New BCI approaches: Selective Attention to Auditory and Tactile Stimulus Streams

N. Jeremy Hill,[1] Cornelius Raths[2]

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]University of Southern California, Los Angeles, USA

with thanks to Ryota Tomioka at FIRST/TU-Berlin

Overview

Long-term goal: develop and optimize new paradigms that are suitable for people in the "completely locked-in" state (CLIS).

Overview

Long-term goal: develop and optimize new paradigms that are suitable for people in the "completely locked-in" state (CLIS).

- Enable the patient to make a binary decision.

Overview

Long-term goal: develop and optimize new paradigms that are suitable for people in the “completely locked-in” state (CLIS).

- Enable the patient to make a binary decision.
- Use signals elicited auditory and tactile stimuli, measureable by EEG.

Overview

Long-term goal: develop and optimize new paradigms that are suitable for people in the "completely locked-in" state (CLIS).

- Enable the patient to make a binary decision.
- Use signals elicited auditory and tactile stimuli, measureable by EEG.
- Use machine-learning algorithms to classify the signals.

Overview

Long-term goal: develop and optimize new paradigms that are suitable for people in the "completely locked-in" state (CLIS).

- Enable the patient to make a binary decision.
- Use signals elicited auditory and tactile stimuli, measureable by EEG.
- Use machine-learning algorithms to classify the signals.

This presentation concerns the results from preliminary experiments with healthy subjects.

Why Non-Visual?

Why Non-Visual?

- In the CLIS state, patients are functionally blind:

Why Non-Visual?

- In the CLIS state, patients are functionally blind:
  - eyes cannot be opened at will;
  - eyes may move involuntarily (often rolling up);
  - lens cannot be refocused or gaze directed;
  - no microsaccades, so images fade out (Troxler effect);
  - no saccades, so no integration of visual scenes: the fovea images a fixed 2 deg. spot, and resolution is very low in most of the visual field;
  - long immobility of the eye often leads to infections;

Why Non-Motor?

- Motor imagery-based BCI shows promising results with normal subjects, and patients with extensive paralysis (Kübler et al 2005, Neurology 10). So far it has not worked with patients in CLIS. Why?
  - Can the patient still imagine movement?

Why Non-Motor?

- Motor imagery-based BCI shows promising results with normal subjects, and patients with extensive paralysis (Kübler et al 2005, Neurology 10). So far it has not worked with patients in CLIS. Why?
  - Can the patient still imagine movement?
  - Can the motor and premotor cortex still produce ERD/ERS during motor imagery?

Why Non-Motor?

- Motor imagery-based BCI shows promising results with normal subjects, and patients with extensive paralysis (Kübler et al 2005, Neurology 10). So far it has not worked with patients in CLIS. Why?
  - Can the patient still imagine movement?
  - Can the motor and premotor cortex still produce ERD/ERS during motor imagery?
  - (...and are these in fact the same question?)

Why Non-Motor?

- Motor imagery-based BCI shows promising results with normal subjects, and patients with extensive paralysis (Kübler et al 2005, Neurology 10). So far it has not worked with patients in CLIS. Why?
  - Can the patient still imagine movement?
  - Can the motor and premotor cortex still produce ERD/ERS during motor imagery?
  - (...and are these in fact the same question?)
  - Are they still intact enough to (relearn to) do so?
    - ⋆ EEG is still the most attractive technology for clinical BCI.
    - ⋆ Most of the EEG signal comes from pyramidal neurons.
    - ⋆ ALS kills the pyramidal neurons of the motor cortex.

Auditory and Tactile BCI

Exogenous (i.e. stimulus-driven) BCI's rely on the conscious direction of the user's *attention*.

Auditory and Tactile BCI

Exogenous (i.e. stimulus-driven) BCI's rely on the conscious direction of the user's *attention*.

For paralysed users, this means *covert* attention.

Auditory and Tactile BCI

Exogenous (i.e. stimulus-driven) BCI's rely on the conscious direction of the user's *attention*.

For paralysed users, this means *covert* attention.

Does covert attention affect auditory ERPs?

Auditory and Tactile BCI

Exogenous (i.e. stimulus-driven) BCI's rely on the conscious direction of the user's *attention*.

For paralysed users, this means *covert* attention.

Does covert attention affect auditory ERPs?

Yes, e.g.:

- Hillyard et al. 1974, *Science* 182.
- Näätänen 1990, *Behavioral and Brain Sciences* 13.
- Schröger and Wolff 1998 *Cognitive Brain Research* 7.

Auditory and Tactile BCI

Exogenous (i.e. stimulus-driven) BCI's rely on the conscious direction of the user's *attention*.

For paralysed users, this means *covert* attention.

Does covert attention affect tactile ERPs?

Yes, e.g.:

- Desmet et al. 1977, *Journal of Physiology* 271.
- García Lorrea 1995, *Psychophysiology* 32.
- Eimer et al. 2003 *Experimental Brain Research* 151.

Auditory and Tactile BCI

Exogenous (i.e. stimulus-driven) BCI's rely on the conscious direction of the user's *attention*.

For paralysed users, this means *covert* attention.

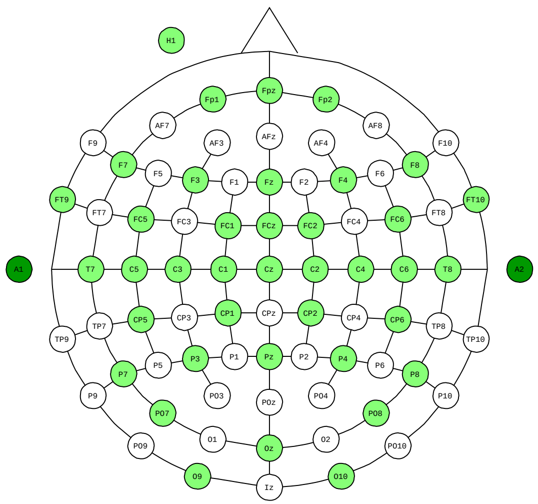
Does covert attention affect tactile ERPs?

Yes, e.g.:

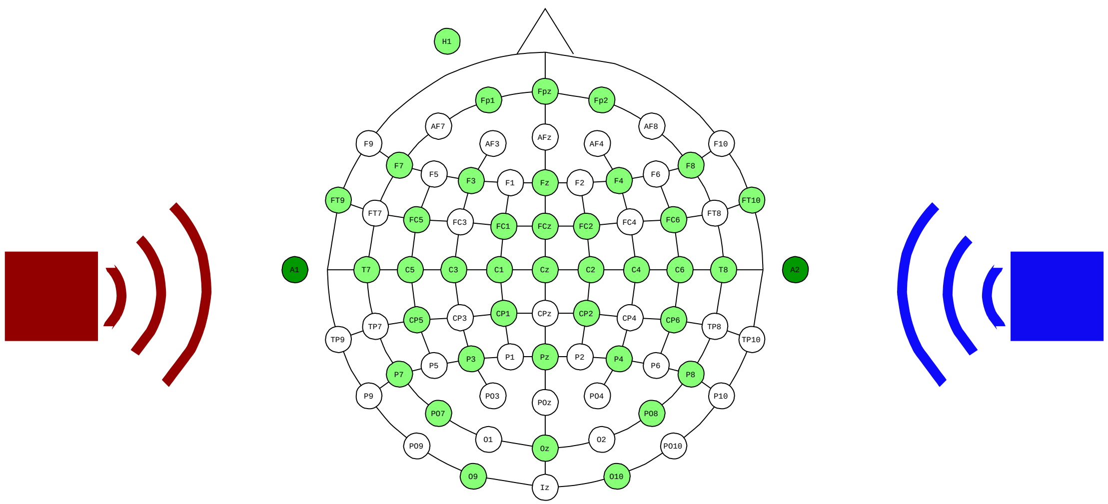
- Desmet et al. 1977, *Journal of Physiology* 271.
- García Lorrea 1995, *Psychophysiology* 32.
- Eimer et al. 2003 *Experimental Brain Research* 151.

...at least, when you average hundreds of trials. Can we obtain a reliable effect on a timescale suitable for BCI?

