

BRainAnalyzer 解説書

内容

1. WAVELET	3
1.1 Wavelet Denoising (Matlab)の流れ	7
1.2 分解	7
1.3 閾値処理	8
1.4 再構築	10
2. 仮説検定	12
多重比較検定の補正	16
p 値について	17
仮説検定における 2 種類の過誤	18
3. PCA 主成分分析	20
4. ICA (Independent Component Analysis, 独立成分分析)	27
5. Filter	31
7. 積分・重心計算	38

GLM については BRANL(GLM)解説書を参照願います。

version 1.3

ビー・アール・システムズ株式会社

2022 年 6 月 13 日



更新履歴

更新日付	版数	内容
2013/11/14	1.00	初版作成
2013/12/2	1.02	WAVELET 内容の追加
2014/2/19	1.03	WAVELET 内容(疑似周波数)の追加
2014/9/16	1.04	GLM 機能の追加
2014/10/16	1.06	t 検定「多重比較検定の補正」の項目追加
2015/11/13	1.07	「仮説検定における 2 種類の過誤」、「Filter」を追加
2016/3/17	1.08	「Filter 設定値に関する情報」を追加
2016/4/2	1.09	「積分・重心」を追加
2017/8/14	1.10	「t 検定」を「仮説検定」に変更し、「p 値について」の追加、 2 種の過誤の表変更
2017/11/13	1.11	「Filter」に関し、パラメータの説明を追加
2019/5/16	1.12	ローパスフィルタ機能と見た移動平均のカットオフ周波数の近似式
2019/7/12	1.2	・WAVELET の全面改訂 ・PCA 3.4 項の追記
2022/6/13	1.3	・2 種の過誤の表追加(p18) ・「Filter 設定値に関する情報」情報の追加(p36)

1. WAVELET

(1) 時間と周波数領域のデータ解析手法

時間と周波数領域のデータ解析手法として、次の 3 手法があります。

① フーリエ変換

フーリエ変換は正弦波による周期信号の展開をベースとしています。時間領域で定常と仮定しています。すべての波は正弦波という前提です。

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt$$

② 短時間フーリエ変換

実際の信号は時間的に変動します。この影響を考慮するため、時刻 $t=b$ まわりで局在する窓関数 $w(t-b)e^{-j\omega t}$ を用いて修正したフーリエ変換を短時間フーリエ変換と言います。窓幅は一定。

$$X(\omega, b) = \int_{-\infty}^{+\infty} x(t) w(t-b) e^{-j\omega t} dt$$

③ ウェーブレット変換

さらに、時刻ごとに周波数特性を抽出する手法をウェーブレット変換と言います。

ウェーブレット変換は、窓関数と信号との相互相関関数として考えられます。

周波数に応じて窓幅を変化させるウェーブレット変換には、連続ウェーブレット変換と離散ウェーブレット変換があります。

窓（母）関数は、連続ウェーブレット変換： $\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right)$

離散ウェーブレット変換： $\varphi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} \varphi\left(\frac{t-nb_0 a_0^m}{a_0^m}\right)$

$\tilde{x}(a, b) = \int_{-\infty}^{+\infty} x(t) \varphi_{a,b}^*(t) dt$ (continuous wavelet transform)

a : 伸長パラメータ、 b : 位置パラメータ

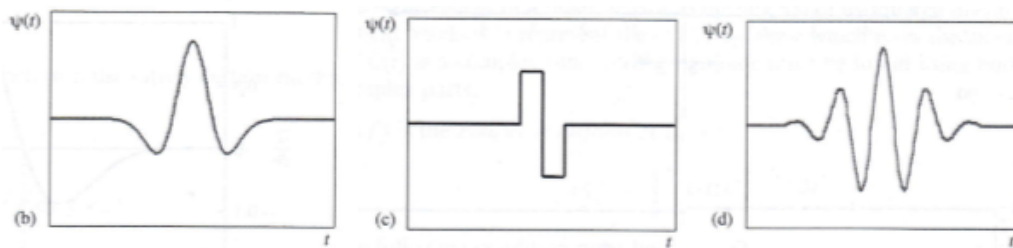
時間軸上の b を順次増加していき、位置 b を中心とした a のすべてのスケールの範囲に対しコヒーレント構造を同定します。

窓関数 φ (Mexican hat, Harr, Morlet, - -) は、信号の性質と窓関数の特性の両方に依存して選択されます。 $\varphi(t)$ の例、

$$\text{Mexican hat: } \varphi(t) = (1 - t^2)e^{-\frac{t^2}{2}}$$

$$\text{Harr: } \varphi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{elsewhere} \end{cases}$$

$$\text{Morlet: } \varphi(t) = \frac{1}{\pi^{1/4}} \exp(i2\pi f_0 t) \exp(-t^2/2)$$



Mexican hat

Harr

Morlet

(2) 離散ウェーブレット変換 信号の分解 (近似と詳細)

離散ウェーブレットでは、原信号を近似と詳細に分割していきます。

近似 $x_m(t)$ と詳細 $d_m(t)$ の関係は、

$$x_m(t) = x_{m-1}(t) - d_m(t)$$

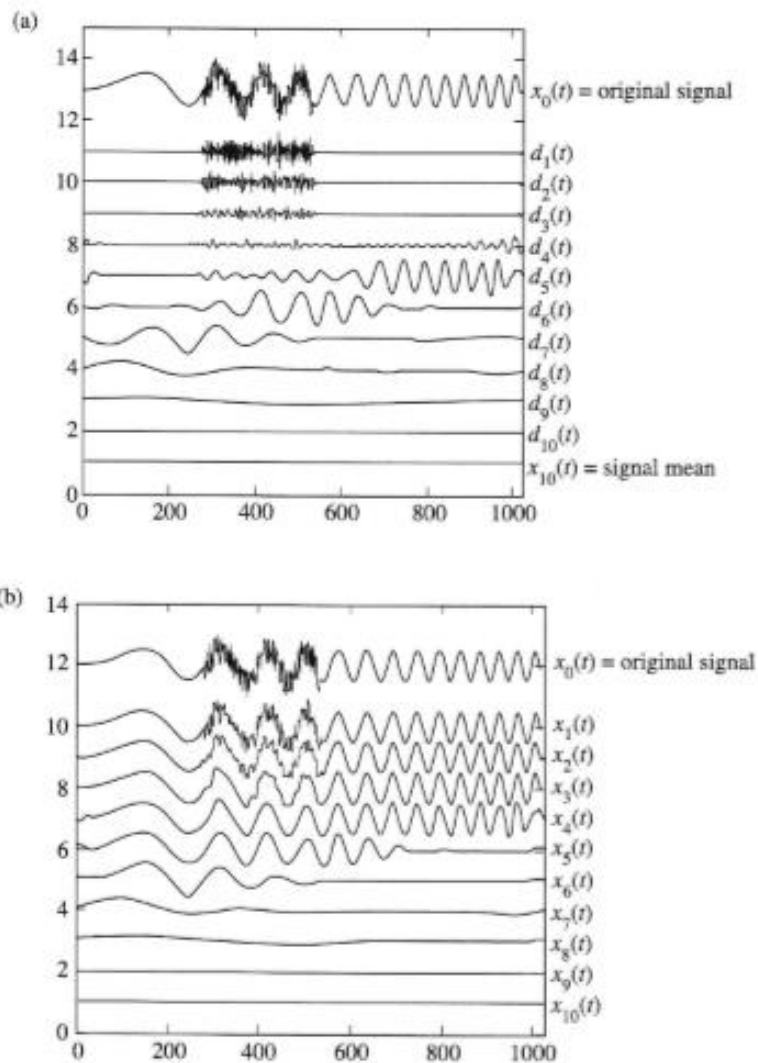
と表されます。下図の(a)は詳細、(b)は近似です。 t は時間、 m はレベル。

例えば、

$m = 1$ の場合、

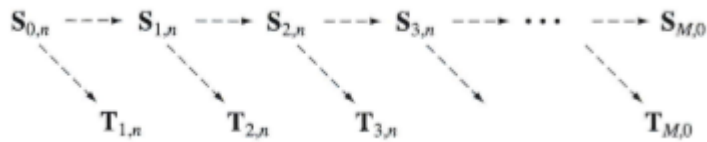
$$x_1(t) = x_0(t) - d_1(t) \quad \rightarrow \quad x_0(t) = x_1(t) + d_1(t)$$

オリジナル信号は近似 $x_1(t)$ に詳細の $d_1(t)$ を加算した信号となります。



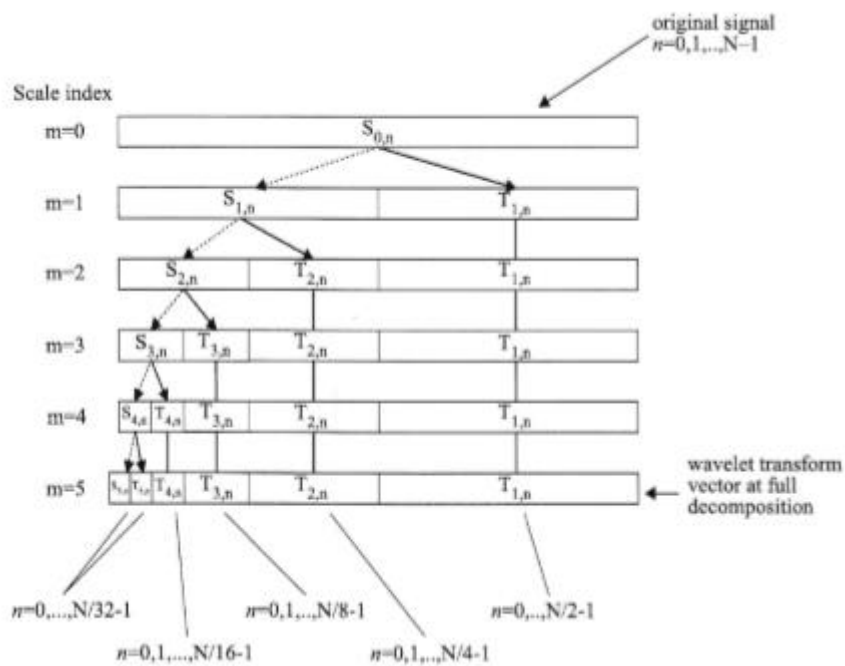
出典 : The Illustrated Wavelet Transform Handbook, Figure 3.5, page 80

