RESEARCH ARTICLE

A General Statistical Analysis for quantifying laterality using functional Transcranial Doppler Ultrasound

Paul A. Thompson | Dorothy V.M. Bishop | Kate E. Watkins | Zoe V.J. Woodhead*

[1]Department of Experimental Psychology, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK.

Correspondence
*Zoe Woodhead, Department of Experimental Psychology, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK. Email: zoe.woodhead@psy.ox.ac.uk

Present Address
Department of Experimental Psychology, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK.

We propose a method to bring analysis of functional laterality using transcranial Doppler ultrasound data in line with modern statistical methods typically used in fMRI. The method modifies the best practice analyses used in fMRI to allow a greater degree of accuracy and sensitivity to fTCD analysis, and increases flexibility to analyse more complex designs that are currently beyond the scope of fTCD. The method has two main implementations including the general linear modelling framework using generalised least squares to allow for temporal autocorrelation, and secondly, consideration of the haemodynamic response function at the subject-specific level.

KEYWORDS:
General Linear Model; fTCD; fMRI; Laterality; Generalised Least Squares; Nonlinear regression.

1 | INTRODUCTION

Lateralisation of brain activity (functional lateralisation) is most commonly measured using fMRI, which benefits from high spatial resolution, flexible experimental designs and well established analysis procedures. Functional transcranial Doppler sonography (fTCD) is less widely used, but has a number of practical advantages over fMRI, such as lower cost, better portability, fewer contraindications and less sensitivity to head motion. While fMRI is often considered the 'gold standard' in functional laterality research, the practical advantages of fTCD make it preferable for large-scale exploratory studies, studies involving special populations such as children, or studies involving speech production. Both modalities quantify functional lateralisation using a laterality index (LI), but the analytical approach used to calculate LI differs greatly between the two. This document explores whether the general linear modelling (GLM) approach used in fMRI may be applied to fTCD data, with the goals of harmonising the two modalities (making it easier to compare results from fMRI and fTCD studies directly), and increasing analytical flexibility in fTCD over and above what is possible with the current methods.

A basic description of fTCD and its similarities and differences to fMRI are presented below. (For more comprehensive details [1,2,3,4,5]).

1.1 | Physiological Basis

FTCD uses ultrasound probes to detect changes in cerebral blood flow velocity (CBFV) in response to a task or stimulus. The change in CBFV (relative to a resting baseline) is used as a proxy measure for brain activity. Most commonly, CBFV is recorded from the middle cerebral artery (MCA), which has good coverage of lateral temporal, frontal and inferior parietal cortical areas (Figure 1); but it is also possible to record from other cerebral arteries. By comparison, fMRI detects the changes in blood oxygenation levels (the blood oxygen level dependent [BOLD] response). The two methods are similar, therefore, in

that they both use indirect measures of brain activity that rely on neurovascular coupling, but whereas fTCD taps perfusion over a widespread (arterial) territory, fMRI detects local changes (i.e. at the spatial scale of individual voxels).

** INSERT FIGURE 1 HERE**

As arterial blood flow and local blood oxygenation are tightly coupled, it is anticipated that they should show similar time courses in their response to brain activity[7,8]. The time course of the BOLD response following a brief experimental stimulus (known as an impulse response) has been well characterised e.g.[9,10]. It increases reach a peak at around 6 seconds post-stimulus, then declines dropping below baseline levels to reach the trough of the undershoot at around 16 seconds, before gradually returning to resting levels. This shape can be closely approximated by a combination of two gamma functions, one modelling the initial peak and one modelling the undershoot[11]. The precise shape of the response is likely to differ for different individuals and brain regions; furthermore, the double gamma function may be insufficient for modelling sustained neural responses to more complex tasks[12].

The time course of the response in fTCD is less well characterised. Recordings with fTCD of the posterior cerebral artery's CBFV in response to prolonged visual stimulation[13,14] show that (similar to fMRI) the response peaks within four to five seconds. Both studies reported adaptation of CBFV during the prolonged stimulation (i.e. a slow decline in CBFV). Conrad and colleagues reported a brief 'off-reaction' (an increase in blood flow) at stimulus offset; and after a 20 second stimulus, Aaslid and colleagues reported that CBFV returned to baseline levels around 6 seconds after stimulus offset. Conrad and colleagues reported an undershoot effect after stimulus offset, similar to that seen in fMRI. Our own investigations with fTCD (Figure 1) that using a 2 second hand movement task (N=1) demonstrated an impulse response similar to that seen in fMRI, with an intial peak at ~4 seconds, and an undershoot which resolved to baseline levels by around 15 seconds. When a 10 second stimulus was used, the response continued to build until a peak at around ~5 seconds, then there was adaptation in the response during the remainder of the task. After task cessation, the signal dropped to baseline by 20 seconds, followed by an undershoot. We infer from this preliminary data that the responses observed in fTCD are likely to have a similar shape to those in fMRI, but that the time to peak and to return to baseline may be slightly faster.

** INSERT FIGURE 2 HERE**

1.2 | Measurement

The goal of fTCD analysis is normally to assess functional lateralisation, i.e. differences between the right and left hemispheres; unlike fMRI, it is not common to measure activation strength per se, i.e. the degree of change in blood flow during task relative to rest. The CBFV recorded in fTCD is measured in meaningful units (cm/s), but the signal amplitude is dependent on both the actual speed of blood flow and the angle of insonation (the position of the ultrasound probe relative to the artery). Positioning of the probe is subject to operator error[15], and some movement of the probe is likely over the course of an experiment. Hence, there is likely to be a difference in the angle of insonation between the hemispheres. The raw recorded values are therefore unreliable on their own, but must be interpreted relative to a baseline condition.

1.3 | Experimental Design

FTCD studies typically use experimental designs similar to a block design in fMRI. The length of the block is designed to allow the cerebral blood flow in response to the task to reach a plateau and thus maximise the signal to noise ratio. Similarly, sufficient rest time is allowed between blocks to let the cerebral blood flow return to baseline levels. In fTCD, resting baselines are the norm, but the problems with resting baselines have been well documented in the fMRI literature[16,17]. Stark and Squire observed that activity in areas such as the medial temporal lobe were higher during 'rest' than during alternative active baseline tasks such as odd / even decision tasks. This activity is thought to arise from unconstrained thoughts. Binder and colleagues demonstrated that contrasting a semantic decision task with an active baseline (tone decision) rather than rest revealed stronger left hemisphere activity, presumably because certain left-lateralised semantic processes were engaged by spontaneous thoughts during rest. They observed stronger activations during rest in the medial temporal lobe, but also in left angular gyrus, left dorsal prefrontal cortex, left orbital frontal cortex, left pars orbitalis, left temporal pole and posterior cingulate gyrus. The same problems with resting baselines are also likely to affect fTCD, as activity in these areas during rest would presumably cause stronger left hemisphere CBFV in the MCA. If the resting baseline CBFV is stronger in the left hemisphere, baseline correction will lead to an underestimation of left lateralisation during the task of interest.

As in fMRI, the block design used in fTCD is efficient in terms of signal detection, but has some known issues. Firstly, some tasks are unsuited to presentation in blocks, for instance, if they relate to infrequent, brief or unpredictable events. Secondly, performing the same task for the length of a block can allow attention to wander, which is particularly a problem for low-level perceptual tasks that are not sufficiently engaging. In fMRI, these problems are typically avoided by using an event-related design, where tasks are presented in individual trials (rather than blocks of trials), with rapid alternation between different tasks of interest. What this type of design lacks in signal to noise ratio, it makes up for in the number of trials that can be presented in a fixed amount of time (**http://imaging.mrc-cbu.cam.ac.uk/imaging/DesignEfficiency**). Such a design may theoretically be possible in fTCD, but has not been attempted to date due to the nature of the data analysis.