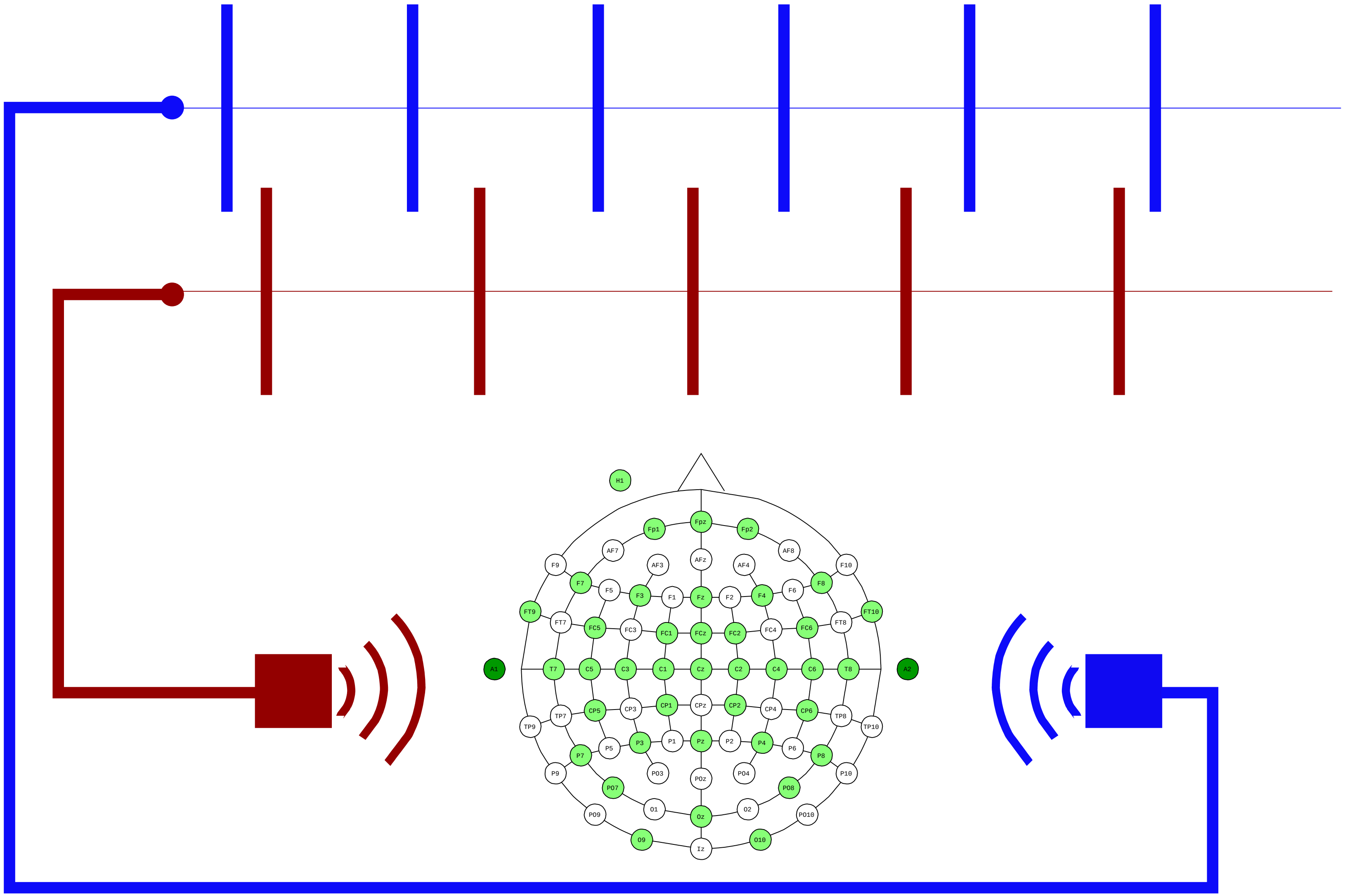
I: Auditory stimulation in EEG



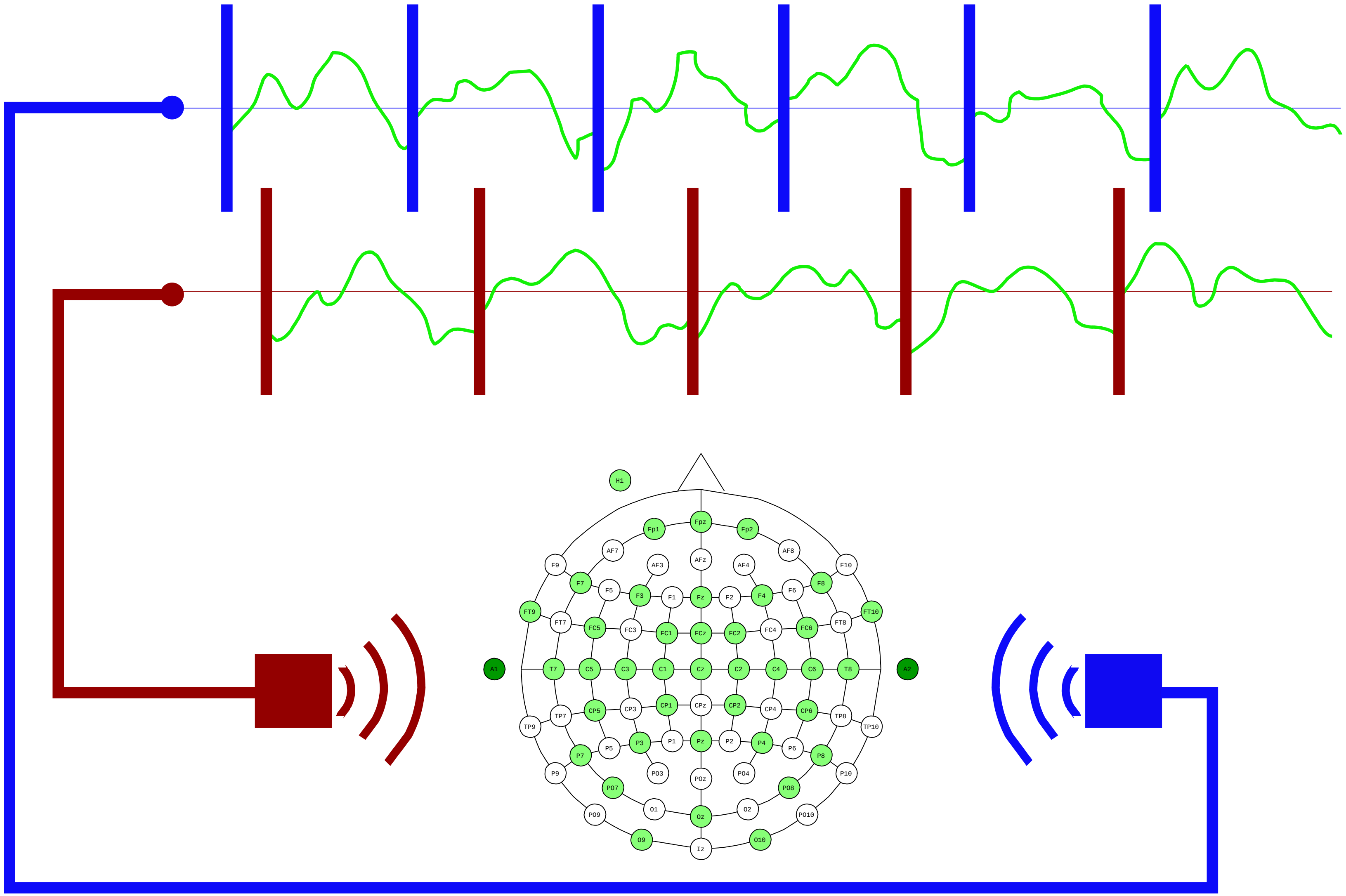
I: Auditory stimulation in EEG



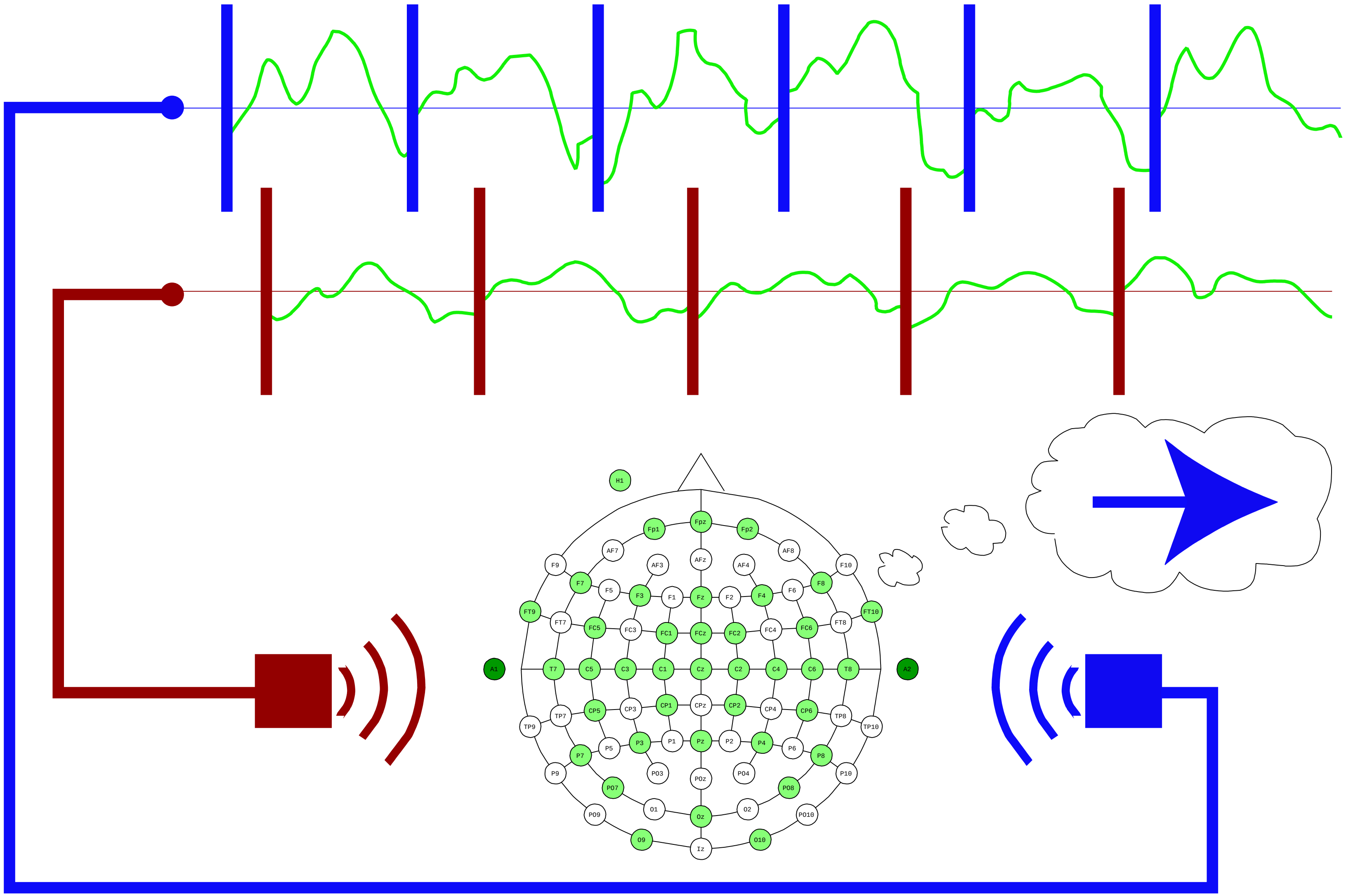
I: Auditory stimulation in EEG



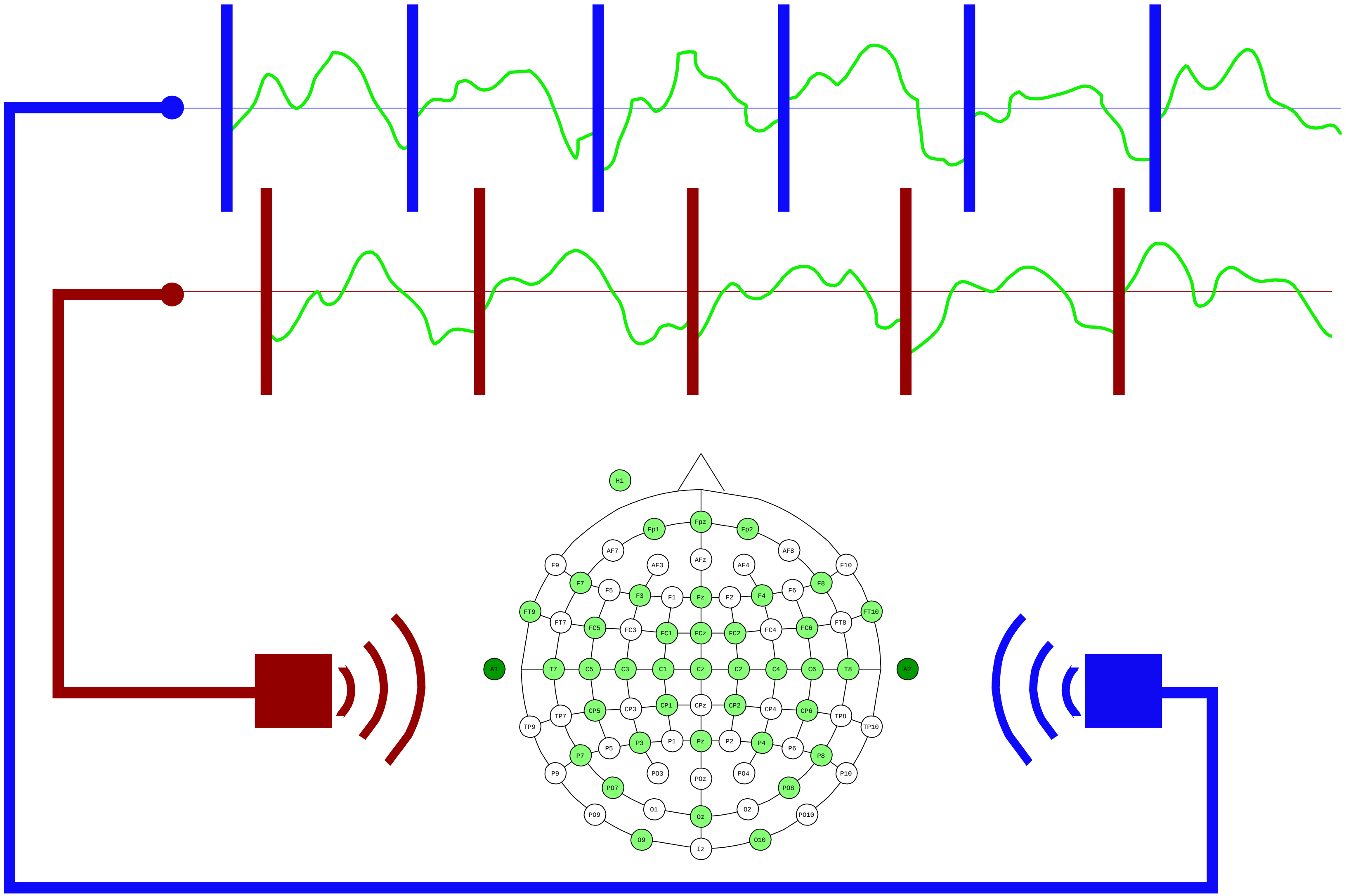
I: Auditory stimulation in EEG



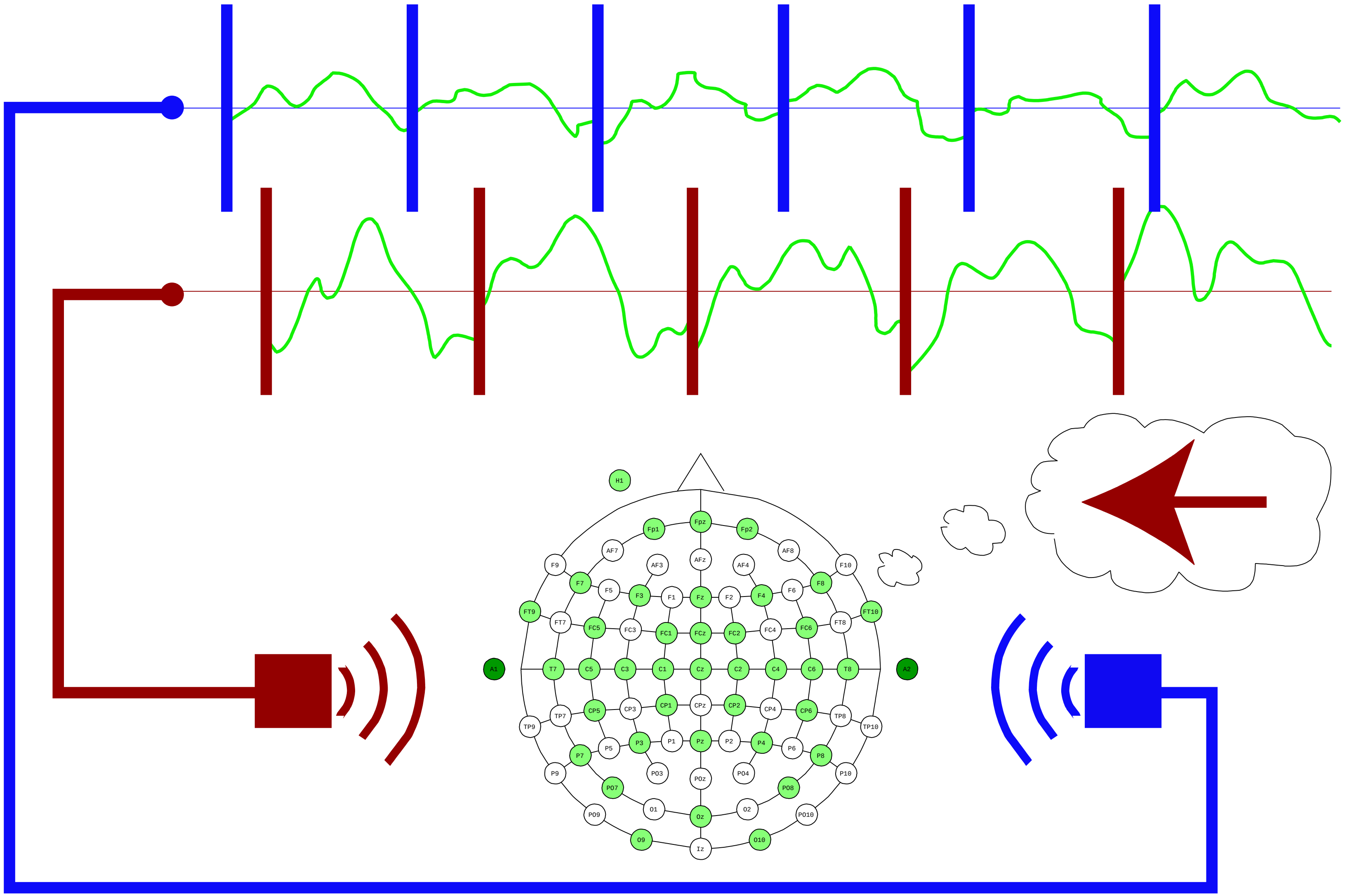
I: Auditory stimulation in EEG



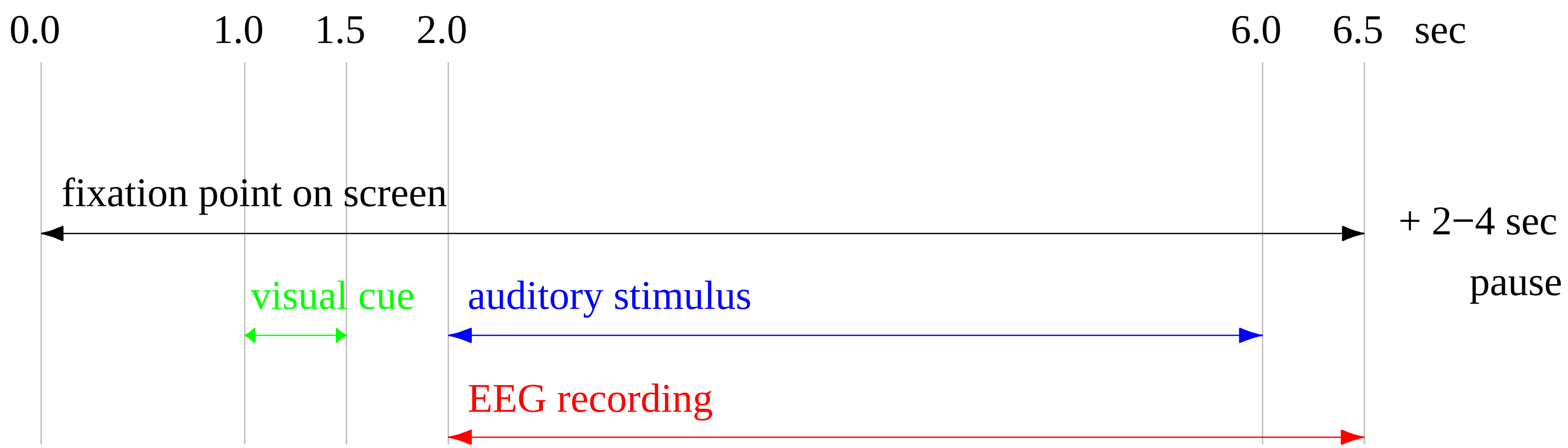
I: Auditory stimulation in EEG



I: Auditory stimulation in EEG



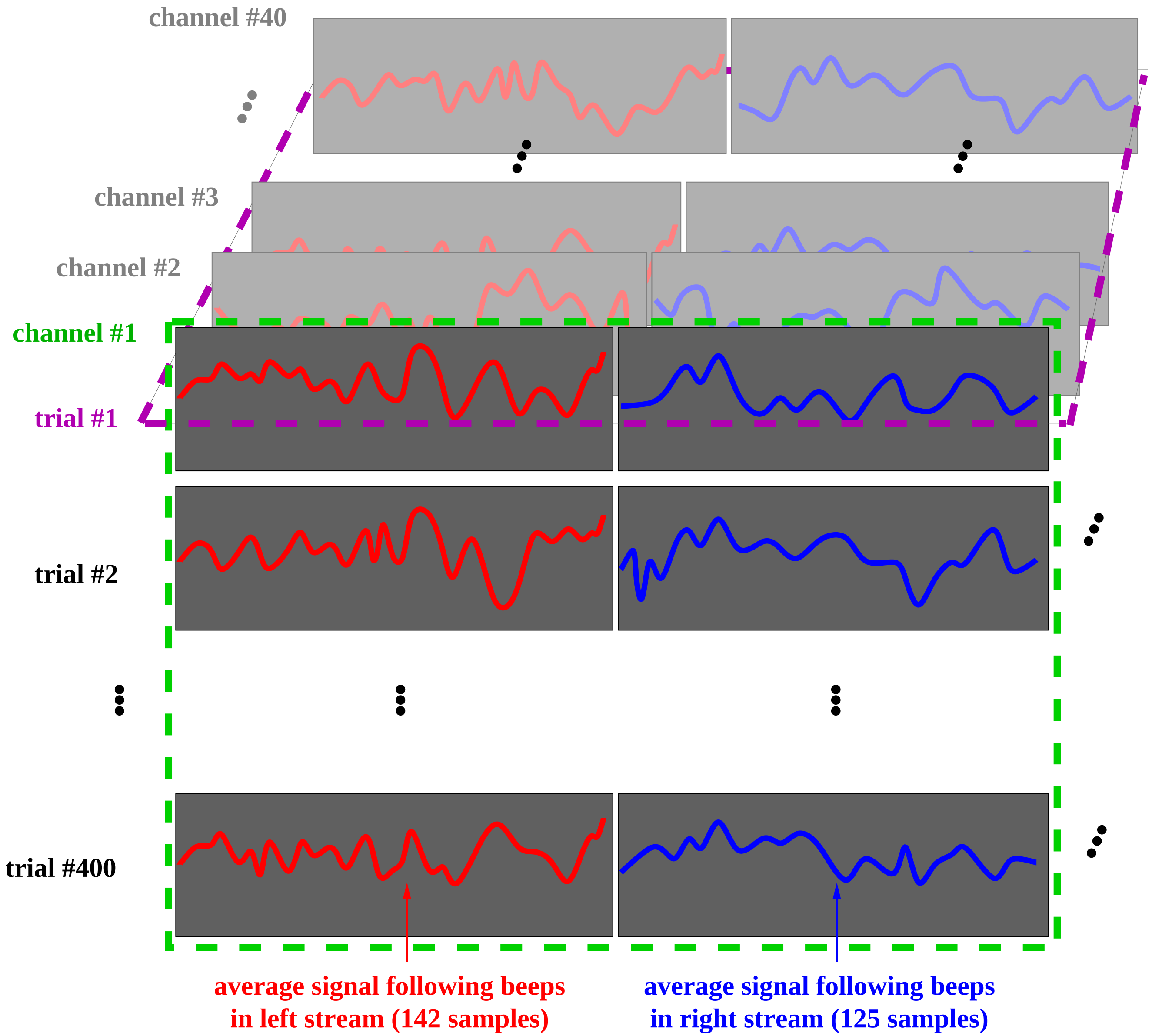
Trial structure



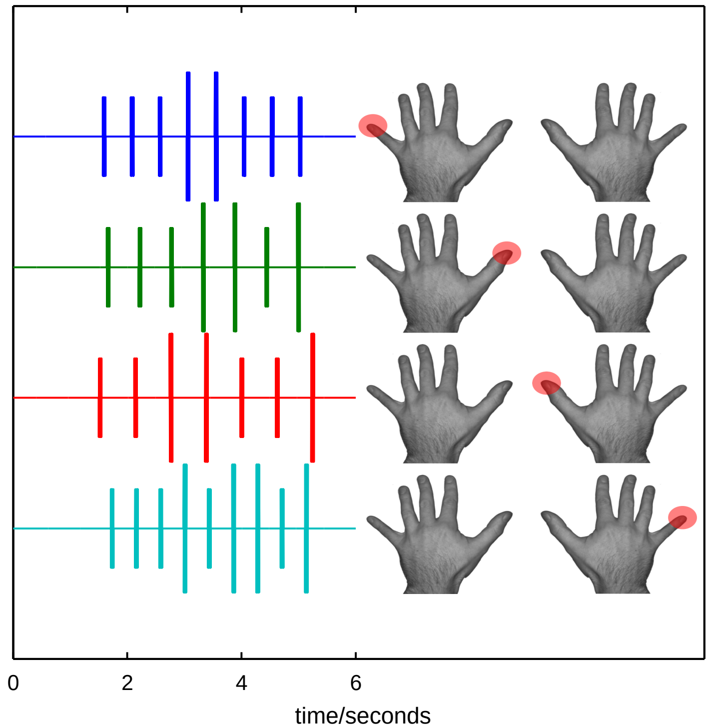
16 subjects, 400 trials each in one 3-hour session.

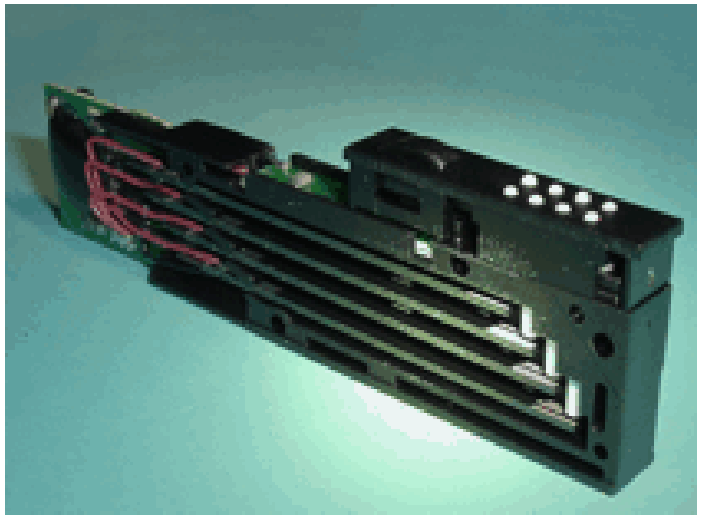
Cued direction of attention without feedback.

Data structure



II: Tactile stimulation in MEG

