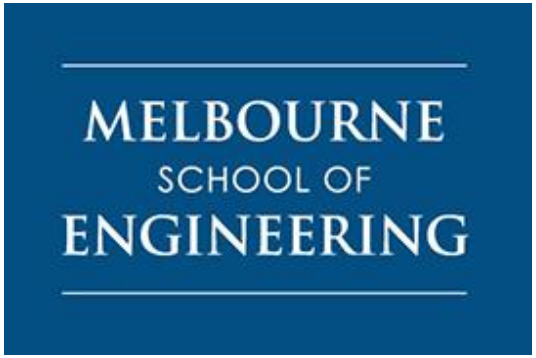
ROBUST HRF ESTIMATION AND ACTIVITY DETECTION FROM FNIRS SIGNALS

Karim Seghouane &
Davide Ferrari



Outline

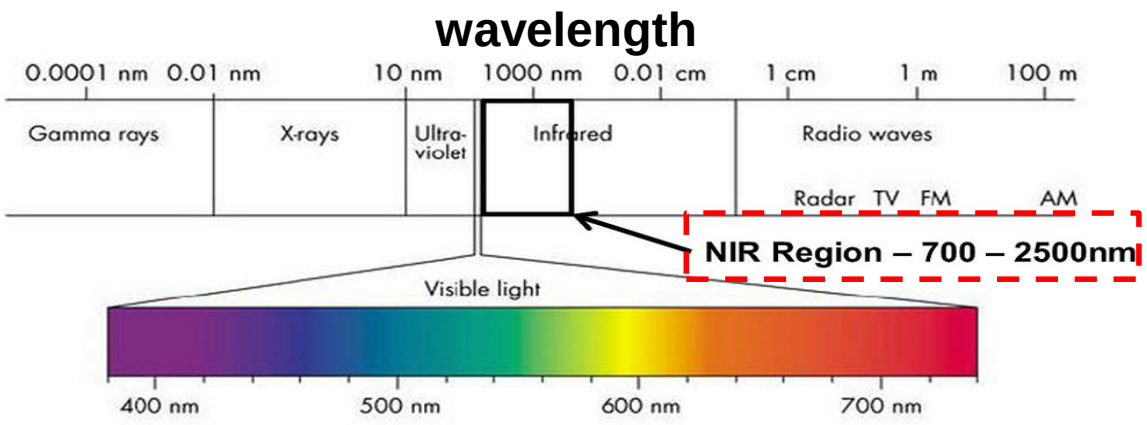
- Functional near-infrared spectroscopy (fNIRS)
- Hemodynamic response function and model
- Robust estimation
- Experimental results

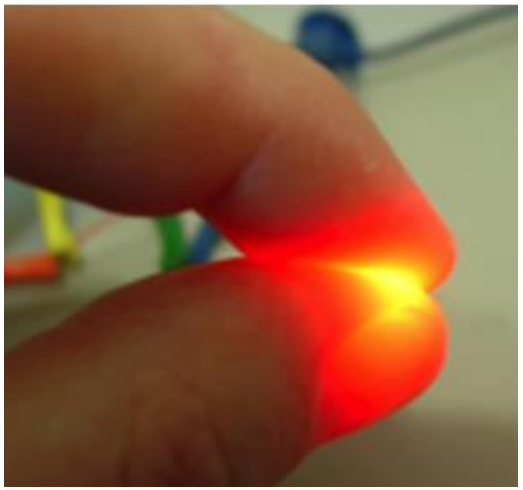
Optical Spectroscopy

Near-Infrared (NIR) is the spectral region slightly above visible light.

NIR has low **Scattering** and **Absorption** in human tissue.

- Spectral absorption has a local minimum at ~800nm
- Scattering is lower in NIR compared to visible light.
- These properties make it attractive for biomedical applications.





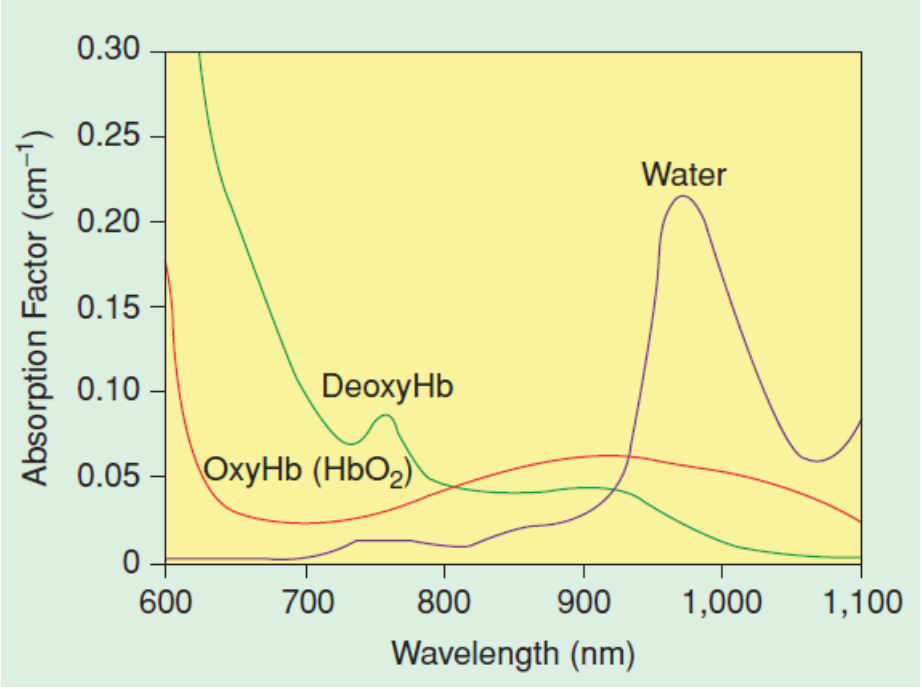
Haemoglobin Absorption

NIR is useful for Haemoglobin analysis, due to:

- Minimal absorption/scattering in tissue.
- Negligible impact of other molecules.
- distinguishable behaviour of the two primary blood components (HbO and HbR).
  - HbO : Oxygenated
  - HbR (Hb) : Deoxygenated

NIR region is high for HbR and HbO and low for other molecules (e.g., water)

NIR absorption in blood



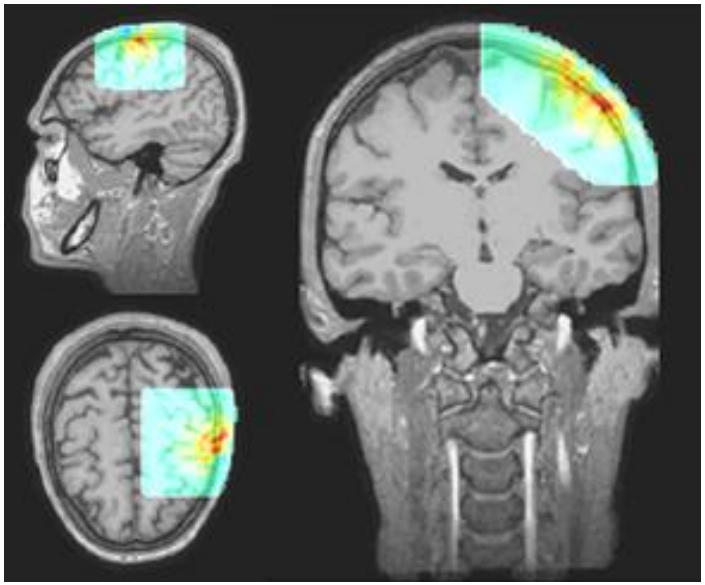
functional NIRS

fNIRS is a method for neuroimaging based on optical spectroscopy.