1.4 | fTCD Data Analysis

The time series data recorded in fTCD is similar to that of fMRI, but with only one time series per hemisphere (versus than the many voxels per hemisphere in fMRI) and with a higher temporal resolution (fTCD data is acquired at ~100Hz and usually down-sampled for convenience to ~25Hz, whereas fMRI is acquired at approximately one image every 2-3 seconds). However, they are conventionally analysed in very different ways. In fTCD, the time series is epoched into individual trials that are time-locked to stimulus onset, then all trials are averaged to reduce the impact task-irrelevant noise (this is similar to analysis of evoked response potentials in electroencephalography). A functional laterality index (LI) is then computed by simply taking the difference between left CBFV and right CBFV over a user-defined period of interest in this averaged epoch. The fTCD time course is thus reduced to a single number representing the difference between left and right hemispheres.

In contrast, fMRI uses a general linear modelling (GLM) approach to data analysis[18]. The time series of blood-oxygen-level dependent (BOLD) signal change in each voxel is analysed separately in a mass univariate analysis. The BOLD signal is entered into the GLM as the dependent variable, with explanatory variables such as the time course of the onset of trials within the experiment (convolved by the expected shape of the haemodynamic response) as the predictors. This tests how well the time course of the task predicts the time course of signal change in the voxel. The output of the GLM is an estimated beta value for each explanatory variable at that voxel. When a resting baseline is used, the significance of the beta for the task of interest is simply converted to a t-statistic to test whether it is significantly greater than zero. Alternatively, when an active baseline is used, a contrast can be performed to comparing the beta for the task of interest to the beta for the baseline task, and again, computing a t-statistic to indicate the strength (and statistical significance) of the difference.

1.5 | Aims

The main aim of the paper is to translate the GLM analysis framework used in fMRI for use with fTCD data. The implementation is not straightforward as the data has a few major structural differences, namely: there are only two time series of data (one per hemisphere) rather than one per voxel; and each time series has around 20,000 observations compared to approximately 100 observations in fMRI time series. The secondary aim was to demonstrate that the new analysis method permits a greater range of experimental designs and potential hypotheses to be tested than is currently available. Several examples are given demonstrating the application of the method to novel fTCD experimental designs.

2 | METHODS

The model development is split into two sections; i) modelling individual-specific haemodynamic response function[19] in fTCD, and ii) modelling fTCD data using the general linear model framework. We make certain adjustments to the application of the techniques to allow for the different structure of the data between the modalities. More specifically, the fMRI data consists of thousands of shorter time series (one per voxel) per individuals, whereas fTCD only contains two time series (left and right middle cerebral arteries) but the sampling rate is far higher (20,000 unique observations per time series). After substantial testing of simulation evaluation timings, we decided to try further downsampling to 5Hz as the estimation was becoming unfeasible. Efforts were made to ameliorate the problem by speeding up the execution using faster computational routines or parallelization of the code.

2.1 | Modelling the haemodynamic response function

The haemodynamic response function in fMRI research has been comprehensively studied and refined, typically, opting for a choice of canonical, gamma or boxcar functions. Specifically, we have used the fMRI HRF parameterisation presented by Proulx et al.[19]. More complex modelling using sets of basis functions was proposed by Lindquist[20]. We make the assumption that the voxel-level HRF and the MCA HRF are correlated; hence, we tested the commonly used HRFs from fMRI research on laterality data collected using fTCD[13,14].

We implement three choices of haemodynamic response function in fTCD data, i) a fixed parameterisation of the single gamma or canonical HRF; ii) empirically-estimated canonical (double gamma) function[19]; and iii) the set of basis functions approach[20]. The three approaches can be selected according to the type of stimulus used. A point stimulus that elicits a transient haemodynamic response would be best modelled by the single gamma or canonical function, whereas a longer task or stimulus with sustained response would use the basis functions approach.

2.1.1 | Point stimulus: Fixed parameterisation single gamma or canonical HRF

The first option is a fixed parameterisation of the most commonly used HRFs in fMRI analyses, single gamma and canonical[21]. The default parameters define the standardized shape of the HRF, but some scaling is applied to allow for differing durations in the experimental design. These parameters are taken directly from the R package 'fMRI'[22].

2.1.2 | Point stimulus: empirical canonical HRF

Increasing the sensitivity of the model using the subject-specific haemodynamic response parametrisation[19] reduces the false positive rate and provides a greater degree of sensitivity. Each individual's time series data is split by epochs, each containing one trial of the experimental task, then all epochs are averaged to calculate the individual's average haemodynamic response. This is then used to model an individual's specific HRF. Due to the form of the haemodynamic response (double gamma), each individual's HRF data are modelled using a nonlinear regression (i.e. not linear in the parameters rather than the variables). The nonlinear model's parameter estimates can be directly related to specific dimensions of the standard shape of the canonical (double gamma) functional form; see Figure 3 and equations 1-6, reproduced from Proulx et al.[19].

** INSERT FIGURE 3 HERE**

The canonical HRF is the difference between two gamma density functions, γ and its magnitude is weighted by an amplitude parameter, *amp*,

$$HRF(t) = amp \times [\gamma_1(t) - ratio \times \gamma_2(t)] \tag{1}$$

where t is the repeated observations across time, *ratio* is an adjustment parameter to scale γ_2 in relation to γ_1. The standard gamma function γ has the following form,

$$\gamma_k(t) = \left(\frac{t}{peak_k}\right)^{a_k} \times \exp^{-\left(\frac{t-peak_k}{b_k}\right)}, \qquad k = 1, 2. \tag{2}$$

The parameters b_k and a_k can be rewritten to correspond to quantities from Figure 3,

$$
\begin{aligned}
a_1 &= 8log(2) \times \frac{peak_1^2}{fwhm_1^2}, \\
b_1 &= \frac{fwhm_1^2}{8log(2) \times peak_1}, \\
a_2 &= \frac{8log(2) \times peak_2^2}{fwhm_2^2}, \\
b_2 &= \frac{fwhm_2^2}{8log(2) \times peak_2}.
\end{aligned}
\tag{3}
$$

The nonlinear model is fitted using nonlinear least squares which is an iterative procedure. Least squares requires the minimisation of the sum of squares for both standard linear regression and its non-linear counterpart. However, nonlinear least squares

often has a problem that there can potentially be multiple minima and therefore, multiple potentially unique solutions. The nonlinear least squares has various algorithms available, but here we use the Levenberg–Marquardt algorithm. This algorithm overcomes the minimisation complication by using starting values. Conditional on the starting values, the Levenberg-Marquardt algorithm should only converge on a global minimum given that it is in close proximity to the starting values[23]. This ensures that the iterative procedure does not converge on a non-optimal solution or outside reasonable parameter limits.

The choice of starting values for model parameters, $peak_1$, $peak_2$, $fwhm_1$, $fwhm_2$, amp, $ratio$ are approximated either from estimates or visually examining the data averaged across all individuals' epoched data (see Table 1 for details).