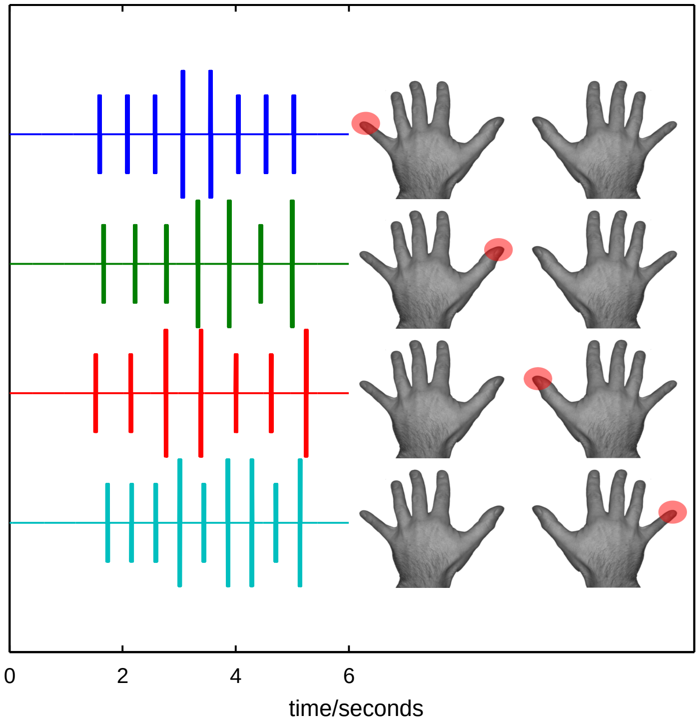


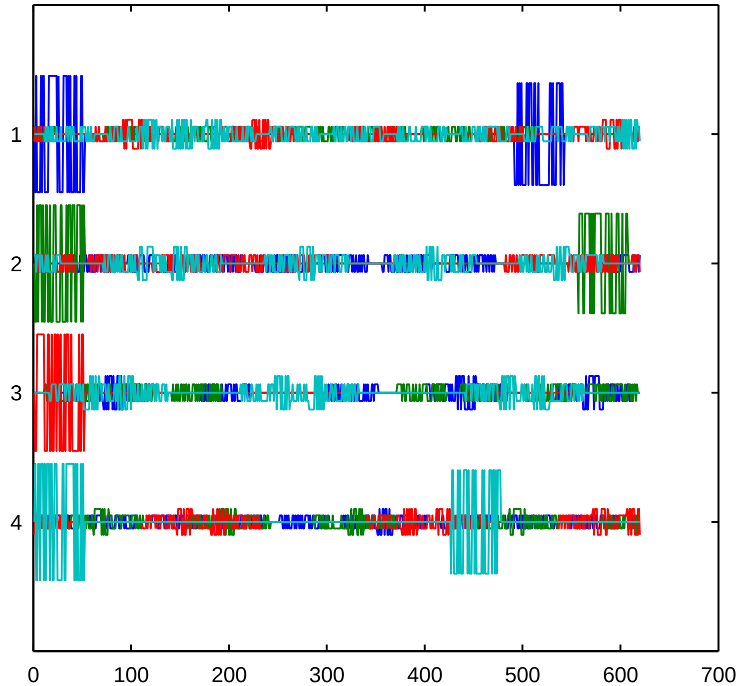


9 subjects, each 200 cued trials without feedback.

5 classes including the no intentional control (NIC) state.

II: Tactile stimulation in MEG

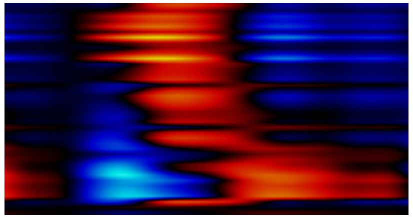


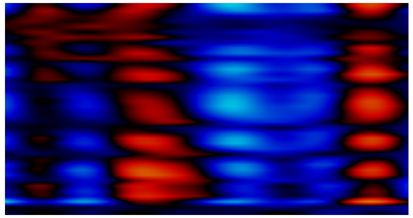


9 subjects, each 200 cued trials without feedback.

5 classes including the no intentional control (NIC) state.

Classification



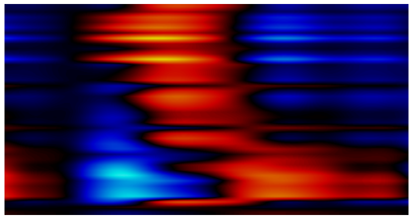


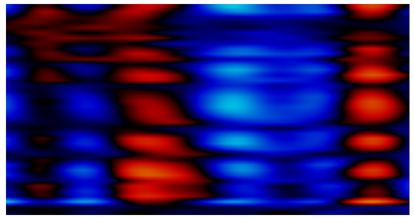
$$f(\mathrm{X}) = \sum_{\text{sensors}} \sum_{\text{time}} \Big(\underset{S \times T}{\mathrm{G}} \;.*\; \underset{S \times T}{\mathrm{X}} \Big) + b$$

G is the weight "vector" found by some classifier (SVM, LR, LDA. . .)