- It measures changes in light absorption along time.
- NIRS can estimate cortical hemodynamic responses of the brain.

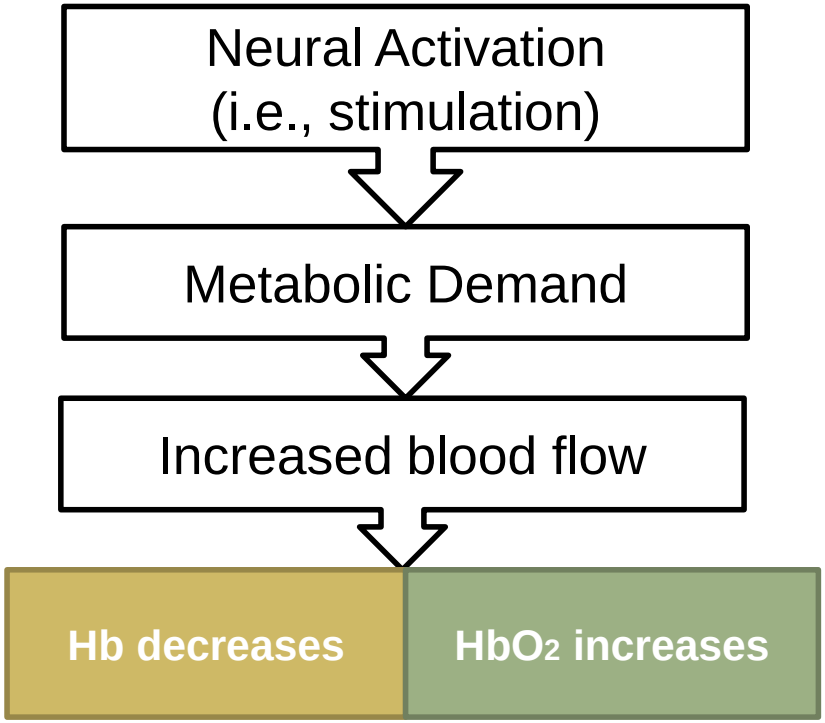
Compared to other modalities, fNIRS is useful because it is:

- fast
- non-invasive

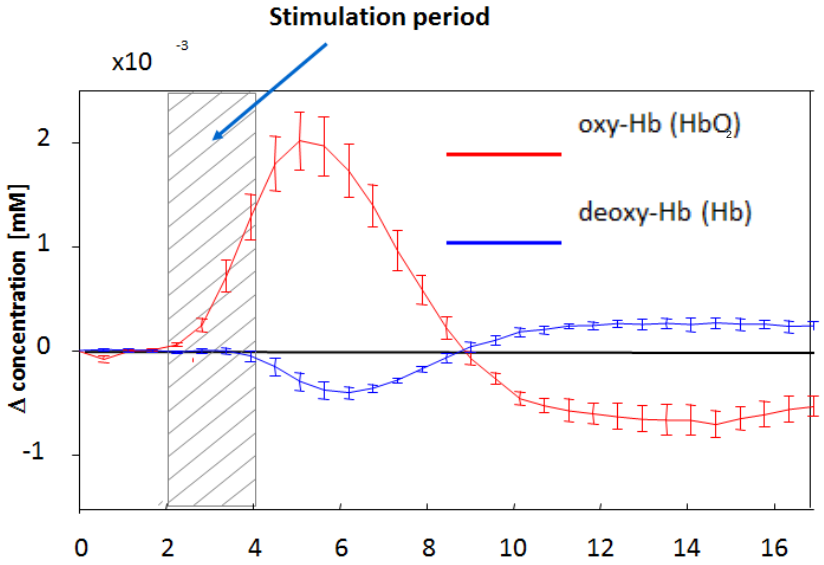


fNIRS Model

- fNIRS studies the relative changes in Hemodynamic responses.



Hemodynamic response function



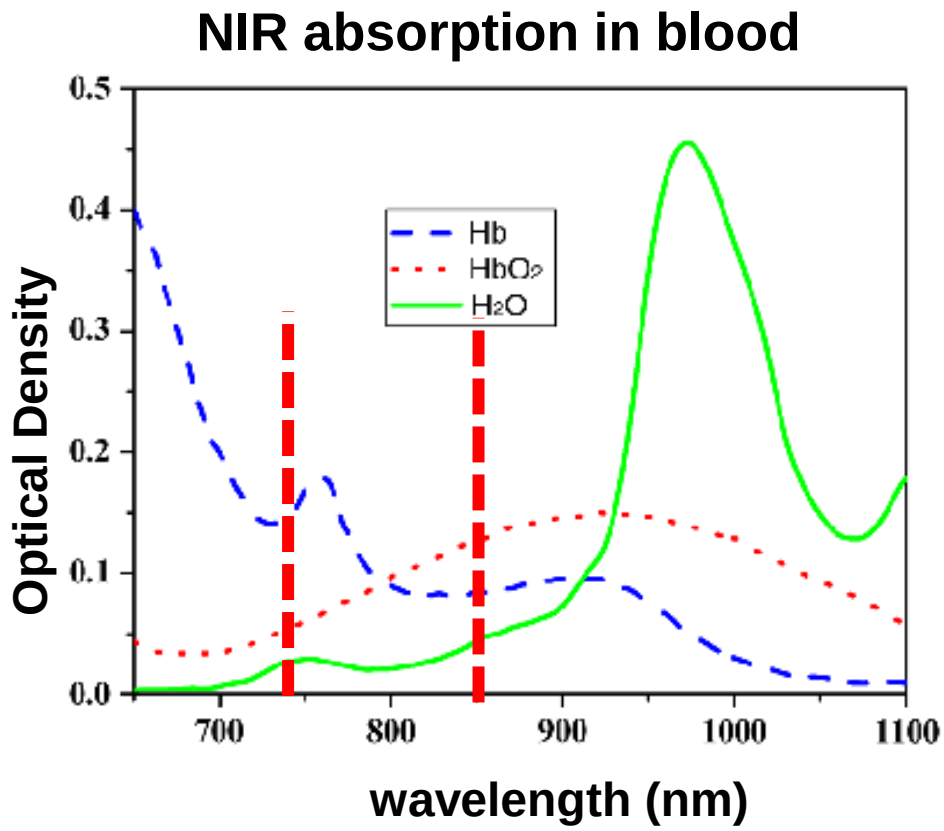
These changes alter the probability of light traveling through tissue

Kohl, M. et al. (2000), “Physical model for the spectroscopic analysis of cortical intrinsic optical signals,” Phys. in Med. & Bio.