TABLE 1 Parameter's initial value descriptions

Parameter	Description
$peak_1$	Estimated for each individual by extracting the time corresponding to the maximum of the mean blood flow (averaged between the two time series across both left and right sensors across all trials) within the window of stimuli activation.
$peak_2$	Estimated for each individual by extracting the time corresponding to the minimum of the mean blood flow (averaged between the two time series across both left and right sensors across all trials) within the window of stimuli activation.
$fwhm_k$, $k = 1, 2$	Given the starting value 5
amp	Estimated by the maximum of the mean blood flow (averaged between the two time series across both left and right sensors across all trials) within the window of stimuli activation
$ratio$	The ratio of the magnitude of the mean blood flow (averaged between the two time series across both left and right sensors across all trials) between peak 1 and peak 2

2.1.3 | Sustained stimulus: basis functions

Our second approach to mapping the haemodynamic response function is tailored to allow for a sustained response to experimental stimulus. More specifically, if the task at each repetition contains multiple responses, then we observed through experimental data that the haemodynamic response plateaus at a saturation point until the end of the task period where it drops off as usually observed in point stimulus tasks. Figure 4 shows an example of an fTCD response to sustained response stimulus. The blood flow surges as usual, but the peaks and is sustained to maintain blood flow volume for the duration of the sustained response. The typical HRF do not fit this sustained response, so we need to use a non-parametric function to track the time series of blood flow response.

** INSERT FIGURE 4 HERE**

The set of basis functions are constructed following the method described by Lindquist[20]. Basis functions are used in non-parametric regression and are an extension to parametric regression. Typically, parametric linear regression will model an outcome measure, y, against some linear combination of regressors, $\mathbf{X}\boldsymbol{\beta}$. The linear combination can be written as follows:

$$g(x) = \mathbf{X}\boldsymbol{\beta} \tag{4}$$

where $\boldsymbol{\beta}$ are the vector of unknown parameter estimates corresponding to the elements in the design matrix, X. We can specify an alternative non-parametric formulation using the basis functions approach[20],

$$g(x) = \sum_j \delta_j \beta_j(x) \tag{5}$$

where δ_j are the corresponding set of basis functions.

2.2 | General linear modelling

As previously discussed, fMRI analysis typically uses the general linear model framework (GLM) to model fMRI time series data for each voxel [18,21], then combines this information to indicate activity in regions of spatially correlated voxels. The GLM framework is adopted here to model hemospheric time series (one per hemisphere per individual). Specifically, generalised least squares is used to estimate the parameters and account for temporal autocorrelation of the residuals, i.e. serial observations of blood flow from the middle cerebral arteries from each individual. We begin with a definition of ordinary least squares, then discuss the adaptation to the residuals structure of generalised least squares.

2.2.1 | Least Squares fitting

Least squares is the default method in regression that finds the best fit of a regression line to the data. Specifically, least squares estimates some set of values for the parameters that minimise the sum of the squared deviations between the observed data and their expected values. If we consider the standard linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{6}$$

where $\mathbf{y}$ is the vector of responses or outcomes, $\mathbf{X}$ is called the design matrix and corresponds to an $n \times p$ matrix with p parameters in the regression, $\boldsymbol{\beta}$ is a vector of p model parameters, and ϵ is a vector of residuals or errors. The model parameters are estimated by ordinary-least squares (OLS [23]) as follows:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{7}$$

with covariance matrix,

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{8}$$

We make the assumption in ordinary-least squares that errors are independent and have the same variance,

$$\epsilon_{ols} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{9}$$

where σ^2 is the error variance, and $\mathbf{I}$ is the identity matrix. Therefore, the covariance matrix $\Sigma_{ols} = \sigma^2\mathbf{I}$ can be expanded,

$$\Sigma_{ols} = \sigma^2\mathbf{I} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \tag{10}$$

However, ordinary least squares can introduce bias into the estimates of the variance due to the temporal correlation between estimates being ignored. Generalised least squares adjusts the error structure by introducing an autocorrelated structure,

$$\epsilon_{gls} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\rho}) \tag{11}$$

where $\boldsymbol{\rho}$ is the correlation matrix defining the lagged correlations between errors. Hence, the resulting covariance matrix is as follows,

$$\Sigma_{\mathbf{gls}} = \sigma^2\boldsymbol{\rho} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix} \tag{12}$$

We make the assumption that the the errors are serially correlated with an autoregressive structure with one lagged period (AR(1)). Therefore, the parameter estimation can be reformulated as,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y} \tag{13}$$

The estimation of the parameters requires computationally intensive linear algebra which is substantially increased when including the huge covariance matrix due to the large number of observations. To alleviate some of the computational load, we down sampled to 5Hz. This permitted the computations to be within a pragmatic time frame while maintaining reasonable parameter accuracy when using the **gls** function from the **R** package 'nlme'[24]. Simulations documenting the comparison of etimators will be presented in full later in the article. The GLS-Rcp implementation used the **C++ Armadillo** linear algebra library via the R packages **Rcpp** and **RcppArmadillo**[25,26]. This substantially reduced computational time for the matrix algebra required to estimate the model parameters. Details of the functions can be found here: **INSERT LINK or ADD APPENDIX**.

2.3 | Extracting laterality and contrasting tasks

Calculating the measured laterality using fTCD is typically assessed by averaging over epoched CBFV data and simply taking the difference between the left and right hemisphere signals across a specific period of interest. This method has some disadvantages including: i) evaluating more complex testing paradigms is difficult or in some cases intractable; ii) accuracy and magnitude of the laterality is assessed without consideration of the uncertainty in the estimates built in; and iii) takes no account of the sensitivity of the haemodynamic response.

The GLM model can remedy these problems simply by using all available observed data rather than relying on averaged data, and building in uncertainty estimation for parameter estimates. The GLM can allow for different designs such as blocking of tasks by the addition of parameters for each task in the GLM. The GLM framework is the same as in fMRI but it differs in that its intended use is for assessing laterality (the difference in signal strength between left and right hemispheres). In fMRI, a GLM is fitted at each voxel. From each model, a single parameter estimate is extracted that related to the effect of interest. The t-statistic for this parameter is then assessed to see if it falls above or below a pre-specified t threshold. If the voxel has a t-statistic that exceeds the threshold, this is taken as a sign of significant voxel activation. The subset of t-statistics that exceed the threshold are collected together for each hemisphere and either summed or counted to give a relative proportion of activity in each hemisphere. This proportion forms the basis of the laterality index and has the benefit of highlighting specific brain regions that activate under certain tasks.

Our application of the GLM method differs somewhat as we only fit two models per individual (left and right hemispheres), so we cannot calculate the laterality index based on the number of t-statistics per hemisphere exceeding a threshold. However, we can use the substantially greater number of observations per time series to estimate the laterality using an interaction term in the model.

2.3.1 | Measuring laterality using an interaction

Suppose we have a single task experiment assessing laterality of basic motor skills. The GLM can be setup to include

$$y_i = \beta_{i,0} + \beta_{i,1}x_{Stim} + \underbrace{\beta_{i,2}x_{Signal}}_{\text{hemisphere index}} + \underbrace{\beta_{i,3}t + \beta_{i,4}t^2 + \beta_{i,5}t^3}_{\text{Drift terms}} + \underbrace{\beta_{i,6}(x_{Signal} \times x_{Stim})}_{\text{Laterality estimate}} + \epsilon_i \tag{14}$$

where i denotes individual, but does not indicate that this is a multilevel model, instead indicates separate models are fitted for each individual and $Stim$ is the individual specific HRF convolved with stimuli boxcar function as a predictor variable. We also include drift terms, similarly to fMRI GLM, which account for any trends across the repeated measures and a x_{Signal} main effect which indicates whether the measurements refer to left or right middle cerebral artery signals. Laterality is derived from the magnitude of the interaction between the x_{Stim} and x_{Signal} which gives us an estimate of the difference in x_{Stim} response between the two signals (left and right hemispheres). Lastly, we make the assumption that the error term follows an autocorrelated structure (specifically AR(1) following from equation 14), and we also assume that x_{Signal} is a fixed effect rather than random effect despite the dependence as we cannot accurately estimate the variance with only two levels[27].