Classification

$$f(\mathrm{X}^{(1)}) =$$

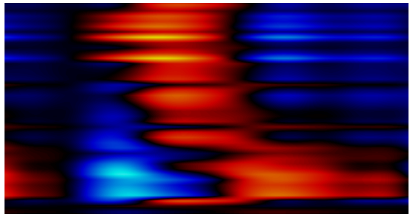




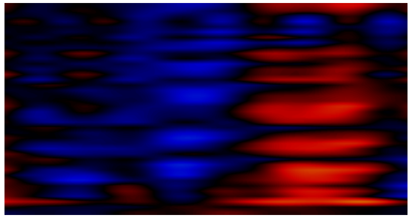
$$\sum_{\text{sensors}} \sum_{\text{time}} \left(\underset{S \times T}{\mathrm{G}} \;.*\; \underset{S \times T}{\mathrm{X}^{(1)}} \right) + b$$

Classification

$$f(\mathrm{X}^{(2)}) \quad = \quad \sum_{\text{sensors}} \sum_{\text{time}} \Bigg($$



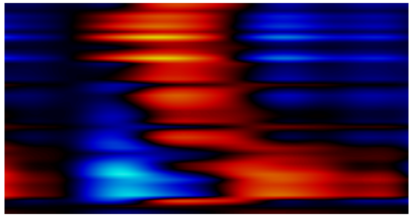
$$\underset{S \times T}{\mathrm{G}} \quad .* \quad \underset{S \times T}{\mathrm{X}^{(2)}}$$

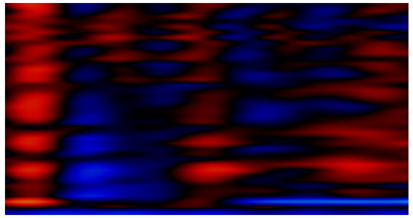


$$\Bigg) + b$$

Classification

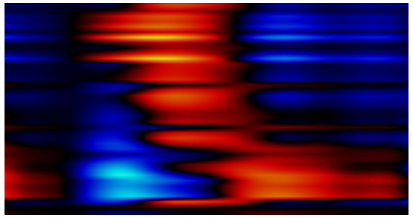
$$f(\mathrm{X}^{(3)}) =$$

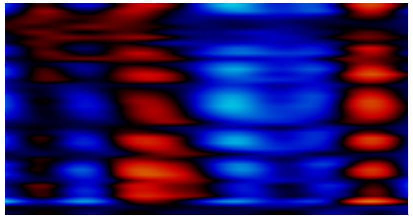




$$\sum_{\text{sensors}} \sum_{\text{time}} \Big(\underset{S \times T}{\mathrm{G}} \;.*\; \underset{S \times T}{\mathrm{X}^{(3)}} \Big) + b$$

Classification



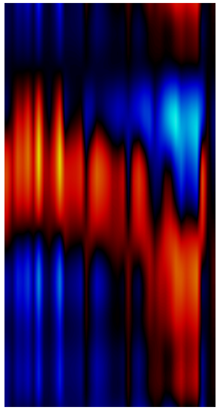


$$f(\mathrm{X}) = \sum_{\text{sensors}} \sum_{\text{time}} \left(\underset{S \times T}{\mathrm{G}} \; .* \; \underset{S \times T}{\mathrm{X}} \right) + b$$

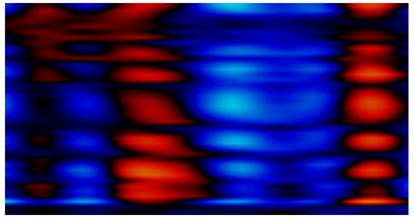
Classification

$$f(\mathrm{X}) \quad = $$

$$\mathrm{trace}\Bigg($$



$$\underset{T \times S}{\mathrm{G}^\top}$$



$$\underset{S \times T}{\mathrm{X}}$$

$$\Bigg) + b$$

Classification

$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$

Classification

$f(\mathrm{X}) = \mathrm{tr}\left[\mathrm{G}^\top \mathrm{X}\right]$

It can be helpful to assume that G consists of only a small number (let's say 2) of relevant spatio-temporal features—

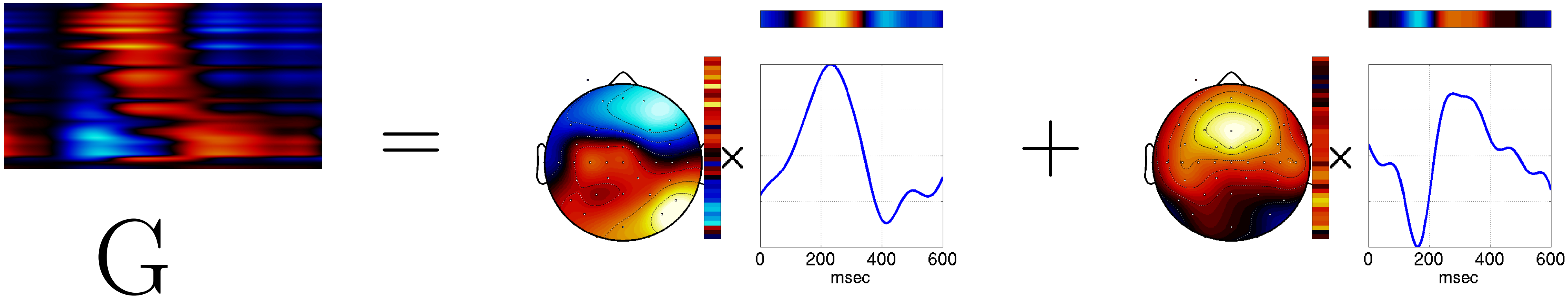
- whose spatial characteristics are stationary in time;
- whose temporal characteristics are stationary in space:

Classification

$f(\mathrm{X}) = \mathrm{tr}\ [\mathrm{G}^\top \mathrm{X}]$

It can be helpful to assume that G consists of only a small number (let's say 2) of relevant spatio-temporal features—

- whose spatial characteristics are stationary in time;
- whose temporal characteristics are stationary in space:

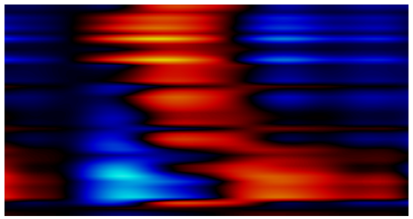


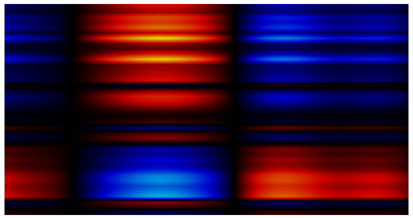
Classification

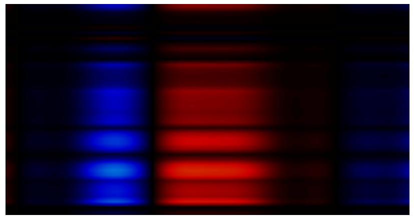
$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$

It can be helpful to assume that G consists of only a small number (let's say 2) of relevant spatio-temporal features—

- whose spatial characteristics are stationary in time;
- whose temporal characteristics are stationary in space:







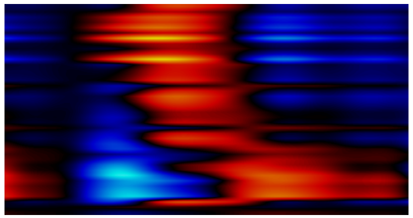
$$\mathrm{G} = \mathbf{W}_{\mathrm{s}1}\,\mathbf{W}_{\mathrm{t}1}^\top + \mathbf{W}_{\mathrm{s}2}\,\mathbf{W}_{\mathrm{t}2}^\top$$

Classification

$f(\mathrm{X}) = \mathrm{tr}\ [\mathrm{G}^\top \mathrm{X}]$

It can be helpful to assume that G consists of only a small number (let's say 2) of relevant spatio-temporal features—

- whose spatial characteristics are stationary in time;
- whose temporal characteristics are stationary in space:



$=$

$$\mathrm{G} \qquad = \qquad \mathrm{W_s} \ \mathrm{W_t}^\top$$

Classification

$$
\begin{aligned}
f(\mathrm{X}) &= \mathrm{tr}\left[\mathrm{G}^\top \mathrm{X}\right] \\
&= \mathrm{tr}\left[\left(\mathrm{W_s W_t^\top}\right)^\top \mathrm{X}\right]
\end{aligned}
$$

Classification

$$f(\mathrm{X}) = \mathrm{tr}\left[\mathrm{G}^\top \mathrm{X}\right]$$

$$= \mathrm{tr}\left[\mathrm{W}_t \mathrm{W}_s^\top \mathrm{X}\right]$$

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t} \mathrm{W_s}^\top \mathrm{X}] \;\; = \mathrm{tr}\left[\begin{array}{ccc} \mathrm{W_s}^\top & \mathrm{X} & \mathrm{W_t} \\ 2 \times S & & T \times 2 \end{array}\right]$$

(because arguments inside a trace operator may be cyclically reordered)

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\left[\mathrm{G}^\top \mathrm{X}\right]$$

$$= \mathrm{tr}\left[\mathrm{W_t W_s^\top X}\right] = \mathrm{tr}\left[\underset{2\times S}{\mathrm{W_s^\top}} \quad \mathrm{X} \quad \underset{T\times 2}{\mathrm{W_t}}\right] = \mathrm{tr}\left[\underset{2\times 2}{\mathrm{L}}\right]$$

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t} \mathrm{W_s}^\top \mathrm{X}] \;= \mathrm{tr} \left[\underset{2 \times S}{\mathrm{W_s}^\top} \; \mathrm{X} \; \underset{T \times 2}{\mathrm{W_t}} \right] \;= \mathrm{tr} \left[\underset{2 \times 2}{\mathrm{L}} \right]$$

By making G low- instead of full-rank, we are assuming that there is a low-dimensional subspace onto which we can project X without loss (and perhaps with improvement) in performance of $f(\mathrm{X})$. We assert that, for classification, each data point $\mathrm{X}^{(i)}$ can be sufficiently represented by a small number of coefficients—two, in our example: $\left(L_{11}^{(i)}, L_{22}^{(i)}\right)$.

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\left[\mathrm{G}^\top \mathrm{X}\right]$$

$$= \mathrm{tr}\left[\mathrm{W_t} \mathrm{W_s}^\top \mathrm{X}\right] = \mathrm{tr}\left[\underset{2 \times S}{\mathrm{W_s}^\top} \; \mathrm{X} \; \underset{T \times 2}{\mathrm{W_t}}\right] = \mathrm{tr}\left[\underset{2 \times 2}{\mathrm{L}}\right]$$

This implies a basis of spatial and temporal features $\mathrm{A_s}$ and $\mathrm{A_t}$ such that

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}), \quad \text{where } \tilde{\mathrm{X}} = \underset{S \times 2}{\mathrm{A_s}} \; \overset{\ddots}{\mathrm{L}} \; \underset{2 \times T}{\mathrm{A_t}^\top} \; .$$

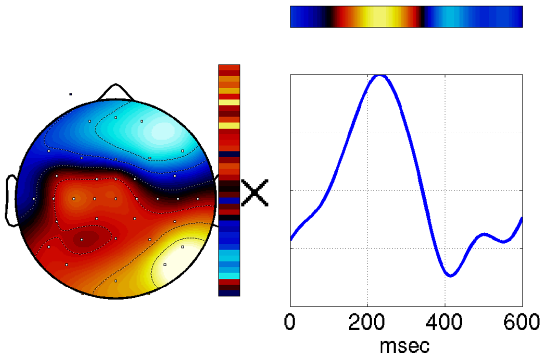
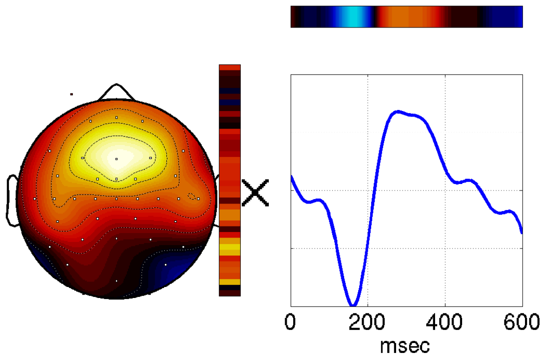
Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t W_s^\top X}] \;=\; \mathrm{tr}\left[\underset{2\times S}{\mathrm{W_s^\top}}\;\;\mathrm{X}\;\;\underset{T\times 2}{\mathrm{W_t}}\right] \;=\; \mathrm{tr}\left[\underset{2\times 2}{\mathrm{L}}\right]$$

This implies a basis of spatial and temporal features $\mathrm{A_s}$ and $\mathrm{A_t}$ such that

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}), \quad \text{where } \tilde{\mathrm{X}} = \underset{S\times 2}{\mathrm{A_s}}\;\overset{\ddots}{\mathrm{L}}\;\underset{2\times T}{\mathrm{A_t^\top}}\;.$$

$\tilde{\mathrm{X}} = \quad L_{11}$  $+ \quad L_{22}$ 

Bilinear Discriminant Analysis

$$
\begin{aligned}
f(\mathrm{X}) &= \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}] \\
&= \mathrm{tr}\,[\mathrm{W_t W_s^\top X}] \;= \mathrm{tr}\left[\underset{2\times S}{\mathrm{W_s^\top}}\ \mathrm{X}\ \underset{T\times 2}{\mathrm{W_t}}\right] \;= \mathrm{tr}\left[\underset{2\times 2}{\mathrm{L}}\right]
\end{aligned}
$$

This implies a basis of spatial and temporal features $\mathrm{A_s}$ and $\mathrm{A_t}$ such that

$$
f(\mathrm{X}) = f(\tilde{\mathrm{X}}), \quad \text{where } \tilde{\mathrm{X}} = \underset{S\times 2}{\mathrm{A_s}}\ \overset{\ddots}{\mathrm{L}}\ \underset{2\times T}{\mathrm{A_t^\top}}\ .
$$

$$
\mathrm{tr}\,[\mathrm{L}] = \mathrm{tr}\left[\mathrm{W_s^\top A_s}\ \overset{\ddots}{\mathrm{L}}\ \mathrm{A_t^\top W_t}\right]
$$

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t}\mathrm{W_s}^\top \mathrm{X}] \;= \mathrm{tr}\left[\underset{2\times S}{\mathrm{W_s}^\top}\;\mathrm{X}\;\underset{T\times 2}{\mathrm{W_t}}\right] \;= \mathrm{tr}\left[\underset{2\times 2}{\mathrm{L}}\right]$$

This implies a basis of spatial and temporal features $\mathrm{A_s}$ and $\mathrm{A_t}$ such that

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}), \;\; \text{where } \tilde{\mathrm{X}} = \underset{S\times 2}{\mathrm{A_s}}\;\overset{\ddots}{\mathrm{L}}\;\underset{2\times T}{\mathrm{A_t}^\top}\;.$$

$$\mathrm{tr}\,[\mathrm{L}] = \mathrm{tr}\left[\mathrm{W_s}^\top \mathrm{A_s}\,\overset{\ddots}{\mathrm{L}}\,\mathrm{A_t}^\top \mathrm{W_t}\right], \;\; \text{satisfied if } \mathrm{W_s}^\top \mathrm{A_s} = \mathrm{W_t}^\top \mathrm{A_t} = \underset{2\times 2}{\mathrm{I}}$$

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\left[\mathrm{G}^\top \mathrm{X}\right]$$

$$= \mathrm{tr}\left[\mathrm{W_t} \mathrm{W_s}^\top \mathrm{X}\right] = \mathrm{tr}\left[\underset{F \times S}{\mathrm{W_s}^\top} \; \mathrm{X} \; \underset{T \times F}{\mathrm{W_t}}\right] = \mathrm{tr}\left[\underset{F \times F}{\mathrm{L}}\right]$$

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}) \quad \text{where} \quad \tilde{\mathrm{X}} = \mathrm{A_s} \overset{\ddots}{\mathrm{L}} \mathrm{A_t}^\top \quad \text{such that} \quad \mathrm{W_s}^\top \mathrm{A_s} = \mathrm{W_t}^\top \mathrm{A_t} = \underset{F \times F}{\mathrm{I}}$$

