IMAGEN

fMRI data analysis : methods notes

Revision 2. Jul. 2010

Content

1. Introduction

1. Foreword

This document describes the decisions we made for the intra-subjects preprocessings and first level analysis of all the IMAGEN fMRI datasets performed at Neurospin and available in the central database. The four functional tasks are Global Cognitive Assessment (GCA), Modified Incentive Delay (MID), Stop Signal Task (SST), and Face Task (FT). This report reflects the choices made for the modelling of the fMRI protocols and therefore is the results of the work or input of many individual. We would like to acknowledge in particular Eva Loth, Mira Buehler, Frauke Nees, Uli Bromberg, Christian Buechel, Hugh Garavan, and many others.

1.2. Information common to all fMRI Protocols

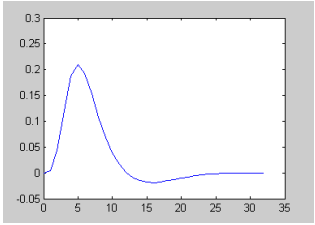
All fMRI analyses share the following framework:

fMRI pre-processings

The pre-processing of the EPI data are done within SPM8 (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm/). Time series data are first **corrected for slice-timing**, then corrected for movement (spatial **realignment**), non-linearly warped on the **MNI space** (using a **custom EPI template**), and gaussian-**smoothed at 5mm-FWHM.**

fMRI first level analyses

Activation maps are computed with SPM8, and regressed using a general linear model (**GLM**) with **AR** noise model (spm default) against a design-matrix built from the informations contained in the Imagen Behavioural Files (csv files). **Estimated movement** was added to the design matrix in the form of 18 additional columns (3 translations, 3 rotations, 3 quadratic and 3 cubic translations, 3 translations shifted 1 TR before, and 3 translations shifted 1 TR later). The regressors modeling the experimental conditions are convolved using **SPM's default HRF** (Hemodynamic Response Function). The estimated model ("beta") parameters maps are linearly-combined to yield contrasts maps and significance maps, hereafter respectively called "**con maps**" and "**spmT maps**", while the residual variance of the model fit is stored as an additional map.



SPM default hrf

Additional Informations

See also the document titled "Processing FAQ" for complementary technical informations about the preprocessings.

Although the data were computed with the SPM software, it was a major design goal to make every files useable in any context, which means at no point should the SPM software itself be required to get access to major parameters. If some information seems to be missing, please contact imagendatabase@cea.fr (before contacting directly Benjamin Thyreau (benjamin.thyreau@cea.fr) or Jean-Baptiste Poline (jbpoline@gmail.com)).

Naming scheme note:

Among a single protocol, we tried to standardize file sets and results across subjects, e.g. the contrast map file named `con_OONN.nii.gz` always refers to the same contrast across subjects. It is however always advisable to have a quick look to every files before processing to an analysis.

Currently the following potential caveat are to be carefully checked:

- Some contrasts for some subjects are not estimable ; there can be several causes for that, but the most common case is that a condition is never present/recorded for some subjects, for instance, a Left-Right motor contrasts concerning an acquisition run whose push-button mechanically failed. Those **non estimable** contrasts contains the string "`unestimable`" in the "description" field of the Nifti image header. That information should also be available directly from the Imagen database frontend.
- Although efforts were made to manually recover from some common errors, subjects datasets which failed to be processed for **any** reason may simply not be available in the database. Reasons may include, but not limited to: bad or incomplete fMRI runs, questionable values in the behavioural-csv files, algorithmic failure of some previous step (e.g. preprocessing, dicom file-conversion), missing runs, etc.

The following is task-specific references notes.

2. GCA – Global Cognitive Assessment

In this Global Cognitive Assessment tasks, the subjects undergo a set of stimuli intended to highlight some specific brain areas. Events are modeled as zero duration Dirac, convolved with a standard Hemodynamic Response Function.

Except for the recorded motor responses, all other design-matrix columns should be fairly similar between subjects.

2.1. Behavioural-files input available:

4D fMRI sequence of 140 volumes, rtime 2.2s.
Behavioural files (cga_*.csv) first few lines are pasted below:

```
GLOBAL_COGNITIVE_ASSESSMENT_TASK task   01.01.1900 19:46:55     Subject ID: 000014185243        Task type:
Scanning
Trial   Trial Category  Trial Start Time (Onset)        Pre-determined onset    Stimulus Presented
Stimulus presentation time      Response made by subject        Pre-determined Jitter   Time response made
Scanner Pulse
1       VISUAL_MATHS    0       0       Rechne elf minus drei   10      B B B C C B C
CGLOBAL_COGNITIVE_ASSESSMENT_TASK task    01/01/00 19:46  Subject ID: 000014185243        Task type:
Scanning
Trial   Trial Category  Trial Start Time (Onset)        Pre-determined onset    Stimulus Presented
Stimulus presentation time      Response made by subject        Pre-determined Jitter   Time response made
Scanner Pulse
1       VISUAL_MATHS    0       0       Rechne elf minus drei   10      B B B C C       200     0       2187
2       VISUAL_MATHS    2853    2400    Rechne sechzehn minus zwei      2860            1100    0       4390
3       REST    5700    5700                            800     0       6581
4       HORIZONTAL_CHECKERBOARD 8700    8700            8704            500     0       11000
5       PRESS_RIGHT_AUDITORY    11400   11400   click3Right.wav 11413           1400    0       17989
6       LISTEN_SENTENCE 17989   15000   3.wav   18002           800     0       17989
7       PRESS_RIGHT_VISUAL      18020   18000   Druecke dreimal die rechte Taste        18028   C C C   500
20813   19798
```

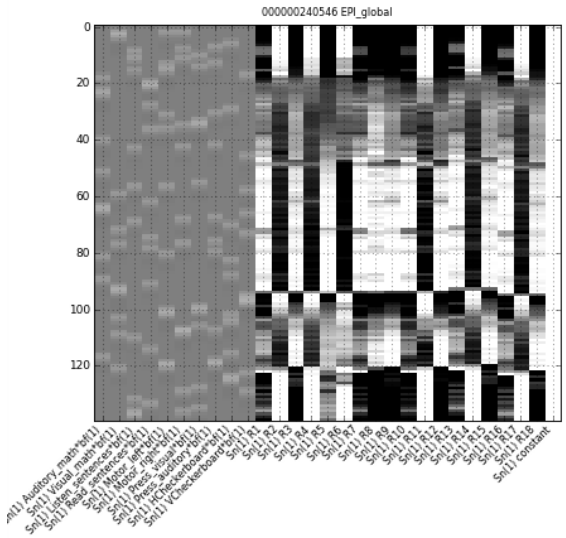
Note : Time column may restart from 0, in this case everything before is assumed to be pre-tests and discarded. This also applies to other runs.

2.2. Model

The model is straightforward, using one regressor per condition type. The events starting time are the one corresponding to the Stimulus Presentation, except for motor events which have explicit recorded values.

```
StimTime = Stimulus_presentation_time / 1000.
ResponseTime = Time_response_made / 1000.

Auditory_math = StimTime[Trial_Category == 'AUDITORY_MATHS']
Visual_math = StimTime[Trial_Category == 'VISUAL_MATHS']
Listen_sentences = StimTime[Trial_Category == 'LISTEN_SENTENCE']
Read_sentences = StimTime[Trial_Category == 'READ_SENTENCE']
Motor_left = ResponseTime[left_press]
Motor_right = ResponseTime[right_press]
Press_visual = StimTime[(Trial_Category == 'PRESS_LEFT_VISUAL') | (Trial_Category == 'PRESS_RIGHT_VISUAL')]
Press_auditory = StimTime[(Trial_Category=='PRESS_LEFT_AUDITORY') |
(Trial_Category=='PRESS_RIGHT_AUDITORY')]
HCheckerboard = StimTime[(Trial_Category == 'HORIZONTAL_CHECKERBOARD')]
VCheckerboard = StimTime[(Trial_Category == 'VERTICAL_CHECKERBOARD')]
```

In case an event is not estimable (no occurrence of the condition), the contrast is replaced by a dummy file to maintain correct file ordering, and the header's “description” field is updated to “unestimable”, as described in the overview section before.