近似係数を S 、詳細係数を T としますと、下図の様に分解されて行きます。



ここに、 M は分解レベル、 n は信号のサンプル数

各レベルの分解図を下図に示します。

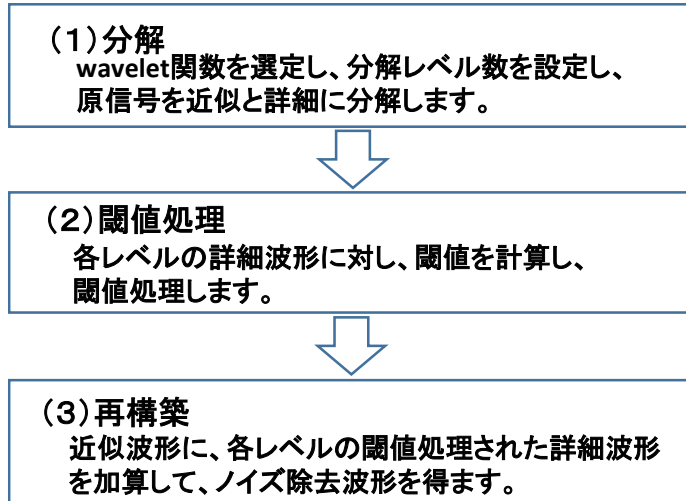


出典：The Illustrated Wavelet Transform Handbook, Figure 3.6, page 82

各レベルで変換係数ベクトル W_i^{scale} (wavelet係数) が得られます。

1.1 Wavelet Denoising (Matlab)の流れ

BRainAnalyzer の Wavelet ノイズ除去機能は、表示形式は異なりますが、Matlab と同様な処理をしています。



1.2 分解

1.1 で信号を分解しますと、下図の様に近似波形とレベル数に応じた詳細波形が得られます。

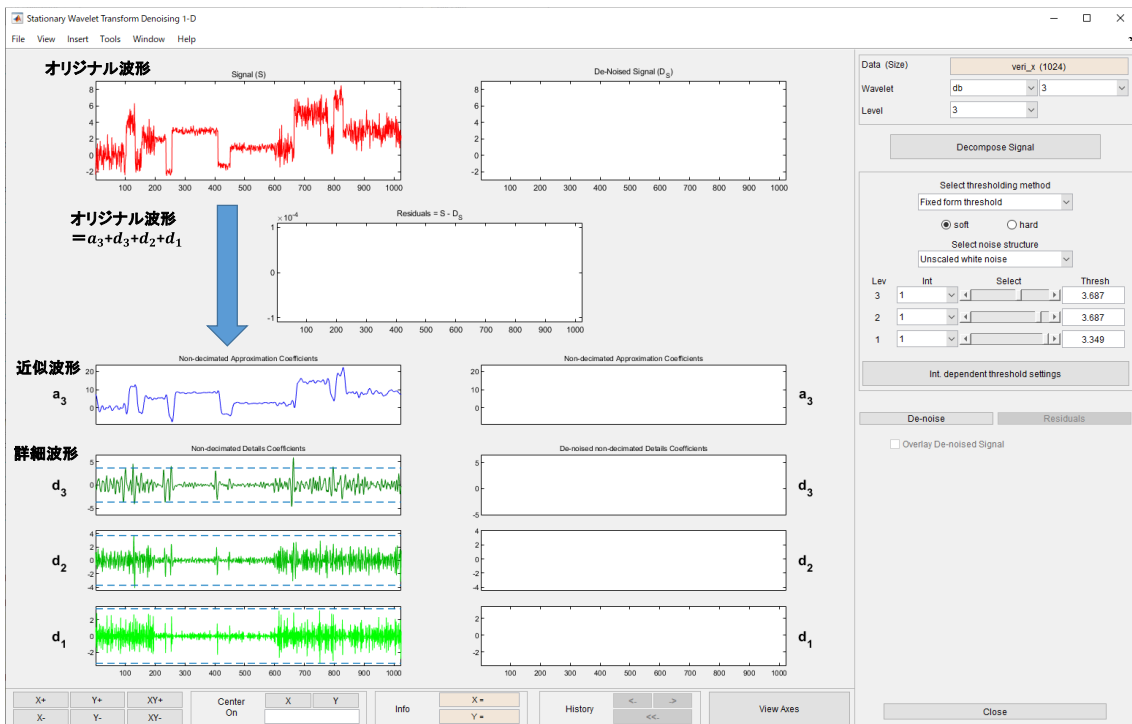


図 1.1

1.3 閾値処理

続いて、信号からノイズを除去します。重要度の低い成分を除去していきます。そのため、wavelet 係数を閾値処理します。ある値より小さい部分はゼロにする操作です。

閾値の値決めは重要となります。

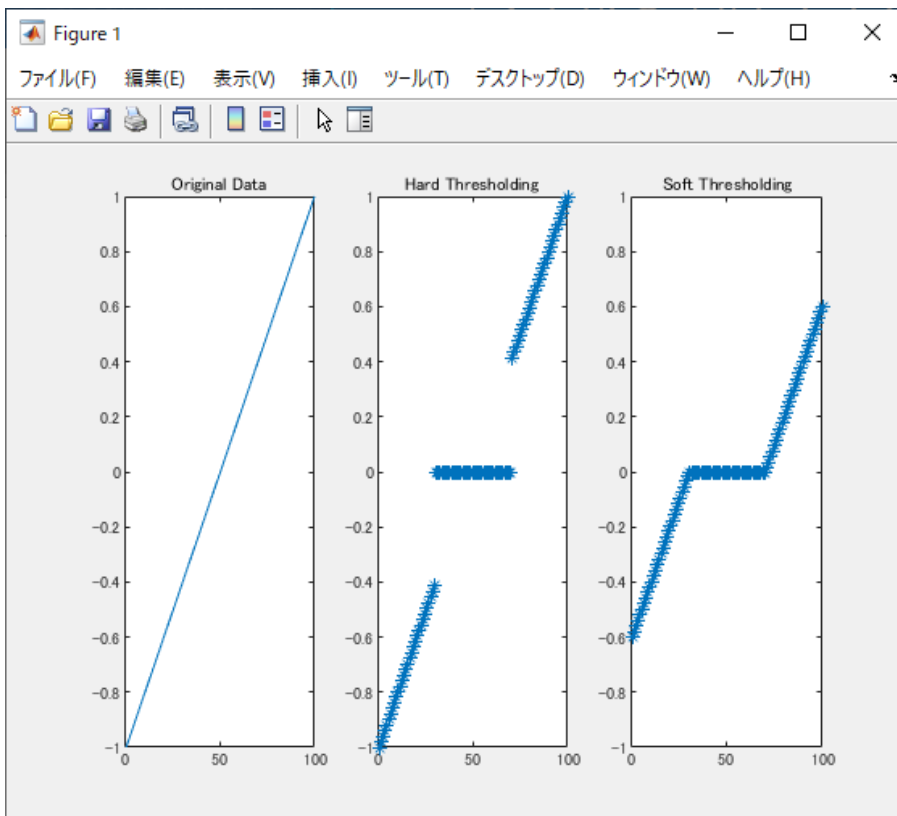
閾値処理方法として、ハード閾値処理とソフト閾値処理の 2 つの方法があります。

$$W_i^{hard} = \begin{cases} 0 & |W_i| < \lambda \\ W_i & |W_i| \geq \lambda \end{cases}$$

$$W_i^{hard} = \begin{cases} W_i - \lambda & W_i > \lambda \\ 0 & |W_i| \leq \lambda \\ W_i + \lambda & W_i < -\lambda \end{cases}$$

処理結果の例を下図に示します。

縦軸は閾値、横軸はデータ。



この閾値 λ を計算するオプションとして、

sqtwolog, rigrsure, minimaxi を採用しています。

sqtwolog: uses a fixed form threshold yielding minimax performance multiplied by a small factor proportional $\log(\text{length}(n))$

rigrsure: uses for the soft threshold estimator a threshold selection rule based on Stein's Unbiased Estimate of Risk(quadratic loss function).

minimaxi: uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics to design estimators. Since the denoised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the option that realizes the minimum, over a given set of functions, of the maximum mean square error.

また、ここでノイズ構造の選択が有ります。

one: unscaled white noise

sln: scaled white noise

mln: non-white noise

•**'one'** corresponds to the basic model.

In general, you can ignore the noise level and it must be estimated. The detail coefficients $cD1$ (the finest scale) are essentially noise coefficients with standard deviation equal to σ . The median absolute deviation of the coefficients is a robust estimate of σ . The use of a robust estimate is crucial for two reasons. The first one is that if level 1 coefficients contain f details, then these details are concentrated in a few coefficients if the function f is sufficiently regular. The second reason is to avoid signal end effects, which are pure artifacts due to computations on the edges.

•**'sln'** handles threshold rescaling using a single estimation of level noise based on the first-level coefficients.

When you suspect a nonwhite noise e , thresholds must be rescaled by a level-dependent estimation of the level noise. The same kind of strategy as in the previous option is used by estimating lev level by level.

This estimation is implemented in M-file `wnoisest`, directly handling the wavelet decomposition structure of the original signal s .

・'**mln**' handles threshold rescaling using a level-dependent estimation of the level noise.

生体信号の特性に拠り、選択する母関数、レベル数、閾値処理は変わりますが、

1.4 再構築

近似波形に、各レベルの閾値処理された詳細波形を加算して、ノイズ除去波形を得ます。

・閾値処理及び再構築を図 1.2（次頁）で解説します。

以上

[参考文献]

1. 「Wavelet Toolbox 概説書」、The MathWorks, Inc. 2009
2. 「図説 ウェーブレット変換ハンドブック」、新 誠一、中野和司 監訳、朝倉書店、2005
3. "Wavelet Denoising and Nonparametric Function Estimation", Mathworks