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t} \mathrm{W_s}^\top \mathrm{X}] = \mathrm{tr}\left[\underset{F \times S}{\mathrm{W_s}^\top} \; \mathrm{X} \; \underset{T \times F}{\mathrm{W_t}}\right] = \mathrm{tr}\left[\underset{F \times F}{\mathrm{L}}\right]$$

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}) \quad \text{where} \quad \tilde{\mathrm{X}} = \mathrm{A_s}\, \overset{\ddots}{\mathrm{L}}\, \mathrm{A_t}^\top \quad \text{such that} \quad \mathrm{W_s}^\top \mathrm{A_s} = \mathrm{W_t}^\top \mathrm{A_t} = \underset{F \times F}{\mathrm{I}}$$

How do we obtain $\mathrm{W_s}, \mathrm{W_t}$? Simple idea: feed X into a classifier and look at the **singular value decomposition** of the resulting G:

$$\mathrm{G} = \mathrm{R_s}\, \overset{\ddots}{\mathrm{S}}{}^{2}\, \mathrm{R_t}^\top, \qquad \text{where} \quad \mathrm{R_s}^\top \mathrm{R_s} = \mathrm{R_t}^\top \mathrm{R_t} = \mathrm{I}$$

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t W_s^\top X}] \;=\; \mathrm{tr}\left[\underset{F\times S}{\mathrm{W_s^\top}}\;\mathrm{X}\;\underset{T\times F}{\mathrm{W_t}}\right] \;=\; \mathrm{tr}\left[\underset{F\times F}{\mathrm{L}}\right]$$

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}) \quad \text{where} \quad \tilde{\mathrm{X}} = \mathrm{A_s}\,\overset{\ddots}{\mathrm{L}}\,\mathrm{A_t^\top} \quad \text{such that} \quad \mathrm{W_s^\top A_s} = \mathrm{W_t^\top A_t} = \underset{F\times F}{\mathrm{I}}$$

How do we obtain $\mathrm{W_s}, \mathrm{W_t}$? Simple idea: feed X into a classifier and look at the **singular value decomposition** of the resulting G:

$$\mathrm{G} = \mathrm{R_s}\,\overset{\ddots}{\mathrm{S}}{}^{2}\,\mathrm{R_t^\top}, \qquad \text{where} \quad \mathrm{R_s^\top R_s} = \mathrm{R_t^\top R_t} = \mathrm{I}$$

Thus

$$\mathrm{W_s} = \mathrm{R_s}\,\overset{\ddots}{\mathrm{S}}\,, \qquad \mathrm{W_t} = \mathrm{R_t}\,\overset{\ddots}{\mathrm{S}}\,,$$

$$\mathrm{A_s} = \mathrm{R_s}\overset{\ddots}{\mathrm{S}}{}^{-1}, \qquad \mathrm{A_t} = \mathrm{R_t}\overset{\ddots}{\mathrm{S}}{}^{-1}.$$

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t W_s^\top X}] \;= \mathrm{tr}\left[\underset{F\times S}{\mathrm{W_s^\top}} \;\; \mathrm{X} \;\; \underset{T\times F}{\mathrm{W_t}}\right] \;= \mathrm{tr}\left[\underset{F\times F}{\mathrm{L}}\right]$$

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}) \quad \text{where} \quad \tilde{\mathrm{X}} = \mathrm{A_s}\,\overset{\ddots}{\mathrm{L}}\,\mathrm{A_t^\top} \quad \text{such that} \quad \mathrm{W_s^\top A_s} = \mathrm{W_t^\top A_t} = \underset{F\times F}{\mathrm{I}}$$

How do we obtain $\mathrm{W_s}, \mathrm{W_t}$? Simple idea: feed X into a classifier and look at the **singular value decomposition** of the resulting G:

$$\mathrm{G} = \mathrm{R_s}\,\overset{\ddots}{\mathrm{S}}{}^{2}\,\mathrm{R_t^\top}\,, \qquad \text{where} \;\; \mathrm{R_s^\top R_s} = \mathrm{R_t^\top R_t} = \mathrm{I}$$

If you want to reduce the rank, you can throw away columns of $\mathrm{R_s}$ and $\mathrm{R_t}$ and recompute G.

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t W_s^\top X}] \;= \mathrm{tr}\left[\underset{F\times S}{\mathrm{W_s^\top}} \;\; \mathrm{X} \;\; \underset{T\times F}{\mathrm{W_t}}\right] \;= \mathrm{tr}\left[\underset{F\times F}{\mathrm{L}}\right]$$

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}) \quad \text{where} \quad \tilde{\mathrm{X}} = \mathrm{A_s}\,\overset{\ddots}{\mathrm{L}}\,\mathrm{A_t^\top} \quad \text{such that} \quad \mathrm{W_s^\top A_s} = \mathrm{W_t^\top A_t} = \underset{F\times F}{\mathrm{I}}$$

How do we obtain $\mathrm{W_s}, \mathrm{W_t}$? Simple idea: feed X into a classifier and look at the **singular value decomposition** of the resulting G:

$$\mathrm{G} = \mathrm{R_s}\,\overset{\ddots}{\mathrm{S}}{}^{2}\,\mathrm{R_t^\top}, \qquad \text{where} \;\; \mathrm{R_s^\top R_s} = \mathrm{R_t^\top R_t} = \mathrm{I}$$

Alternatively: obtain a low-rank G in the first place, by using a classifier which is regularized by penalizing the rank of G:

Tomioka and Aihara (ICML 2007) give a (convex!) formulation for LR.

Bilinear Discriminant Analysis

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G}^\top \mathrm{X}]$$

$$= \mathrm{tr}\,[\mathrm{W_t} \mathrm{W_s}^\top \mathrm{X}] \;=\mathrm{tr}\left[\underset{F\times S}{\mathrm{W_s}^\top}\;\mathrm{X}\;\underset{T\times F}{\mathrm{W_t}}\right] \;=\mathrm{tr}\left[\underset{F\times F}{\mathrm{L}}\right]$$

$$f(\mathrm{X}) = f(\tilde{\mathrm{X}}) \quad \text{where} \quad \tilde{\mathrm{X}} = \mathrm{A_s}\, \overset{\ddots}{\mathrm{L}}\, \mathrm{A_t}^\top \quad \text{such that} \quad \mathrm{W_s}^\top \mathrm{A_s} = \mathrm{W_t}^\top \mathrm{A_t} = \underset{F\times F}{\mathrm{I}}$$

How do we obtain $\mathrm{W_s}, \mathrm{W_t}$? Simple idea: feed X into a classifier and look at the **singular value decomposition** of the resulting G:

$$\mathrm{G} = \mathrm{R_s}\, \overset{\ddots}{\mathrm{S}}{}^{2}\, \mathrm{R_t}^\top , \qquad \text{where} \;\; \mathrm{R_s}^\top \mathrm{R_s} = \mathrm{R_t}^\top \mathrm{R_t} = \mathrm{I}$$

…but neither approach performs at its best if G is found directly in the space of X.

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top \mathrm{X}\,,\quad$ with P such that $\mathrm{E}\{\mathrm{P}^\top\mathrm{X}(\mathrm{P}^\top\mathrm{X})^\top\} = \mathrm{I}$

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top\,\mathrm{X}\,,$ with P such that $\mathrm{E}\{\mathrm{P}^\top\mathrm{X}\mathrm{X}^\top\mathrm{P}\} = \mathrm{I}$

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top\,\mathrm{X}\,,$ with P such that $\mathrm{P}^\top \mathrm{E}\{\mathrm{X}\mathrm{X}^\top\}\mathrm{P} = \mathrm{I}$

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top \mathrm{X}\,,$ with P such that $\mathrm{P}^\top \Sigma \mathrm{P} = \mathrm{I}$

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top\, \mathrm{X}\,,\quad$ with P such that $\Sigma = \mathrm{P}^{-\top}\mathrm{P}^{-1}$

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top\, \mathrm{X}\,,\quad$ with (e.g.) $\mathrm{P} = \Sigma^{-\frac{1}{2}}$

Spatial Whitening in ERP classification

Spatial whitening (decorrelating) transformations are very common in EEG analysis, as a partial way of undoing the volume conduction effect which causes all sensor outputs to be highly correlated. Examples:

- surface Laplacian filters (approximate)
- ICA
- CSP and friends (esp. for extraction of bandpower features)

Generally: $\mathrm{X_P} = \mathrm{P}^\top \, \mathrm{X}\,, \quad$ with (e.g.) $\mathrm{P} = \Sigma^{-\frac{1}{2}}$

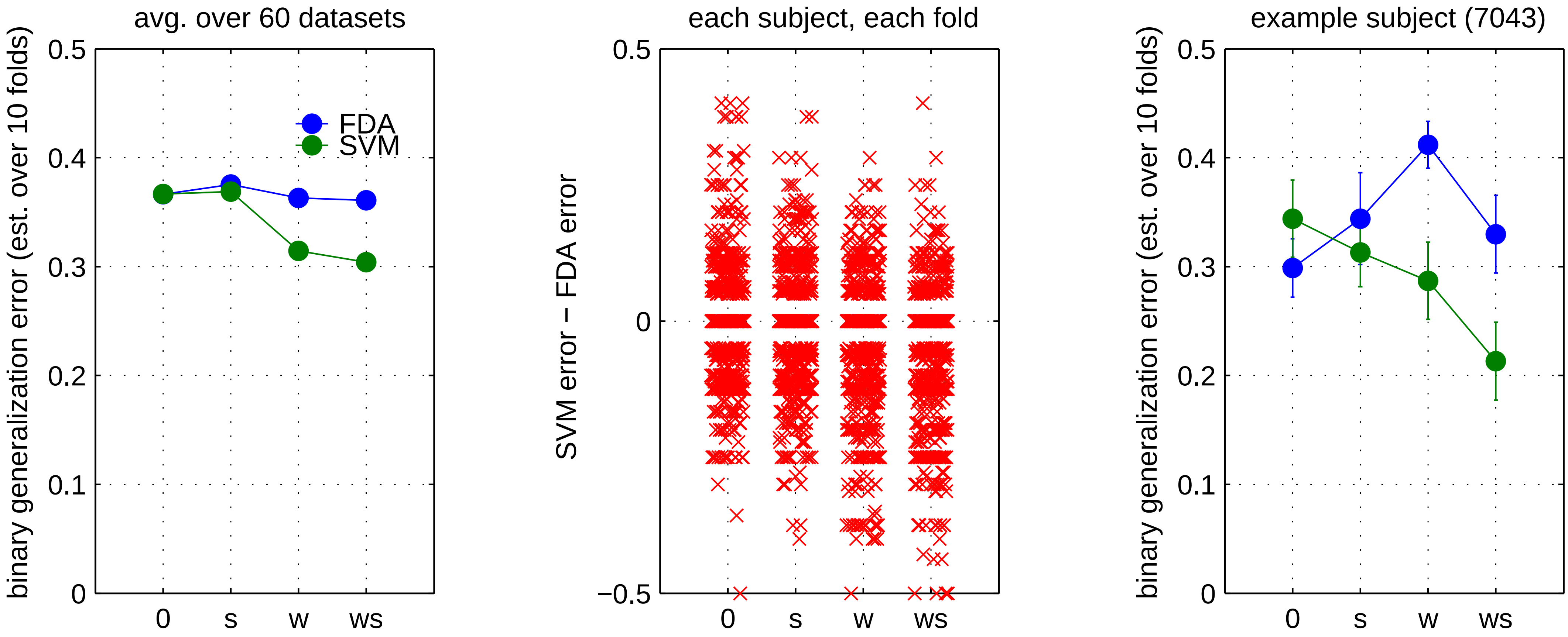
Even in linear classification, non-orthonormal transformations of the data affect the classifier's **regularization environment** and can lead to very different results.

Spatial Whitening in ERP classification

Krusienski et al. (J. Neural Eng. 2006) report in visual ERP classification (in a grid-speller) that Fisher's (unregularized) Linear Discriminant Analysis performs as well as, or better than, Support Vector Machines.

We suggest that this is because no decorrelation was performed. The lack of whitening masked the potential benefits of regularization:

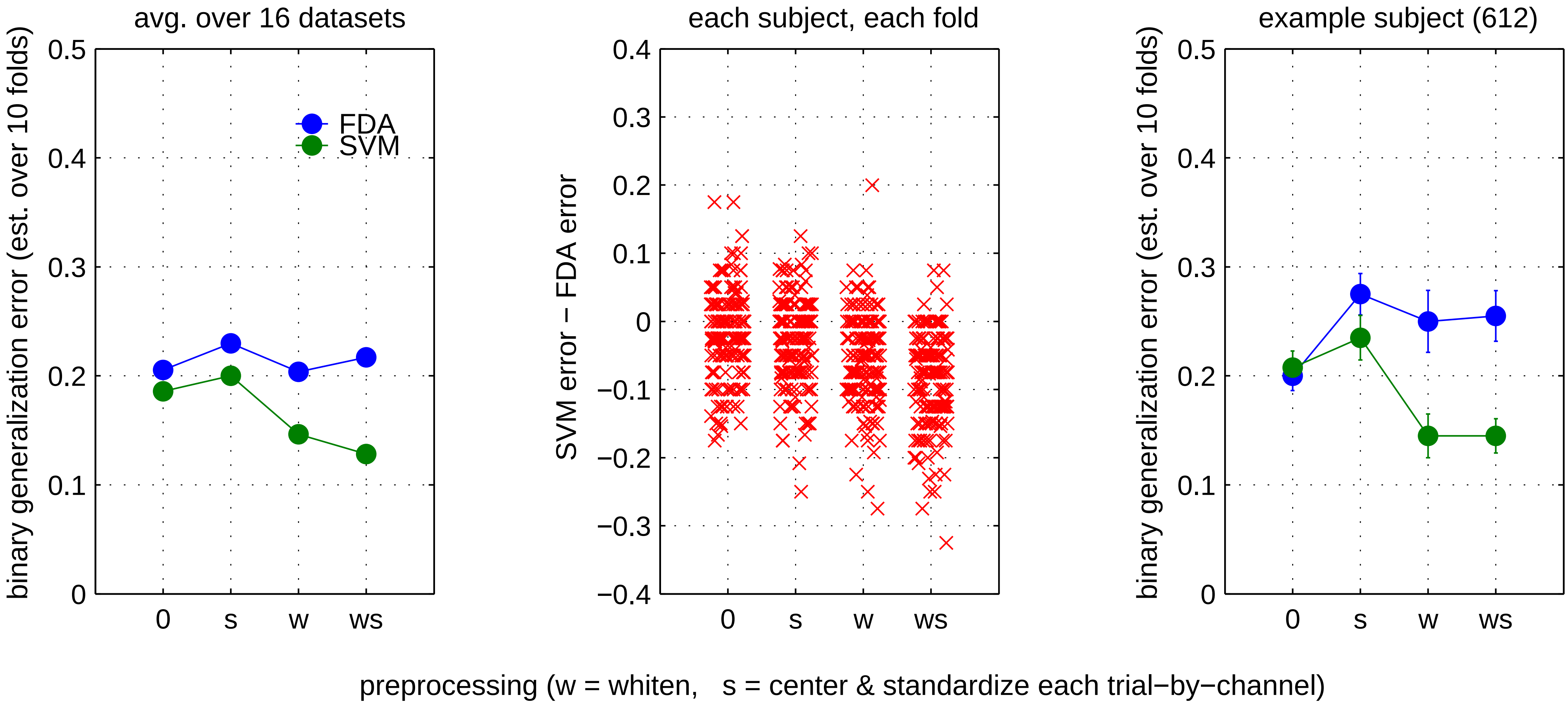
Spatial Whitening in ERP classification



preprocessing (w = whiten, s = center & standardize each trial−by−channel)

