterization of Radiographic Observer Performance, *SPIE*, **127**, 124/135, (2000).
- 15) R. Campanini, D. Donquivanni, E. Iampieri, M. Masetti, G. Palermo, A. Riccardi, and M. Roffilli: A Novel Featureless Approach to Mass Detection in Digital Mammograms based on Support Vector Machines, *Phys. Med. Biol.*, **49-6**, 961/975, (2004).
- 16) N. H. Eltonsy, G. D. Tourassi, and A. S. Elmaghraby: A Concentric Morphology Model for the Detection of Masses in Mammography, *IEEE Trans. Med. Imag.*, **26-6**, 880/889, (2004).

# 深層学習による乳房 X 線画像上の腫瘍鑑別

○鈴木真太郎 (東北大学) 張曉勇 (仙台高等専門学校) 佐々木拓也 (京都大学)  
本間経康 市地慶 魚住洋佑 高根侑美 柳垣聡 川住祐介 石橋忠司 吉澤誠 (東北大学)

## Classification of Mammographic Masses by Deep Learning

\*S. Suzuki (Tohoku University), X. Zhang (Sendai National College of Technology),  
T. Sasaki (Kyoto University), N. Homma, K. Ichiji, Y. Uozumi, Y. Takane, S. Yanagaki,  
Y. Kawasumi, T. Ishibashi, and M. Yoshizawa (Tohoku University)

**Abstract**— Classification of benign and malignant masses in mammogram is one of the most difficult tasks in development of mammographic computer-aided diagnosis (CAD) system. This paper presents a deep learning-based method that utilizes a deep convolutional neural network (DCNN) to classify mammographic masses into two classes: benign and malignant masses. In order to train the DCNN to mass classification, a transfer learning strategy which pre-trains the DCNN on a large-scale natural image database and subsequently fine-tunes the DCNN on a relative small-scale mammogram database is used in this study. We test our method on the mammogram database and evaluate the classification performance using a receiver operating characteristic (ROC) curve. The experimental results demonstrate the area under the curve (AUC) of ROC is about 0.8 that is closed to the performance of radiologists.

**Key Words:** Deep convolutional neural network (DCNN), Computer-aided diagnosis (CAD), Mammogram

## 1 序論

乳がんは日本人女性における主な死亡原因のひとつであり、死亡率の低減と良好な予後のためには、その早期発見が極めて重要である<sup>1)</sup>。乳房 X 線撮影 (マンモグラフィ) は、乳がんの早期発見に有効な方法のひとつとして、乳がん検診に広く用いられている。乳房 X 線画像の診断には、高い専門性と長年の経験が必要とされる。しかし、乳がん検診の普及とともに、大量の乳房 X 線画像の読影業務が発生し、医師にとって大きな負担となっている。このような背景から、画像上より病変候補を検出・鑑別し、第二の意見として医師に提示することで、医師の負担軽減と読影精度の向上を目的とする、計算機支援画像診断システム (computer-aided diagnosis: CAD) システムの開発が進められている<sup>2, 3)</sup>。

乳房 X 線画像の読影において、事前に検出された腫瘍陰影の良悪性鑑別は、読影医にとって最も難しい作業のひとつである。良性および悪性の腫瘍の例を Fig. 1 に示す。良性の腫瘍は、明瞭な境界を持ち、腫瘍領域内で輝度はほぼ一定であることが多い。対して悪性の腫瘍は、不明瞭な境界を持つことが多く、その周辺にはしばしばスピキュラと呼ばれる放射線状の構造が観察される<sup>4)</sup>。

従来提案されてきた手法の多くは、これらの特徴を抽出、定量化することにより腫瘍の良悪性を鑑別するものである<sup>4, 5, 6, 7, 8)</sup>。しかしながら、腫瘍の特徴は多種多様であり、それらの良悪性を包括的に鑑別できる特徴の設計は非常に困難である。加えて、乳房 X 線画像上の腫瘍と周辺組織とのコントラストは非常に低いことも多く、形状等の特徴を正確に抽出することも極めて難しい。

近年、自然画像認識においては、深層学習 (deep learning) の一種である深層畳み込みニューラルネットワーク (deep convolutional neural network: DCNN) が非常に優れた性能をもつことが明らかになった<sup>9)</sup>。

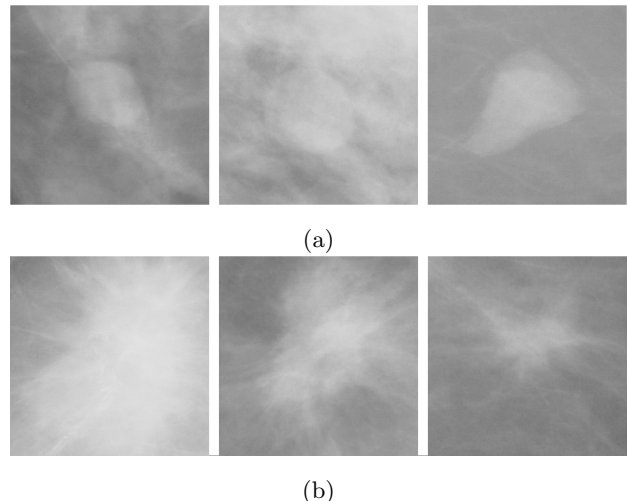


Fig. 1: Examples of benign and malignant masses in mammogram regions of interest (ROIs). (a) Benign masses with a well-defined and sharp boundary. (b) Malignant masses with blurred and spiculated boundary.

DCNN の特徴は、大量の画像データを用いて学習を行うことで、対象の識別に必要な特徴や識別規則を自動的に獲得する点である。

そこで本研究では、DCNN の優れた性能を乳房 X 線画像の CAD システムへ応用することを目指し、DCNN を用いて腫瘍の良悪性鑑別を試みた。DCNN の学習の際には、転移学習<sup>10)</sup>を用いることにより、自然画像の識別で獲得した知識 (特徴量や識別規則) を腫瘍の良悪性鑑別に転用することを試みた。臨床画像を用いた実験の結果、提案法の腫瘍鑑別性能は実際の医師による鑑別性能に匹敵することが示されたので、報告する。

## 2 提案法

本研究では、*Alexnet*<sup>9)</sup> と呼ばれる既存の DCNN を用いて腫瘍の鑑別を試みた。*AlexNet* は、Fig. 2 に示す

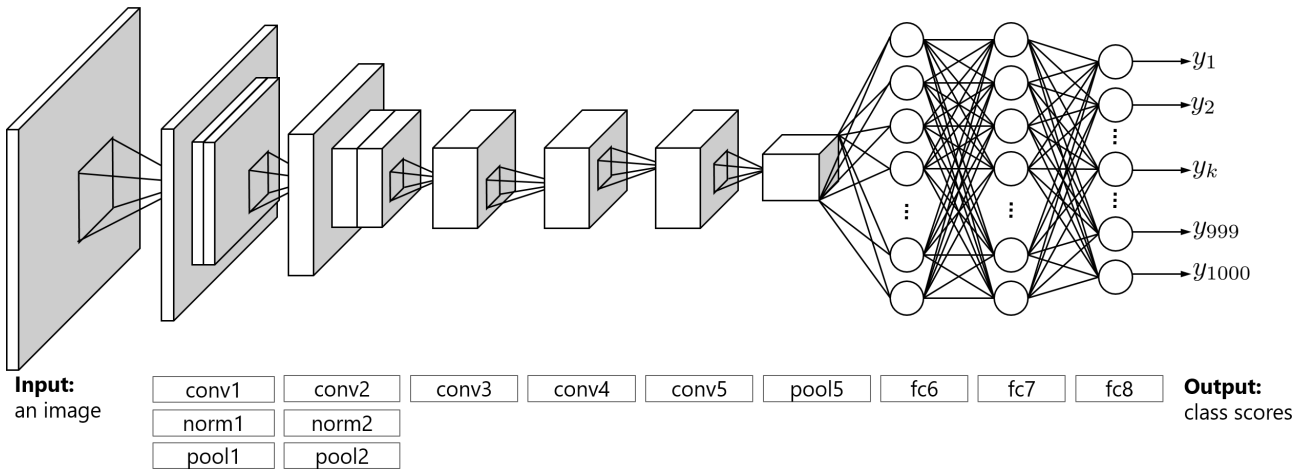


Fig. 2: Architecture of the *AlexNet*<sup>9)</sup> which consists of 5 convolutional layers (conv1~5), 2 normalization layers (norm1, norm2), 3 pooling layers (pool1, pool2, and pool5), and 3 fully-connected layers (fc6~8).