Calculating Haemoglobin Concentrations

Calculating Hb/HbO_2 concentration from light intensity uses the Beer Lambert Law (BLL).

- BLL takes advantage of the different absorption signatures of Hb and HbO_2.



Calculating Haemoglobin Concentrations

Beer Lambert Law (BLL):

- Blood molecular concentration is a weighted some of individual concentrations:

$$C = \varepsilon_{Hb} C_{Hb} + \varepsilon_{HbO} C_{HbO_2} + \ \ldots$$

- In NIR, HbO_2 and Hb are considered to be the dominating molecules:

$$C = \varepsilon_{Hb} C_{Hb} + \varepsilon_{HbO} C_{HbO_2}$$

- The BLL states the relationship between light intensity and C:

$$\log\left(\frac{I}{I_0}\right) = e \times C \times d \times DPF + constant$$

I :	light intensity
I_0:	baseline intensity
e :	absorptivity coefficient (depends on wave-length)
d :	light path length
DPF:	differential path length factor

Calculating Haemoglobin Concentrations

Modified Beer Lambert Law (MBLL):

- fNIRS uses the MBLL
- Using two measurements from two wavelength, λ_1 and λ_2, the changes in optical density, OD , between two time points t_1 and t_2 are:

$$\begin{bmatrix} \Delta[Hb] \\ \Delta[HbO_2] \end{bmatrix} = (d)^{-1} \begin{bmatrix} \varepsilon_{HbO_2,\lambda_1} & \varepsilon_{HbO,\lambda_1} \\ \varepsilon_{Hb,\lambda_2} & \varepsilon_{HbO,\lambda_2} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\Delta OD(\Delta t,\lambda_1)}{DPF(\lambda_1)} \\ \dfrac{\Delta OD(\Delta t,\lambda_2)}{DPF(\lambda_2)} \end{bmatrix}$$

where

$$OD(t,\lambda) = \log\left(\frac{I(t,\lambda)}{I_0}\right)$$

$$\Delta t = t_2 - t_1$$

Cortical Response

Light source/detector pairs are used to measure absorption.

- Sources generate optical pulse.
- Detectors measure light intensity.
- Light path forms a "**Photon Banana**"
- Source-Detector pairs are called Channels
- Channel depth is proportional to its length.
  - ~3cm channel → 1-2 cm depth

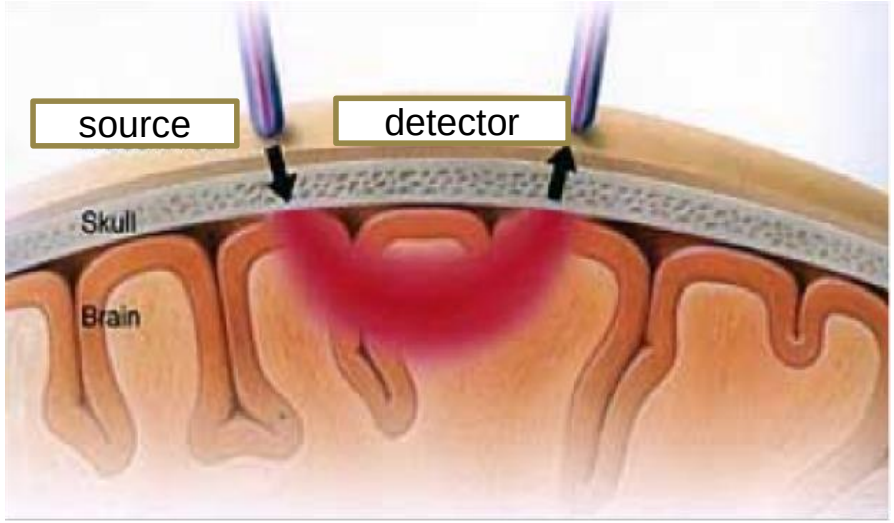
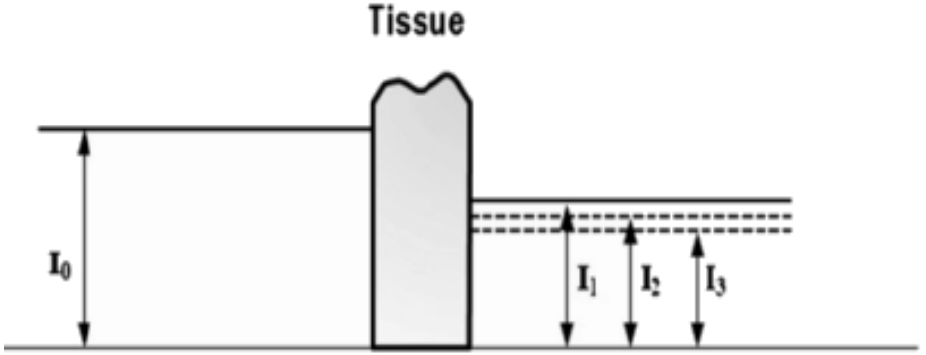


Diagram of Continuous Wave (CW) NIRS system



Note:

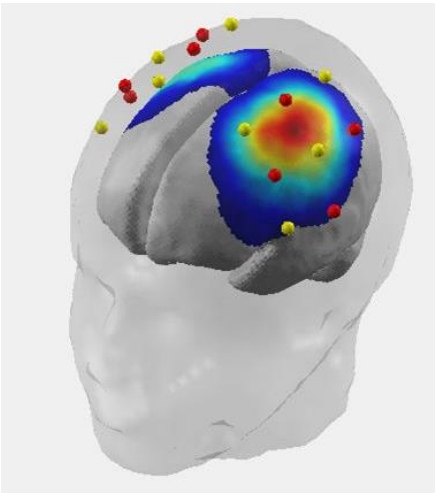
There are other methods for NIRS.
CW is a straightforward approach for modelling the source-detector relationship.

Cortical Response

Probe schematic:

- For a wide cortical response, a spatially distributed array of LED sources and detectors is used.