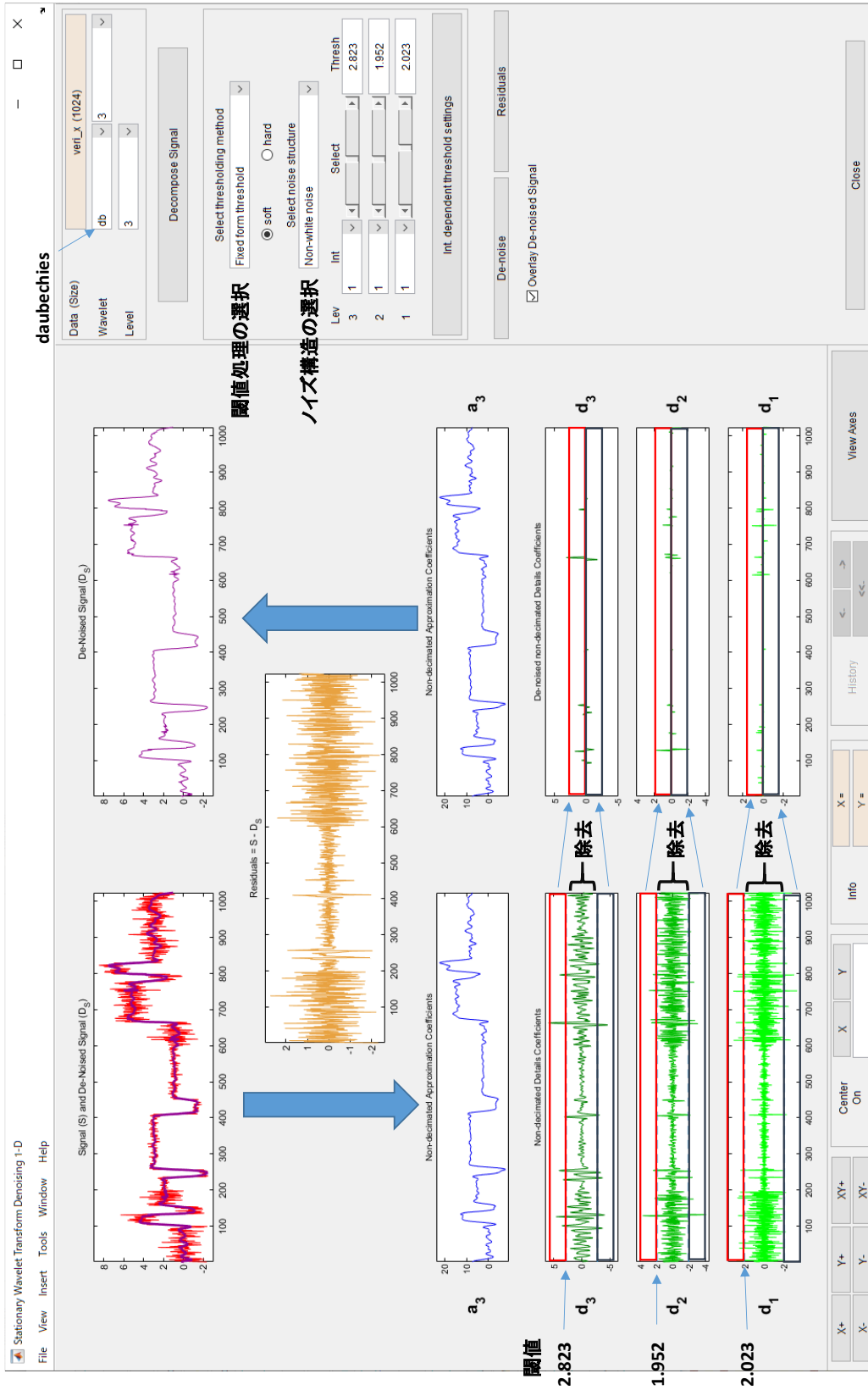


図 1.2

2. 仮説検定

2つの差の検定の分類

	標本標準偏差	サンプル数	標本分散	適用分布	注記
対応の有る データ	—	$n \geq 100$	—	z 分布	
		$n < 100$	—	t 分布	
対応の無い データ	既知	$n_1 + n_2 \geq 100$	—	z 分布	
	不明	$n_1 + n_2 < 100$	等分散	t 分布	
			非等分散	ウェルチ t 分布	
		$n < 100$	—	t 分布	

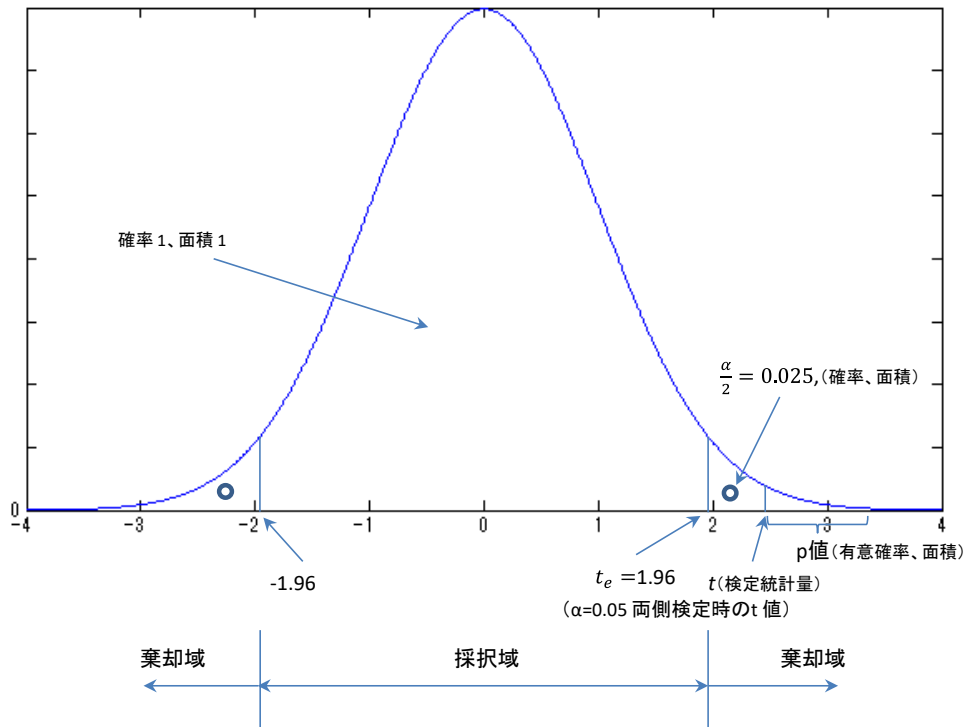
(t 検定は、外れ値が有る場合には適さないと言われています。)

標本：母集団から部分的に抽出されたデータの集まりです。

α ：有意水準 p ：有意確率

t 分布 (自由度が100以上になるとz分布に近づく。)

$\alpha=0.05$ 両側検定



グループ統計量

- ① 標本数： データ数 n
- ② 標本の平均値：
データの合計をデータ数 n で割った値
- ③ 標本の標準偏差 s
分散 s^2 の平方根が s

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

自由に選べるデータの個数、自由度は $(n-1)$

(We don't actually have n independent pieces of information about the variance.)

- ④ 標本平均の標準誤差

$$\frac{s}{\sqrt{n}}$$

等分散性の検定

- ① F 値 (検定統計量)

2つの標本分散 s_A^2, s_B^2 を基に、母分散 σ_A^2 と σ_B^2 の等分散性を判定します。

$$F = \frac{s_A^2}{s_B^2}$$

F 分布表から α 、自由度に対応する F_e を算出し、 $F_e < F$ の場合は不等分散と判定します。

- ② 有意確率、p 値 (後述の「p 値について」を参照願う。)

検定統計量の外側の確率を意味します。

2つの母平均の差の検定

- ① H

帰無仮説(2つの平均に差は無いという仮説)ベースで、 $H=1$ では棄却される (有意差が有る)。 $H=0$ では棄却されない (有意差が有るとは言えない)。

- ② 平均値の差

グループ1の標本数 n_1 とグループ2の標本数 n_2 とすると、

グループ1の平均値 \bar{x}_1 とグループ2の平均値 \bar{x}_2 の差

- ③ 標準偏差

$$\text{偏差平方和} : S = \sum(x_i - \bar{x})^2$$

$$\text{標本分散} : s^2 = \frac{\text{偏差平方和}}{(n-1)}$$

グループ1の標本標準偏差 s_1

グループ2の標本標準偏差 s_2

④ 標本平均の標準誤差

グループ 1 の標本平均の標準誤差 $\frac{s_1}{\sqrt{n_1}}$

グループ 2 の標本平均の標準誤差 $\frac{s_2}{\sqrt{n_2}}$

⑤ 信頼区間 (区間推定、two-sample)

例えば、母集団の真の母数が含まれる確率が、一定の確率以上と保証される区間を信頼区間といいます。

対応有り	対応無し	
	等分散 ($\sigma_1^2 = \sigma_2^2$)	非等分散 (Welch の検定)
$\mu_d \sim \bar{d} \pm t\left(f, \frac{\alpha}{2}\right) \frac{s_d}{\sqrt{n}}$ $\bar{d} = \frac{\sum d}{n}$ $\sum d = \sum (x_i - y_i)$ <p> x_i : データ1 y_i : データ2 s_d : 対になったデータの差の標準偏差 f : 自由度 t : t 分布表 </p>	$(\mu_1 - \mu_2) \sim (\bar{x}_1 - \bar{x}_2) \pm t\left(n_1 + n_2 - 2, \frac{\alpha}{2}\right)$ $* \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \left\{ \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} \right\}}$ <p> n_1, n_2 : データ数 μ_1, μ_2 : 母平均 \bar{x}_1, \bar{x}_2 : 標本平均 </p>	$(\mu_1 - \mu_2) \sim (\bar{x}_1 - \bar{x}_2) \pm$ $t\left(f, \frac{\alpha}{2}\right) * \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$ <p>ここに、自由度 f は、</p> $f = \frac{1}{\frac{C^2}{n_1 - 1} + \frac{(1 - C)^2}{n_2 - 1}}$ $C = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

⑥ t 値 (= (標本平均の差) / (標本平均の差の標準誤差))

対応有り	対応無し	
	等分散	非等分散 (Welch の検定)
$t = \frac{\frac{\sum (x_i - y_i)}{n}}{\frac{s_d}{\sqrt{n}}}$ <p> ・2つの対応有るデータ x_i と y_i ・s_d : 対応の有るデータの差の標準偏差 </p>	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} - \frac{1}{n_2}} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

対応の有る t 検定では差の標準誤差は小さくなり、対応の無い t 検定では差が大きくなるため標準誤差も大きくなります。

標準誤差が小さくなるという事は、信頼区間の幅が狭くなり、真の値を推測できます。

⑦ 自由度 m

対応有り	対応無し	
	等分散	非等分散 (Welch の検定)
$m = n_1 - 1 = n_2 - 1$	$m = n_1 + n_2 - 2$	$m = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$

多重比較検定の補正

多重比較の問題点は、検定数が増えると type I エラーの確率が増えることです。「表. 2 種類の過誤」を参照願います。

このエラー回避のため、多重比較検定の補正を行います。

fMRI での補正方法 (357 ページ、文献 4) :

①有意水準 $\alpha=0.05$ で t 検定しますと、スクリーンには脳以外の場所でも有意差のポイントが検出されます。これはおかしい検定です (独立した検定数は、voxel 数と対応します)。

②一つの断面が $64 \times 64 \times 34$ (slices) で $140,000$ voxels が有ったとします。刺激が無い状態でも、 $\alpha=0.05$ の場合、 $140,000 \times 0.05$ で 6800 個の仮説検定における第 1 種の誤りが発生する確率が有ります。

第 1 種の誤りが発生しない確率 p (脳全体での検定で、1 voxel たりとも間違いでない確率) は、

$$p(\text{no Type I error}) = (1 - \alpha)^N \quad \text{ここに、} N \text{ は voxel 数.}$$

例えば、single slice 断面で、 $\alpha=0.01$ で $N=4096$ の場合、確率 p は 1.3×10^{-18} となり、全脳の偽陽性出現率は、ほぼ 100% となります。

③第 1 種の誤りを補正するために、単純な Bonferroni correction を適用しますと、 $\alpha=\alpha/V$ で、例えば、 $0.01/4096=0.000002$ となり、第 1 種の誤りは補正されます。が、逆に第 2 種の誤りの確率が増大します。第 2 種の誤りは、被験者の症状を見落とす深刻な問題となります。