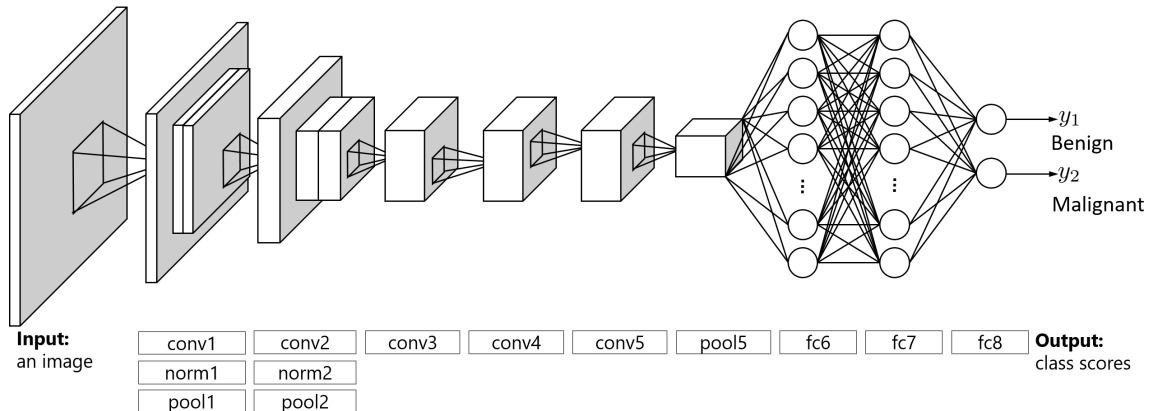


Fig. 3: The pre-trained *AlexNet* in which the last fully-connected layer (fc8) is modified by two neurons corresponding to the benign and malignant masses.

ように、5層の畳み込み層と3層の全結合層をもつDCNNである。入力としてR,G,Bの3チャンネルを持つ画像を与えると、入力が *ImageNet*<sup>11)</sup> の1,000種類の各クラス(カテゴリ)に属する確率を表すクラススコアを出力する。なお、*AlexNet*の詳細は参考文献<sup>9)</sup>を参照されたい。

病変鑑別の学習の前に、約120万枚の自然画像からなる *ImageNet*<sup>11)</sup> データセットを用いてDCNNの事前学習を行う。*AlexNet*のうち、浅い(入力に近い)層ではエッジや色といった単純で普遍的な特徴が抽出されることが知られており、これらの画像特徴は他の様々な対象の識別にも有効であることが明らかになっている<sup>13)</sup>。したがって本研究では、Fig.3に示すように、自然画像で事前学習済みの *AlexNet* の1,000ニューロンからなる出力層を、良性および悪性に対応する2つのニューロン(出力はそれぞれ  $y_1, y_2$ ) からなる新たな全結合層で置き換え、その他の層には手を加えず病変鑑別に転用する。その後、乳房X線画像の臨床データベースより作成した腫瘍画像を用いて、改変した *AlexNet* に対し腫瘍鑑別の学習(fine-tuning)を行った。なお、事前学習およびfine-tuningはいずれも誤差逆伝播法(back-propagation method)<sup>9)</sup>により行った。

Table 1: Experimental data.

	Training data	Test data	Total
Benign	1075	120	1195
Malignant	961	107	1068
Total	2036	227	2263

### 3 実験結果

学習(fine-tuning)および識別テストでは、乳房X線画像のデータベースである *Digital database for Screening Mammography (DDSM)*<sup>12)</sup>より作成した2,263枚のROI画像を用いた。良性、悪性腫瘍ともに医師の指定した領域を中心として  $454 \times 454$ 画素で切り出したのち、DCNNの入力サイズに合わせるため  $227 \times 227$ 画素に縮小した。Table 1に示すように、良性1,075枚、悪性961枚、計2,036枚のROI画像を用いてDCNNを学習(fine-tuning)し、良性120枚、悪性107枚、計227枚のROI画像で鑑別性能を評価した。Fig.4に示したのは、学習の進行に伴う学習データ・テストデータそれぞれに対する鑑別誤差の推移である。学習データに対する鑑別誤差は学習の進行とともに低下し、50

Table 2: Confusion matrix of mass classification.

		Prediction outcome		Total
		Positive	Negative	
True diagnosis	Malignant	True positive (TP) 77	False negative (FN) 30	107
	Benign	False positive (FP) 39	True negative (TN) 81	120
Total		116	111	227

Sensitivity =  $\frac{TP}{TP+FN} \approx 72.0\%$ ; Specificity =  $\frac{TN}{TN+FP} = 67.5\%$

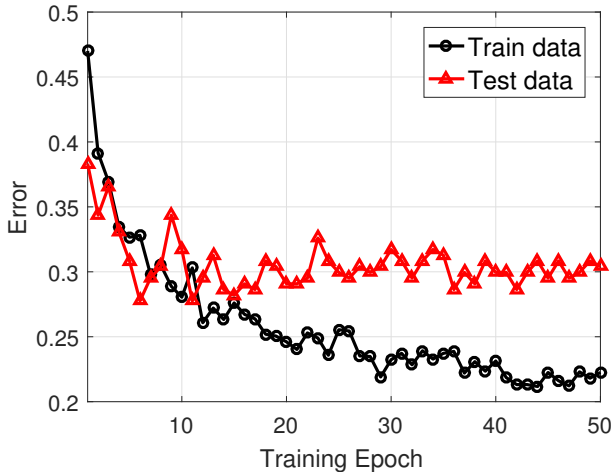


Fig. 4: Error rates of mass classification in the training phase and test phase.

回学習後には 20% 近くまで低下した。一方で、テストデータに対する鑑別誤差は 20 回学習後に 30% 程度まで低下した。

テストデータに対する鑑別結果の混合行列 (confusion matrix) を Table 2 に示す。なお, positive および negative はそれぞれ良性と悪性に相当する。この結果より, 腫瘍鑑別の感度 (sensitivity) は 72.0%, 特異度 (specificity) は 67.5% となった。

Fig. 5 は, 良悪性判定の閾値を変化させたときの真陽性率と偽陽性率の変化を表した受信者操作特性 (receiver operating characteristic: ROC) 曲線である。曲線下面積 (area under the curve: AUC) はおよそ 0.8 となり, これは医師による鑑別に匹敵する性能である<sup>6)</sup>。

Fig. 6 は, DCNN により高いクラススコア  $y_1$  で正しく鑑別された良性の ROI 画像, 対して Fig. 7 は同じく高いスコア  $y_2$  で正しく鑑別された悪性画像の例である。これらの画像より, 境界が明瞭で腫瘍領域内の輝度がほぼ一定の画像が良性と判定されていることが分かる。一方で, 悪性と判定された画像は不整形であるものや放射状のスピキュラを持つものなど, 悪性腫瘍に特有の特徴を持つことが読み取れる。これらの結果から, 提案法により学習された DCNN は腫瘍の良悪性鑑別に有効な特徴を獲得することができ, 腫瘍を適切に鑑別可能であると言える。

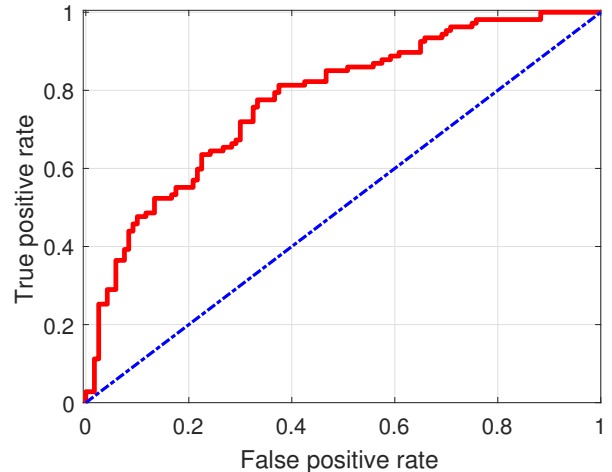


Fig. 5: The ROC curve of the classification of malignant mass.