Example showing NIRS cap
Sources: Red
Detectors: Green



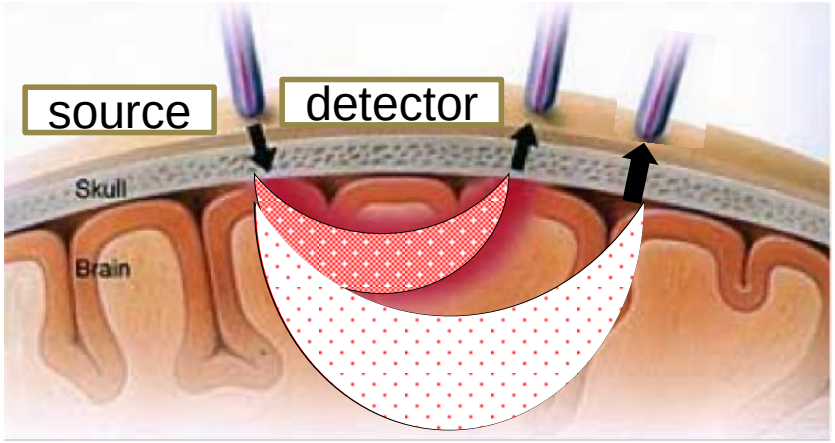
Cortical Response

Light can penetrate up to a certain depth.

- Deeper paths have low signal to noise ratio.

Compromise

- ~3cm distance convention
- Hence, cortical Analysis

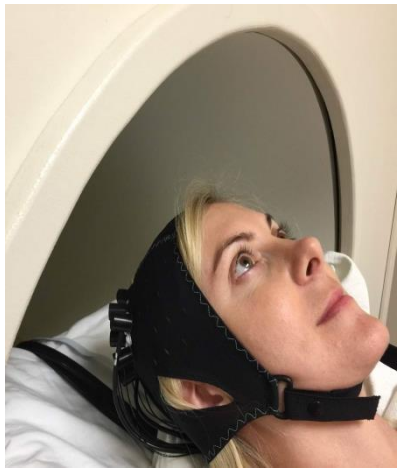


Adding more channels improves SNR and cover a wider cortical region.

fNIRS compared to other techniques

Advantages of NIRS

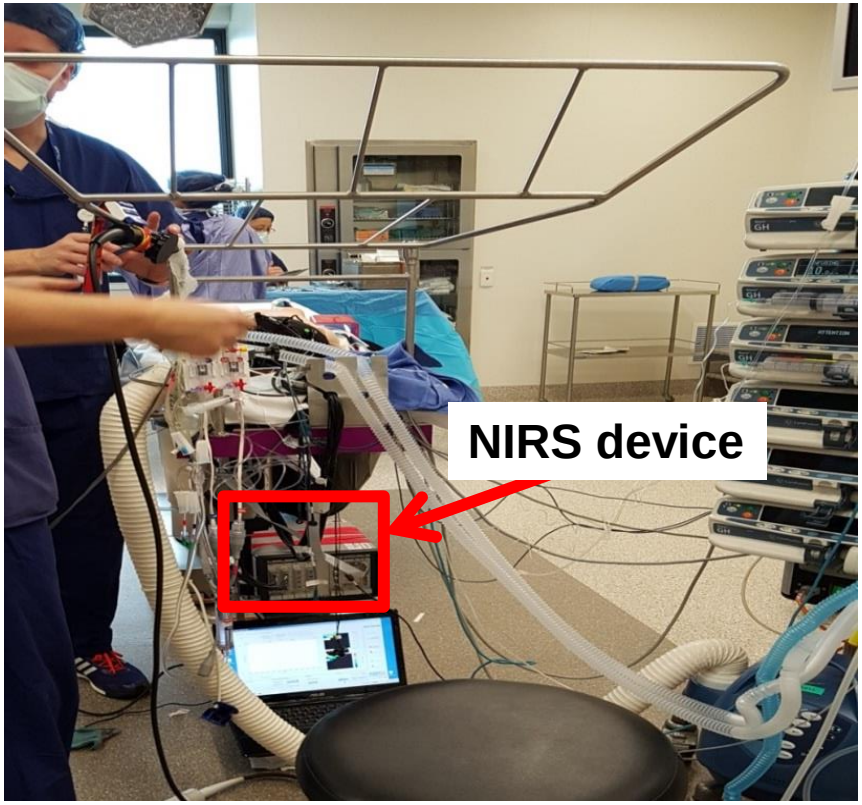
- Portability
- Minimal interference in Magnetic and Electric fields
- Hyper-scanning
- Easier to use for newborns:
  - MRI causes claustrophobia and is unsuitable for children
- No maintenance cost



Field experiments:
In the operating room

We have consent for the Operating Room:

- This is possible due to high portability of the equipment.
- Allows studying the brain under anaesthesia.
- NIRS device was placed out the surgical team's way.

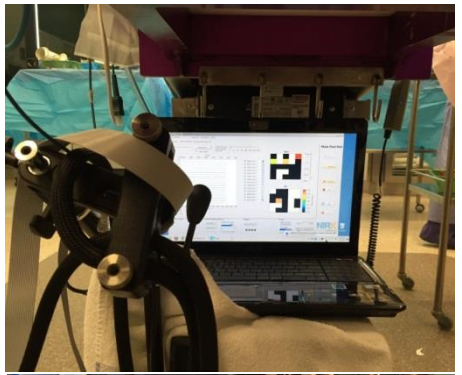


NIRS cap on patient during surgery

Field experiments:
In the operating room

placed under
operating bed

Laptop recording NIRS data

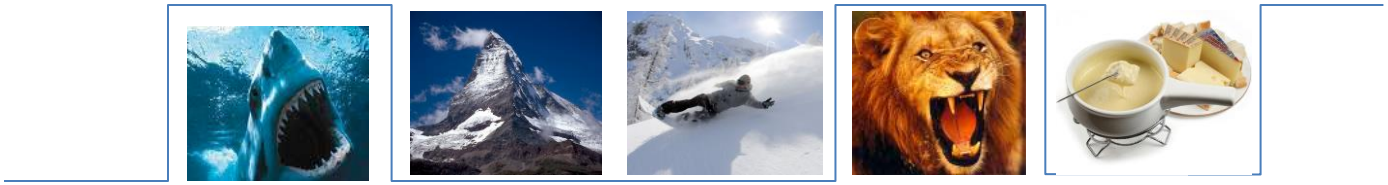


Controlled experiments in fNIRS

Block/epoch designs – periodic stimulus (activation detection)

Unpleasant pleasant Unpleasant

Event-related designs – random stimulus (Hemodynamic response estimation)



Unpleasant pleasant pleasant Unpleasant pleasant

Model for the fNIRS signal

In general, standard analysis of functional hemodynamic changes in neuroimageing techniques such as fNIRS or fMRI is based on an assumption of the linear model.