2.3. Contrasts

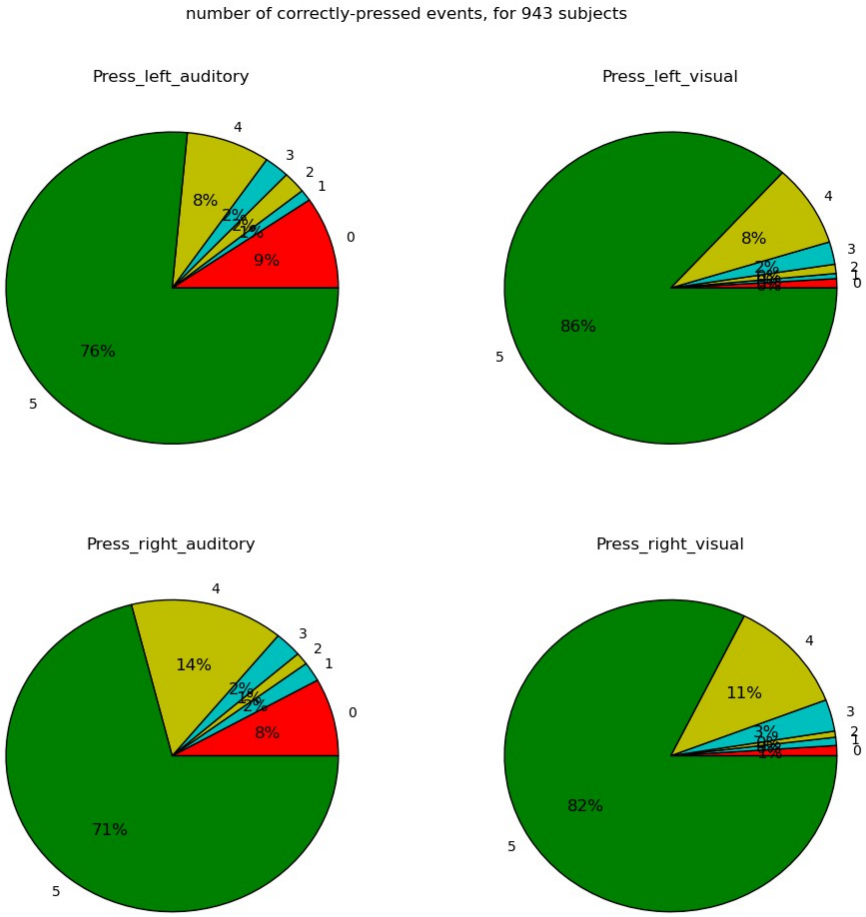
The following contrast maps are currently available:

1. 'Auditory_math'
2. 'Visual_math'
3. 'Listen_sentences'
4. 'Read_sentences'
5. 'Motor_left'
6. 'Motor_right'
7. 'Press_visual'
8. 'Press_auditory'
9. 'HCheckerboard'
10. 'VCheckerboard'
11. 'Auditory' *('Auditory_math' + 'Listen_sentences' + 'Press_auditory')*
12. 'Visual' *('Visual_math' + 'Read_sentences' + 'Press_visual' + 'HCheckerboard' + 'VCheckerboard')*
13. 'Auditory - Visual' *('Auditory_math' + 'Listen_sentences' + 'Press_auditory' - 'Visual_math' - 'Read_sentences' - 'Press_visual' - 'HCheckerboard' - 'VCheckerboard')*
14. 'Visual - Auditory' *(- 'Auditory - Visual')*
15. 'Motor L - R' *('Motor_left' - 'Motor_right')*
16. 'Motor R - L' *(- 'Motor L - R')*
17. 'Motor L + R' *('Motor_left' + 'Motor_right')*
18. 'Computation' *('Auditory_math' + 'Visual_math')*
19. 'Sentences' *('Listen_sentences' + 'Read_sentences')*
20. 'Sentences - Computation' *('Listen_sentences' + 'Read_sentences' - 'Auditory_math' - 'Visual_math')*
21. 'Computation - Sentences' *(- 'Sentences - Computation')*
22. 'AudioComputation - AudioSentences' *('Auditory_math' - 'Listen_sentences')*
23. 'VisualComputation - VisualSentences' *('Visual_math' - 'Read_sentences')*
24. 'Motor - Cognitive' *('Motor L + R' - 'Sentences' - "'Computation')*
25. 'Cognitive - Motor' *(- 'Motor - Cognitive')*
26. 'Checkerboard H + V' *('HCheckerboard' + 'VCheckerboard')*
27. 'Checkerboard H - V' *('HCheckerboard' - 'VCheckerboard')*
28. 'Checkerboard V - H' *(- 'Checkerboard H - V')*
29. 'Visual - Checkerboard' *('Visual_math' + 'Read_sentences' + 'Press_visual' - 'Checkerboard H + V')*
30. 'Read - Checkerboards' *('Read_sentences' - 'HCheckerboard' - 'VCheckerboard')*
31. 'Read - HCheckerboards' *('Read_sentences' - 'HCheckerboard')*

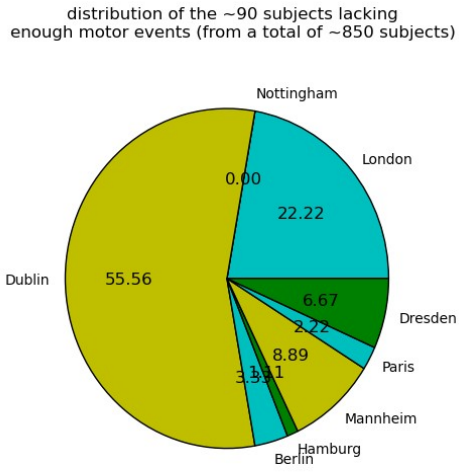
2.4. Comments

The only potentially unestimable regressors are the motor responses. The pie chart below shows the number of recorded motor responses during the task, across ~943 subjects, for the four conditions involving motor condition. The corresponding regressor gets unestimable if 0 event occurs, and prevent computation of all contrast maps which depend on that regressor. (i.e. not limited to “pure motor” only ; e.g. a “Visual” map involves *Press_visual* which is a motor event). The pie-chart colors highlight the best cases (Green, meaning all 5 events were recorded) and worse cases (Red, meaning 0 events have been recorded) proportions. As it shows, auditory motor

stimuli seems to contain more failure cases than Visual motor ones.



Subjects missing enough recorded motor events to fully estimate the model seems to be unevenly distributed per acquisition centre. We're in the process of investigating it, but it might reflect bad auditory device in the most affected centres.



2.5. subject-specific notes or workaround

Manually-fixed csv:

000022053782
000054552397
000079848243
000086071132

Empty csv:

000027074970
000012809392
000078056005

Not yet fixed:

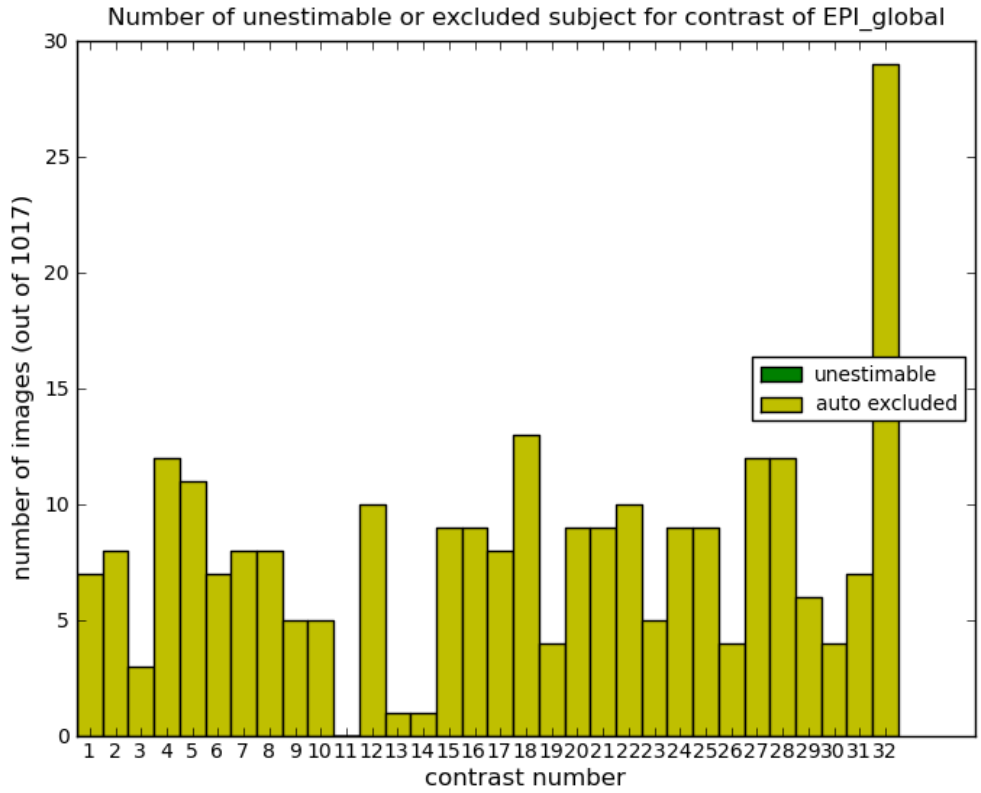
000083358101

Response Time is always 0 :

000099677574
000031171219
000087296615

2.6. Unestimable or Excluded contrast maps

A Group-wide distribution (computed on 1017 subject) of unestimable contrasts or excluded subjects. *Unestimable* covers contrasts whose lack of events prevented the estimation of the corresponding parameter value, while *excluded* covers images which were dropped due to a too abnormal activation profile computed with the automatic QC procedure.



(Note that only contrast 1-31 are relevant on this picture – contrast 32 refers to a testing case)

3. SST – Stop and Signal Task (SST)

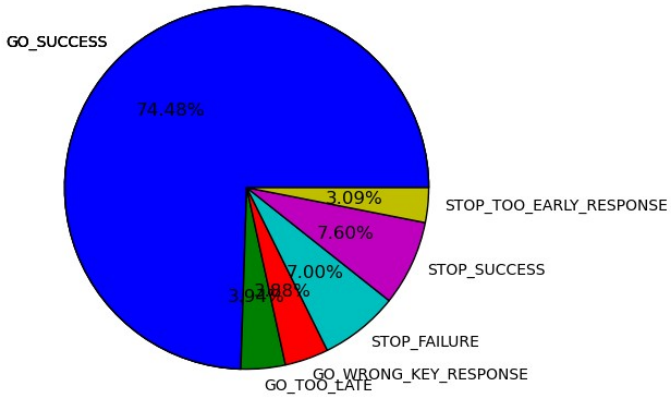
This protocols is derived from a Go/No-go response inhibition paradigm, measuring motor responses speed after stimuli with varying jitter time. STOP signal occurs on average 20% of the time.

Data : 4D fMRI sequence of 444 volumes, rtime 2.2s.
This protocols has 480 events. SPM's Standard Hemodynamic Function is used.

The csv relevant columns includes “Trial Category”, “Stimulus Presented” and “Outcome”, with related timings. All the possible combination found (reviewing ~900 subjects csv files) are:

```
GO      LeftArrow       GO_SUCCESS
GO      LeftArrow       GO_TOO_LATE
GO      LeftArrow       GO_WRONG_KEY_RESPONSE
STOP_VAR        LeftArrow       STOP_FAILURE
STOP_VAR        LeftArrow       STOP_SUCCESS
STOP_VAR        LeftArrow       STOP_TOO_EARLY_RESPONSE
GO      RightArrow      GO_SUCCESS
GO      RightArrow      GO_TOO_LATE
GO      RightArrow      GO_WRONG_KEY_RESPONSE
STOP_VAR        RightArrow      STOP_FAILURE
STOP_VAR        RightArrow      STOP_SUCCESS
STOP_VAR        RightArrow      STOP_TOO_EARLY_RESPONSE
```

The distribution of those 480 events across the Behavioural files of about ~900 subjects from all centres are depicted on the pie chart below.

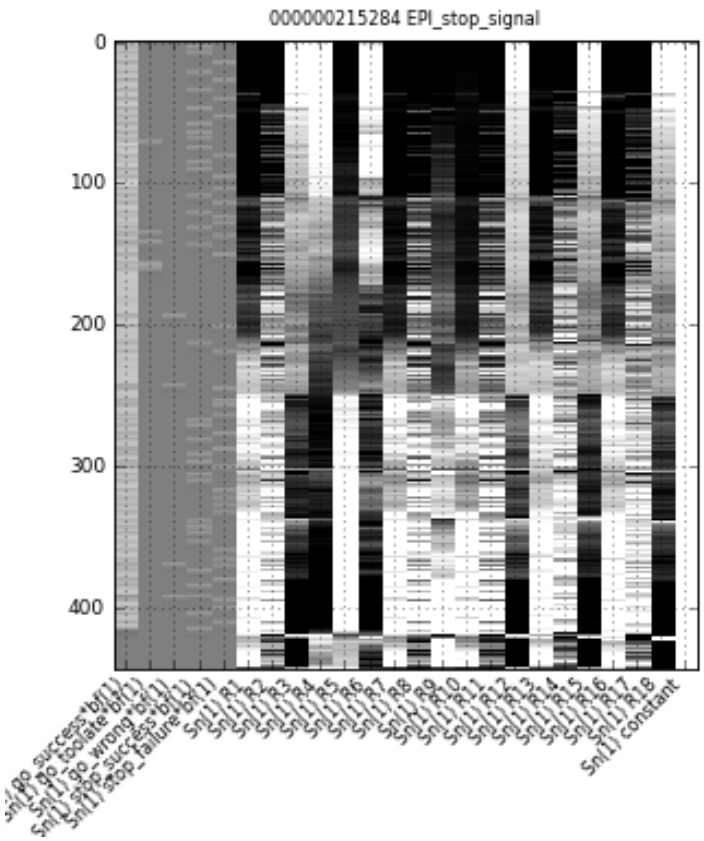


As in other protocols csv files, time column values for some subjects may reset from 0 at some point, in this case everything before is assumed to be pre-tests and discarded.

3.1. Model

With GoTime = Go_Stimulus_Presentation_Time / 1000.

```
go_success = GoTime[(Response_Outcome=='GO_SUCCESS') | (Response_Outcome=='STOP_TOO_EARLY_RESPONSE')]
go_toolate = GoTime[Response_Outcome == 'GO_TOO_LATE']
go_wrong = GoTime[Response_Outcome == 'GO_WRONG_KEY_RESPONSE']
stop_success = GoTime[Response_Outcome == 'STOP_SUCCESS']
stop_failure = GoTime[Response_Outcome == 'STOP_FAILURE']

namelist = ["go_success", "go_toolate", "go_wrong", "stop_success", "stop_failure"]
```



3.2. Contrasts (model version 1.0)

The following contrast maps are available:

1. 'go_success'
2. 'go_toolate'
3. 'go_wrong'
4. 'stop_success'
5. 'stop_failure'
6. 'stop_success - go_success' *('stop_success' - 'go_success')*
7. 'go_success - stop_success' *(- 'stop_success - go_success')*
8. 'stop_success - stop_failure' *('stop_success' - 'stop_failure')*
9. 'stop_failure - stop_success' *(- 'stop_success - stop_failure')*
10. 'go_success - stop_failure' *('go_success' - 'stop_failure')*
11. 'stop_failure - go_success' *(- 'go_success - stop_failure')*
12. 'go_wrong - go_success' *('go_wrong' - 'go_success')*
13. 'go_success - go_wrong' *(- 'go_wrong - go_success'")*