#### 4 おわりに

本稿では, 深層学習を応用した乳房 X 線画像上の腫瘍鑑別法を提案した。医用画像データ数の不足を補うため, 自然画像認識を学習済みの DCNN に対し転移学習を行うことで, 腫瘍の良悪性鑑別の学習を試みた。臨床画像を用いた実験の結果, 提案法が医師に匹敵する腫瘍鑑別性能を達成可能であることが確認され, CAD システムとしての有効性が示唆された。

今後は, *GoogLeNet*<sup>14)</sup> のようなより大規模な DCNN モデルを用いることや, 学習データの拡張を行うことで腫瘍鑑別性能のさらなる向上を目指す。

#### 5 謝辞

本研究は JSPS 科研費 15K20852 ならびに 25293258 の助成を受けたものです。

#### References

- 1) R. E. Bird, T. W. Wallace, and B. C. Yankaskas: Analysis of cancers missed at screening mammography, *Radiology*, **184**-3, 613/617, (1992).
- 2) J. Tang, R. M. Rangayyan, J. Xu, I. E. El Naqa, and Y. Yang: Computer-aided detection and diagnosis of breast cancer with mammography: recent advances, *IEEE Transactions on Information Technology in Biomedicine*, **13**-2, 236/251, (2009).
- 3) X. Zhang, N. Homma, S. Goto, Y. Kawasumi, T. Ishibashi, M. Abe, N. Sugita, M. Yoshizawa: A hybrid image filtering method for computer-aided detection of microcalcification clusters in mammograms, *Journal of Medical Engineering*, **2013**, 1/8, (2013).

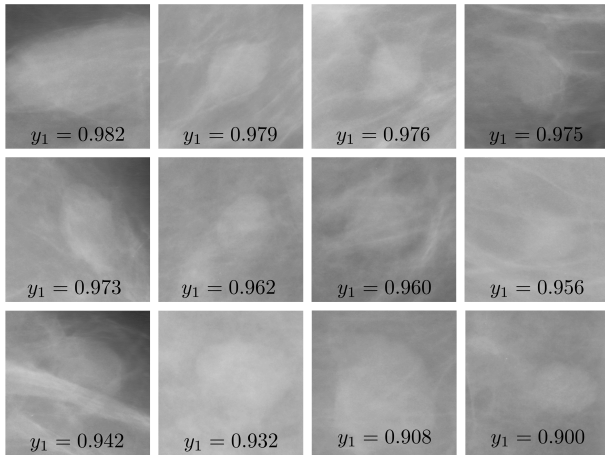


Fig. 6: True negative classification: the benign masses are correctly diagnosed as the negative.

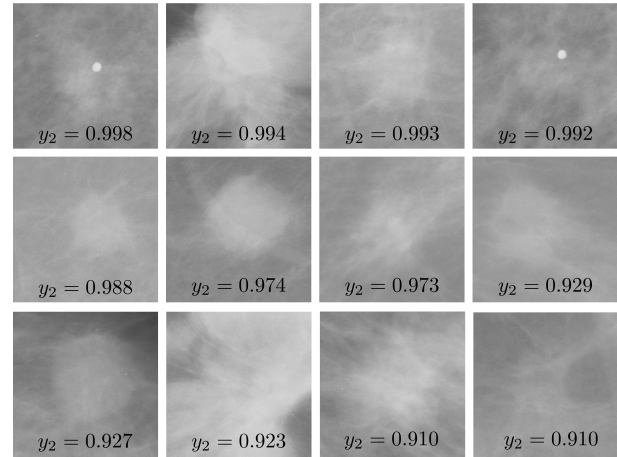


Fig. 7: True negative classification: the malignant masses are correctly diagnosed as the positive.

- 4) N. Mudigonda, R. Rangayyan, J. Desautels: Gradient and texture analysis for the classification of mammographic masses, *IEEE Trans. on Med. Img.*, **19**-10, 1032/1043, (2000).
- 5) X. Liu, J. Tang: Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method, *IEEE Systems Journal.*, **8**-3, 910/920, (2014).
- 6) Z. Huo, *et al.*: Analysis of spiculation in the computerized classification of mammographic masses, *Medical Physics*, **22**-10, 1569/1579, (1995).
- 7) H. Cheng, *et al.*: Approaches for automated detection and classification of masses in mammograms, *Pattern Recognition*, **39**, 646/668, (2006).
- 8) F. Moayedi, *et al.*: Contourlet-based mammography mass classification using the SVM family, *Computers in Biology and Medicine*, **40**, 373/383, (2010).
- 9) A. Krizhevsky, I. Sutskever, and G. Hinton: Imagenet classification with deep convolutional neural networks, *In Advances in neural information processing systems*, 1097/1105, (2012).
- 10) S. J. Pan and Q. Yang: A survey on transfer learning, *IEEE Transaction on Knowledge and Data Engineering*, **22**-10, 1345/1359, (2010).
- 11) J. Deng, *et al.*: ImageNet: a large-scale hierarchical image database, *In Proceedings of CVPR 2009*, 248/255, (2009).
- 12) M. Heath, *et al.*: The digital database for screening mammography, *In Proceedings of the Fifth International Workshop on Digital Mammography*, 212/218, (2001).
- 13) J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrel: DeCAF: A deep convolutional activation feature for generic visual recognition, *In Proceedings of JMLR 2014*, 647/655, (2014).
- 14) C. Szegedy, *et al.*: Going deeper with convolutions, *In Proceeding of CVPR 2015*, 1/9, (2015).

# Design of an EEG-based brain-computer interface using motor imagery for wheelchair control

\*P. Roussel and Y. Mitsukura (Keio University)

**Abstract**— Brain-computer interfaces (BCI) offer to some severely disabled individuals the possibility to interact with their environment by using their brain activity as a control signal for an external device. In particular, people unable to use conventional physical interfaces would benefit from brain-controlled assistive mobility devices such as electric-powered wheelchairs. Motor imagery can be performed by disabled individuals and produces distinct brainwave patterns that can be noninvasively measured using electroencephalography and used to produce a control signal for a mobile device. In this paper, a BCI specifically designed for the control of wheelchair is proposed. It aims to achieve reliability by restricting the command sequence and using an associated classification procedure that enables to differentiate two non-control states.

**Key Words:** Brain-computer interface (BCI), electroencephalography (EEG), motor imagery (MI), support vector machine (SVM), principal component analysis (PCA), wheelchair control

## 1 Introduction

Electroencephalography (EEG)-based brain-computer interfaces (BCI) for the control of mobile devices can particularly benefit to severely disabled individuals. While many projects could achieve very satisfying results by using external stimuli to facilitate the translation of brain signals into a control signal, the realization of an stimuli-independent, self-paced BCI achieving a good enough accuracy to be used in real environment remains challenging. Motor imagery (MI) can be performed by people with severe motor disabilities and seems a logical choice for a BCI related to mobility. MI triggers frequency changes in the EEG – namely a decrease in the spectral powers of the  $\mu$  (7.5-12.5 Hz) and  $\beta$  (12.5-30 Hz) bands. The reliability of MI-based BCI depends on its ability to distinguish the different commands but also to identify the non-control (NC) state during which the user should be able to focus on other activities without involuntarily triggering a command.

It is important to define a robust NC state in order to design a BCI suitable for the control of a physical device<sup>1)</sup>. Some previous studies aim to solve the problem by using a distinctive "switch" command for activating the BCI system<sup>2)</sup>. If the NC state can be distinguished accurately enough from the control states, it allows the user to use the BCI in a spontaneous self-paced way. In the case of a BCI operated by using MI, it seems that the NC state is distinguishable enough to allow the user to go forward and stop in a self-paced way in a virtual environment<sup>3)</sup>.

However, in order to facilitate the detection of the transitions from NC to MI and from MI to NC, it can be useful to take into account the multiple physiological phenomena that this NC state covers. Indeed, the two types of transitions provoke specific types of frequency changes in the EEG respectively known as event-related desynchronization (ERD) and synchronization (ERS)<sup>4)</sup>. Consequently, the NC states before and right after performing MI show different characteristics. The ERS shows a transient power increase and a sustained high average value in the beta and mu bands and the ERD consists in a transient decrease and a sustained low value in these band powers<sup>5)6)</sup>. These phenomena can be seen in the average spectrogram in Fig. 1.

The following study therefore aims to evaluate

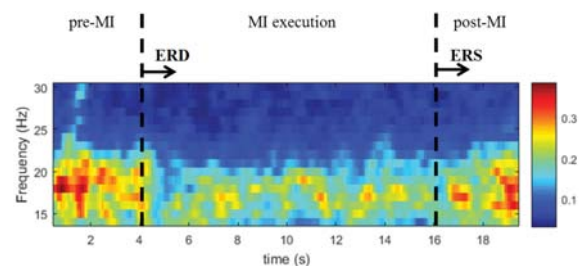


Fig. 1: Average spectrogram during a MI task on all the experimental data for the electrode C3. The first dotted line corresponds to the stimulus asking to the subject to practice one of three type of MI and the second one corresponds to the stimulus indicating to stop MI (see Fig. 3).

whether the consideration of the NC state as two distinct states and to can benefit to the realization of a brain-actuated wheelchair within a particular control strategy.