A standard tool for analyzing fMRI data is some variant of the linear regression model fitted by least-squares to channel (pair of source-detector).

In this approach, the brain region of interest is modeled as a stationary system characterized by its impulse response, the hemodynamic response function (HRF) .

Under this assumption, the measured hemodynamic changes are modelled as a convolution of the stimulus function by the HRF of a linear time invariant system (LTI).

Model for the fNIRS signal

An fNIRS signal measured in channel C_i $\{x(t_i), y(t_i)\}, i = 1, \dots, N$ for a given subject is be represented by

$$y(t) = x(t) * h(t) + f(t) + \epsilon(t), \qquad t = t_1, \dots, t_N$$

$y(t)$: measured noisy fMRI signal

$x(t)$: external input stimulus,
$x(t) = 1\ or\ 0$ indicates the presence or absence of a stimulus
$h(t)$: HRF

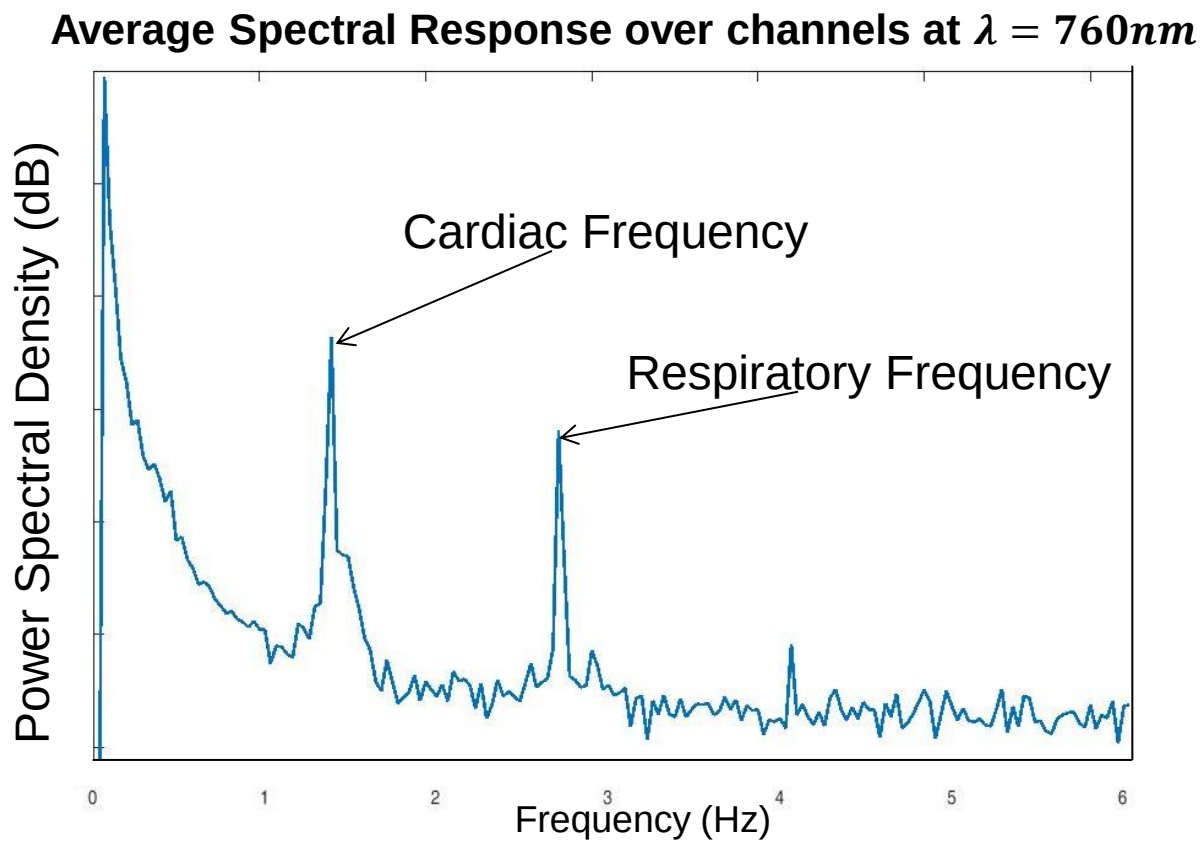
$f(t)$: drift (cardiac pulsation, respiration, mean blood pressure, Mayer waves, low oscillations)

$\epsilon(t)$: physiological noise

Physiological Artefacts or drift

Raw NIRS measurements contain unwanted frequency components:

- High frequency: heart-rate, respiratory
- Low frequency: Mayer waves



HRF estimation

- Assume we have the observation $(y_j, x_j, t_j), j = 1, \dots, N$, with t_i and x_i nonrandom. Furthermore we assume $t_1 < t_2 < \cdots < t_N$. For $j = 1, \dots, N-1$

- $$y_{j+1} - y_j = f(t_{j+1}) - f(t_j) + (x_{j+1} - x_j)\boldsymbol{h} + e_j$$

- where $e_j = \epsilon_{j+1} - \epsilon_j$. If $f(t)$is a smooth function of t, Yatchew (1997) suggested to use

- $$f(t_{j+1}) - f(t_j) \approx 0$$

- when $f(t_j)$ is close to $f(t_{j+1})$ and further advocated using ordinary least squares to estimate $\boldsymbol{h}$

A. Yatchew (1997), “An elementary estimator of the partial linear model,” Economics letters, pp. 135-143.

HRF estimation

- Here we assume that $e_j = \epsilon_{j+1} - \epsilon_j$ is a stochastic error term with unkonwn distribution such that $E(e_i) = 0$ and $Var(e_i) = \sigma^2$.

- Motivated by the ordinary least squares which corresponds to the minimization of the Kullback-Leibler divergence under the Gaussian assumption

- The Kullback-Leibler divergence can be generalized using the α-logarithm function

$$\log_\alpha(x) = \frac{x^{(1-\alpha)} - 1}{1 - \alpha}$$

- and is $log(x)$ for $\alpha = 1$

HRF estimation

- The family of power divergence of $f(x;\theta)$ with respect to $f(x;\theta^*)$

$$D_\alpha\left(f(\mathbf{x};\theta) \parallel f(\mathbf{x};\theta^*)\right) = -\frac{1}{\alpha}\int \log_\alpha\left(\frac{f(\mathbf{x};\theta)}{f(\mathbf{x};\theta^*)}\right) dF(x,\theta^*)$$

- By varying α we obtain different well known distances; for example for $\alpha = \frac{1}{2}$ we have the Hellinger distance

$$D_{1/2}\left(f(\mathbf{x};\theta) \parallel f(\mathbf{x};\theta^*)\right) = 2\int \left(\sqrt{f(\mathbf{x};\theta)} - \sqrt{f(\mathbf{x};\theta^*)}\right)^2 d\mathbf{x}$$

HRF estimation

- For $\alpha = 2$ we have Pearson Chi-square distance

$$D_2\left(f(\mathbf{x};\theta) \parallel f(\mathbf{x};\theta^*)\right) = \frac{1}{2}\int \frac{\left(f(\mathbf{x};\theta) - f(\mathbf{x};\theta^*)\right)^2}{f(\mathbf{x};\theta^*)} d\mathbf{x}$$

- For $\alpha = 0$ we have the reserse Kullback-Leibler divergence and for $\alpha = -1$ we have the Neyman Chisquare distance.