The addition of the autoregressive residual structure increases the computational time considerably, and we have made some headway to improving this using more efficient computation linear algebra in **C++**. However, Mumford & Nichols[28,29] show that parameter estimates from the GLM using ordinary least squares are unbiased. Using ordinary least squares assumes an

independent error structure and reduces the computation load to a reasonable and pragmatic level. However, we must note that the estimate of the standard errors is likely to inaccurate and potentially over or underestimated.

2.3.2 | Task contrasts using interactions

The GLM framework has the added benefit that is can incorporate multiple tasks into the model and test a specific hypothesis. Using a contrast in the GLM allow us to test whether one task is more lateralised than the other. We can estimate the contrast directly by reformulating the model (see Koh et al. [30] for details).

Suppose we have a hypothesis looking to test whether one task is more lateralised than the other. Specifically, we need to test whether our interaction between Stimulus 1 (x_1) and Signal(S) is bigger than the interaction between Stimulus 2 (x_2) and Signal(S). We can write this more formally as,

$$H_0 : x_1.S = x_2.S$$
$$H_1 : x_1.S > x_2.S \tag{15}$$

If we now take the model formulation (equation 17),

$$
\begin{aligned}
y_i &= \beta_{i,0} + \beta_{i,1}x_1 + \beta_{i,2}x_2 + \beta_{i,3}S + \beta_{i,4}t + \beta_{i,5}t^2 + \beta_{i,6}t^3 + \beta_{i,7}(x_1.S) + \beta_{i,8}(x_2.S) + \epsilon_i \\
&= \beta_{i,0} + \beta_{i,1}x_1 + \beta_{i,2}x_2 + \beta_{i,3}S + \beta_{i,4}t + \beta_{i,5}t^2 + \beta_{i,6}t^3 + \beta_{i,7}(x_1.S + x2.S - x_2.S) + \beta_{i,8}(x_2.S) + \epsilon_i \\
&= \beta_{i,0} + \beta_{i,1}x_1 + \beta_{i,2}x_2 + \beta_{i,3}S + \beta_{i,4}t + \beta_{i,5}t^2 + \beta_{i,6}t^3 + \beta_{i,7}(x_1.S + x_2.S) + \underbrace{(\beta_{i,8} - \beta_{i,7})(x_2.S)}_{\text{Contrast}} + \epsilon_i
\end{aligned}
\tag{16}
$$

The contrast $(\beta_{i,8} - \beta_{i,7})$ is then estimated in the GLM as one parameter and can be interpreted in the usual way for significance to test the hypothesis (18).

2.3.3 | Parametric modulation in fTCD

** INSERT FIGURE 5 HERE**

Alternatively, we can use contrast coding when generating our stimulus variables. Changing the contrast coding for stimulus from the default 'treatment' coding can permit investigation of other experimental designs, such as parametric modulation [31]. For example, let us assume that we have a task that is not just on or off but each trial iteration has a different level or magnitude called Parametric modulation. Figure 5 shows the example stimulus trace prior to convolving with the hemodynamic response function.

The GLM can incorporate this design by adding an additional predictor to the model as a parametric modulator. The parametric modulator is assigned contrast weights to the different factor levels of the stimulus, so that when convolved together the fitted response accurately reflects the stepped design allowing for different levels within each stimulus.

$$
\begin{aligned}
y_i = \beta_{i,0} + \beta_{i,1}x_{Stim} + \beta_{i,2}x_{modulator} + \underbrace{\beta_{i,3}x_{Signal}}_{\text{hemisphere index}} + \underbrace{\beta_{i,4}t + \beta_{i,5}t^2 + \beta_{i,6}t^3}_{\text{Drift terms}} \\
+ \underbrace{\beta_{i,7}(x_{Signal} \times x_{modulator})}_{\text{difference in modulation LR}} + \epsilon_i
\end{aligned}
\tag{17}
$$

An estimate laterality is different at different levels of the stimuli can be quantified using an interaction between the parametric modulator and the hemisphere signal variable.

3 | SENSITIVITY ANALYSIS USING MONTE CARLO SIMULATION: TESTING PARAMETER ESTIMATORS

To determine the optimal model parameter estimators, we perform a sensitivity analysis via simulation to justify our choice of estimator. The observed fTCD data has a temporally-correlated error structure (i.e. a time series of sequentially collected recordings from an individual) that unless accounted for in the model estimation can, potentially, add increased uncertainty. Three

parameter estimation techniques are tested on simulated data to control the test process including: ordinary least squares (OLS) and generalised least squares (GLS; once using nlme::gls() and secondly, an implementation using RCPP and linear algebra in C++ Armadillo). The second GLS estimator's implementation was motivated by speed of evaluation on larger time series. We present results from simulated data using the single gamma and canonical (double gamma) HRFs. We do not, however, present a simulation for the fitted empirical canonical HRF method as this requires empirical data to estimate the HRF.

TABLE 2 Simulated data - linear model parameters

Parameter Name	Value
Intercept (β_0)	100
Stimulus 1 (β_1)	2.26
Stimulus 2 (β_2)	-2.18
$Time$ (β_3)	0.00002
$Time^2$ (β_4)	-0.00000004
$Time^3$ (β_5)	-0.000000005
Signal (β_6)	-0.5
$Signal * Stimulus1$ (β_7)	-0.5
lag 1 autocorrelation (ρ)	0.7

Simulated data was generated to have temporally-correlated errors and a specific linear model structure (see Table 2 for model parameters for simulation). Figure 6 shoes an example simulated data set using the Canonical HRF function. The red and green points indicated the different signals from left and right hemispheres respectively.

** INSERT FIGURE 6 HERE**

The same data was fitted using each model estimator (OLS, GLS-nlme, GLS-Rcpp) and the process was repeated for 100 iterations to provide bootstrap confidence intervals (CIs) for each set of estimates. The uncertainty for each estimator can be assessed by looking at both how closely the estimates are to the 'true' parameters (black dashed vertical line) and secondly, the magnitudes of the CIs. Figures 7 and 8 show the estimates from two simulations for data with single gamma and canonical (double gamma) respectively.

** INSERT FIGURE 7 HERE**

** INSERT FIGURE 8 HERE**

The estimators all perform relatively similarly with non-distinct differences in the parameter estimates and similar magnitudes of the bootstrap intervals. The differnce between simulations using different HRFs is negligible as expected as each simulations generates data accordsing to the target HRF for each stimuli. In practice, the difference in HRF can be assessed for model fit using a likelihood ratio test. Overall, the GLS-nlme estimator appears to have the smallest intervals on average across the parameter estimates and in both simulations using either the canonical HRF or the single gamma HRF. As a result, the GLS-nlme estimator will be used in all real data example applications in the next section.

4 | PRACTICAL EXAMPLES

4.1 | Measuring laterality

Functional transcranial doppler ultrasound can be used as a course approximate measure of neural activity in language-related areas of the brain via the blood flow in the right and left hemispheres' MCAs. In this example application of our implementation of the GLM method for fTCD data, we use data from Bruckett *et al.*[32] which investigated whether FTCD ultrasonography could reliably assess functional lateralization across different language processes. The data are re-analysed using the GLM based method and laterality estimates from fMRI, traditional fTCD analysis and new fTCD GLM analysis techniques are compared using correlation.