## 2 Methods

### 2.1 The control strategy

In accordance with the objective of controlling a wheelchair, the control signal produced by the BCI system is considered as a 4Hz sequence of commands corresponding to the wheelchair movements "turn left" (L), "turn right" (R) and "go forward" (F). These commands are respectively associated to left, right and both hands MI. When no MI is detected (NC state), either "stop" (S) or "stay still" (SS) should be issued depending on the situation. The NC states before and after MI are respectively called pre-MI and post-MI.

Instead of solely realizing a 4-class classification procedure which can lead to insufficient reliability, a different strategy has been considered. In constraining the user to use the "stop" command between two different types of commands, the system is allowed to take advantage of the particular nature of the post-MI NC state. Fig. 2 shows the two types of classifiers being used; the first type distinguishes the 4 classes whereas the second one only discriminates the post-MI state against MI. This new paradigm is believed to improve the detection of the "stop" command while the wheelchair is moving and to make it easier for the user to keep moving in a selected direction.



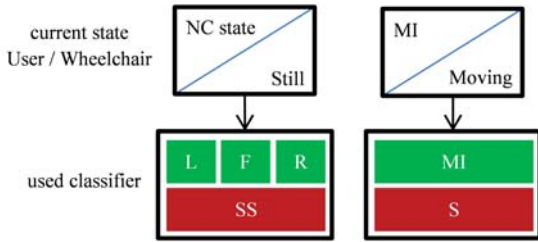


Fig. 2: States and associated classifiers.

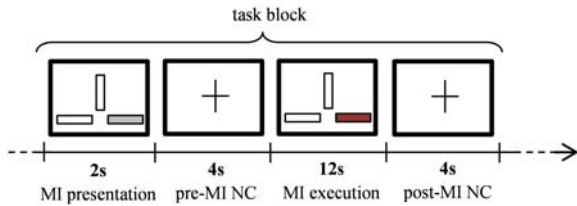


Fig. 3: Cues forming an experimental task block.

The present study therefore exposes a method to determine whether the distinction of the NC states can be beneficial to the classification accuracy of the MI and NC data. In order to focus on the NC state detection, the three types of MI are considered in this study as a unique class.

## 2.2 Data acquisition

The experimental data was acquired on a healthy subject using a g.USBamp recording device. The 9 channels of the 10-20 electrode placement system F3, F4, P3, P4, T3, T4, C3, C4, and Cz as well as electrooculogram (EOG) and both forearms electromyogram (EMG) were recorded at 512Hz and highpass-filtered at 0.1Hz. EMG signals were checked during the experiments in order to ensure that no real motor action was performed.

## 2.3 Experimental protocol

The subject had to perform specific motor imagery tasks based on cues displayed on a screen. As shown on Fig. 3, a basic task block consists in a presentation of a specific MI task, a pre-MI NC state, a MI execution period and finally a post-MI NC state. The presentation step is realized in order to obtain a reaction time as short as possible when the MI execution stimulus is shown. The subject was told to avoid moving during the whole experiment, to perform left, right or both hands kinesthetic MI during the execution time according to the direction of the pointer and to relax during the other steps. A session was composed of 9 task blocks, 3 for each type of MI in a random order. 2 experiments were carried out on 2 consecutive days, each consisting of 4 sessions.

## 2.4 Data processing

Large Laplacian filtering was applied on C3 and C4 signals using the data of the surrounding electrodes in order to partially remove the external noise. The power spectral densities of the two signals were computed on 1-second Hamming windows with 87.5% overlap in order to obtain an 8Hz sequence. Only the frequency bins comprised between 7 and 41Hz were used in the classification process. The data was normalized by computing the standard score for each frequency bin power and a logarithm was applied on the obtained values.

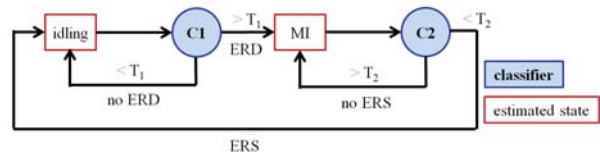


Fig. 4: Sequential classification process using classifiers  $C_{ERD}$  and  $C_{ERS}$  and thresholds  $T_1$  and  $T_2$ .

Based on the timestamp of its last sample, each window was labeled –pre-MI NC, post-MI NC, F, L, R– depending on the cue displayed during the experiment. However, in order to study the classification of NC states, the three types of MI were simply labeled MI. Three 2-class classifiers were considered:

- $C_{ERD}$ : pre-MI vs MI
- $C_{ERS}$ : post-MI vs MI
- $C_0$ : (pre-MI + post-MI) vs MI

$C_{ERD}$  represents a simplified version of the left classifier in Fig. 2 and  $C_{ERS}$  is the right one.  $C_0$  does not distinguish between the two types of NC states.

As initial features, the frequency powers of the associated 1-second time window (delay  $d = 0$ ) as well as those of the preceding 1-second window were considered ( $d = 1$ ). A process similar to the one described in a previous study<sup>7)</sup> was used to select features and perform dimensionality reduction. For each frequency bins  $f$  of the present and preceding time windows of the class  $i$ , the mean  $\mu_{i,f}$  and the standard deviation  $\sigma_{i,f}$  were computed. For each classifier the Fisher ratio was used to rank the frequency bins according to their class separability (CS):

$$F_{f,d} = \frac{(\mu_{1,f,d} - \mu_{2,f,d})^2}{\sigma_{1,f,d}^2 + \sigma_{2,f,d}^2} \quad (1)$$

The  $N_{CS}$  best ranked were kept and a principal component analysis (PCA) was performed on the corresponding features. After the PCA, the  $N_{PC}$  most important components were kept for the classification process. Finally, all the features were normalized by means of standard score (zero mean and unit standard deviation for each component).

## 2.5 Machine learning

The classification was implemented by using the Support Vector Machine (SVM) with a radial basis function (RBF) kernel provided by the LIBSVM<sup>8)</sup> algorithm. The two parameters of this model are the penalty parameter  $C$  and the RBF parameter  $\gamma$ . Only the data of the first experiment was used in this step. For various parameters  $N_{CS}$  and  $N_{PC}$  the following classification 3-step process was applied on the data of the first experiment:

1. Computing a feature set by using parameters  $N_{CS}$  and  $N_{PC}$
2. Finding the parameters  $C$  and  $\gamma$  that maximize accuracy
3. Training on 75% of the features and testing on the remaining part

The classifiers having the best performance on the second experiment (see Tab. 1) were then selected in order to ensure their generalization capabilities. The final classification scheme uses the two classifiers  $C_{ERD}$  and  $C_{ERS}$  in a sequential process showed in Fig. 4. The values of the thresholds  $T_1$  and  $T_2$  were

Table 1: Accuracy and parameters of the best classification achieved on the second experiment for each type of classifier. All classifiers were trained on the first experiment.

Classifier	Accuracy	$N_{CS}$	$N_{PC}$
$C_{ERD}$	69.3%	10	3
$C_{ERS}$	75.1%	8	1
$C_0$	71.9%	8	6

Table 2: Accuracy and error rate during NC state for both the sequential classification process and  $C_0$  with optimized thresholds.

Classifier	Accuracy	$NC_{error}$
$C_{seq}$	66.1%	14%
$C_0$	66.2%	18%

optimized in order to minimize the ratio of misclassified samples during the NC state  $NC_{error}$  while maintaining the average accuracy of the sequential process above 66%. In order to compare this process with a single classifier scheme, the same process was applied to  $C_0$  with a simple threshold.

### 3 Results

The accuracy of the selected classifiers (see Tab. 1) is satisfying given the nature of the classes and the length of the time windows use for feature computation. It is also interesting to notice that the classifiers which perform the best in generalization use a small number of frequency bins.

The results of the final evaluation can be seen in Tab. 2. Even though the gap does not allow for categorical conclusion, the fact that  $C_{seq}$  obtains a lower error during NC state than  $C_0$  tends to confirm the possible benefit of the proposed scheme.