④Bonferroni では、すべての検定が独立していると仮定しています。しかし実際は異なります。実際の image は空間的に相関しています。つまり、しばしば voxels 毎に独立しておらず同じ voxel 族です。そこで SPM では random field theory に基づいた補正 FWE (family wise error) が選択されます。

⑤補正手法として他に、FDR, small-volume correction, permutation, cluster-size thresholding, GRF (Gaussian random field) が有ります。

OEG-16/SpO2 での補正方法 (案)

OEG-16/SpO2 での t 検定評価についても、上記 fMRI の多重検定と同じ問題が存在します。

そこで、その対処として、次の案が有ります。

①Bonferroni をそのまま適用しますと、 α は $\alpha/16$ で、低い厳しい閾値となり、実際に有意差が有る賦活を捉えない場合が有ります。

②Bonferroni の代替案は数多く提案されています。e. g. Dubey/Armitage-Parmar algorithm

③ここでは、1 例として下記手法を提案します。

i) 賦活の無いチャンネルは無視するというスクリーニングをまず行います。

ii) 続いて、有意な賦活が認められたチャンネルを ROI とします。例えば、CH2, 3, 13, 14 の 4 チャンネルに賦活が見られたとした場合、有意水準 $\alpha/4$ という値で検定します。

p 値について

- p 値は inference であって、estimation では無い。
- p 値は、測定データから得られる確率（有意確率）
- α (False positive)値は、経験則から得られる値（有意水準）。（例えば、 $\alpha = 0.05$ or 0.01 ）
- β (False negative)値は、経験則から得られる値。
- $(1-\beta)$ の値は、power と言われ、80%か 90%にセットされるところ（文献 5 pp76）。
- 信頼区間は信頼度に関係しています。信頼区間が狭いほど信頼度は増します。
例えば、標準正規分布の信頼度 95%の区間は（one sample、両側検定）、
母平均 μ 、母標準偏差 σ 、標本平均 \bar{x} 、サンプル数 n 標準誤差 SE

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \rightarrow \bar{x} - 2 * SE < \mu < \bar{x} + 2 * SE$$
 これは、 $\bar{x} \pm 2 * SE$ の範囲に母平均 μ がある確率が 95%となることを示しています。
- データのばらつき（parameters）を表す標準偏差 SD については次図を参照願います。

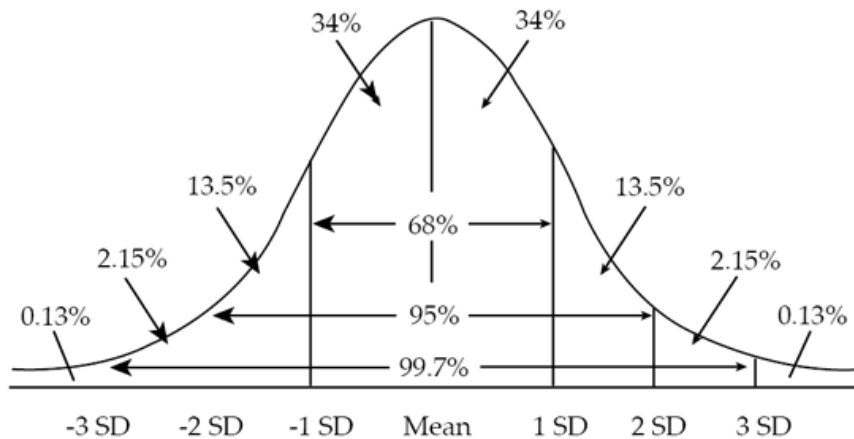


Fig. 1: Normal distribution

(Normal Distribution, "p" Value and Confidence Intervals NJ Gogtay, SP Deshpande, UM Thatte)

- 2つの標本（two sample）の母平均の差の信頼区間は、前の⑤表に記載されています。
- 標準誤差は信頼区間と関係しています。

$$SE = \frac{Variation}{\sqrt{Sample\ size}}$$

精度を上げる一つのやり方は、sample 数を増やすことが、この式より判ります。

仮説検定における 2 種類の過誤

仮説検定は、確率的な事象に基づいているため、誤った判断も有り得ます。2 種類の過誤について下図に説明します。

直観的把握用の図

		True Classes	
		<i>True Positive</i>	<i>False Positive</i>
Predicted Classes	case	Cyber Attack	No Cyber Attack
	model	Alarm Activated	Alarm Activated
	result	You saved your server	You didn't lose anything but got tensed
		<i>False Nrgative</i>	<i>True Negative</i>
	case	Cyber Attack	No Cyber Attack
	model	Alarm NOT Activated	Alarm NOT Activated
	result	You lost your data	All good

“Are you evaluating your model’ s performance correctly?”

Kartikeya Rana, May3, 2020

表. 2 種類の過誤 (文献 5)

帰無仮説 H0 「○○は□□である」

		Truth (仮説の真偽)	
		Null hypothesis false H1 (active) (偽の仮説)	Null hypothesis true H0 (inactive) (正しい仮説)
Result of statistical test (検定に拠る判定)	Statistically significant: reject null hypothesis Reject H0(active) (仮説を棄却)	True positive $(1 - \beta)$:power	False positive Type I error α (有意水準) 偽陽性
	Not statistically significant: fail to reject null hypothesis Accept H0(inactive) (仮説を棄却しない)	False negative Type II error β 偽陰性	True negative $(1 - \alpha)$

(注) この $(1 - \alpha)$ が、多重比較検定の補正の説明と対応します

(注) 陽性とは、例えばインフルエンザの検査を受けて、インフルエンザに感染しているケース

[参考文献]

1. 「入門はじめての統計解析」石村貞夫、東京図書（2013）
2. 「44 の例題で学ぶ統計的検定と推定の解き方」上田拓治、オーム社（2009）
3. 「Modern Statistics for the Life Sciences」,Alan Grafen & osieHails,Oxford ,2002
4. 「Functional Magnetic Resonance Imaging」、Huettel SA,Song AW,McCarthy G,
Sinauer Associates Inc,2nd edition,2009
5. 「What is a p-value」, Andrew Vickers,Addison-Wesley,2010
6. 「SPM12 STARTER' GUIDE」, Erno Hermans

3. PCA 主成分分析

3.1 主成分分析の目的

平均偏差（変数と平均の差）の分散が最も大きくなるようにしたものによって、評価したい時。

例えば、

- ① 生理学的な systemic noise を低減し、activation pattern を抽出します。
- ② 他のバイオマーカー測定値と相関する成分を抽出し、波形を復元します。
- ③ 同一被験者の複数タスク間の相似性、非相似性を確認します。

3.2 分析手法（分散の最大化）

手法としては 3 種類あります。特異値分解法(SVD, Singular Value Decomposition、文献 1)、共分散行列の固有値分解法(EIG, Eigenvalue Decomposition of the covariance matrix、文献 2)と交互最小 2 乗法(ALS, Alternating Least Squares) です。

EIG は、信号数 p よりサンプル数 n が大きい場合、SVD より処理速度が速くなりますが、精度は落ちると言われています。

本ツールでは、SVD を採用します。

3.3 SVD 法

「線形代数 基礎と応用、新井仁之、日本評論社」のページ 326 から抜粋しています。

x_{ij} は n 行 p 列の測定データ (n はサンプル数、 p は信号チャネル数) とし、

w_{ij} は係数（方向比）とすると、

次の式であらわされる z を主成分とします。 m は主成分の数 ($m \leq p < n$) 。

$$z_1 = w_{11}x_{11} + w_{12}x_{12} + \dots + w_{1j}x_{1j} + \dots + w_{1p}x_{1p}$$

$$z_2 = w_{21}x_{21} + w_{22}x_{22} + \dots + w_{2j}x_{2j} + \dots + w_{2p}x_{2p}$$

$$z_i = w_{i1}x_{i1} + w_{i2}x_{i2} + \dots + w_{ij}x_{ij} + \dots + w_{ip}x_{ip}$$

$$z_m = w_{m1}x_{m1} + w_{m2}x_{m2} + \dots + w_{mj}x_{mj} + \dots + w_{mp}x_{mp}$$

z の分散が最大化する（データのばらつきを最も良く表す）係数 w_{ij} を算出すれば、主成分が求まります。

次の平均偏差行列

$$\tilde{X} = \begin{bmatrix} x_{11} - \frac{\sum_{i=1}^n x_{i1}}{n} & \cdots & x_{1p} - \frac{\sum_{i=1}^n x_{ip}}{n} \\ \vdots & \ddots & \vdots \\ x_{n1} - \frac{\sum_{i=1}^n x_{i1}}{n} & \cdots & x_{np} - \frac{\sum_{i=1}^n x_{ip}}{n} \end{bmatrix}$$

の特異値分解をします。 \tilde{X} の特異値を

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_p \geq 0$$

とします。

この時、特異値分解

$$\tilde{X} = \sum_{j=1}^p \sigma_j \mathbf{u}_j \mathbf{w}_j^T$$

を与えるような実数 R の R^n の正規直交基底 $\{u_1, \dots, u_n\}$ と R^p の正規直交基底 $\{w_1, \dots, w_p\}$ が存在します。

w_j を第 j 主成分ベクトル、 $\langle x_n - \bar{x}, w_j \rangle$ を個体 n の第 j 主成分得点

(平均偏差行列) * (固有ベクトル) = (主成分得点) と言います。

3.4 任意の主成分を除去した後の波形復元

任意の主成分を除去したい場合、主成分得点の図から除去したい成分を選択しますと、その主成分得点の列の値は 0 となります。

(修正後の波形) = (修正後の主成分得点) * (固有ベクトル) の逆行列を計算することに拠り、任意の主成分を除去した後の波形復元が可能となります。

実際の matlab 計算では、

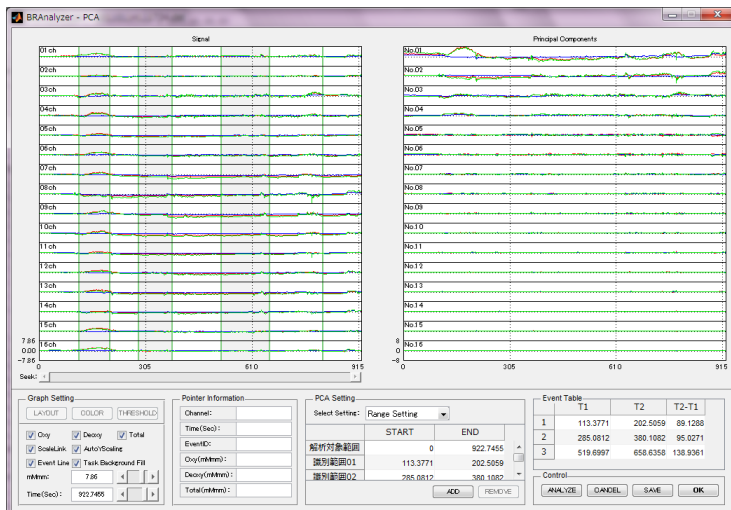
- ① 特異値分解や固有値分解を計算する前に、列平均 μ を引くことでセンタリングを行っています。
- ② 生のデータの主成分をそのまま計算することなく、列に拠って分散が大きく異なる場合を想定して、データを標準化します。そのため各列を標準偏差 std で割って標準化します。
- ③ 主成分分析は Z に対して行われます。

$$Z = \frac{X_{act} - \mu}{std} \quad \text{ここに、} X_{act} \text{ は実際の値、} \mu \text{ は平均値、} std \text{ は標準偏差、}$$