4.2 | Using contrasts to assess laterality between tests

In this example, we demonstrate the use of constrasts in the GLM method for fTCD data. The GLM contrast facilitates the comparison of magnitude of laterality between experimental stimuli (different tests). We use data from Woodhead *et al.*[33] as they present three tasks: word generation (WG), sentence generation (SG) and list generation (LG). The experimental design of that study provides an ideal example application of the GLM contrasts method to assess laterality as the tasks are interleaved in a pre-determined pseudorandomised order.

The contrast tests the hypothesis that the sentence Generation task is more lateralised than the list generation task. Laterality in each task is estimated as the difference in stimulus in left vs right signal calculated via the individual interactions, then the interactions are contrasted.

4.3 | Parametric modulation

HELP ZOE, need something sensible here about how Holly's data fits with this paradigm.

5 | DISCUSSION

6 | CONCLUSIONS

Potential benefits of employing a GLM analysis in fTCD:

- More flexible options for study design, e.g. event related designs or parametric modulators

- The possibility of performing contrasts between tasks, e.g. to use active rather than resting baselines

- Biologically plausible modelling of the neurovascular response

- More direct comparability to fMRI analysis methods

ACKNOWLEDGEMENT(S)

DISCLOSURE STATEMENT

No competing interests were disclosed.

FUNDING

References

1. Badcock N. A., Groen M. A.. What can functional Transcranial Doppler Ultrasonography tell us about spoken language understanding?. *Language, Cognition and Neuroscience.* 2017;32(7):818–828.

2. Bishop Dorothy V. M., Badcock Nicholas A, Holt Georgina. Assessment of cerebral lateralization in children using functional transcranial Doppler ultrasound (fTCD).. *Journal of Visualized Experiments : JoVE.* 2010;(43):e2161.

3. Deppe M, Ringelstein E. Bernd, Knecht Stefan. The investigation of functional brain lateralization by transcranial Doppler sonography. *NeuroImage.* 2004;21(3):1124–1146.

4. Deppe M, Knecht Stefan, Lohmann Hubertus, Ringelstein E Bernd. A method for the automated assessment of temporal characteristics of functional hemispheric lateralization by transcranial Doppler sonography. *Journal of Neuroimaging.* 2004;14(3):226–30.

5. Lupetin A R, Davis D A, Beckman I, Dash N. Transcranial Doppler sonography. Part 1. Principles, technique, and normal appearances.. *Radiographics.* 1995;15(1):179–191.

6. Kim Dong-Eog, Park Jong-Ho, Schellingerhout Dawid, et al. Mapping the Supratentorial Cerebral Arterial Territories Using 1160 Large Artery Infarcts. *JAMA Neurology.* 2019;76(1):72.

7. Buxton Richard B., Wong Eric C., Frank Lawrence R.. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine.* 1998;39(6):855–864.

8. Gagnon Louis, Sakadžić Sava, Lesage Frédéric, et al. Quantifying the microvascular origin of bold-fMRI from first principles with two-photon microscopy and an oxygen-sensitive nanoprobe. *Journal of Neuroscience.* 2015;35(8):3663–3675.

9. Boynton Geoffrey M., Engel Stephen A., Glover Gary H., Heeger David J.. Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience.* 1996;16(13):4207–4221.

10. Glover Gary H.. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage.* 1999;9(4):416–429.

11. Friston K. J., Fletcher P., Josephs O., Holmes A., Rugg M. D., Turner R.. Event-related fMRI: Characterizing differential responses. *NeuroImage.* 1998;7(1):30–40.

12. Chapter 40 - Analysis of fMRI Time Series: Linear Time-Invariant Models, Event-Related fMRI, and Optimal Experimental Design. In: Frackowiak Richard S.J., Friston Karl J., Frith Christopher D., et al. , eds. *Human Brain Function (Second Edition)*, Burlington: Academic Press second edition ed.2004 (pp. 793 - 822).

13. Aaslid Rune. Visually evoked dynamic blood flow response of the human cerebral circulation. *Stroke.* 1987;18(4):771–775.

14. Conrad B., Klingelhofer J.. Dynamics of regional cerebral blood flow for various visual stimuli. *Experimental Brain Research.* 1989;77(2):437–441.

15. McMahon C. J., McDermott P., Horsfall D., Selvarajah J. R., King A. T., Vail A.. The reproducibility of transcranial Doppler middle cerebral artery velocity measurements: Implications for clinical practice. *British Journal of Neurosurgery.* 2007;21(1):21–27.

16. Binder Jeffrey R., Swanson Sara J., Hammeke Thomas A., Sabsevitz David S.. A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia.* 2008;49(12):1980–1997.

17. Stark C E, Squire Larry R. When zero is not zero: the problem of ambiguous baseline conditions in fMRI.. *Proceedings of the National Academy of Sciences of the United States of America.* 2001;98(22):12760–12766.

18. Worsley K.J., Liao C.H., Aston J., et al. A General Statistical Analysis for fMRI Data. *NeuroImage.* 2002;15(1):1 - 15.

19. Proulx Sébastien, Safi-Harb Mouna, LeVan Pierre, An Dongmei, Watanabe Satsuki, Gotman Jean. Increased sensitivity of fast BOLD fMRI with a subject-specific hemodynamic response function and application to epilepsy. *NeuroImage.* 2014;93:59 - 73.

20. Lindquist Martin A., Loh Ji Meng, Atlas Lauren Y., Wager Tor D.. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage.* 2009;45(1, Supplement 1):S187 - S198. Mathematics in Brain Imaging.

21. Lindquist Martin A.. The Statistical Analysis of fMRI Data. *Statistical Science.* 2008;23(4):439–464.

22. Tabelow Karsten, Polzehl Jörg. Statistical Parametric Maps for Functional MRI Experiments in R: The Package fmri. *Journal of Statistical Software.* 2011;44(11):1–21.

23. Draper N.R., Smith H.. *Applied regression analysis*. No. v. 1 in Wiley series in probability and statistics: Texts and references sectionWiley; 1998.

24. Pinheiro Jose, Bates Douglas, DebRoy Saikat, Sarkar Deepayan, R Core Team . nlme: Linear and Nonlinear Mixed Effects Models2019. R package version 3.1-140.

25. Eddelbuettel Dirk, Balamuta James Joseph. Extending R with C++: A Brief Introduction to Rcpp. *The American Statistician.* 2018;72(1):28-36.

26. Eddelbuettel Dirk, Sanderson Conrad. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics & Data Analysis.* 2014;71:1054 - 1063.

27. Gelman Andrew, Hill Jennifer. *Data analysis using regression and multilevel hierarchical models*. New York: Cambridge University Press; 2007.

28. Mumford J. A., Nichols T.. Modeling and inference of multisubject fMRI data. *IEEE Engineering in Medicine and Biology Magazine.* 2006;25(2):42-51.

29. Mumford Jeanette A., Nichols Thomas. Simple group fMRI modeling and inference. *NeuroImage.* 2009;47(4):1469 - 1475.

30. Koh Natalie, Lee Adel, McLaughlin PhD, Grabowski Thomas J, Madhyastha Tara M. *A Tutorial on Modeling fMRI Data using a General Linear Model*. 2017.

31. Mumford Jeanette A., Poline Jean-Baptiste, Poldrack Russell A.. Orthogonalization of Regressors in fMRI Models. *PLOS ONE.* 2015;10(4):1-11.