### 4 Discussion

The present study gives a positive insight on the proposed method. In order to draw a more rigorous conclusion, the number of subject should be increased and a third experiment should be added in the protocol in order to test the generalization capabilities of the sequential classification process.

The recorded EOG may also be used to confirm that ocular artifacts do not influence the classification. If they do, it could be used for further denoising by using adaptive filtering.

### 5 Conclusion

This work showed that the assumption of a dual nature of the NC state in a BCI operated by MI might improve the reliability of the control of a mobile device. It also shows that having a sequential classification process as proposed in this paper allows for a tuning of the trade-off between the maximization of the global classification accuracy and the minimization of the misclassification during NC state.

### 6 Acknowledgment

This work was supported by JSPS KAKENHI (Scientific Research A) Grant Number YYK5A05.

### References

- 1) B. He, B. Baxter, B. J. Edelman, C. C. Cline and W. W. Ye, "Noninvasive brain-computer interfaces based on sensorimotor rhythms," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 907–925, 2015.
- 2) S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 10, pp. 1297–1307, Oct. 2000.
- 3) R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-Paced (Asynchronous) BCI Control of a Wheelchair in Virtual Environments: A Case Study with a Tetraplegic," *Computational Intelligence and Neuroscience*, pp. 1–8, 2007.
- 4) G. Pfurtscheller, C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neuroscience Letters*, vol. 239, no. 2–3, pp. 65–68, Dec. 1997.
- 5) G. Pfurtscheller, A. Jr Stancák, C. Neuper, "Post-movement beta synchronization. A correlate of an idling motor area?," *Electroencephalography and Clinical Neurophysiology*, vol. 98, no. 4, pp. 281–93, Apr. 1996.
- 6) G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregegnier, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 6, pp. 642–651, Dec. 1997.
- 7) N.F. Ince, A.H. Tewfik, S. Arica, "Extraction subject-specific motor imagery time-frequency patterns for single trial EEG classification," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 499–508, Apr. 2007.
- 8) C.C. Chang and C.J. Lin, "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

# 前頭前野における脳波を用いた耳鳴の検出および苦痛度の推定

○松岡将司 満倉靖恵 (慶應義塾大学)

## Tinnitus Detection and Distress Estimation Using Prefrontal Cortex Electroencephalogram

\* M. Matsuoka and Y. Mitsukura (Keio University)

**Abstract**— In this paper, we aim to classify tinnitus patients and healthy controls, and estimate tinnitus distress using prefrontal cortex electroencephalogram (EEG). Conventional methods of investigating tinnitus are subjective and need long time. In this research, we used prefrontal cortex EEG because the relationship between tinnitus and prefrontal cortex has been revealed. EEG data and questionnaire about distress were obtained from 40 tinnitus and 40 controls. We classified tinnitus and controls with amplitude spectra calculated EEG data and achieved 70% classification accuracy. In addition, we found 9 Hz spectra was related to distress. As a result, there is some possibility of investigating tinnitus using EEG.

**Key Words:** Tinnitus, Distress, EEG, Prefrontal cortex

### 1 Introduction

The purpose of this paper is to classify tinnitus patients and healthy controls and estimate tinnitus distress level using prefrontal cortex electroencephalogram (EEG). Tinnitus is a phenomenon that a person hears sounds even though there is no sound source. These sounds can only be heard by the patients. The prevalence of tinnitus in the population is estimated at 10–20% [1]. Tinnitus distress is related to emotional health as depression and anxiety. Depression and anxiety were enhanced at tinnitus onset in patients with high distress [1]. Therefore, tinnitus should be treated because of the negative effects.

There are two kind of tinnitus that are objective tinnitus and subjective tinnitus. Other people can hear objective tinnitus and they can't hear subjective tinnitus. Objective tinnitus is caused by ears and mouth muscles being cramped or hearing blood vessel noise. This is thought to be caused by a part not related to the brain and hearing sense. On the other hand, it is said that subjective tinnitus is caused by an abnormality of the ear and the brain, and mainly by damage of the inner ear due to aging and noise. A person with this tinnitus hears the continuous sounds which are like pure tones. It is difficult for the doctor to grasp tinnitus at the clinical site because other people can't hear subjective tinnitus. In this research, we focused on subjective tinnitus.

At the clinical site, Questionnaires and examination equipment are used to evaluate the tinnitus symptoms (distress, loudness and tone) at the clinical site. Latter method is used to recognize pitch of tinnitus. Patients compare their tinnitus tones with pure tones coming from an audiometric equipment. Then, they choose one pure tone which is the closest to their tinnitus tones. In clinical practice, doctors usually diagnose the state of tinnitus. However, these conventional methods are subjective and it takes about 2 hours to investigate tinnitus symptoms. So these inspection is a burden to patients with tinnitus.

Therefore, our final goal is constructing a simple and objective investigation method for tinnitus using prefrontal cortex EEG. Tinnitus may be caused by brain abnormalities and related to prefrontal cortex [3, 4]. Furthermore, the pre-

Table 1: THI score and tinnitus distress level

THI score	0-16	18-36	38-56	58-100
Distress level	non-distress	mild	moderate	severe
Patients number	7	9	11	13

frontal cortex is also associated with tinnitus distress [4,5]. Hence, prefrontal cortex EEG may be an effective indicator to evaluate tinnitus. We use a simple electroencephalograph (Figure 1) to measure the EEG signal. Since we can measure EEG data with this device easily, we thought that it is suitable for practice use. As the first step, we aim to classify tinnitus patients and healthy controls and estimate tinnitus distress level.

### 2 Experiments

40 patients with tinnitus patients (22 males, 18 females; age:  $60.5 \pm 10.3$  years) and 40 healthy controls (22 males, 18 females; age:  $60.4 \pm 9.62$  years) participated in this study. They sat on a chair and wore the EEG measurement device (Figure 1). Then, we recorded the resting state EEG signals for 30 seconds with their eyes closed. EEG measurement was conducted twice including 30 seconds break. The measurement point was Fp1 based on the international 10-20 system (Figure 2). The sampling frequency was 1024 Hz. After measuring EEG, we made questionnaire survey to patients with tinnitus. We used tinnitus handicap inventory (THI). The questionnaire can evaluate the degree of tinnitus distress experienced in everyday life with tinnitus. THI score is given on a scale of 0 to 100 because there are 25 items in this questionnaire and people answer each question in three stages 0, 2, 4. Distress are divided to 4 level by THI score. Table 1 shows the relationship between THI score and the distress level of tinnitus. From the experiments, we obtained EEG data and questionnaire about the degree of tinnitus distress.

**Table 2: The classification accuracy between tinnitus and controls**

Comparison	Significant difference	Classification accuracy [%]	Precision [%]	
			Tinnitus	Controls
Tinnitus vs Controls	13,14 Hz	71.3	85.0	57.5
Mild or more symptoms tinnitus vs Controls	13 Hz	67.1	84.8	57.5
Moderate or more symptoms tinnitus vs Controls	9,13 Hz	70.3	79.2	65.0
Severe symptoms tinnitus vs Controls	9,13,21 Hz	86.8	69.2	92.5

### 3 Analysis methods

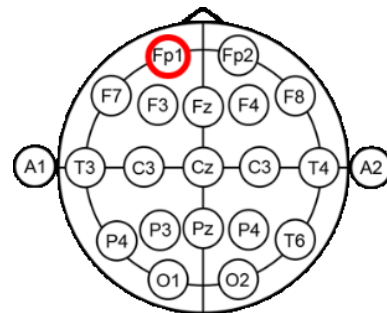
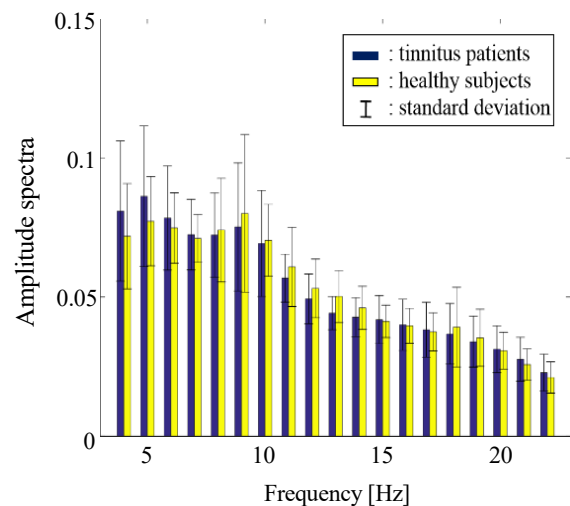
In the analysis, we used 60 seconds EEG data and THI score. EEG data was low-pass filtered below 30 Hz and fast Fourier transform with hamming window was applied per second. Then, the amplitude spectra were obtained every 1 Hz and we got 60 data from each subject. In this paper, we focus on the 19 frequency bands (4-22 Hz per 1 Hz). After frequency transformation, the frequency bands were normalized for each subject so that the sum of spectra is 1. Finally, we classified tinnitus patients and healthy controls using Support Vector Machine (SVM). To investigate the difference of EEG data due to tinnitus distress, we divided tinnitus patients to 2 groups by distress and classified them using SVM. We used regression analysis to investigate the correlation between tinnitus distress and EEG data. In this paper, we use the Leave-One-Subject-Out Cross-Validation (LOSOVCV) method to evaluate classifiers. One subject data was taken evaluation data and the other data was taken learning data. This process was repeated until each subject data has been taken for evaluation. After this method, the classification accuracy and precision were obtained.