- ④ 拠って修正後の波形 X_{mod} は、

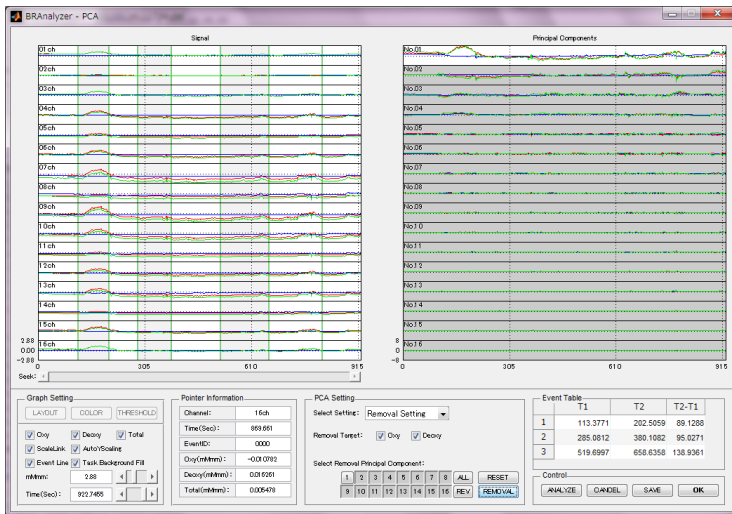
$$X_{mod} = Z * std + \mu$$

で復元されます。

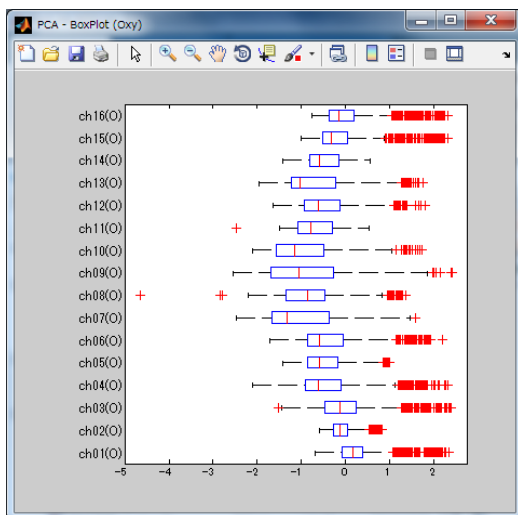


左図は原波形で、右図は主成分得点の図です。

例えば、1 次的主成分得点のみ残し、他成分を削除しまして波形を復元しますと、下図の波形を得ます。



5. PCA 出力の説明



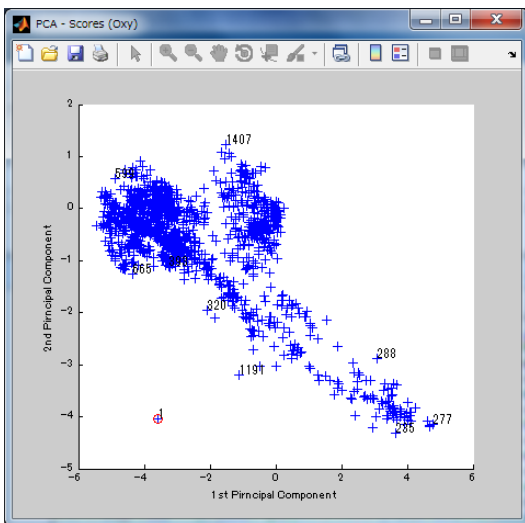
(1) ボックスプロット図

関数 $\text{boxplot}(X)$ は、 X のデータの箱ひげ図を作成します。 X が行列の場合は、列ごとに 1 つのボックスがあり、 X がベクトルの場合は、ボックスが 1 つだけです。各ボックスにおいて、中心の印は中央値で、ボックスのエッジは 25 と 75 の百分位数です。ひげはアルゴリズムが外れ値でないと考慮する最極のデータ点に伸びます。

	1st	2nd	3rd	4th	5th
CH01	0.0200	-0.3946	0.4089	-0.2494	0.0340
CH02	0.0668	-0.0688	0.1869	-0.0342	0.0164
CH03	0.1033	-0.3217	0.1180	0.7416	-0.2474
CH04	0.2325	-0.1896	0.1513	0.2767	0.5514
CH05	0.2081	0.0070	0.0566	0.1432	0.2325
CH06	0.2241	-0.2071	-0.0359	-0.1587	0.2883
CH07	0.4166	0.0639	-0.3497	-0.1471	0.3090
CH08	0.2471	0.5599	0.5235	-0.0554	-0.0944
CH09	0.4000	-0.0714	-0.4932	-0.0989	-0.3330
CH10	0.3852	0.0799	-0.0165	0.2422	-0.2025
CH11	0.2547	0.1645	0.0762	0.1484	0.0174
CH12	0.2437	-0.1001	0.0703	-0.1313	-0.4131
CH13	0.3169	0.0068	0.1312	-0.1950	0.0979
CH14	0.1868	0.2032	0.1746	-0.0267	-0.0891
CH15	0.1502	-0.3279	0.0873	-0.0923	-0.0365
CH16	0.1082	-0.3771	0.2252	-0.2908	-0.2238

(2) 主成分係数ベクトル

主成分係数ベクトル (固有ベクトル、直交行列) の表です。



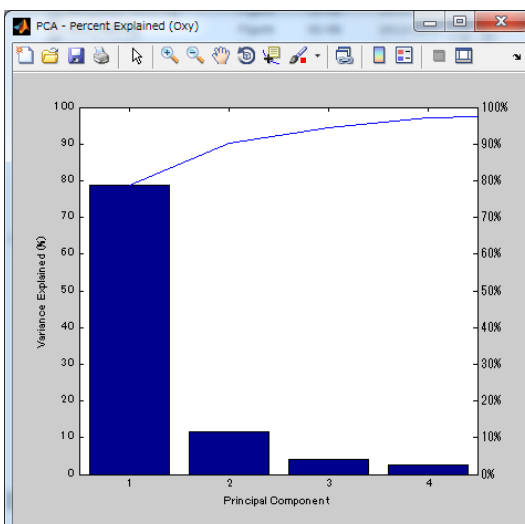
(3) 主成分得点図

サンプル数 n の最初の点から最後の点までの、各サンプルの 1 次主成分得点と 2 次主成分得点の点を図示します。カーソルを持っていきますと、何番目かの番号が表示されます。赤丸は最初の点を表します。

	1st	2nd	3rd	4th	5th
1	-0.0012	-0.0041	0.0953	0.1065	0.1791
2	0.0505	-0.0430	-0.0516	0.0600	0.0888
3	0.1127	-0.0498	0.0172	0.0855	0.0135
4	0.2407	0.0036	-0.0174	-0.0618	-0.0799
5	0.1355	-0.1094	0.1285	0.0584	0.1169
6	0.0797	-0.0268	0.0746	0.0570	0.0248
7	-0.2382	0.0874	-0.0519	-0.0399	-0.3148
8	-0.1133	0.0760	-0.1551	-0.1446	-0.1302
9	-0.1387	0.1013	-0.0377	-0.0370	0.0435
10	-0.1147	-0.0441	0.0029	-0.0775	0.0684
11	-0.1475	0.0288	0.0245	-0.1864	-0.1743
12	0.0635	-0.0322	-0.0447	-0.0729	0.0341
13	-0.0258	-0.1021	0.0331	0.1618	0.2004
14	-0.0359	0.0135	-0.0684	-0.0559	-0.0304
15	0.0708	-0.0431	-0.0043	-0.0901	0.1973
16	-0.0774	0.0183	-0.0041	-0.1251	0.0777
17	-0.0848	0.0670	-0.1284	-0.0451	-0.0622
18	-0.1664	-0.0448	-0.0376	-0.0572	0.0035
19	-0.1250	0.0313	-0.0951	-0.0573	0.2760
20	-0.3752	0.1312	-0.0974	-0.0641	-0.1468
21	-0.2721	0.0673	-0.0441	-0.1036	0.0508

(4) 主成分得点の表

1 次から 16 次までの主成分得点の表です。横方向が次数で、縦方向がサンプル数です。



(5) 主成分の割合図

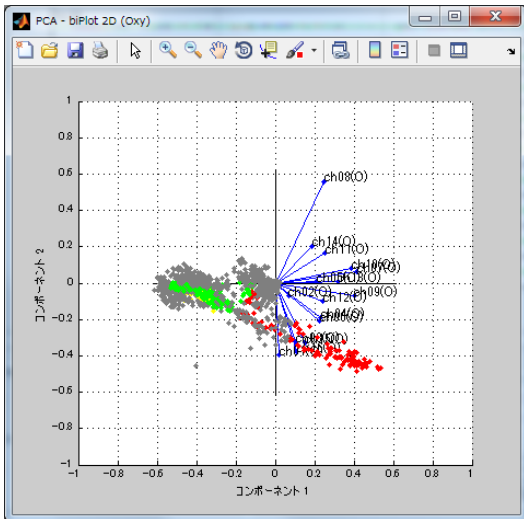
各主成分の寄与率の図です。

PCA - Components Variances (Oxy)

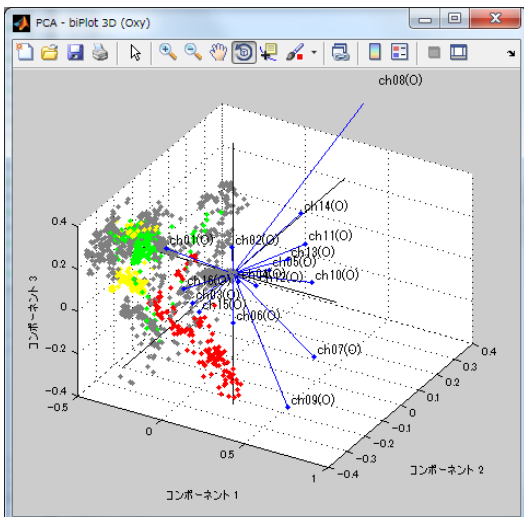
	1
1st	9.7655
2nd	1.4194
3rd	0.5194
4th	0.3161
5th	0.1154
6th	0.0811
7th	0.0436
8th	0.0348
9th	0.0219
10th	0.0204
11th	0.0138
12th	0.0131
13th	0.0097
14th	0.0055
15th	0.0053
16th	0.0037