32. Bruckert Lisa. Is language laterality related to language abilities ? ThesisUniversity of Oxford2016.

33. Woodhead Z.V.J., Rutherford H.A., Bishop D.V.M.. Measurement of language laterality using functional transcranial Doppler ultrasound: a comparison of different tasks. *Wellcome Open Res 2018.* 2018;3(104).
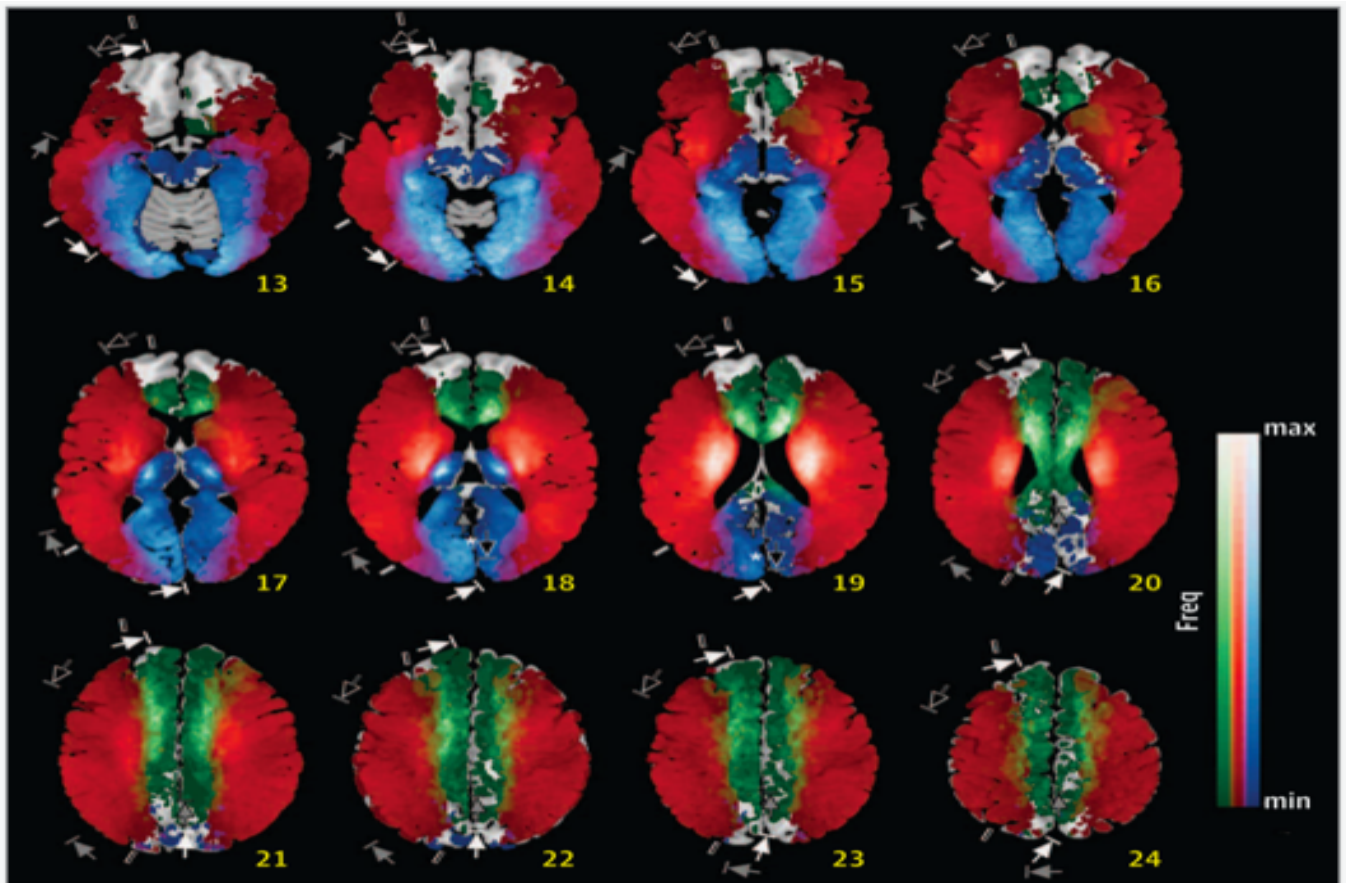
FIGURE 1 Approximation of the MCA territory (red) based on lesion extents in 1160 acute stroke patients. Reproduced from [6].
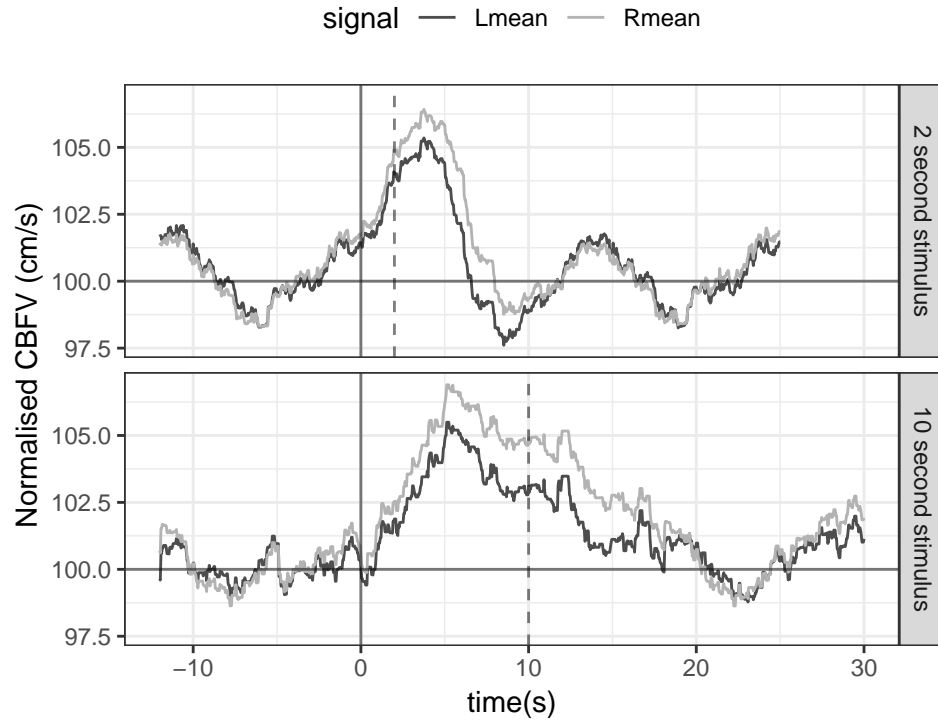
FIGURE 2 Example fTCD responses from the left hemisphere (dark grey) and the right hemisphere (light grey) to a 2 second (top) and 10 second (bottom) movement task using the left hand (N=1, averaged over multiple trials).
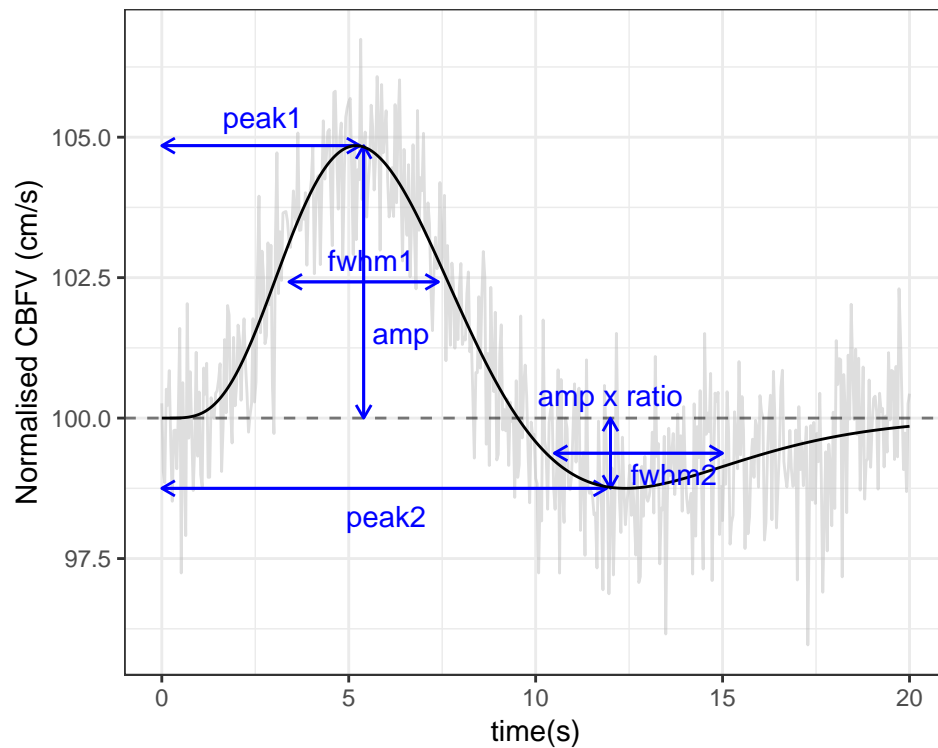