### 4 Results

In this section, we show the results of significance test and classification.

#### 4.1 Classifying tinnitus and healthy controls

Figure 3 shows mean amplitude spectra for tinnitus patients and healthy controls. We found significant differences at 13 and 14 Hz amplitude spectra (significance level: 10%) and the classification accuracy was 71.3%. We divided tinnitus patients at the stage of severity and examined significant differences between mild or more symptoms tinnitus patients and healthy controls, moderate or more symptoms tinnitus patients and healthy controls, and severe symptoms tinnitus patients and healthy controls. We found significant differences at 13 Hz, 9 and 13 Hz, and 9, 13, and 21 Hz (significance level: 10%). The classification accuracies were 67.1%, 70.3%, 86.8% respectively. Table 2 shows these results and each precision. The higher distress patients feel, the higher classification accuracy was. As a result, we thought the higher the distress was, the larger the difference in EEG between tinnitus patients and healthy controls.

**Fig. 1: EEG measurement device****Fig. 2: International 10-20 system****Fig. 3. Mean amplitude spectra of tinnitus and controls**

**Table 1: The classification accuracy between 2 groups of tinnitus**

Comparison (Tinnitus)	Selected feature	Classification accuracy [%]	Precision [%]	
			Distress: low	Distress: high
non-distress vs The other symptoms	9 Hz	92.1	80.0	93.9
Mild or less symptoms vs Moderate or more symptoms	9 Hz	76.3	57.1	87.5
Severe vs The other symptoms	9 Hz	65.8	85.2	15.4

#### 4.2 Estimating tinnitus distress level

We divided the tinnitus patients into 2 groups according to the tinnitus distress level in order to compare 2 groups (Table 3). We found a significant difference in 9 Hz amplitude spectrum between B: mild or less symptoms and moderate or more symptoms, and C: moderate or less symptoms and severe symptoms tinnitus patients. Figure 4 shows mean amplitude spectra of mild or less symptoms and moderate or more symptoms tinnitus patients. We applied simple regression analysis with 9 Hz amplitude spectra as the objective variable and THI score as the explanatory variables. Then, we found negative correlation between them (Figure 5). We also classified tinnitus patients between A: non-distress and the other symptoms, B: mild or less symptoms and moderate or more symptoms, and C: severe and the other symptoms using SVM. The classification accuracies were A: 92.1%, B: 76.3% and C: 65.8% respectively. Table 4 shows these results and each precision.

### 5 Discussion

In this section, we discuss the results.

#### 5.1 Classifying tinnitus and healthy controls

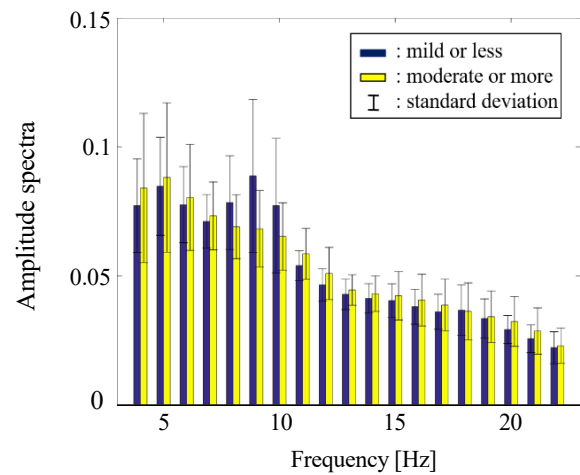
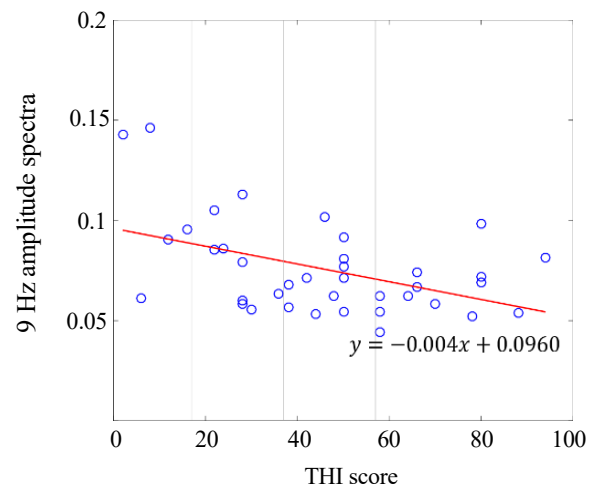
There were differences in EEG data between tinnitus and controls. The prefrontal cortex is related to cognition and emotion of sound [4]. We considered that tinnitus changed brain activity and it appeared in EEG. The classification accuracy between tinnitus patients and healthy controls was 71.3%. This is not enough considering actual use. So it is necessary to improve the classification accuracy. It is extremely dangerous to misidentify tinnitus patients as healthy controls in the clinical site as it leads to misdiagnosis. Therefore, even if a high classification accuracy is obtained, it is hard to say that it is a good classification when the precision of tinnitus patients is low. In this paper, the precision of tinnitus patients and healthy controls were 85.0% and 57.5%. Therefore, although the classification accuracy was not sufficient, we thought that it was a good because the precision of tinnitus patients was high.

#### 5.2 Estimating tinnitus distress level

There were no significant differences in comparison between A: tinnitus patients with non-distress and the other

**Table 2: Comparison**

A	Non-distress vs The other symptoms
B	Mild or less symptoms vs Moderate or more symptoms
C	Severe vs The other symptoms

**Fig. 4. Mean amplitude spectra of tinnitus****Fig. 5. Relation between 9 Hz spectra and THI**

symptoms, but we obtained high classification accuracy. Although there were significant differences in comparison between C: tinnitus patients with severe and the other symptoms, we obtained low classification accuracy. We consider that one of the causes of this difference is the dis-

persion of data. As can be seen from figure 5, the dispersion is large in tinnitus patients with low distress (THI: 0 - 40), but the distribution of the amplitude spectra is uniform in tinnitus patients with a high distress (THI: 40-100) and the dispersion is small. We considered that it made the classification accuracy different that different dispersion for each distress level. In this verification, we divided tinnitus patients to two groups and classified them. There was a difference in the number of people in the compared group. Therefore, it is necessary to increase the number of tinnitus patients and verify the results again.

## 6 Conclusions

In this paper, we aimed to classify tinnitus patients and healthy controls and estimate tinnitus distress level using prefrontal cortex EEG. As a result, we were able to classify tinnitus patients and healthy controls about 70% accuracy. Furthermore, we found a negative correlation between 9 Hz amplitude spectrum and tinnitus distress. From these results, there is some possibility of investigating tinnitus using EEG.

## Acknowledge

This study was supported by JSPS KAKENHI (Scientific Research (A)) Grant Number YYK5A05.