(6) 各主成分の分散表

各主成分によって説明される分散の表です。



(7) 2次元座標（1次、2次）表示の各信号(ch)の主成分係数と全測定値の主成分得点を表示



(8) 3次元座標（1次、2次、3次）表示の各信号(ch)の主成分係数と全測定値の主成分得点を表示

各タスクで色を変えていますので、タスク間の比較が可能です。

以上

[参考文献]

1. 「線形代数 基礎と応用」、新井仁之、日本評論社 (2006)
2. 「入門はじめての多変量解析」、石村貞夫、東京図書 (2007)
3. 「主成分分析」、MATLAB (2013a)

4. ICA (Independent Component Analysis, 独立成分分析)

1. 独立成分分析の目的

多変量データの成分の中から、他から独立した成分を抽出することが目的です。

2. 手法の概要

ICA の手法として、FastICA を採用しています (文献 1,2)。

FastICA Algorithm を次ページに示します。

FastICA では、最適化法として不動点法を用いて収束解を求めます。非線形性の評価関数として、kurtosis (pow3)、negentropy(tanh)、等を選択します。

3. 課題 (文献 3)

計算された独立成分の信頼性に関し、次の課題があります。

① 初期値

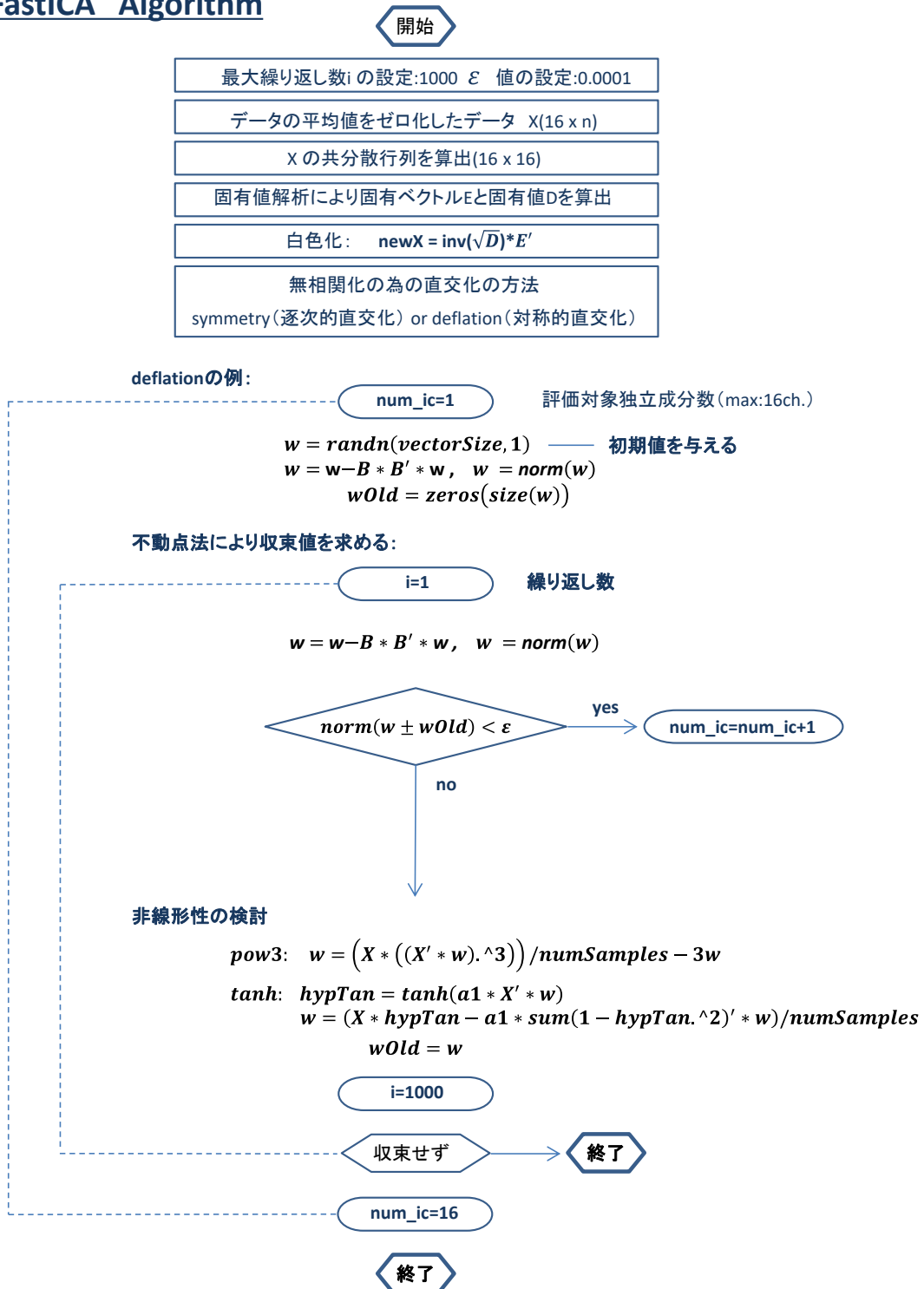
ICA では最適化法のため、初期値を設定します。この初期値を変えますと、その結果に影響を与えます。本来は、異なる初期値を与え、どの初期値からでも現れる成分を評価すべきですが、現在の本ツールでは、計算結果の再現性の観点から初期値を 0 に固定しています。

② アルゴリズム

採用するアルゴリズム (最適化法、非線形性の評価関数) によって、その結果が変動します。

例えば、計算が収束しなかった場合、直交化の手法や評価関数を変えますと収束する場合も有ります。

FastICA Algorithm



3. 任意の成分を除去した後の波形復元

$$x = As$$

$$s = Wx$$

x : 測定値 s : 原信号 A : 混合行列 W :分離行列

任意の独立成分を除去したい場合、図 1 右側の独立成分の図から除去したい成分を選択しますと、混合行列 A の選択された列の値は 0 となり修正混合行列 A' が得られます。

$$(\text{修正後の波形}) = A' * W * (\text{原波形})$$

を計算することに拠り、任意の独立成分を除去した後の波形復元が可能となります。

図 1 は原波形です。図 1 右側の独立成分結果から 1,2,3 次成分を除去し、波形を復元した結果が図 2 左側に再生されます。図 1 に表れています雑音成分が低減されたことが確認できます。

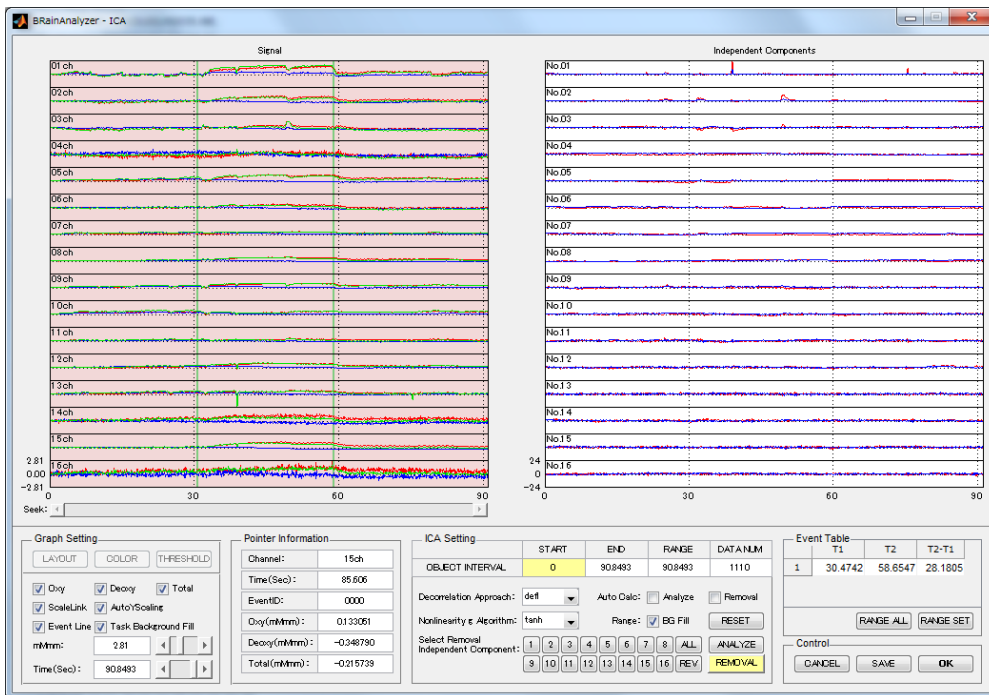


図 1

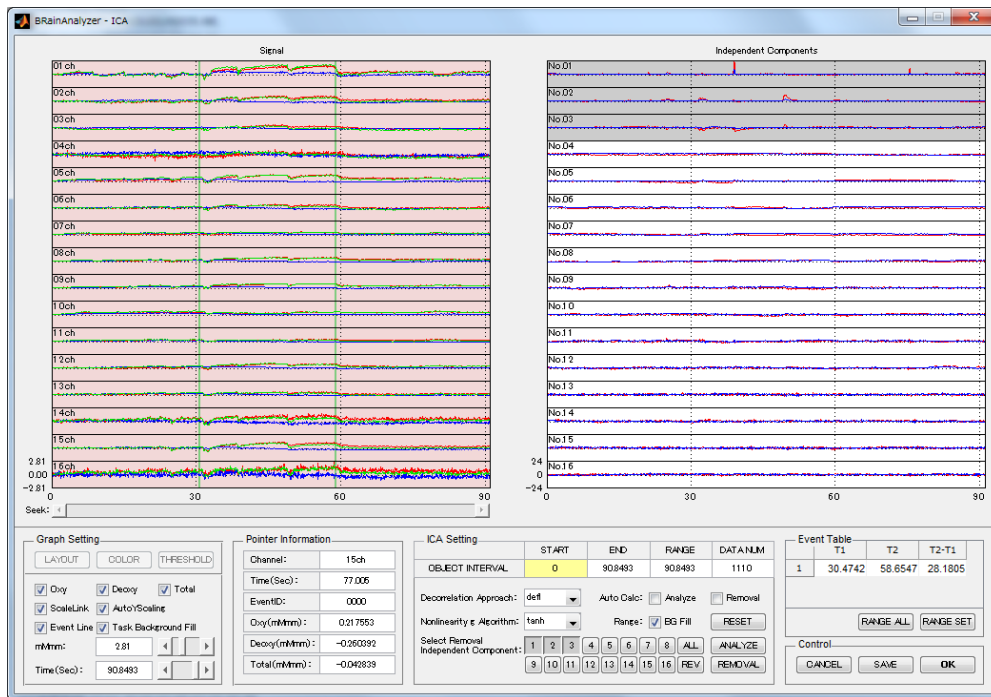


図 2

以上

[参考文献]

1. 「INDEPENDENT COMPONENT ANALYSIS」、Aapo Hyvarinen, Juha Karhunen, Erkki Oja、Wiley, ew York (2001)
2. 「入門 独立成分分析」、村田昇、東京電機大学出版局 (2009)
3. “Validating the independent components of neuroimaging time series via clustering and visualization”, Aapo Hyvarinen, et al., NeuroImage 22(2004) 1214-1222