FIGURE 3 Typical haemodynamic response function with illustrations of the parameter relationships. Grey line indicates the raw data, the bold black line indicates the modelled response using the canonical haemodynamic response function, and the grey dashed line is a reference baseline to peaks 1 and 2.
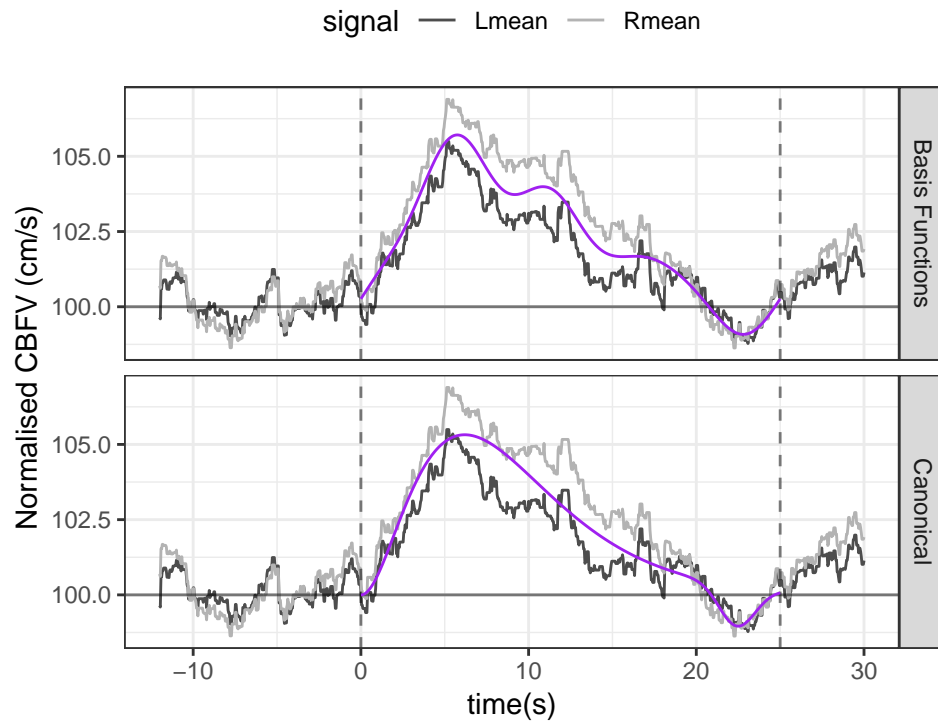
FIGURE 4 Example haemodynamic response data from one individual for a sustained response task from the left hemisphere (dark grey) and the right hemisphere (light grey). The purple lines in each panel are the model fits using canonical HRF (bottom) and basis functions (top). The grey, vertical dashed lines indicate the start and end of the task stimulus.
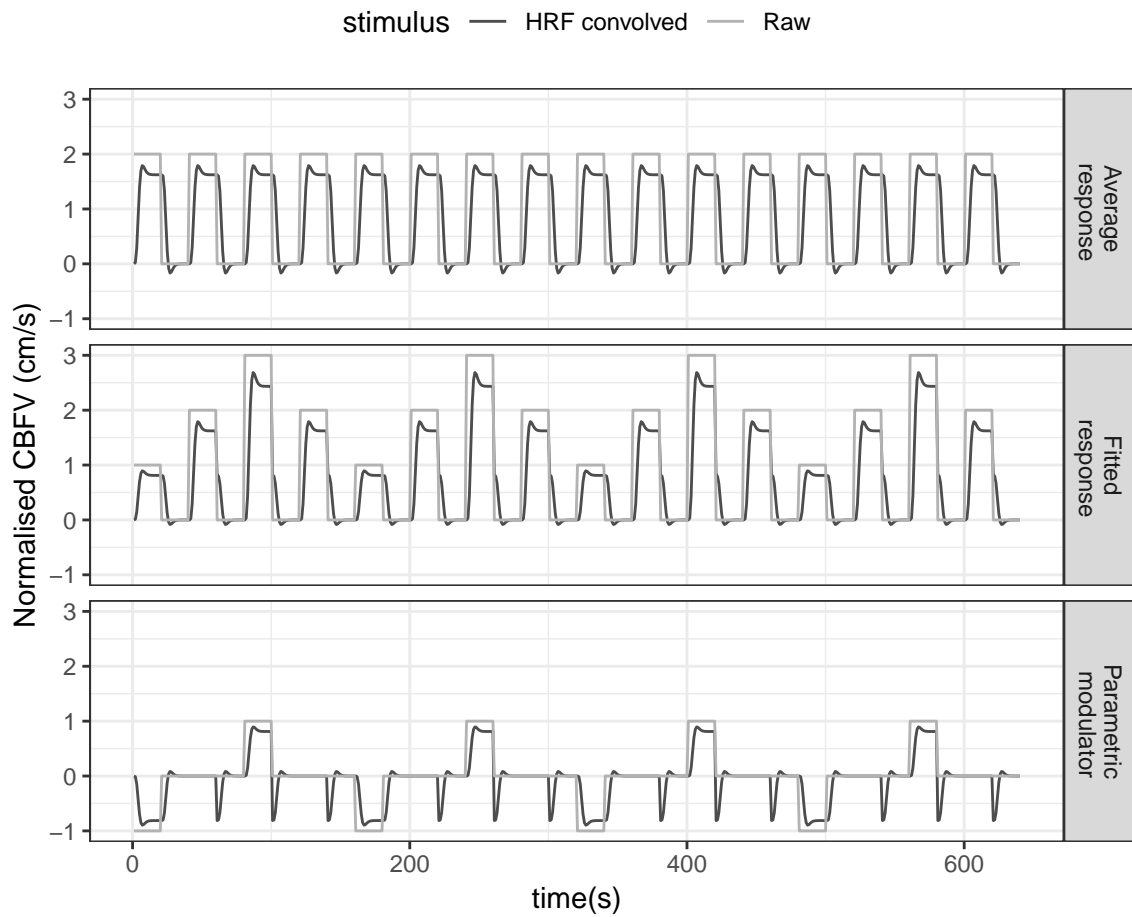
FIGURE 5 Example of a stimulus trace using parametric modulation. The grey and black lines are the raw and convolved stimuli signals for the average response, parametric modular and the combined stimuli signal.