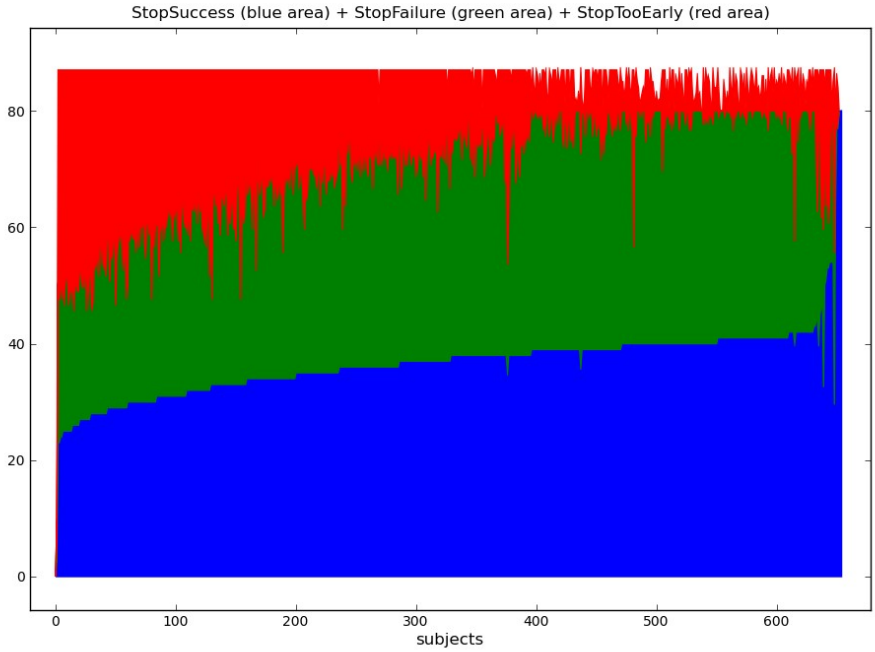
3.3. Comments

Although not fully necessary, we modeled the “Go Success” condition explicitly in the hope to improve estimation accuracy and to have an explicit baseline.

A stop condition for which the “stop” stimulus hasn't been shown yet is considered the same as the “go success” condition in the model.

No motor responses are modeled since the timings involved wouldn't yield a regressor orthogonal enough to the others stimuli.

Repartition of the different outcome of Stop condition, for ~650 Behavioural files :

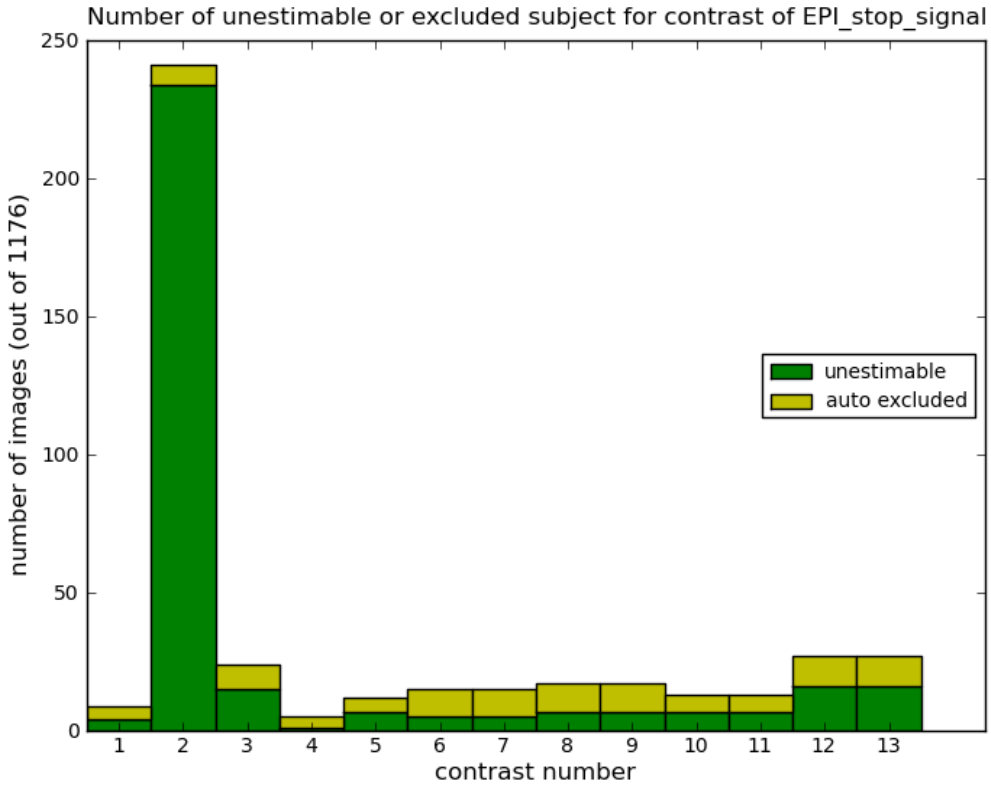


3.4. subject-specific notes or workaround

```
empty files:
ss_000008633229.csv
ss_000061630843.csv
ss_000049850700.csv

Non-480 events:
ss_000049850700.csv has 4 events
ss_000018750014.csv has 354 events
ss_000027718055.csv has 472 events

comma manually changed to tabulations:
ss_000083358101.csv

manually fixed bad quoting:
ss_000093240591.csv
ss_000033775752.csv
ss_000075438006.csv:

incorrect image files dimensions (Standard is (64, 64, 40, 444))
000023079648 has dim (64, 64, 40, 438)
000075465686 has dim (64, 64, 40, 403)
000099550415 has dim (64, 64, 40, 126)
000083308215 has dim (64, 64, 40, 34)

000023431130 has dim (64, 64, 11688)
000063770152 has dim (64, 64, 14588)
000083358101 has dim (64, 64, 36, 444)
```

3.5. Unestimable or Excluded contrast maps

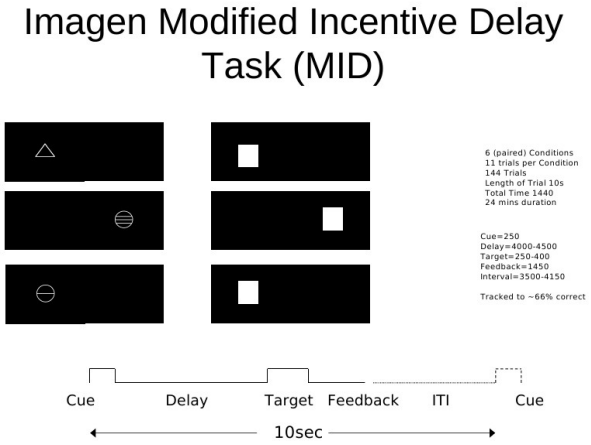
A Group-wide distribution (computed on 1176 subject) of unestimable contrasts or excluded subjects. *Unestimable* covers contrasts whose lack of events prevented the estimation of the corresponding parameter value, while *excluded* covers images which were dropped due to a too abnormal activation profile, computed with the automatic QC procedure.



Note that contrast 2 models an error condition (*GO TOO_LATE*) which might not occur during a normal run, thus explaining the relatively high number of unestimable images for this contrast.

4. MID – Modified Incentive Delay

The Modified Incentive Delay task shows the subjects sequences of clues, target, feedback phase (REF). The clue indicates the amount of the gain and the subjects are instructed to respond when the Target shows, roughly ~4 sec after the clue, after which a ~1.5sec feedback messages tells the subject about the win or loss of the trial.