Felix Biessmann's visual speller data (10 subjects x 6 stimulus conditions), offline analysis

Spatial Whitening in ERP classification

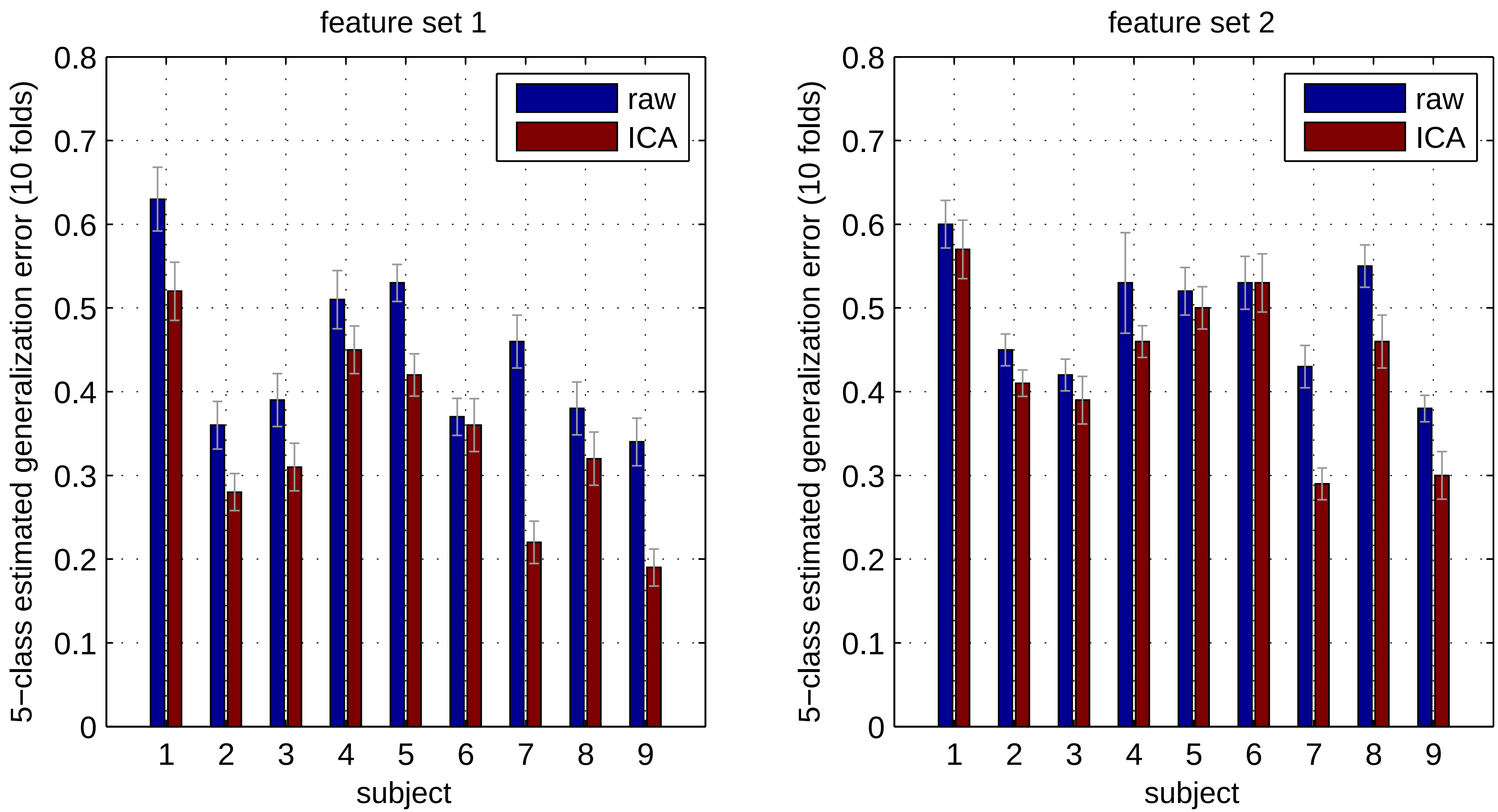


auditory ERP data, offline analysis

Spatial Whitening in ERP classification

SVM results on tactile MEG data

(Cornelius Raths' Diploma Thesis 2007, paper in preparation):



Filters and Patterns

Learn the classifier weights on the *preconditioned* data $X_P = P_s^\top X P_t$:

$$
\begin{aligned}
f(X) &= \operatorname{tr}\left[G_P^\top \; X_P\right] &&= \operatorname{tr}\left[G^\top \; X\right] \\
&= \operatorname{tr}\left[G_P^\top \; P_s^\top X P_t\right] &&= \operatorname{tr}\left[P_t G_P^\top P_s^\top \; X\right]
\end{aligned}
$$

Filters and Patterns

Learn the classifier weights on the *preconditioned* data $\mathrm{X_P} = \mathrm{P_s^\top X P_t}$:

$$
\begin{aligned}
f(\mathrm{X}) &= \mathrm{tr}\,[\mathrm{G_P^\top}\ \mathrm{X_P}] &&= \mathrm{tr}\,[\mathrm{G^\top}\ \mathrm{X}] \\
&= \mathrm{tr}\,[\mathrm{G_P^\top}\ \mathrm{P_s^\top X P_t}] &&= \mathrm{tr}\,[\mathrm{P_t G_P^\top P_s^\top}\ \mathrm{X}]
\end{aligned}
$$

...and decompose $\mathrm{G_P}$ instead of G:

$$\mathrm{G_P} = \mathrm{R_s}\ \mathrm{S}^{2}\ \mathrm{R_t^\top}$$

Learn the classifier weights on the *preconditioned* data $\mathrm{X_P} = \mathrm{P_s^\top X P_t}$:

$$\begin{aligned} f(\mathrm{X}) &= \mathrm{tr}\,[\mathrm{G_P^\top}\ \ \mathrm{X_P}] &&= \mathrm{tr}\,[\mathrm{G^\top}\ \ \mathrm{X}] \\ &= \mathrm{tr}\,[\mathrm{G_P^\top}\ \ \mathrm{P_s^\top X P_t}] &&= \mathrm{tr}\,[\mathrm{P_t G_P^\top P_s^\top}\ \ \mathrm{X}] \end{aligned}$$

...and decompose $\mathrm{G_P}$ instead of G:

$$\mathrm{G_P} = \mathrm{R_s}\ \mathrm{S}^{\ddots\,2}\ \mathrm{R_t^\top}$$

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G_P^\top}\ \ \mathrm{P_s^\top X P_t}] = \mathrm{tr}\left[\ \overset{\ddots}{\mathrm{S}}\ \mathrm{R_s^\top P_s^\top}\ \mathrm{X}\ \mathrm{P_t R_t}\ \overset{\ddots}{\mathrm{S}}\ \right]$$

Filters and Patterns

Learn the classifier weights on the *preconditioned* data $\mathrm{X_P} = \mathrm{P_s^\top X P_t}$:

$$\begin{aligned} f(\mathrm{X}) &= \mathrm{tr}\,[\mathrm{G_P^\top \;\; X_P}] &&= \mathrm{tr}\,[\mathrm{G^\top \;\; X}] \\ &= \mathrm{tr}\,[\mathrm{G_P^\top \;\; P_s^\top X P_t}] &&= \mathrm{tr}\,[\mathrm{P_t G_P^\top P_s^\top \;\; X}] \end{aligned}$$

…and decompose $\mathrm{G_P}$ instead of G:

$$\mathrm{G_P} = \mathrm{R_s} \, \overset{\ddots}{\mathrm{S}}{}^{2} \, \mathrm{R_t^\top}$$

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{G_P^\top \;\; P_s^\top X P_t}] = \mathrm{tr}\left[\; \overset{\ddots}{\mathrm{S}} \, \mathrm{R_s^\top P_s^\top \; X \; P_t R_t} \, \overset{\ddots}{\mathrm{S}} \;\right]$$

So

$$\mathrm{W_s} = \mathrm{P_s R_s} \, \overset{\ddots}{\mathrm{S}} \,, \qquad \mathrm{W_t} = \mathrm{P_t R_t} \, \overset{\ddots}{\mathrm{S}} \,,$$

$$\mathrm{A_s} = \mathrm{P_s^{-\top} R_s} \overset{\ddots}{\mathrm{S}}{}^{-1} \,, \qquad \mathrm{A_t} = \mathrm{P_t^{-\top} R_t} \overset{\ddots}{\mathrm{S}}{}^{-1} \,.$$

Preconditioning and Regularization

Learn the classifier weights on the *preconditioned* data $\mathrm{X_P} = \mathrm{P_s^\top X P_t}$:

$$
\begin{aligned}
f(\mathrm{X}) &= \mathrm{tr}\left[\mathrm{G_P^\top} \ \ \mathrm{X_P}\right] &&= \mathrm{tr}\left[\mathrm{G^\top} \ \ \mathrm{X}\right] \\
&= \mathrm{tr}\left[\mathrm{G_P^\top} \ \ \mathrm{P_s^\top X P_t}\right] &&= \mathrm{tr}\left[\mathrm{P_t G_P^\top P_s^\top} \ \ \mathrm{X}\right]
\end{aligned}
$$

Preconditioning and Regularization

Learn the classifier weights on the *preconditioned* data $\mathrm{X_P} = \mathrm{P_s^\top X P_t}$:

$$\begin{aligned} f(\mathrm{X}) &= \mathrm{tr}\,[\mathrm{G_P^\top}\ \mathrm{X_P}] &&= \mathrm{tr}\,[\mathrm{G^\top}\ \mathrm{X}] \\ &= \mathrm{tr}\,[\mathrm{G_P^\top}\ \mathrm{P_s^\top X P_t}] &&= \mathrm{tr}\,[\mathrm{P_t G_P^\top P_s^\top}\ \mathrm{X}] \end{aligned}$$

the classifier (LR, SVM,...) optimizes $\mathrm{G_P}$ instead of G. This does not change $f(\mathrm{X})$, so the loss term is unaffected. However, assuming an L2 regularizer, the classifier's objective becomes:

$$\lambda\,\mathrm{tr}\,[\mathrm{G_P^\top G_P}] + \sum_i \mathcal{L}\left\{ y^{(i)} f(\mathrm{X}^{(i)}) \right\},$$

Preconditioning and Regularization

Learn the classifier weights on the *preconditioned* data $\mathrm{X_P} = \mathrm{P_s^\top X P_t}$:

$$\begin{aligned} f(\mathrm{X}) &= \mathrm{tr}\left[\mathrm{G_P^\top}\ \mathrm{X_P}\right] &= \mathrm{tr}\left[\mathrm{G^\top}\ \mathrm{X}\right] \\ &= \mathrm{tr}\left[\mathrm{G_P^\top}\ \mathrm{P_s^\top X P_t}\right] &= \mathrm{tr}\left[\mathrm{P_t G_P^\top P_s^\top}\ \mathrm{X}\right] \end{aligned}$$

the classifier (LR, SVM,...) optimizes $\mathrm{G_P}$ instead of G. This does not change $f(\mathrm{X})$, so the loss term is unaffected. However, assuming an L2 regularizer, the classifier's objective becomes:

$$\lambda\ \mathrm{tr}\left[\mathrm{G_P^\top G_P}\right] + \sum_i \mathcal{L}\left\{y^{(i)} f(\mathrm{X}^{(i)})\right\},$$

in which further expansion, substitution and reordering gives

$$\mathrm{tr}\left[\mathrm{G_P^\top G_P}\right] = \mathrm{tr}\left[\mathrm{S}^4\ \mathrm{A_t^\top P_t P_t^\top A_t}\quad \mathrm{S}^4\ \mathrm{A_s^\top P_s P_s^\top A_s}\right].$$

Preconditioning and Prior knowledge

The regularization term can be written as

$$\mathrm{tr}\left[\mathrm{S}^4\ \mathrm{A}_\mathrm{t}^\top \Sigma_\mathrm{t}^{-1} \mathrm{A}_\mathrm{t} \quad \mathrm{S}^4\ \mathrm{A}_\mathrm{s}^\top \Sigma_\mathrm{s}^{-1} \mathrm{A}_\mathrm{s}\right] \qquad \text{where} \qquad \Sigma_\mathrm{t}^{-1} = \mathrm{P}_\mathrm{t}\mathrm{P}_\mathrm{t}^\top \ \text{ and } \ \Sigma_\mathrm{s}^{-1} = \mathrm{P}_\mathrm{s}\mathrm{P}_\mathrm{s}^\top\ .$$

It contains terms which look a little like logged Gaussian prior probabilities over the spatial and temporal patterns, (not incorporated in the usual way, though: cross-terms for $F > 1$, and multiplication of spatial and temporal terms).