HRF estimation

Description of the proposed method

- We are interested in finding $\hat{\theta}_n$

$$\hat{\theta}_n = \arg\min_{\theta} D_\alpha \left(f(\mathbf{x};\theta) \parallel f(\mathbf{x};\theta^*) \right)$$

Proposition 1: Assuming that $0 < \int f(\mathbf{x};\theta^*)^{1/\alpha} d\mathbf{x} < 0$ then

$$D_\alpha \left(f(\mathbf{x};\theta) \parallel f(\mathbf{x};\theta^*) \right) = -\xi^{-1} \int \log_\alpha f(\mathbf{x};\theta) dF(\mathbf{x};\theta^*) + C$$

where $\xi = \alpha E_{f(\mathbf{x};\theta^*)} \left[f(\mathbf{x};\theta^*)^{1-\alpha} \right]$ and
$C = \int \log_\alpha \left(f(\mathbf{x};\theta^*)^{1/\alpha} \right) dF(\mathbf{x};\theta^*)$

- We therefore propose the surrogate likelihood function

$$L_\alpha(\theta) = \frac{1}{\alpha(N-1)} \sum_{i=1}^{N} \left(1 - f(\mathbf{x}_i;\theta)^\alpha \right)$$

D. Ferrari and D. La Vecchia, “On robust estimation via pseudo additive information”, Biometrika, pp. 238-244, 2012.

HRF estimation

- Under the assumption that the noise distribution is close to a Gaussian $\widehat{\boldsymbol{h}}_\alpha$ corresponds to a the minimum power divergence estimator.

$$\begin{aligned}\hat{\mathbf{h}}_\alpha &= \underset{\mathbf{h}}{\operatorname{argmin}}\ L_\alpha(\mathbf{h}) \\ &= \underset{\mathbf{h}}{\operatorname{argmin}}\ \frac{1}{N-1}\sum_{j=1}^{N-1}\rho_\alpha\left\{\frac{\tilde{y}_j-\mu_j(\mathbf{h},\mathbf{X})}{\sigma}\right\}\end{aligned}$$

- with $\mu_j(\boldsymbol{h},\boldsymbol{X}) = (\boldsymbol{x}_{j+1}-\boldsymbol{x}_j)\boldsymbol{h}$ and $\tilde{y}_j = y_{j+1}-y_j$ and ρ_α is the robust loss function

$$\rho_\alpha(v) = \alpha^{-1}\left\{1-\exp\left(-\alpha v^2/2\right)\right\},\ \alpha>0.$$

- If $\alpha \to 0$, we have $\rho_\alpha \to \rho_0(v) = v^2/2$.

HRF estimation

- Leading to the estimating equation defined as

$$0 = \sum_{j=1}^{N-1} w_j(\mathbf{h}, \mathbf{X}; \alpha) \frac{d}{d\mathbf{h}} \left(\frac{\tilde{y}_j - \mu_j(\mathbf{h}, \mathbf{X})}{\sigma} \right)^2$$

- where the weights are given by

$$w_j(\mathbf{h}, \mathbf{X}; \alpha) = \frac{1}{2} \exp \left\{ -\frac{\alpha}{2} \left(\frac{\tilde{y}_j - \mu_j(\mathbf{h}, \mathbf{X})}{\sigma} \right)^2 \right\}$$

- in the special case of $\alpha = 0$, we have uniform weights and the corresponding estimator is just least squares.

HRF estimation

- **Properties**
- **Existence and Unicity**: If $\widetilde{\boldsymbol{X}}$ has full rank and $\boldsymbol{h}$ is in some subset of R^p, the minimized $\widehat{\boldsymbol{h}}_\alpha$ exist and is unique for any $0 < \alpha < \left\{\frac{1}{max_j Z_j(h)}\right\}$ and any $\boldsymbol{h}$ in a subset of R^p

- **Consistency:** If $\widetilde{\boldsymbol{X}}$ has full rank and $\boldsymbol{h}$ is in some subset of R^p, the minimized $\widehat{\boldsymbol{h}}_\alpha \rightarrow \boldsymbol{h}_0$ for any $0 < \alpha < \left\{\frac{1}{max_j Z_j(h)}\right\}$ and any $\boldsymbol{h}$ in a subset of R^p (where $\rightarrow$ denotes convergence in probability and $\boldsymbol{h}_0$ is the minimizer of $EL_\alpha(\boldsymbol{h})$).

- **Asymptotic normality:** Let $\widehat{\boldsymbol{h}}_\alpha$ the solution of the estimating equation such that $\widehat{\boldsymbol{h}}_\alpha \rightarrow \boldsymbol{h}_0$ then under appropriate assumptions

$$\sqrt{N}\mathbf{H}_{N,\alpha}\mathbf{K}_{N,\alpha}^{-1/2}\left(\hat{\mathbf{h}}_\alpha - \mathbf{h}_0\right) \xrightarrow{d} \mathcal{N}_m\left(\mathbf{0}, \mathbf{I}_m\right)$$

- where H and K are sensitivity and variability matrices.