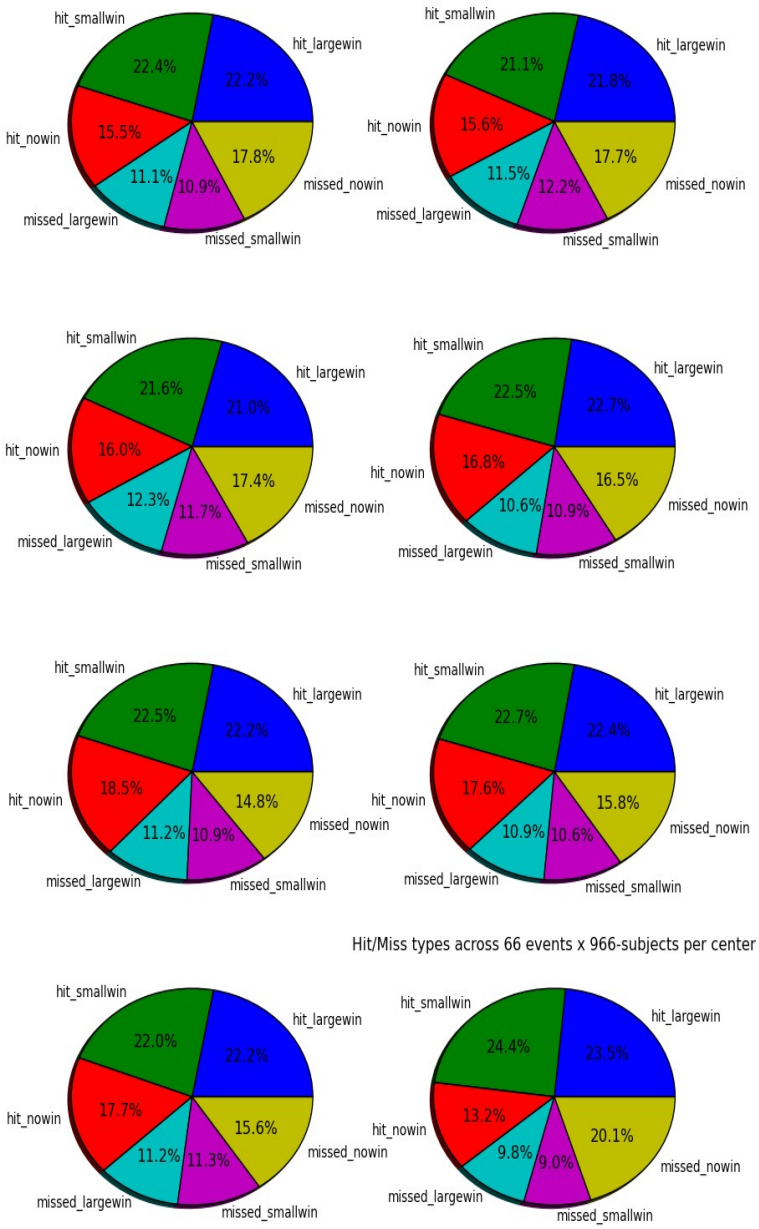


Data : 4D fMRI sequence of 300 volumes, rtime 2.2s.
Columns of interest from the behavioural csv files are *Trial_Category*, *Response_Made_by_Subject*, and *Outcome* which combines the following ways across ~900 subjects:

```
BIG_WIN Left    FAILURE
BIG_WIN Left    SUCCESS
BIG_WIN NO      RESPONSE        FAILURE
BIG_WIN Right   FAILURE
BIG_WIN Right   SUCCESS
BIG_WIN TOO_EARLY:Left  FAILURE
BIG_WIN TOO_EARLY:NO    RESPONSE        FAILURE
BIG_WIN TOO_EARLY:Right FAILURE
BIG_WIN TOO_LATE:Left   FAILURE
BIG_WIN TOO_LATE:NO     RESPONSE        FAILURE
BIG_WIN TOO_LATE:Right  FAILURE
NO_WIN  Left    FAILURE
NO_WIN  Left    SUCCESS
NO_WIN  NO      RESPONSE        FAILURE
NO_WIN  Right   FAILURE
NO_WIN  Right   SUCCESS
NO_WIN  TOO_EARLY:Left  FAILURE
NO_WIN  TOO_EARLY:NO    RESPONSE        FAILURE
NO_WIN  TOO_EARLY:Right FAILURE
NO_WIN  TOO_LATE:Left   FAILURE
NO_WIN  TOO_LATE:NO     RESPONSE        FAILURE
NO_WIN  TOO_LATE:Right  FAILURE
SMALL_WIN       Left    FAILURE
SMALL_WIN       Left    SUCCESS
SMALL_WIN       NO      RESPONSE        FAILURE
SMALL_WIN       Right   FAILURE
SMALL_WIN       Right   SUCCESS
SMALL_WIN       TOO_EARLY:Left  FAILURE
SMALL_WIN       TOO_EARLY:NO    RESPONSE        FAILURE
SMALL_WIN       TOO_EARLY:Right FAILURE
SMALL_WIN       TOO_LATE:Left   FAILURE
SMALL_WIN       TOO_LATE:NO     RESPONSE        FAILURE
SMALL_WIN       TOO_LATE:Right  FAILURE
```

Other important fields are available, such as response times, but they are not taken into account in our model of the BOLD responses.

Overview of the responses:



The subjects outputs were roughly distributed the same across centers, with “hit” occuring slightly more than “missed” (ie. SUCCESS : 60%, FAILURE : 40%).(left-right then up-down: London, Nottingham, Dublin, Berlin, Hamburg, Mannheim, Paris, Dresden, see the pie chart above)

As in other protocols csv files, time column values for some subjects may reset from 0 at some point, in this case everything before is assumed to be pre-tests and discarded.

4.1. Model

The model included:

1- Single events at time, (convolved with SPM's HRF):

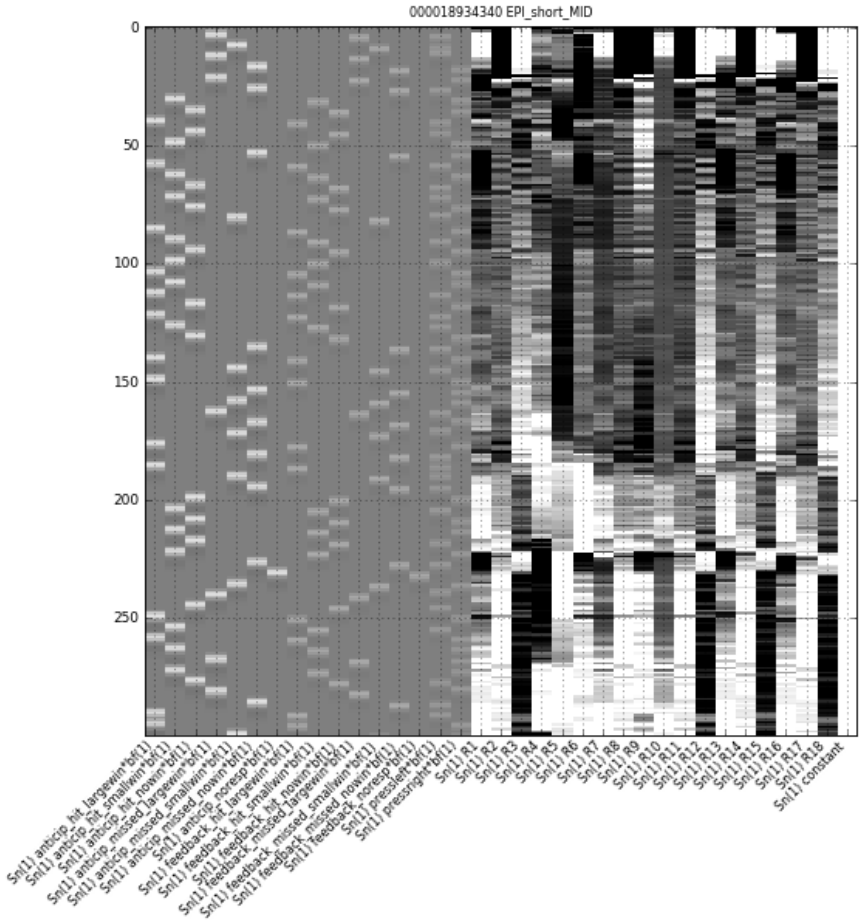
```
AnticipStartTime = Anticipation_Phase_Start_Time / 1000.
ResponseTime = Response_time / 1000.
FeedbackStartTime = (Target_Phase_Start_Time + Target_Phase_Duration) / 1000.
```

2- Events with duration, (convolved with SPM's HRF):

```
Anticip: always modeled for 4s
Feedback: always modeled for 1.5s
```