5. Filter

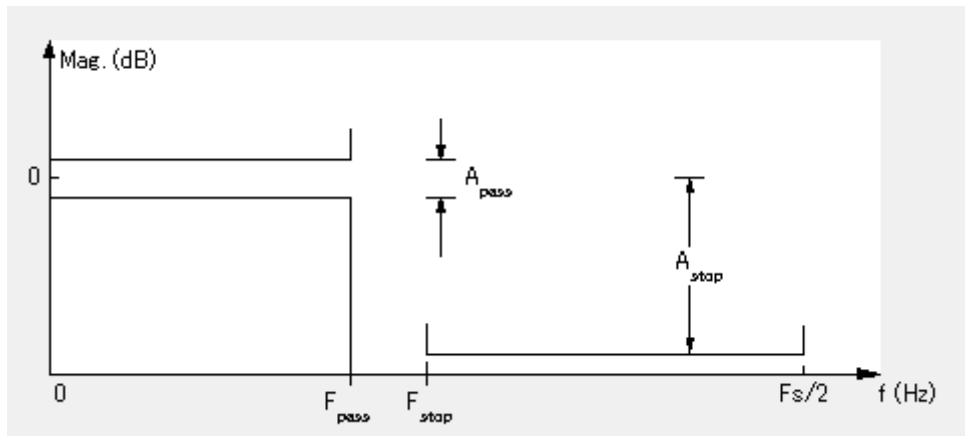
フィルター処理は、源波形から抽出したい信号成分を取り出したり、ノイズを低減するために行います。

本ツールでは、最も一般的で広く使用されています有限インパルス応答 (FIR, Finite Impulse Response) フィルターを用い、フィルター設計法として等リップルを使用しています。

(<http://jp.mathworks.com/help/signal/ug/opening-fdatool.html>)

フィルター処理として、①Low Pass Filter, ②High Pass Filter, ③Band Pass Filter, ④Band Stop Filter の 4 種類、及び⑤Moving Average (移動平均) が有ります。

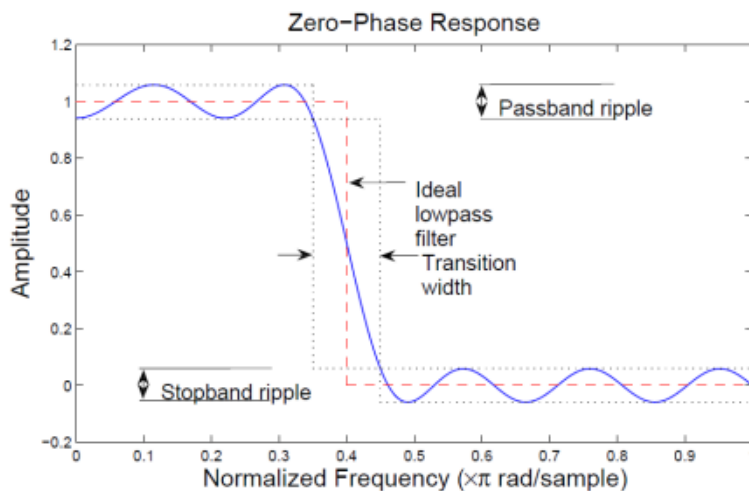
①Low Pass Filter



ここで、BRainAnalyzer で設定します Order, Fpass, Fstop の意味について、Low Pass Filter の例で説明します。

Fpass, Fstop の意味については下図を参照願います。Fpass = 0.37[Hz], Fstop = 0.43[Hz]の例です。

0~0.37[Hz]の範囲で Amplitude を 1 レベルにし、0.43~1[Hz]の範囲で Amplitude を 0 レベルにするフィルターです。



(出典 : https://www.mathworks.com/help/signal/examples/practical-introduction-to-digital-filter-design_ja_JP.html)

Fpass, Fstop の値を設定しますと、0~Fpass[Hz]の範囲で Amplitude を 1 レベルにし、Fstop~Fs/2[Hz]の範囲で Amplitude を 0 レベルにするフィルターを掛けることになります。

Order の意味については、下図を参照願います。Order は減衰傾度に関する設定です。Order の数値が高いほど急峻に減衰します。一方、数値が高いほど、位相差が表れます。fNIRS 信号ではその影響は少ないと判断しています。また、この Order の数値は偶数で設定願います。

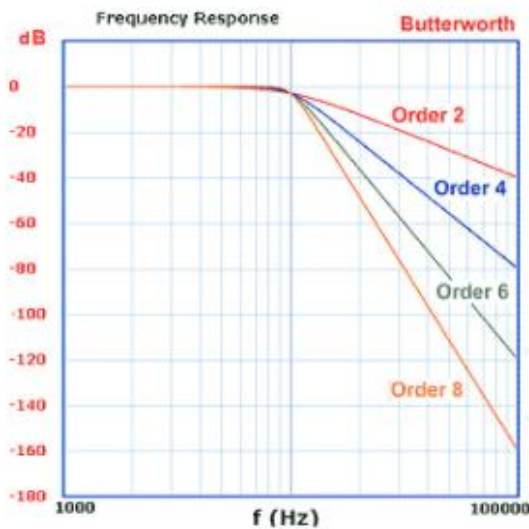
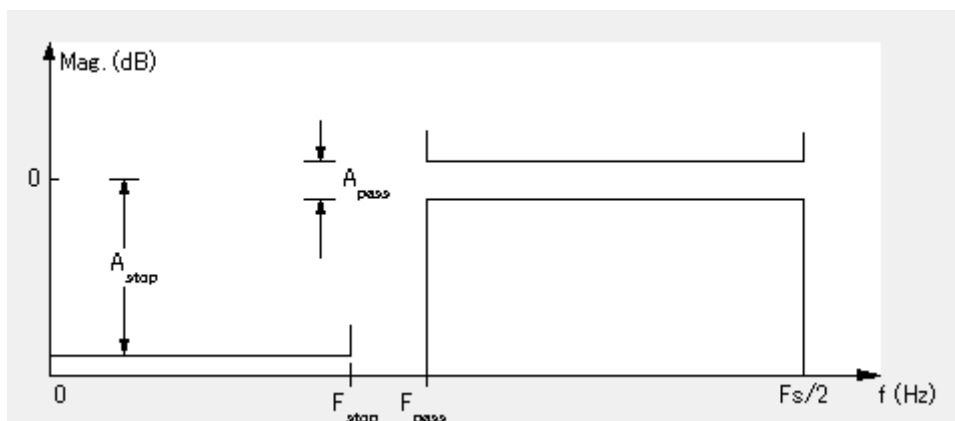


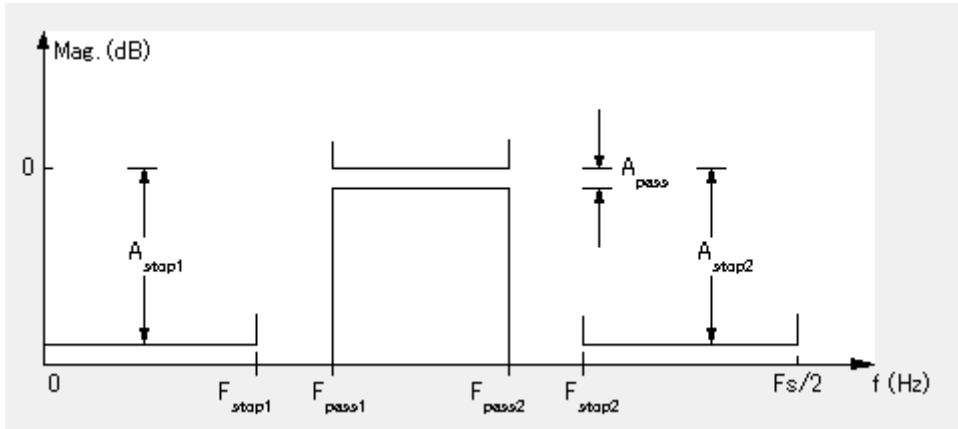
図1 : バタワースフィルタの振幅周波数特性

(出典 : http://www.orixrentec.jp/helpful_info/detail.html?id=43)

② High Pass Filter

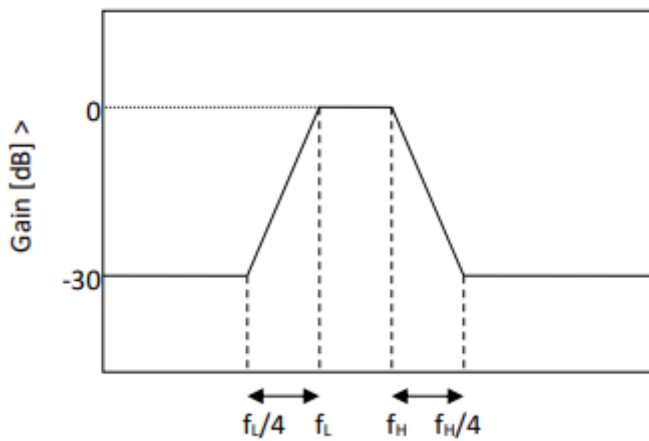


③ Band Pass Filter



$F_{stop1}, F_{pass1}, F_{pass2}, F_{stop2}$ の関係

Artinis Filters の例 : FIR、Gain(attenuation)は 30[dB]



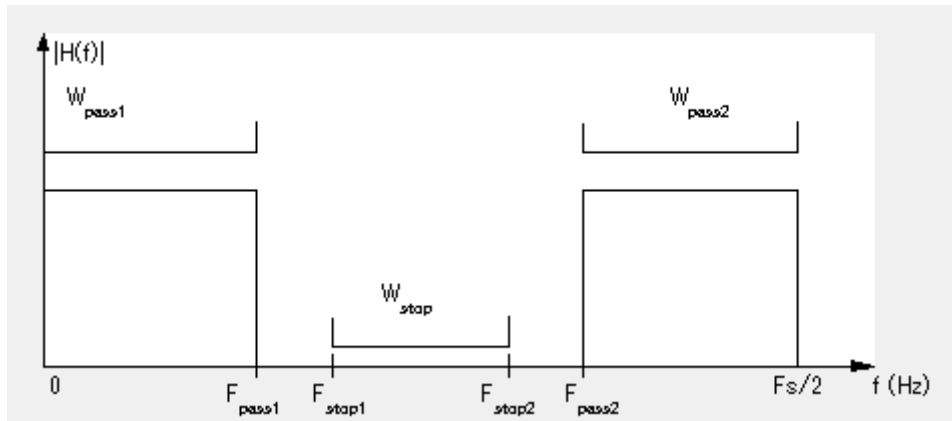
Homer2 の例 :

Homer2 では、IIR/Butterworth を用い次数入力方式で、LowPass の次数は 3、HighPass の次数は 5。

BrainAnalyzer は、IIR/FIR から選択できます。遷移領域、減衰は任意に設定できます。

遷移領域は 20%、減衰は 40[dB]を推奨していますが、遷移領域、減衰が適正かどうかはフィルタ仕様（振幅応答、位相応答、極-零点プロット、等）が表示されますので、試行錯誤で選択願います。