Preconditioning and Prior knowledge

The regularization term can be written as

$$\mathrm{tr}\left[\mathrm{S}^4\ \mathrm{A}_\mathrm{t}^\top \Sigma_\mathrm{t}^{-1} \mathrm{A}_\mathrm{t} \quad \mathrm{S}^4\ \mathrm{A}_\mathrm{s}^\top \Sigma_\mathrm{s}^{-1} \mathrm{A}_\mathrm{s}\right] \qquad \text{where} \qquad \Sigma_\mathrm{t}^{-1} = \mathrm{P}_\mathrm{t}\mathrm{P}_\mathrm{t}^\top \ \text{ and } \ \Sigma_\mathrm{s}^{-1} = \mathrm{P}_\mathrm{s}\mathrm{P}_\mathrm{s}^\top\ .$$

It contains terms which look a little like logged Gaussian prior probabilities over the spatial and temporal patterns, (not incorporated in the usual way, though: cross-terms for $F > 1$, and multiplication of spatial and temporal terms).

Some alternative approaches use direct prior probabilities on the spatial patterns or filters, for example to smooth them (e.g. Dyrholm et al. JMLR 2007, Dyrholm and Parra, Proc. IEEE EMBS 2006, Farquhar et al. Applied Neuroscience Conference 2007).

Similar, though not identical, effects are achieved by suitable choice of Σ_s and Σ_t in the preconditioning approach above, with the advantage that convex formulations can be applied for finding G_P.

Preconditioning and Prior knowledge

The regularization term can be written as

$$\mathrm{tr}\left[\mathrm{S}^4\ \mathrm{A}_\mathrm{t}^\top \Sigma_\mathrm{t}^{-1} \mathrm{A}_\mathrm{t} \quad \mathrm{S}^4\ \mathrm{A}_\mathrm{s}^\top \Sigma_\mathrm{s}^{-1} \mathrm{A}_\mathrm{s}\right] \qquad \text{where} \qquad \Sigma_\mathrm{t}^{-1} = \mathrm{P}_\mathrm{t}\mathrm{P}_\mathrm{t}^\top \ \text{ and } \ \Sigma_\mathrm{s}^{-1} = \mathrm{P}_\mathrm{s}\mathrm{P}_\mathrm{s}^\top \ .$$

It contains terms which look a little like logged Gaussian prior probabilities over the spatial and temporal patterns, (not incorporated in the usual way, though: cross-terms for $F > 1$, and multiplication of spatial and temporal terms).

Some alternative approaches use direct prior probabilities on the spatial patterns or filters, for example to smooth them (e.g. Dyrholm et al. JMLR 2007, Dyrholm and Parra, Proc. IEEE EMBS 2006, Farquhar et al. Applied Neuroscience Conference 2007).

Similar, though not identical, effects are achieved by suitable choice of Σ_s and Σ_t in the preconditioning approach above, with the advantage that convex formulations can be applied for finding G_P.

Choosing Σ_s equal to the EEG or MEG sensor covariance matrix seems sensible, since it means the spatial basis functions by which we represent $\tilde{\mathrm{X}}$ reflect realistic EEG- or MEG-like volume-conduction properties.

A Σ_t may also be chosen as a prior which smooths the temporal patterns, although to save computation time one can also simply smooth and downsample the data.

Preconditioning and Prior knowledge

The regularization term can be written as

$$\mathrm{tr}\left[\mathrm{S}^4\ \mathrm{A}_\mathrm{t}^\top \Sigma_\mathrm{t}^{-1} \mathrm{A}_\mathrm{t} \quad \mathrm{S}^4\ \mathrm{A}_\mathrm{s}^\top \Sigma_\mathrm{s}^{-1} \mathrm{A}_\mathrm{s}\right] \qquad \text{where} \qquad \Sigma_\mathrm{t}^{-1} = \mathrm{P}_\mathrm{t}\mathrm{P}_\mathrm{t}^\top \ \text{ and } \ \Sigma_\mathrm{s}^{-1} = \mathrm{P}_\mathrm{s}\mathrm{R}_\mathrm{s}\mathrm{R}_\mathrm{s}^\top \mathrm{P}_\mathrm{s}^\top\ .$$

It contains terms which look a little like logged Gaussian prior probabilities over the spatial and temporal patterns, (not incorporated in the usual way, though: cross-terms for $F > 1$, and multiplication of spatial and temporal terms).

Some alternative approaches use direct prior probabilities on the spatial patterns or filters, for example to smooth them (e.g. Dyrholm et al. JMLR 2007, Dyrholm and Parra, Proc. IEEE EMBS 2006, Farquhar et al. Applied Neuroscience Conference 2007).

Similar, though not identical, effects are achieved by suitable choice of Σ_s and Σ_t in the preconditioning approach above, with the advantage that convex formulations can be applied for finding G_P.

Choosing Σ_s equal to the EEG or MEG sensor covariance matrix seems sensible, since it means the spatial basis functions by which we represent $\tilde{\mathrm{X}}$ reflect realistic EEG- or MEG-like volume-conduction properties.

A Σ_t may also be chosen as a prior which smooths the temporal patterns, although to save computation time one can also simply smooth and downsample the data.

MAX-PLANCK-GESELLSCHAFT

Preconditioning and Prior knowledge

BIOLOGISCHE KYBERNETIK

The regularization term can be written as

$$\mathrm{tr}\left[\mathrm{S}^4\ \mathrm{A}_\mathrm{t}^\top \Sigma_\mathrm{t}^{-1} \mathrm{A}_\mathrm{t} \quad \mathrm{S}^4\ \mathrm{A}_\mathrm{s}^\top \Sigma_\mathrm{s}^{-1} \mathrm{A}_\mathrm{s}\right] \qquad \text{where} \qquad \Sigma_\mathrm{t}^{-1} = \mathrm{P}_\mathrm{t}\mathrm{P}_\mathrm{t}^\top \ \text{ and } \ \Sigma_\mathrm{s}^{-1} = \mathrm{P}_\mathrm{s}\mathrm{P}_\mathrm{s}^\top\ .$$

It contains terms which look a little like logged Gaussian prior probabilities over the spatial and temporal patterns, (not incorporated in the usual way, though: cross-terms for $F > 1$, and multiplication of spatial and temporal terms).

Some alternative approaches use direct prior probabilities on the spatial patterns or filters, for example to smooth them (e.g. Dyrholm et al. JMLR 2007, Dyrholm and Parra, Proc. IEEE EMBS 2006, Farquhar et al. Applied Neuroscience Conference 2007).

Similar, though not identical, effects are achieved by suitable choice of Σ_s and Σ_t in the preconditioning approach above, with the advantage that convex formulations can be applied for finding G_P.

Choosing Σ_s equal to the EEG or MEG sensor covariance matrix seems sensible, since it means the spatial basis functions by which we represent $\tilde{\mathrm{X}}$ reflect realistic EEG- or MEG-like volume-conduction properties.

A Σ_t may also be chosen as a prior which smooths the temporal patterns, although to save computation time one can also simply smooth and downsample the data.

Note that an arbitrary invertible $F \times F$ matrix B may be applied after the filters have been found:

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{B}\ \ \mathrm{S}\mathrm{R}_{\mathrm{s}}^{\top}\mathrm{P}_{\mathrm{s}}^{\top}\ \mathrm{X}\,\mathrm{P}_{\mathrm{t}}\mathrm{R}_{\mathrm{t}}\mathrm{S}\ \ \mathrm{B}^{-1}]\ .$$

Note that an arbitrary invertible $F \times F$ matrix B may be applied after the filters have been found:

$$f(\mathrm{X}) = \mathrm{tr}\left[\mathrm{B}\ \ \mathrm{S}\mathrm{R}_{\mathrm{s}}^{\top}\mathrm{P}_{\mathrm{s}}^{\top}\ \mathrm{X}\ \mathrm{P}_{\mathrm{t}}\mathrm{R}_{\mathrm{t}}\mathrm{S}\ \ \mathrm{B}^{-1}\right].$$

This does not change $f(\mathrm{X})$, but it changes the interpretation of the discriminative components:

Now

$$\mathrm{W}_{\mathrm{s}} = \mathrm{P}_{\mathrm{s}}\mathrm{R}_{\mathrm{s}}\ \overset{\ddots}{\mathrm{S}}\ \mathrm{B}^{\top}, \qquad \mathrm{W}_{\mathrm{t}} = \mathrm{P}_{\mathrm{t}}\mathrm{R}_{\mathrm{t}}\ \overset{\ddots}{\mathrm{S}}\ \mathrm{B}^{-1},$$

$$\mathrm{A}_{\mathrm{s}} = \mathrm{P}_{\mathrm{s}}^{-\top}\mathrm{R}_{\mathrm{s}}\overset{\ddots}{\mathrm{S}}{}^{-1}\mathrm{B}^{-1}, \qquad \mathrm{A}_{\mathrm{t}} = \mathrm{P}_{\mathrm{t}}^{-\top}\mathrm{R}_{\mathrm{t}}\overset{\ddots}{\mathrm{S}}{}^{-1}\mathrm{B}^{\top}.$$

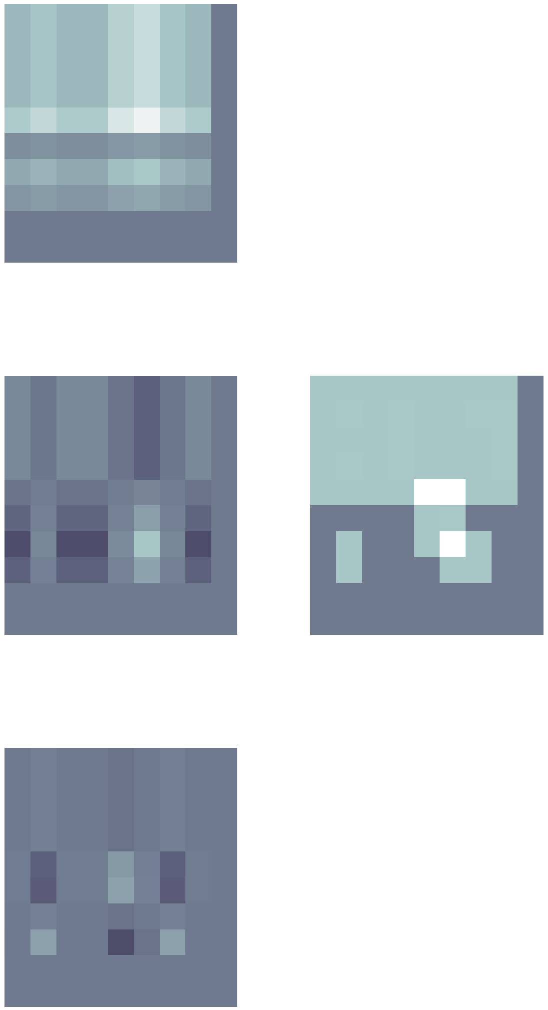
Note that an arbitrary invertible $F \times F$ matrix B may be applied after the filters have been found:

$$f(\mathrm{X}) = \mathrm{tr}\,[\mathrm{B}\ \ \mathrm{S}\mathrm{R}_\mathrm{s}^\top \mathrm{P}_\mathrm{s}^\top\ \mathrm{X}\ \mathrm{P}_\mathrm{t}\mathrm{R}_\mathrm{t}\mathrm{S}\ \ \mathrm{B}^{-1}]\ .$$

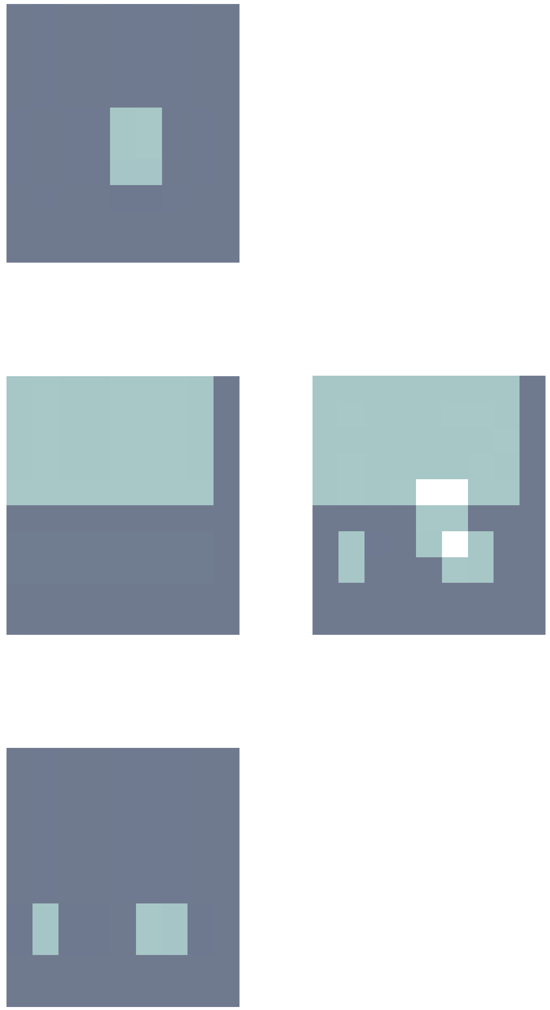
Various approaches are available:

- set B = I (leave filters and patterns to be determined by the interaction between regularizer and preconditioner)
- optimize B such that the $(L_{11}, \ldots, L_{FF})$ are maximally independent across different instances $\mathrm{L}^{(i)}$ (Dyrholm et al., JMLR 2007).
- optimize B according to L1 penalties on A_s and A_t (for sparse basis functions, perhaps confined in space and time) or on W_s and W_t (fewer electrodes to stick).