The design matrix code creation from the csv file is not straightforward, but is copied here for completeness:

```
didpress = np.array([not x.endswith("NO RESPONSE") for x in Response_Made_by_Subject])

anticip_hit_largewin=AnticipStartTime[(Trial_Category == 'BIG_WIN')&(Outcome == 'SUCCESS')]
anticip_hit_smallwin=AnticipStartTime[(Trial_Category == 'SMALL_WIN')&(Outcome == 'SUCCESS')]
anticip_hit_nowin=AnticipStartTime[(Trial_Category == 'NO_WIN')&(Outcome == 'SUCCESS')]

anticip_hit = np.hstack((anticip_hit_largewin, anticip_hit_smallwin, anticip_hit_nowin))
anticip_hit_modgain = np.hstack([[3.]*len(anticip_hit_largewin), [2.]*len(anticip_hit_smallwin),
[1.]*len(anticip_hit_nowin)])

# missed => too late or too early, but never NO RESPONSE
anticip_missed_largewin=AnticipStartTime[(Trial_Category == 'BIG_WIN')&(Outcome == 'FAILURE')& didpress]
anticip_missed_smallwin=AnticipStartTime[(Trial_Category == 'SMALL_WIN')&(Outcome == 'FAILURE')& didpress]
anticip_missed_nowin=AnticipStartTime[(Trial_Category == 'NO_WIN')&(Outcome == 'FAILURE')& didpress]

anticip_missed = np.hstack((anticip_missed_largewin, anticip_missed_smallwin, anticip_missed_nowin))
anticip_missed_modgain = np.hstack([[3.]*len(anticip_missed_largewin), [2.]*len(anticip_missed_smallwin),
[1.]*len(anticip_missed_nowin)])

# This one is the no response
anticip_noresp=AnticipStartTime[(Outcome == 'FAILURE')& (didpress==False)]

allResponsesAtLeft = np.array([x.endswith('Left') for x in Response_Made_by_Subject])
pressleft=ResponseTime[allResponsesAtLeft]
allResponsesAtRight = np.array([x.endswith('Right') for x in Response_Made_by_Subject])
pressright=ResponseTime[allResponsesAtRight]

feedback_hit_largewin=FeedbackStartTime[(Trial_Category == 'BIG_WIN')&(Outcome == 'SUCCESS')]
feedback_hit_smallwin=FeedbackStartTime[(Trial_Category == 'SMALL_WIN')&(Outcome == 'SUCCESS')]
feedback_hit_nowin=FeedbackStartTime[(Trial_Category == 'NO_WIN')&(Outcome == 'SUCCESS')]

feedback_hit = np.hstack((feedback_hit_largewin, feedback_hit_smallwin, feedback_hit_nowin)) #!
feedback_hit_modgain = np.hstack([[3.]*len(feedback_hit_largewin), [2.]*len(feedback_hit_smallwin),
[1.]*len(feedback_hit_nowin)])

feedback_missed_largewin=FeedbackStartTime[(Trial_Category == 'BIG_WIN')&(Outcome == 'FAILURE')& didpress]
feedback_missed_smallwin=FeedbackStartTime[(Trial_Category == 'SMALL_WIN')&(Outcome == 'FAILURE')& didpress]
feedback_missed_nowin=FeedbackStartTime[(Trial_Category == 'NO_WIN')&(Outcome == 'FAILURE')& didpress]

feedback_missed = np.hstack((feedback_missed_largewin, feedback_missed_smallwin, feedback_missed_nowin)) #!
feedback_missed_modgain = np.hstack([[3.]*len(feedback_missed_largewin), [2.]*len(feedback_missed_smallwin),
[1.]*len(feedback_missed_nowin)])

feedback_noresp=FeedbackStartTime[(Outcome == 'FAILURE')& (didpress==False)]

namelist = ["anticip_hit", "anticip_missed", "anticip_noresp", "feedback_hit", "feedback_missed",
"feedback_noresp", "pressleft", "pressright"]
modulationnamelist = ["anticip_hit_modgain", "anticip_missed_modgain", "feedback_hit_modgain",
"feedback_missed_modgain"]
```



As in other protocols csv files, time column values for some subjects may reset from 0 at some point, in this case everything before is assumed to be pre-tests and discarded.

4.2. Contrasts

The associated T-contrasts are numerous :

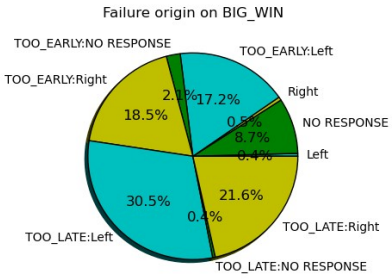
1. "anticip" *('anticip_hit_largewin' + 'anticip_hit_smallwin' + 'anticip_hit_nowin' + 'anticip_missed_largewin' + 'anticip_missed_smallwin' + 'anticip_missed_nowin')*
2. "anticip_hit" *('anticip_hit_largewin' + 'anticip_hit_smallwin' + 'anticip_hit_nowin')*
3. "anticip_missed" *('anticip_missed_largewin' + 'anticip_missed_smallwin' + 'anticip_missed_nowin')*
4. "anticip_noresp" *('anticip_noresp')*
5. "anticip_hit-missed" *('anticip_hit' - 'anticip_missed')*
6. "anticip_missed-hit" *('anticip_missed' - 'anticip_hit')*
7. "anticip_hit-noresp" *('anticip_hit' - 'anticip_noresp')*
8. "anticip_noresp-hit" *('anticip_noresp' - 'anticip_hit')*
9. "anticip_hit_largewin - smallwin" *('anticip_hit_largewin' - 'anticip_hit_smallwin')*
10. "anticip_hit_largewin - nowin" *('anticip_hit_largewin' - 'anticip_hit_nowin')*
11. "anticip_hit_smallwin - nowin" *('anticip_hit_smallwin' - 'anticip_hit_nowin')*
12. "anticip_missed_largewin - smallwin" *('anticip_missed_largewin' - 'anticip_missed_smallwin')*
13. "anticip_missed_largewin - nowin" *('anticip_missed_largewin' - 'anticip_missed_nowin')*
14. "anticip_missed_smallwin - nowin" *('anticip_missed_smallwin' - 'anticip_missed_nowin')*
15. "anticip - anticip_noresp" *('anticip_hit' + 'anticip_missed' - 'anticip_noresp')*
16. "feedback" *('feedback_hit_largewin' + 'feedback_hit_smallwin' + 'feedback_hit_nowin' + 'feedback_missed_largewin' + 'feedback_missed_smallwin' + 'feedback_missed_nowin')*
17. "feedback_hit" *('feedback_hit_largewin' + 'feedback_hit_smallwin' + 'feedback_hit_nowin')*
18. "feedback_missed" *('feedback_missed_largewin' + 'feedback_missed_smallwin' + 'feedback_missed_nowin')*
19. "feedback_hit-missed" *('feedback_hit' - 'feedback_missed')*
20. "feedback_missed-hit" *('feedback_missed' - 'feedback_hit')*
21. "feedback_hit_largewin - smallwin" *('feedback_hit_largewin' - 'feedback_hit_smallwin')*
22. "feedback_hit_largewin - nowin" *('feedback_hit_largewin' - 'feedback_hit_nowin')*
23. "feedback_hit_smallwin - nowin" *('feedback_hit_smallwin' - 'feedback_hit_nowin')*
24. "feedback_missed_largewin - smallwin" *('feedback_missed_largewin' - 'feedback_missed_smallwin')*
25. "feedback_missed_largewin - nowin" *('feedback_missed_largewin' - 'feedback_missed_nowin')*
26. "feedback_missed_smallwin - nowin" *('feedback_missed_smallwin' - 'feedback_missed_nowin')*
27. "press L + R" *('pressleft' + 'pressright')*
28. "press L - R" *('pressleft' - 'pressright')*
29. "press R - L" *(- 'press L – R')*
30. "anticip_hit_somewin - nowin" *('anticip_hit_largewin' + 'anticip_hit_smallwin' - 'anticip_hit_nowin')*
31. "anticip_missed_somewin - nowin" *('anticip_missed_largewin' + 'anticip_missed_smallwin' - 'anticip_missed_nowin')*
32. "feedback_hit_somewin - nowin" *('feedback_hit_largewin' + 'feedback_hit_smallwin' - 'feedback_hit_nowin')*
33. "feedback_missed_somewin - nowin" *('feedback_missed_largewin' + 'feedback_missed_smallwin' - 'feedback_missed_nowin')*
34. "feedback_somewin_hit - missed" *('feedback_hit_largewin' + 'feedback_hit_smallwin' - 'feedback_missed_largewin' - 'feedback_missed_smallwin')*
35. "feedback_somewin_missed - hit" *(- 'feedback_somewin_hit - missed')*
36. "feedback_somewin - nowin" *('feedback_hit_somewin - nowin' + 'feedback_missed_somewin – nowin')*
37. "anticip_hit_largewin"
38. "- anticip_hit_largewin"
39. "feedback_hit_largewin"
40. "- feedback_hit_largewin"
41. "anticip_hit_largewin - feedback_hit_largewin"
42. "feedback_hit_largewin - anticip_hit_largewin"
43. "anticip_hit_nowin - feedback_hit_nowin"
44. "feedback_hit_nowin - anticip_hit_nowin"