④Band Stop Filter



上記フィルターの設計仕様として、

- ・フィルター次数：設定可能
- ・オプション（密度係数、デフォルト=16）、
- ・バンドパスフィルタの周波数に関する設定：設定可能
- ・バンドパスフィルタの振幅仕様（デフォルト値 $W_{pass1,2} = 1, W_{stop1,2} = 5$ ）

次期バージョンでは、密度係数、バンドパスフィルタの振幅仕様も任意に設定できるようにします。

⑤Moving Average (移動平均)

移動平均は、データの平滑化を行います。信号 $x(n)$ のサンプル数 n の平均値を求めることにより達成します。

例えば、fNIRS 装置のサンプリング周期が 0.6553 秒の場合、 $n=7$ では、約 4.6 秒の (0.6553×7) の平均値 $y(n)$ となります。

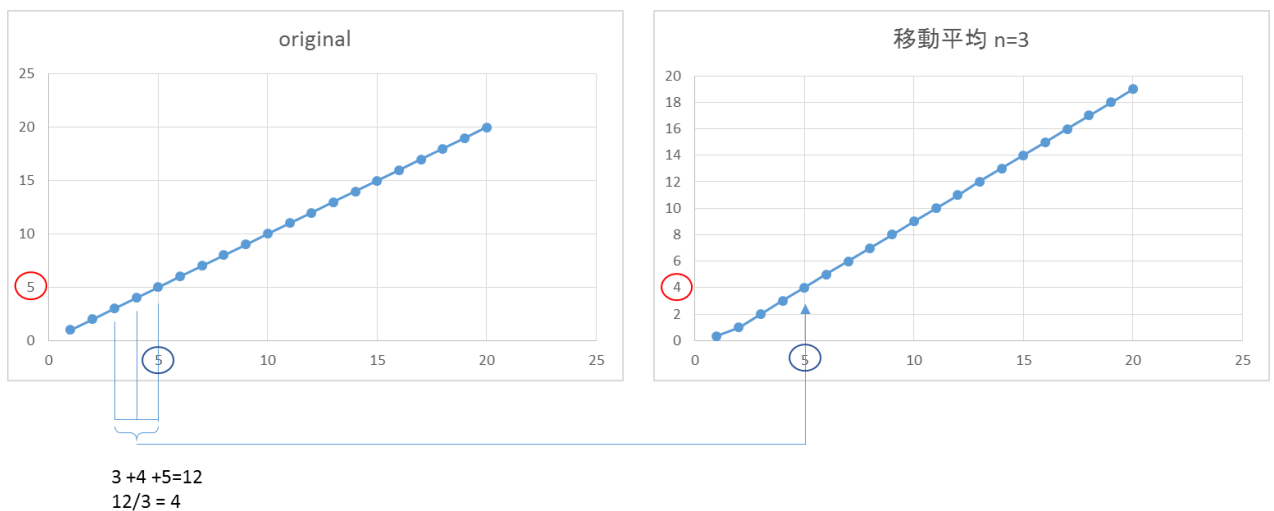
$$y(n) = \frac{x(n)+x(n-1)+x(n-2)\cdots+x(1)}{n}$$

・移動平均データ処理の説明図 $n=3$

original=[1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]

移動平均処理後のデータ

= [0.33 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19]



移動平均は、ローパスフィルタと同じ役割を行います。

F_s : サンプリング周波数 F_c : カットオフ周波数 n : 移動平均数

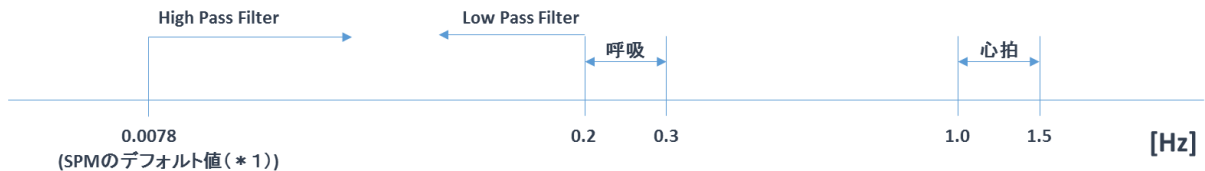
では、近似カットオフ周波数 F_c は、次式で表されます。

$$F_c = \frac{0.442947}{\sqrt{N^2 - 1}} \cdot F_s$$

(StackExchange SIGNAL PROCESSING “What is the cut-off frequency of a moving average filter?”)

Filter 設定値に関する情報 (1 例)

Filter の High Pass の設定値は、実験デザインの周期を考慮して決めます。



- (* 1) タスクが 30 秒、レストが 30 秒の実験デザインの場合、1 周期は 60 秒となります。通常 1 周期の 2 倍程度をカットオフ周期にするとされていますので $60 \times 2 = 120$ 秒となります。SPM のデフォルトは近似値の 128 秒です。周期 128 秒は周波数に換算しますと $1/128 = 0.0078$ [Hz] となります。何故 2 倍程度かは、遮断周波数を達成するためと推定されます。

SPM@JISCMail.AC.UK に、

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=spm;ba86e11a.06>

If you specify a "too short" HP cut-off period compared to your experimental design you will filter away "interesting" experimentally induced variance. The rule of thumb there is still:

$2 \times$ (the longest time between onset of consecutive epochs of the same type).

の記載があります。

- (* 2) ブロックデザインの場合は、上記の考え方で良いですが、event-related の実験デザインの場合は、呼吸の周波数域と重なる場合も有り、Filter の設定は難しくなります。

出典：・ Huettel SA, Song AW, McCarthy G, "Functional Magnetic Resonance Imaging", Sinauer Associates Inc. Second edition, 2009 PP285～

・ 菊池、妹尾、安保、渡邊、米本、「SPM8 脳画像解析マニュアル」、医歯薬出版株式会社、2012

下記文献は、生理学的なノイズを除去するための filter について詳しく調べています。

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3865602/>

fNIRS での関心周波数域は、0.005～0.2Hz で、0.2～0.5Hz を HFOs (HighFrequencyOscillations) 域とし、0.08～0.15Hz を LFOs (LowFrequencyOscillations) 域とし、0.02～0.08Hz を VLFOs

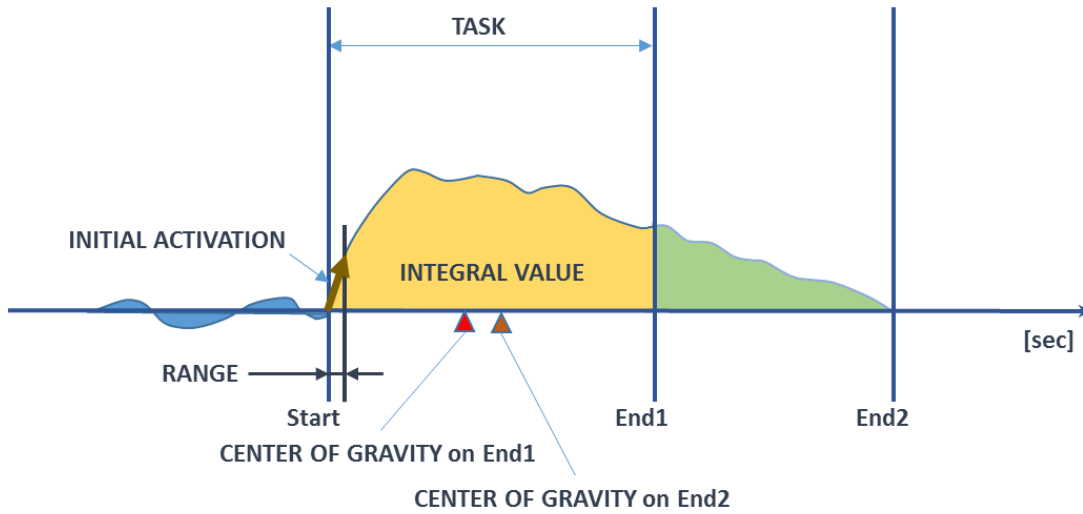
(VeryLowFrequencyOscillations) 域としています。この文献では fNIRS データに加え、mean arterial blood pressure, heart rate, skin blood flow 及び respiration を計測しており、Low pass の遮断

周波数は 0.8Hz, High pass の遮断周波数は 0.008Hz で、5 次の Butterworth filter を使用しているとの事。

fNIRS のブロックデザインの周波数 ($1 / (\text{タスク秒} + \text{レスト秒})$) はしばしば VLFOs 域と重なり、扱いに注意する必要があるとの事。

尚、この文献では 5 個の信号を wavelet coherence analysis と GLM を用いて解析し、生理学的ノイズ除去の手法について述べています。

7. 積分・重心計算



BRain Suite - Analyzer - Integral

Select Display Hb
 Oxy Deoxy Total

Select Display Signal
 Selected Average ⑤

Integral Setting
 Initial Activation: ① 5.24

Select ODE Solver: ② 23

Select Channel (Average)

01	02	03	04	05	06	07	08	09	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	ALL		

③

Range Setting
 Start (sec): 83.64
 End 1 (sec): 113.70
 End 2 (sec): 121.73

<input checked="" type="checkbox"/>	T1	T2	T2-T1
1	22.5280	34.5702	12.0422
2	55.6237	65.6998	10.0762
3	83.6403	113.7050	30.0646
4	129.6794	149.7498	20.0704

④

Signal
 Channel: 13 mMmm: 2.000

⑥

Signal of Target Area
 Channel: 13 mMmm: 2.000

⑦

Calculate Control
 ⑧

Integral Control
 ⑨

①初期賦活値計算時のタスク開始点から何秒までの範囲を設定します。

初期賦活とは、課題開始後の任意時間における反応の速さで、その範囲内の時間の測定点に最小 2 乗法を適用して傾きを算出します。

②ODE Solver

面積は、ODE(Ordinary Differential Equation)45 等の微分方程式を解く関数を用い、その入力引数に関数ハンドルとして指定した方程式を積分して算出します。

重心は、面積を 2 等分する時間で算出しています。

ODE ソルバーの説明は下記サイトを参照願います。

<http://jp.mathworks.com/help/matlab/math/ordinary-differential-equations.html?refresh=true>

ODE ソルバーとして、ODE23,ODE45,ODE113,ODE15s,ODE23s,ODE23tODE23tb を用意しています。デフォルトは ODE45 です。

③積分・重心計算するチャンネルを設定します。

④積分・重心計算する範囲を設定、重心計算については、範囲を 2 個設定可能

⑤積分・重心計算する際、各チャンネル毎か、選択されたチャンネルの加算平均かを選択します。

⑥積分・重心計算するデータを表示します。

⑦計算を実行します。

⑧対象波形の表示と、重心位置を表示します。

⑨計算結果を保存します。

以上