## References

- [1] E. W - Franke, et al. : "Tinnitus-Related Distress and Personality Characteristic Resilience", *Neural Plasticity*, Vol. 2014 (2014)
- [2] J. A. Kaltenbach, et al. : "Neurophysiological Mechanisms of Tinnitus", *JOURNAL-AMERICAN ACADEMY OF AUDIOLOGY*, pp. 125-137 (2000)
- [3] W Schlee, et al. : "Mapping Cortical Hubs in Tinnitus", *BMC Biology*, Vol.7, No.1 (2009)
- [4] N. Weisz, et al. : "Tinnitus Perception and Distress is Related to Abnormal Spontaneous Brain Activity as Measured by Magnetoencephalography", *PLoS medicine*, Vol. 2, No. 6 (2005)
- [5] S. Vanneste, et al. : "The Neural Correlates of Tinnitus-Related Distress", *NeuroImage*, Vol. 52, No. 2, pp. 470-480 (2010).
- [6] T. Ueyama, et al. : "Brain Region Responsible for Tinnitus Distress and Loudness: a Resting-State fMRI Study", *PLoS one*, Vol. 8, No. 6, e67778 (2013)
- [7] K. Takayama and Y. Mitsukura : "Relationship between Discomfort caused by Pure Tones and EEG", *THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS HCG SYMPOSIUM 2012*, pp. 298-301 (2012)

# Improvement of Consumer-Grade Brain-Machine Interface Accuracy through Preliminary Signal Analysis

\*G. A. Goussarov and Y. Mitsukura (Keio University)

**Abstract**– The aim of the present research is to create a brainwave controlled wheelchair based on Electroencephalographic data, with the express intention of using foot motor motion or foot motion imagination as a control scheme, and a consumer-grade electroencephalogram as an input device. For this method to be effective, it is necessary to separate the signals corresponding to the foot motion from other kinds of signals. In this paper, we show that this task is feasible and it is possible to separate signals obtained during foot motion, hand motion, teeth clenching and inaction without the need for an extensive training period.

**Key Words:** Brain-machine interface (BMI), electroencephalography (EEG), Wheelchair Control System (WCS)

## 1 Introduction

In this paper, we evaluate the feasibility of using Event-Related Desynchronisation (ERD) and Event-Related Synchronisation (ERS) related to foot Motor Execution (ME) without subject training as an input method for a Wheelchair Control System (WCS) for patients suffering from spinal cord injury who often lose control over their legs, which results in an inability to move without a wheelchair. While a lot of research has been performed on the subject of WCS's, there are still many issues, including long training time of both the patient and the machine (currently around one to two weeks), and a long period of use is necessary before the WCS becomes easy to use. A reliable and easy-to-use WCS which leaves the user's hands free and does not require extensive training of either the user or machine has yet to be proposed.

We have conducted a number of experiments to collect data as the basis for a future classification algorithm for smooth control of the wheelchair.

## 2 State of the Art

Hands-free wheelchair control can be achieved in a number of ways. Possible options include eye tracking techniques, voice commands<sup>1, 2)</sup>, and Electromyograms which detects the activity of muscles<sup>2)</sup>. However, using these techniques might get in the way of a normal daily life, as they require the user to activate parts of their body which would be used for something else. An interesting alternative method is electroencephalography which measures signals directly from the scalp and does not require equipment which would strongly inconvenience the user. Therefore, an Electroencephalogram (EEG)-based WCS might not interfere with normal daily life. Other schemes for detecting brain activity such magnetoencephalography exist, but they are less portable than EEG's.

The concept of direct Brain-Computer Interfaces (BCI), also known as a Brain-Machine Interfaces (BMIs) has been proposed as early as 1973<sup>3)</sup>. Since then, a large number of publications has been made describing various mechanisms which could be used to control prostheses based on brain-wave analysis. A number of approaches has been proposed to control a wheelchair using EEG-based BMI's. Common techniques<sup>4, 5)</sup> involve the use of Steady State Visually Evoked Potentials (SSVEPs), the P300 Event-Related Potential (ERP), Motor Imagery (MI) or Mental tasks. In many cases, one of these approaches is supplemented by another system, such as a collision

avoidance system. Such systems are referred to as hybrid WCS's. Systems based on SSVEP's or the P300 wave are called synchronous since measured signal occur at a known time after and external stimulus whose onset is known. These have a high accuracy but also a slow response time. On the other hand, MI-based systems are asynchronous as they do not require an external signal but have relatively low accuracy and may require training the user, where the duration has been reported to vary greatly, ranging from minutes to months<sup>5)</sup>. Furthermore, these systems rely primarily on hand MI, which may reduce the usability of such systems since MI and ME activate the same regions of the brain<sup>7)</sup>, which means that using their hands may produce undesired commands.

Another problem of the above systems is that they rely on research-grade EEG systems, which require another person to help the user install the EEG. The alternative is consumer EEG, but these suffer from a reduced number of electrodes, usually located on the forehead.

Taking all this into account, the present research attempts to extract foot ME signals using a small number of electrodes. We show that it is possible to separate signals obtained during foot motion, hand motion, teeth clenching and inaction without the need for an extensive training period.

## 3 Materials and Methods

A total of 13 subjects (8 male, 5 female) participated in the study. Subjects were asked to perform various motions through prompts appearing on a computer screen placed in front of them. A short description of the tasks was given to each participant before the experiment. No further training of any kind was performed. Brainwave data were collected for 45 events of foot motion, 15 events of hand motion and 10 events of jaw clenching for each session. Proper execution of each motion was verified visually. Arm, neck and eye motion signals were also considered, but were not measured after feedback from the first participants, indicating that the sessions were too long, which could have an impact of the quality of the recorded signals, since fatigue affects brainwaves. The number of sample epochs for each kind of signal can be found in table 1.

Data recordings were made using the Guger Technologies (g.tec, <http://www.gtec.at/>) g.USBamp signal amplifier, with 5 sintered Ag/AgCl active electrodes. These were placed at locations Fp1, Fp2, C1, Cz and C2 of the international Ten Percent Electrode

Sample Type	Num. per Session
Foot motion	45
Left	15
Right	15
Both	15
Hand clenching	15
Left	5
Right	5
Both	5
Jaw clenching	10

Table 1: Number of recorded signal epochs by type

System (10-10 system)<sup>8)</sup> (see fig.1) and complemented by 1 reference electrode located at AFz and a ground electrode located on the right earlobe.

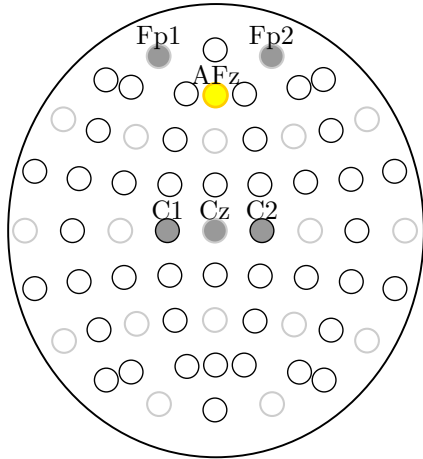


Fig. 1: Available electrode positions on the g.tec cap. Positions used for acquisitions are highlighted and their names are indicated.

The typical session proceeded as follows:

Prior to any experiment, the subject was asked to complete a consent form, informing them to the way data which they provided would be handled. They were then seated in front of the computer screen, and the EEG cap was put on without electrodes. The experimenter then moved the hair below the relevant electrode positions out of the way before inserting the electrodes in the appropriate slots. The electrodes had been pre-emptively connected to the amplifier, which was itself connected to a computer whose screen could not be seen by the subject. This was done to reduce the time subjects spent doing nothing. The experimenter then briefed the subject about what they would do during the experiment if necessary, completing the setup phase.

Each session contained 5 independent trials (see table 2), each corresponding to a specific type of motion. Trials 1, 3 and 4 contained written instructions before the trial itself, while trials 2 and 5 only informed the subject that instructions were identical to trial 1. After this, the subject had to look at a cross for 30 seconds to extract baseline information. Then, a symbol appeared which would prompt the subject to perform one kind of motion, followed by a cross indicating they should stop moving, which is referred to as the 'rest' state. This alternation of motion and rest was then repeated multiple times, until the next trial or the end of the session. A short 1 minute break

where subjects could move however they wanted was present between successive trials.

Trial	Total Duration	Content
Foot motion	~7 min	15 Foot ME events (5 Left, 5 Right, 5 Both)
Foot motion	~6 min	15 Foot ME events (5 Left, 5 Right, 5 Both)
Hand clench	~5 min	15 Hand ME events (5 Left, 5 Right, 5 Both)
Jaw clench	~4 min	10 Jaw Clench Events
Foot motion	~6 min	15 Foot ME events (5 Left, 5 Right, 5 Both)

Table 2: Number of recorded signal epochs by type

It should be noted that participants had to perform the motion as soon as the prompt appeared instead of having some time to prepare. While this causes the actions to begin and end slightly after the recorded time-stamp, it provides a clear separation between epochs of rest and actions. On the other hand, informing the subjects of coming events results in difficulties when classifying measurements.