Experiment results

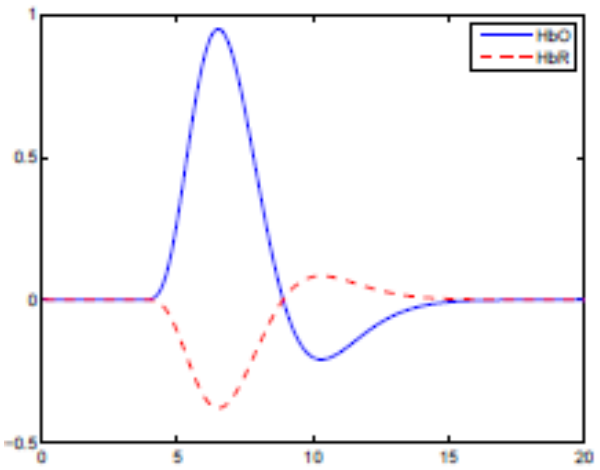
- The controlled evoked hemodynamic fNIRS response signals was generated according to

$$y(t_i) = x(t_i) \star h(t)$$

- with $N = 1000, t_j = \frac{j}{N}, j = 1, \ldots, 1000.$

- Event-related stimuli were generated from indepdendent Bernoulli trials such that $P(x(t_i) = 1) = 0.2$

- The canonical HRF generated



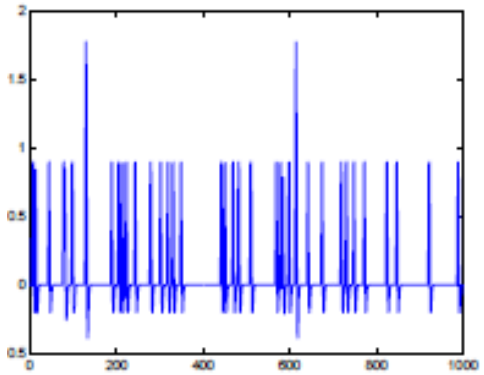
K. J. Friston, “Statistical Parametric Mapping: the Analysis of Function Brain images”, Academic Press, London

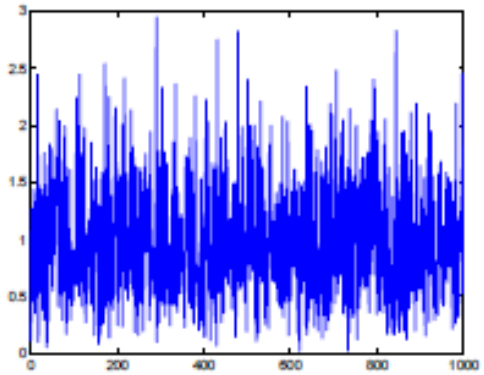
Experiment results

- Performance is assessed using the mean quadratic error $\widehat{MSE} = E\left\{\left\|\hat{\boldsymbol{h}}^{l} - \boldsymbol{h}_{true}\right\|^{2}\right\}$ and estimated emprically by monte Carlo mean

$$\hat{MSE} = \frac{1}{L}\sum_{l=1}^{L}\left\|\hat{\mathbf{h}}_{\alpha}^{(l)} - \mathbf{h}_{true}\right\|^{2}$$

- 100 real fNIRS rest signals where used for simulated drift and noise

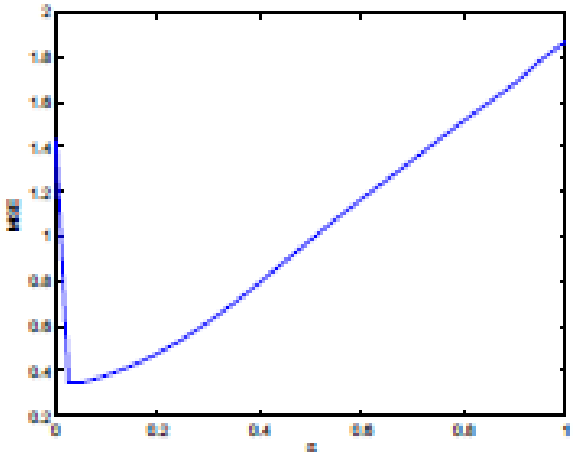


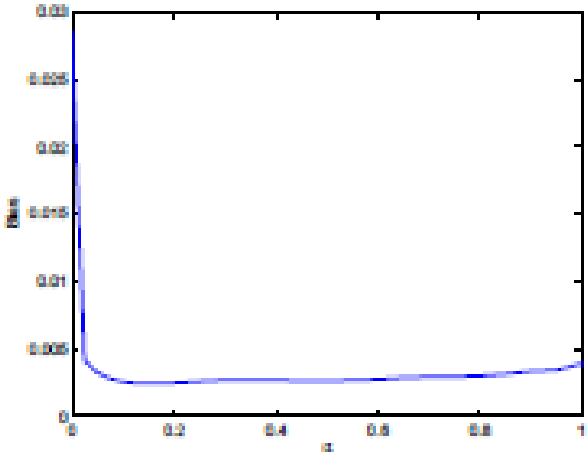


- Example of the evoked response and the obtained fNIRS signal.

Experiment results

- The optimal estimation was obtained for $\alpha = 0.025$ and the associated empirical MSE of 0.347 was obtained compared to 1.428.