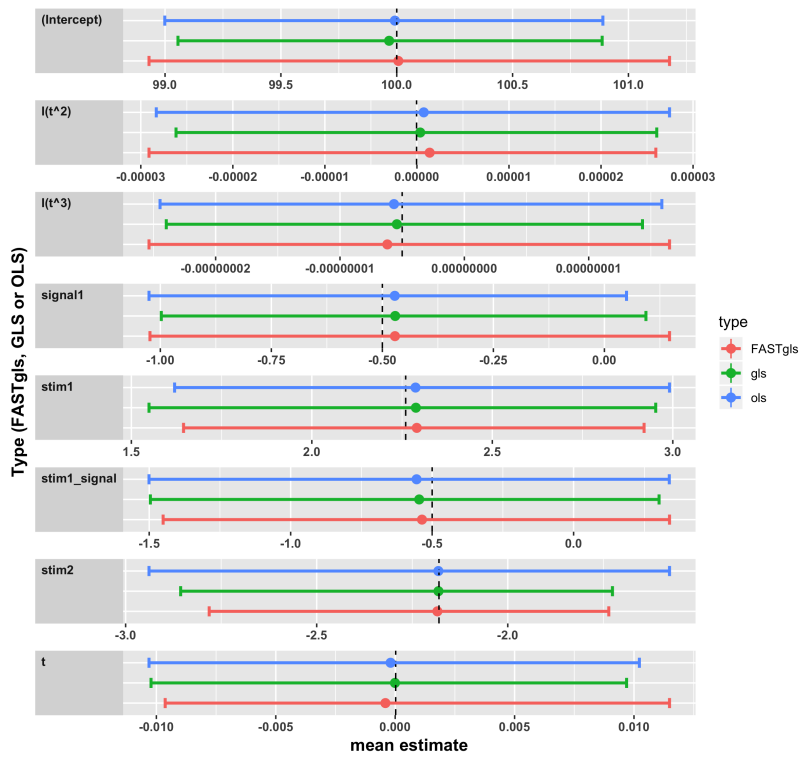**FIGURE 6** Figure showing an example of the Canonical HRF simulated data
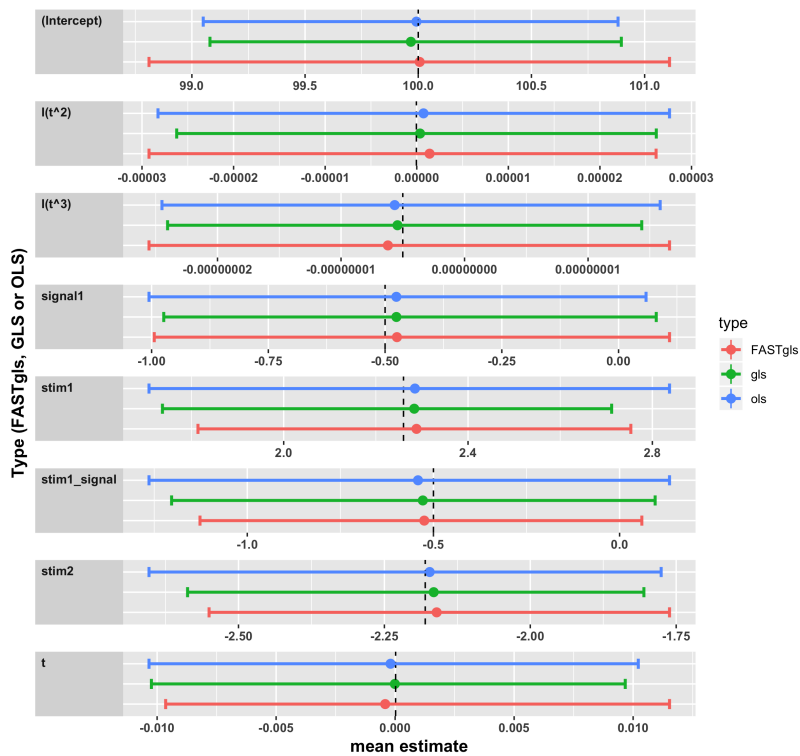
FIGURE 7 Single Gamma HRF to generate stimulus

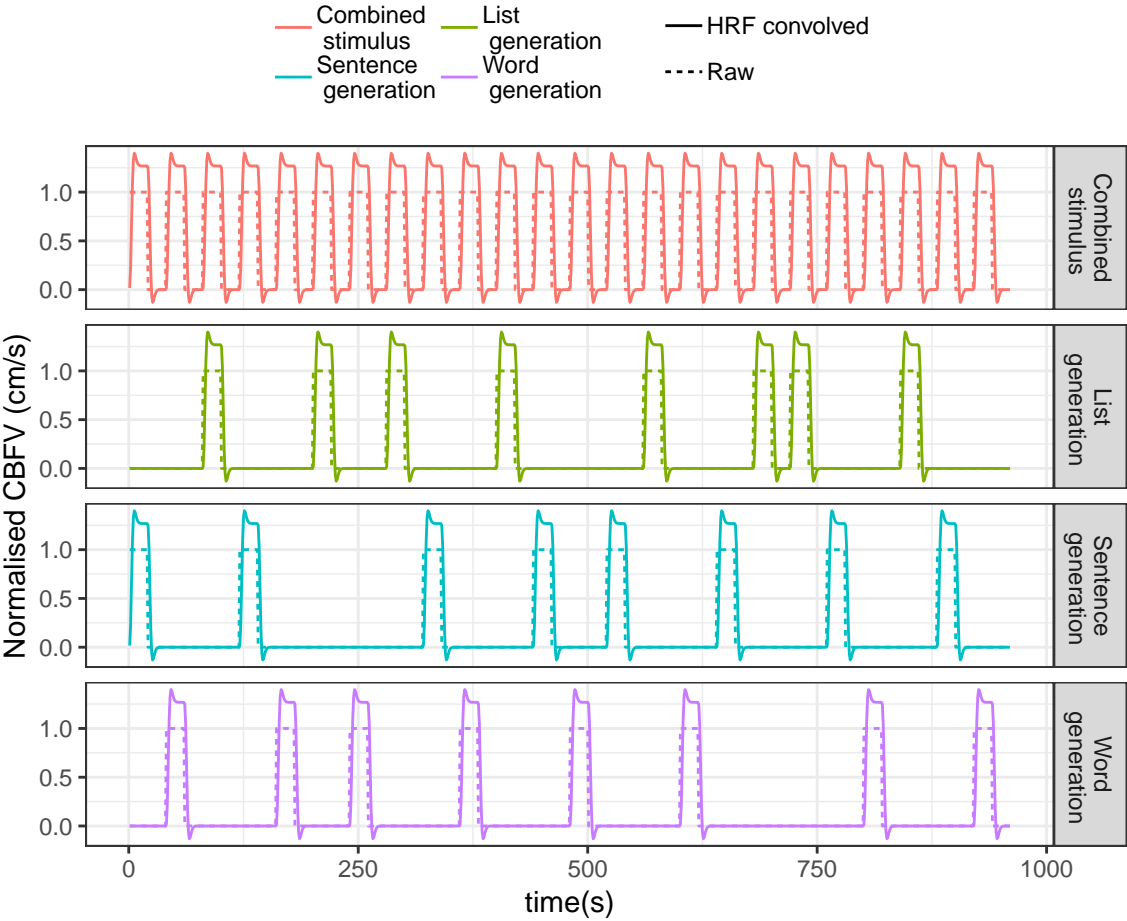**FIGURE 8** Canonical HRF to generate stimulus

FIGURE 9 Example of stimulus trace for each task to demonstrate task contrasts.