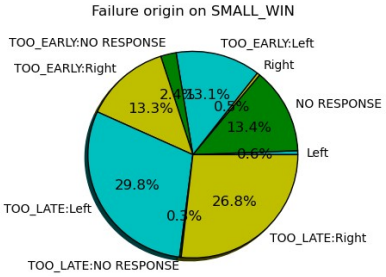
4.3. Comments

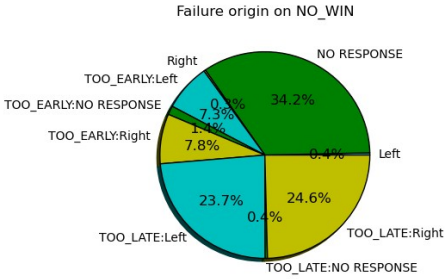
- Events Durations:
    - Although the “Anticipation Phase Duration” is variable between trial (ranging from 4.0s to 4.5s), the model consistently fits it as a 4.0s-long event. This makes the fitted model more suitable at computing accurate contrasts (meaningful difference of estimated parameters). This allows a simpler interpretation of the comparison of the different anticipation periods.

    - The same applies to the “Feedback Phase Duration”, which are consistently modeled as 1.45s-long events.

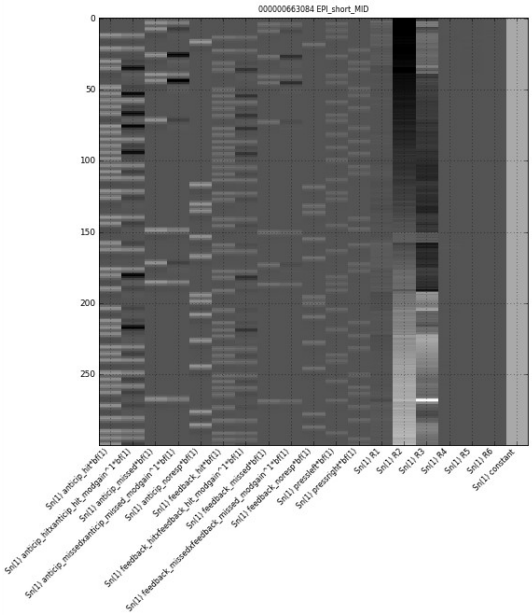
- Motor response
    - The Left or Right motor response are explicitly modeled, as the experiment design allowed it. Obviously, response motors includes both Success and Failure trials.
- Feedback phase timing
    - The beginning of the Feedback phase corresponds to the timing (`Target_Phase_Start_Time` + `Target_Phase_Duration`) from the csv behavioural files. This appeared more reliable than using `Feedback Phase Start Time`, for which some subjects (~8%) had some unexpected values.
    - To model the beginning of the Feedback phase, there is almost no difference in the final contrasts values when using (`Target_Phase_Start_Time` + `Target_Phase_Duration`) or `Target Phase Start Time` either, confirming further the above choice.
- The “FAILURE” response origin
    - When subjects fail at responding on time, a FAILURE outcome is recorded. Yet the cause of the failure may vary and therefore it is possible that different brain networks are involved. The failure may be due to a bad timing (too late/early), or because there is no motor response at all. We therefore split the FAILURE cases in two cases (Failure)Miss and (Failure)Noresponse. **Therefore “missed” anticipation or feedback conditions refers only to events where a motor response was actually recorded**. “no resp” refers to the other failure case (no response).
    - The NO RESPONSE cases interact with the Gain size, as shown on the figure below (and as expected).







- The Gain size (BIG WIN, SMALL WIN, NO WIN) could have been used to modulate the main regressors of interest (Anticipation (hit or miss), Feedback (hit or miss)). In this alternative parametric model (suggested by Christian Buechel at some stage of the interaction), this modulation take the form of additional regressors, which are linear from 1 (No Win) to 3 (Big win).



- The motivation for such a parametric model is in the interpretation of the Gain regressors.
- We collectively decided to keep the categorical model because it was shown that

the difference between Large versus Small was not equal to the difference between Small and No (e.g. for the `anticip_win` condition). The underlying assumption of the linearity effect due to Gain is therefore not entirely valid in all brain regions.

- A few subjects events record file included some “NORESPONSE:TOOLATE” and “NORESPONSE:TOOEARLY” events. It isn't fully clear to me what are those conditions. Yet, because there are very few of them (see pie-chart above) it was considered and modeled the same as usual NO RESPONSE events.

4.4. subject-specific notes or workaround

A potentially severe problem involve the `Response Time` columns. In some subjects (70 from 966 reviewed), the `Reponse Time` value is unrealistically close to the `Target Phase Start Time` on every Success outcome, meaning that one of the value may be wrong or misinterpreted. This may potentially reduce accuracy at fitting the motor response.

- The 70 subjects where (`TargetPhaseStartTime - ResponseTime`) < 10ms on Success, implying either a unrealistically accurate human response at the target, a erroneous value somewhere, or my misinterpretation of the field:

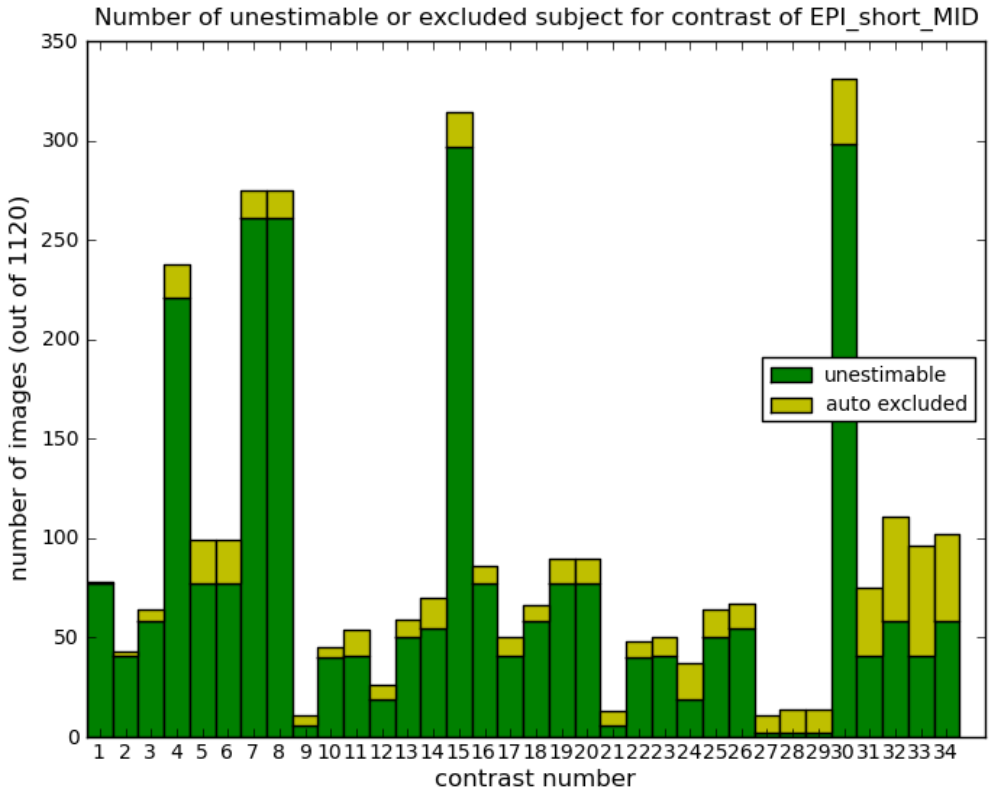
```
mid_000037836103.csv mid_000029280369.csv mid_000064564726.csv mid_000016408573.csv
mid_000013679970.csv mid_000083785552.csv mid_000026045785.csv
mid_000091159380.csv mid_000019025504.csv mid_000032062629.csv mid_000048727429.csv
mid_000087248884.csv mid_000061943676.csv mid_000085498115.csv
mid_000027553119.csv mid_000001380042.csv mid_000073108506.csv mid_000077511665.csv
mid_000023079648.csv mid_000043968174.csv mid_000057133727.csv
mid_000094265316.csv mid_000099561404.csv mid_000047119736.csv mid_000031561113.csv
mid_000099747799.csv mid_000042169862.csv mid_000005182386.csv
mid_000097084811.csv mid_000066378811.csv mid_000046457090.csv mid_000039935082.csv
mid_000007839943.csv mid_000036503289.csv mid_000098516748.csv
mid_000060919135.csv mid_000080303743.csv mid_000025453247.csv mid_000065251291.csv
mid_000078673570.csv mid_000043525440.csv mid_000002296749.csv
mid_000036525962.csv mid_000030978066.csv mid_000066092324.csv mid_000053515504.csv
mid_000013424686.csv mid_000098205180.csv mid_000041224729.csv
mid_000094345446.csv mid_000015398752.csv mid_000098562537.csv mid_000034252547.csv
mid_000071766352.csv mid_000033925730.csv mid_000056595836.csv
mid_000081257501.csv mid_000025229476.csv mid_000026193033.csv mid_000008685800.csv
mid_000004004697.csv mid_000082922411.csv mid_000022456630.csv
mid_000064672179.csv mid_000048616789.csv mid_000063489515.csv mid_000035968441.csv
mid_000028250707.csv
```

- The same subjects as above have the following characteristic: (`Target_Phase_Start_Time + Target_Phase_Duration`) could differ from `Feedback_Phase_Start_Time` for up to 100 ms, whereas it is expected that those values should match, as it is the case in most subjects (the Feedback phase should begin at the end of the Target phase).
- Another characteristic of those 70 subjects is that the time difference between `Feedback Phase Start Time` and `Target Phase Start Time` can get down to less than 10ms. Those are expected to be separated by at least `Target Phase Duration`, as it is the case in other subjects. This may be a side effects of other incorrect values.

`mid_000079848243.csv` has a quote formatting problem, manually fixed.

4.5. Unestimable or Excluded contrast maps

A Group-wide distribution (computed on 1120 subject) of unestimable contrasts or excluded subjects. *Unestimable* covers contrasts whose lack of events prevented the estimation of the corresponding parameter value, while *excluded* covers images which were dropped due to a too abnormal activation profile, computed with the automatic QC procedure.



It should be noted that the relatively high number of unestimable values for some regressors (4, 7,8, 15, 30..) is due to the fact that they model a failure condition (“NO RESPONSE”) which may never occur. Other contrasts with somehow large number of unestimable images (4, 5...) tended to include the “No Gain” event.

5. FT – Face Task

The face tasks is a passive task where 18-seconds blocs of either a Face movie or a Control stimulus are shown. The faces can be either “Angry” or “Neutral”. The third condition is a visual control denoted “Control”.

Example of “ neutral” face

Data : 4D fMRI sequence of 444 volumes, rtime 2.2s.
The csv behavioural file includes the movie played and the start event timing. They should be almost similar across all subjects. It consists of 19 stimuli blocks, 10 of which are faces (angry or neutral) animations and 9 of which are control.

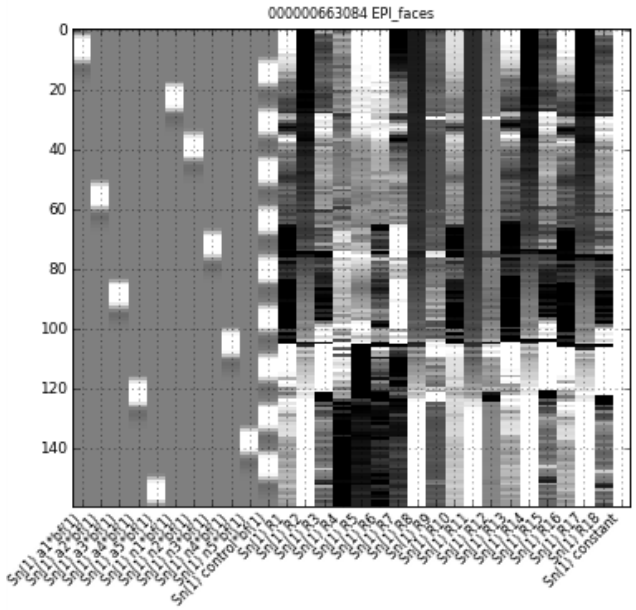
5.1. Model

The design is straightforward:

```
namelist = ["a1", "a2", "a3", "a4", "a5", "n1", "n2", "n3", "n4", "n5", "control"]
```

where “a” and “n” denotes respectively “angry” and “neutral” conditions.

Blocks are 18s long, and convolved SPM's Standard Hemodynamic Function.



5.2. Contrasts

Associated contrasts are:

1. 'control'
2. 'neutral' *('n1' + 'n2' + 'n3'+ 'n4' + 'n5')*
3. 'angry' *('a1' + 'a2' + 'a3' + 'a4' + 'a5')*
4. 'neutral - control' *('neutral' - 'control')*
5. 'control - neutral' *(- 'neutral - control')*
6. 'angry - control' *('angry' - 'control')*
7. 'control - angry' *(- 'angry - control')*
8. 'angry - neutral' *('angry' - 'neutral')*
9. 'neutral - angry' *(- 'angry - neutral')*
10. 'angry+neutral - control' *('angry' + 'neutral' - 'control')*
11. 'control - angry+neutral' *(- 'angry+neutral - control')*

5.3. Comments

Each face is modeled on its separate regressor (a1 to a5 for angry, n1 to n5 for neutral) because although this will not change the amount of the bold signal in the contrasts (averaged across “a” or “n” conditions), it may reduce the noise variance if there is BOLD level differences in the processing of each individual face presentation.

A “face recognition” task is performed outside of the scanner to assert that the subjects did actually watches the faces. This information is not relevant for the intra-subject modeling.

5.4. subject-specific notes or workaround

- `000042694237` (from site 02) : Stimulus occurred more slowly (apparently), so that only 80% stimuli had been done when the run finished (at 160 * 2,2 sec). Did the computer slow down ?
- `000099550415` (from site 07) has been aborted after 35 TR.
- Subjects having an ft_*.csv file, but missing a recog*.csv (recognition task) file, among ~900 subjects :

```
000003970752
000010643071
000006881937
000097982061
000053120454
000013148228
000093240591
000094897310
000011104036
000008633229
000060835696
000035077363
000052045472
000057930100
000045526627
000049321994
000069832522
000025103079
000065032538
000098442422
000035734504
000013679970
000007127936
000023207511
000099875982
000079947544
000057359042
000076310486
000069026991
000074102261
000065871415
000043813881
```

- Weird number of events in .csv file
    - First event doubled

ft_000002371789.csv
ft_000090129274.csv

- All events doubled

ft_000092318687.csv

- Empty:

ft_000048284598.csv
ft_000067498283.csv

5.5. Unestimable or Excluded contrast maps

A Group-wide distribution (computed on 1188 subject) of unestimable contrasts or excluded subjects. *Unestimable* covers contrasts whose lack of events prevented the estimation of the corresponding parameter value, while *excluded* covers images which were dropped due to a too abnormal activation profile, computed with the automatic QC procedure.