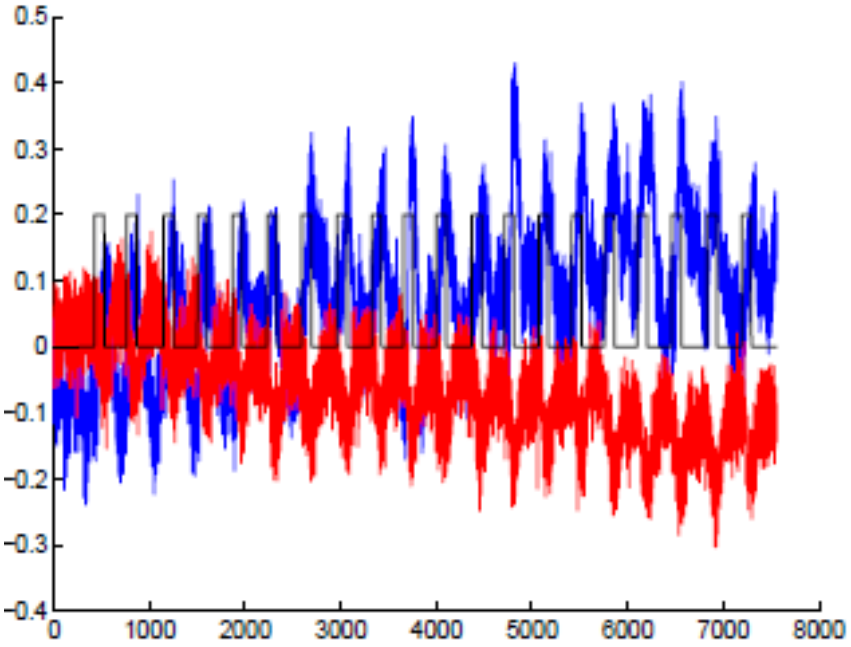


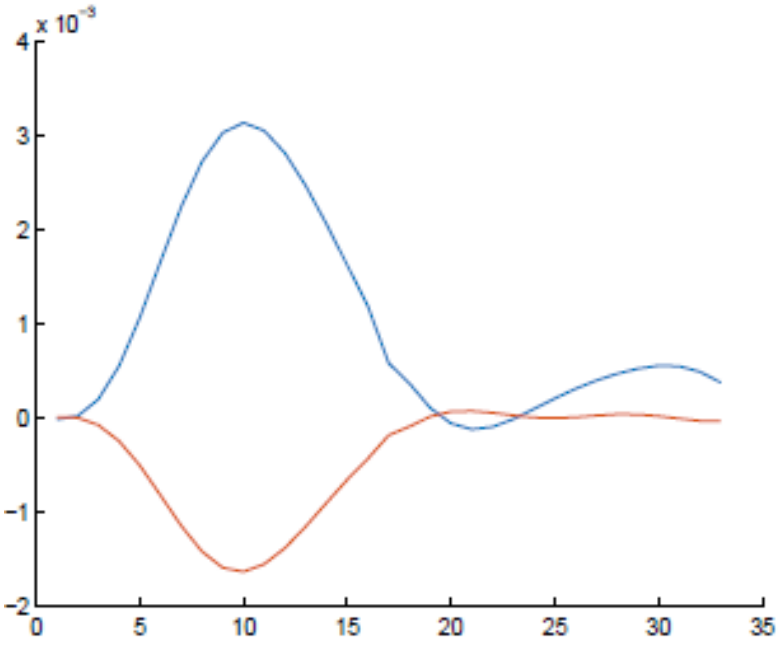
- Empirical MSE as a function of α and the Bias$(\widehat{\boldsymbol{h}}_{\alpha})$. MSE and Bias are large for $\alpha = 0$.

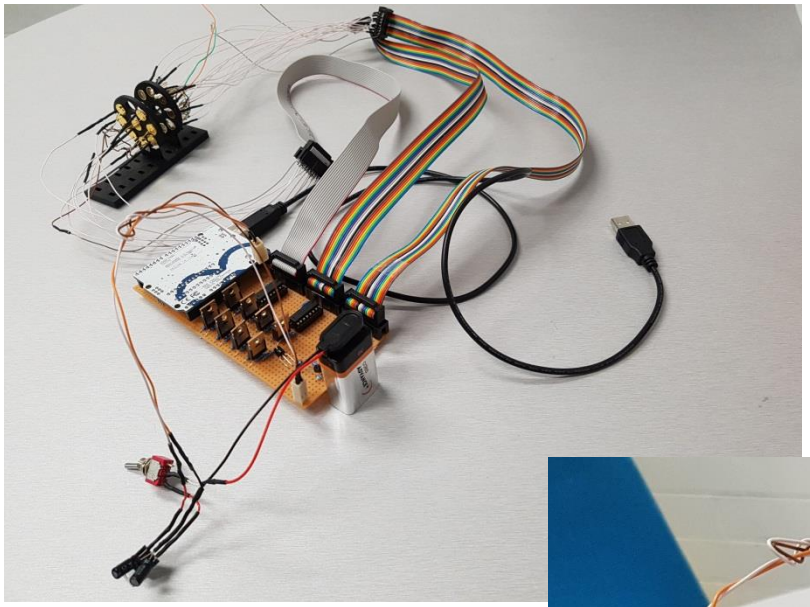
Experiment results

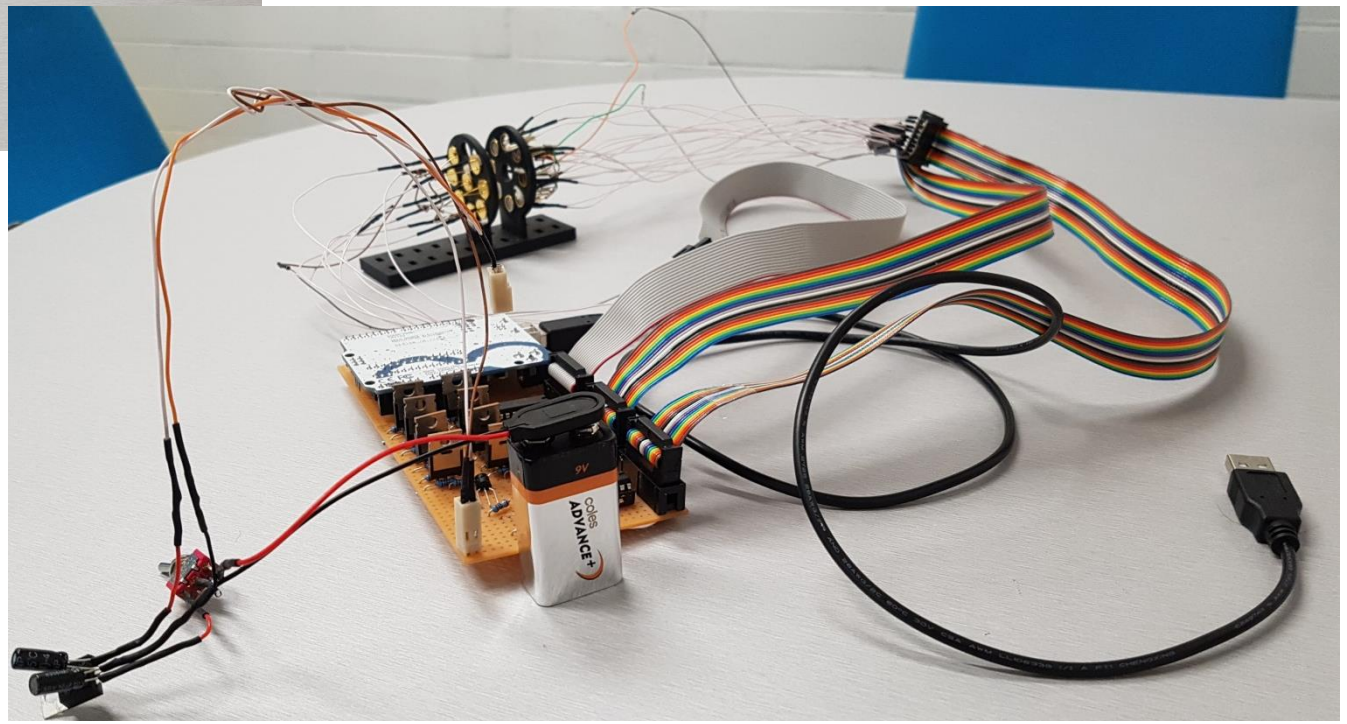
- Finger tapping task with no head movement.
- Investigate the HbO and HbR dynamic during a motor task
- The finger tapping task consisted of 20 alternating tapping and rest epochs.
- Each tapping epoch lasted 10s and each resting epoch lasted 20s
- One couple of HbO and HbR preprocessed signals obtained from the left motor cortex were used for HRF estimation.
- Preprocessing consisted in removing physiological artefacts (cardiac and Mayer frequencies). Motion artefacts were minimal.

Experiment results








9V
coles
ADVANCE+