## 4 Results

The analysis of signals can be separated into two successive analysis steps. The first step was the separation of foot motion signals from artefactual signals, illustrated here by hand motion and jaw clenching, while the second will consist of separating the three kinds of foot motion (left foot, right foot, and both) from each other.

Previous research suggests that detection of ERD's and ERSs should be easily observable by analysing the changes in band power for the alpha (7 to 13Hz) and beta (14 to 28 Hz) brainwaves<sup>9)</sup>. However, our preliminary experiments have shown that while these observations are verified on average, they cannot be used as unequivocal indicators of actual events. Furthermore, the actual frequency ranges affected during motion vary on an individual basis. Therefore, in this research, the logarithm of the power spectral density was used to generate feature vectors for each sample. These feature vectors were then labelled according to the supposed nature of the event they were taken from, and passed through a Linear Discriminant Analysis algorithm to create a classifier for separating the rest state from the motion state.

In order to properly evaluate differences between subjects, the first imperative was to properly annotate the data by detecting the exact moment when the transition between brain states occurs, as the reaction times of subjects, which introduces a delay in relation to the prompts which appear on screen, is not known a priori. These reaction times were estimated using three separate approaches. The first was through observation of the average spectrogram (see fig.2), the second was through an estimate based on the reaction time for jaw clenching which can easily be detected by a simple threshold (results in fig.3), and the third was through the use of a pattern search algorithm which attempted to minimize the number of samples when rest was misclassified as activity or which were ignored due to having an uncertain nature (results in fig.4).



All three methods produced an estimated time delay varying between 0.5 seconds and 1.5 seconds, with an uncertainty period of varying between 0.3 and 1.2 seconds, depending on subject and the method employed, with the pattern search method usually producing lower estimates in terms of delay but high variability in terms of uncertainty period (some subjects had uncertainty periods as short as 0.3 seconds while for others, this value reached 1.2 seconds) compared to the other results. Full results will be presented at the conference.

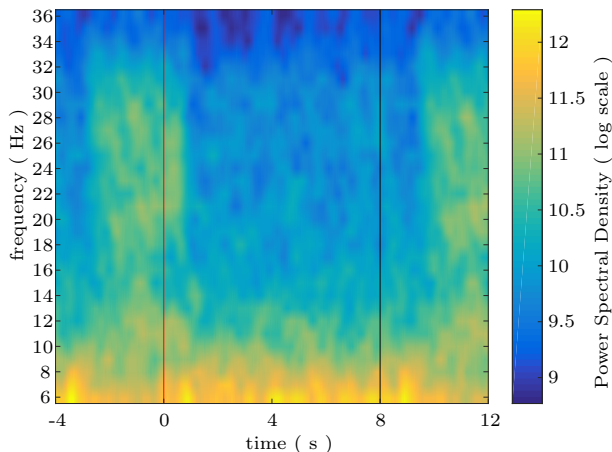


Fig. 2: Average Event spectrogram. The red line indicates when the subject was instructed to begin moving their foot, and the black line indicates when the subject was instructed to stop moving their foot, subject 2

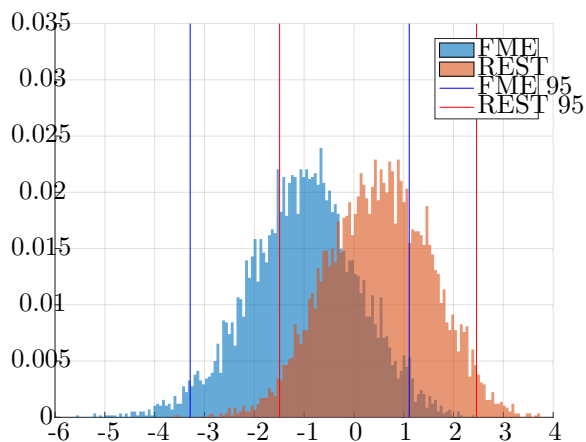


Fig. 3: Distinction between rest and foot ME using the jaw-clench-based time estimation, vertical lines indicate the range containing 95% of samples, subject 2

## 5 Discussion

While the first method, observation of the average spectrogram produced good results with one subject for whom more samples were collected, in general this approach proved to be ineffective. The second method, use of jaw clenching data, was shown to provide decent results. However, the use of the pattern search algorithm illustrated that choice of the lag time and period of uncertainty was not unequivocal (i.e. it is possible to chose an incorrect delay). The number of subjects was also insufficient to properly describe the time delay, and its variance on a session to session

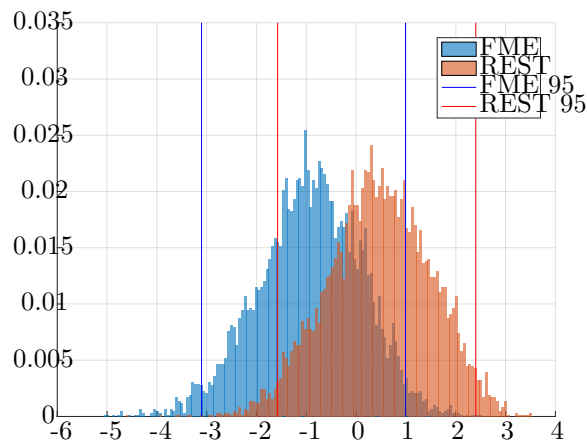


Fig. 4: Distinction between rest and foot ME pattern search, vertical lines indicate the range containing 95% of samples, subject 2

and subject to subject basis. Creating a linear classifier on the whole population sample using fixed delay and uncertainty period has provided results which will not be usable to annotate transitions in individual subjects.

Furthermore, the frequencies selected using the linear discriminant analysis were erratic, making it impossible to understand effectively which frequencies could be used to produce accurate recognition in most subjects. Verification of the classifiers on full EEG data also showed that the output was extremely noisy and would require additional signal processing to be usable. It should however be noted that classification of hand ME ERD's was much more clear, reaching near complete separability from foot ME ERD's. This explains why current systems focus on hand ME or MI, as these signals are much more easily separable from the rest.

Finally, our preliminary testing supported the notion that brain-waves require a long time to stabilize after ME, as suggested by Pfurtscheller et al.<sup>9)</sup>.

## 6 Conclusion

Despite the drawbacks mentioned above, the results show sufficient promise for separability of foot ME ERD and ERS from other states using only few electrodes and a linear classifier to warrant further research into the subject. The resulting classifier does not require training the user.

Further work will focus on expanding the number of samples per subject, as well as adding new ones, as the current data size seems to be insufficient, which results in theoretical results which do not correspond to reality. Attempts will be made to further improve the annotation of the ERD and ERD timestamps by introducing data from previous samples into the sample vector reducing the reliance on absolute frequency values. Once this is complete, ERD and ERS transitions will be annotated with more accuracy and a classifier will be constructed which can detect these transitions in any subject will be constructed. After that, another classifier for distinguishing between the different subclasses of foot ME will be created.

## References

- 1) R. C. Simpson, "Smart wheelchairs: A literature review," *Journal of Rehabilitation Research and Development*, vol. 42, no. 4, pp. 423–36, Jul. 2005.
- 2) M. Hashimoto, K. Takahashi, and M. Shimada, "Wheelchair control using an EOG-and EMG-based

- gesture interface,” in *Advanced Intelligent Mechatronics*, 2009. AIM 2009. IEEE/ASME International Conference on, 2009, pp. 1212–1217.
- 3) J. J. Vidal, “Toward direct brain-computer communication,” *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.
  - 4) G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.
  - 5) L. Bi, X.-A. Fan, and Y. Liu, “EEG-Based Brain-Controlled Mobile Robots: A Survey,” *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 2, pp. 161–176, Mar. 2013.
  - 6) B. He, B. Baxter, B. J. Edelman, C. C. Cline and W. W. Ye, “Noninvasive brain-computer interfaces based on sensorimotor rhythms,” *Proceedings of the IEEE*, vol. 103, no. 6, pp. 907–925, 2015.
  - 7) M. Lotze et al., “Activation of Cortical and Cerebellar Motor Areas during Executed and Imagined Hand Movements: An fMRI Study,” *Journal of Cognitive Neuroscience*, vol. 11, no. 5, pp. 491–501, Sep. 1999.
  - 8) G. E. Chatrian, E. Lettich, and P. L. Nelson, “Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities,” *American Journal of EEG Technology*, vol. 25, no. 2, pp. 83–92, 1985.
  - 9) G. Pfurtscheller and F. L. Da Silva, “Event-related EEG/MEG synchronization and desynchronization: basic principles,” *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.