Sparsification

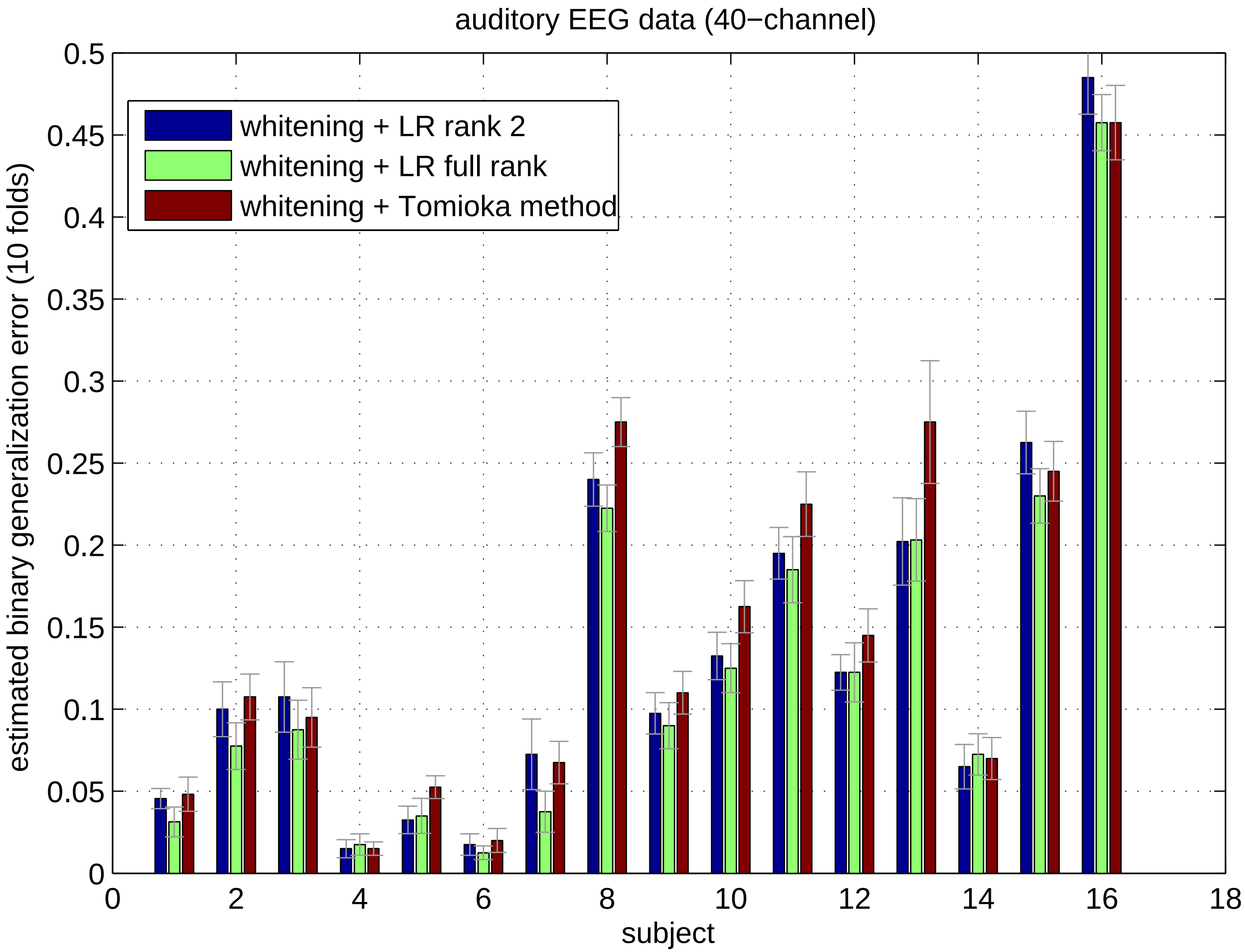


outer products from SVD of composite, $G = USV^\top$

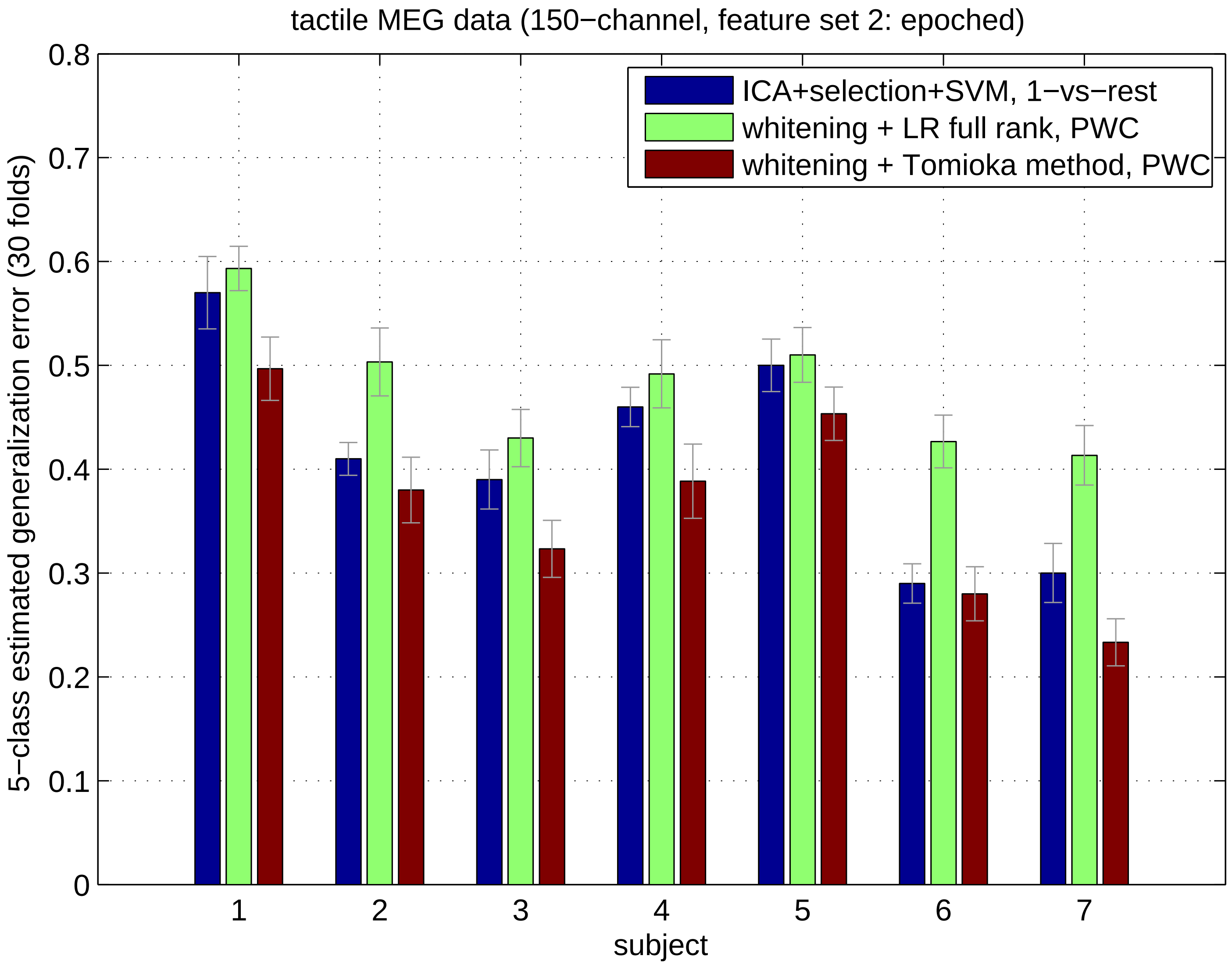


after minimization of L1 norms of $UB^\top$ and VB^{-1}, w.r.t $F \times F$ matrix B

Does a low-rank constraint help?

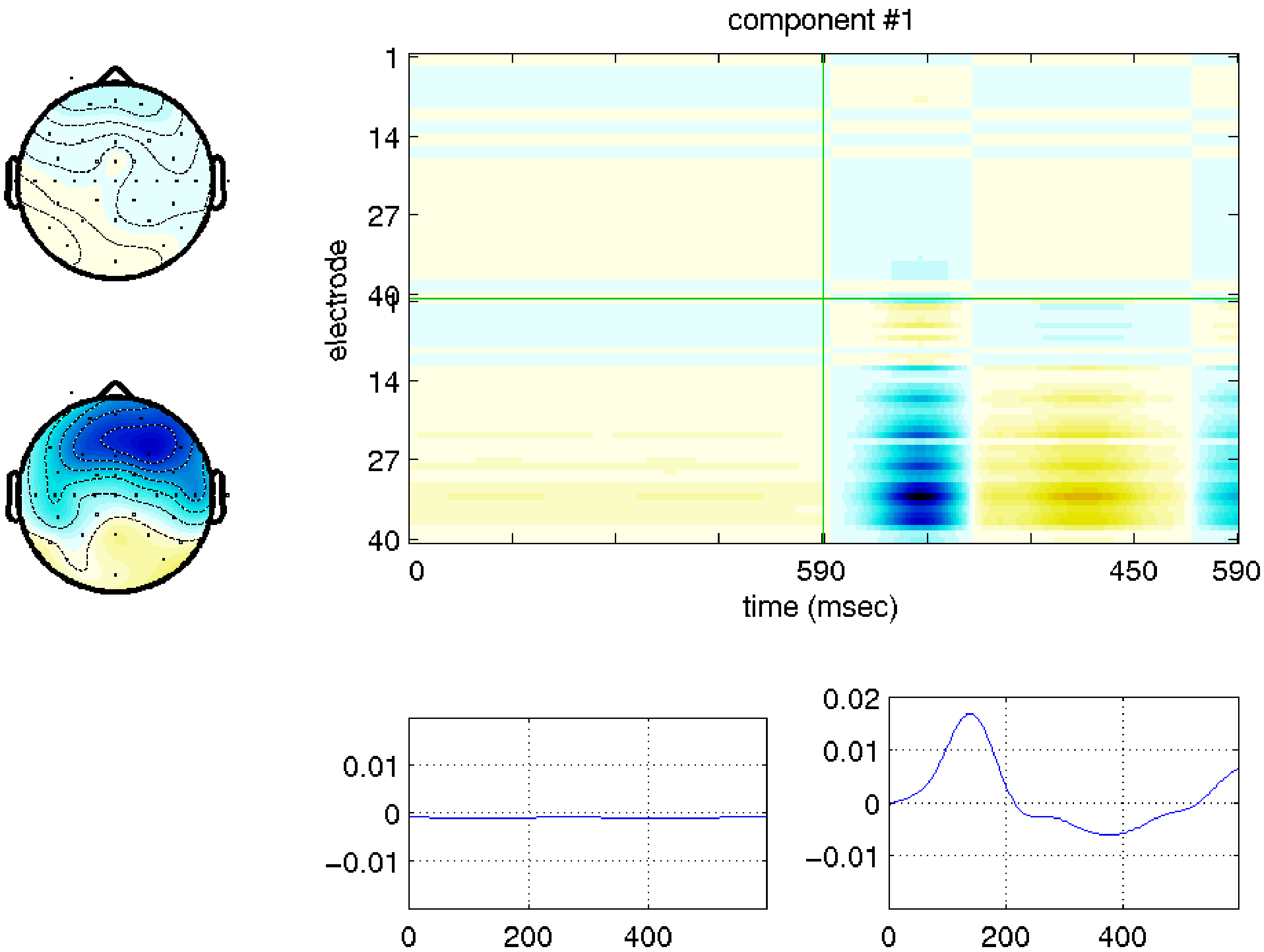


Does a low-rank constraint help?



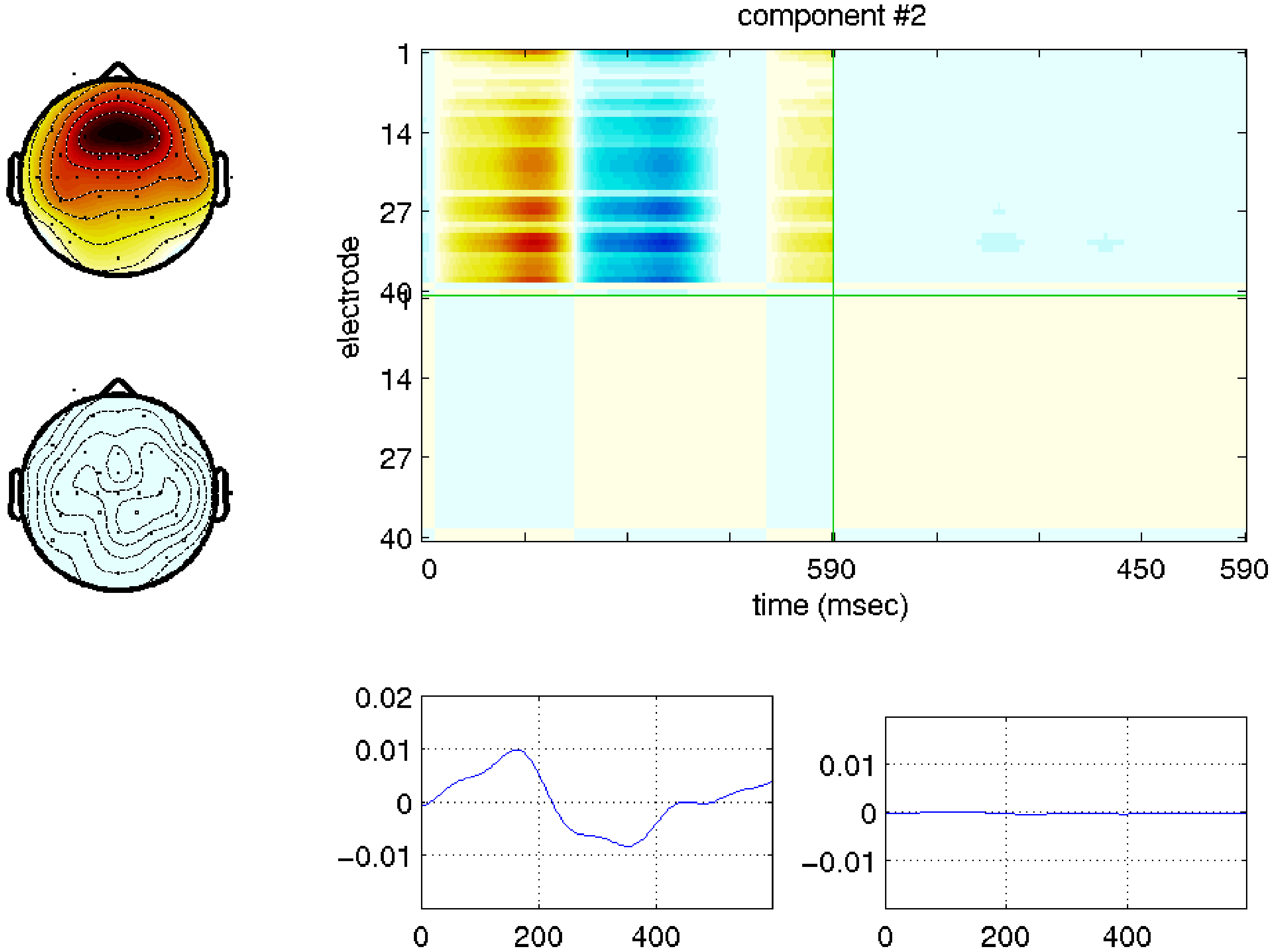
Example decomposition (auditory)

Auditory EEG data:



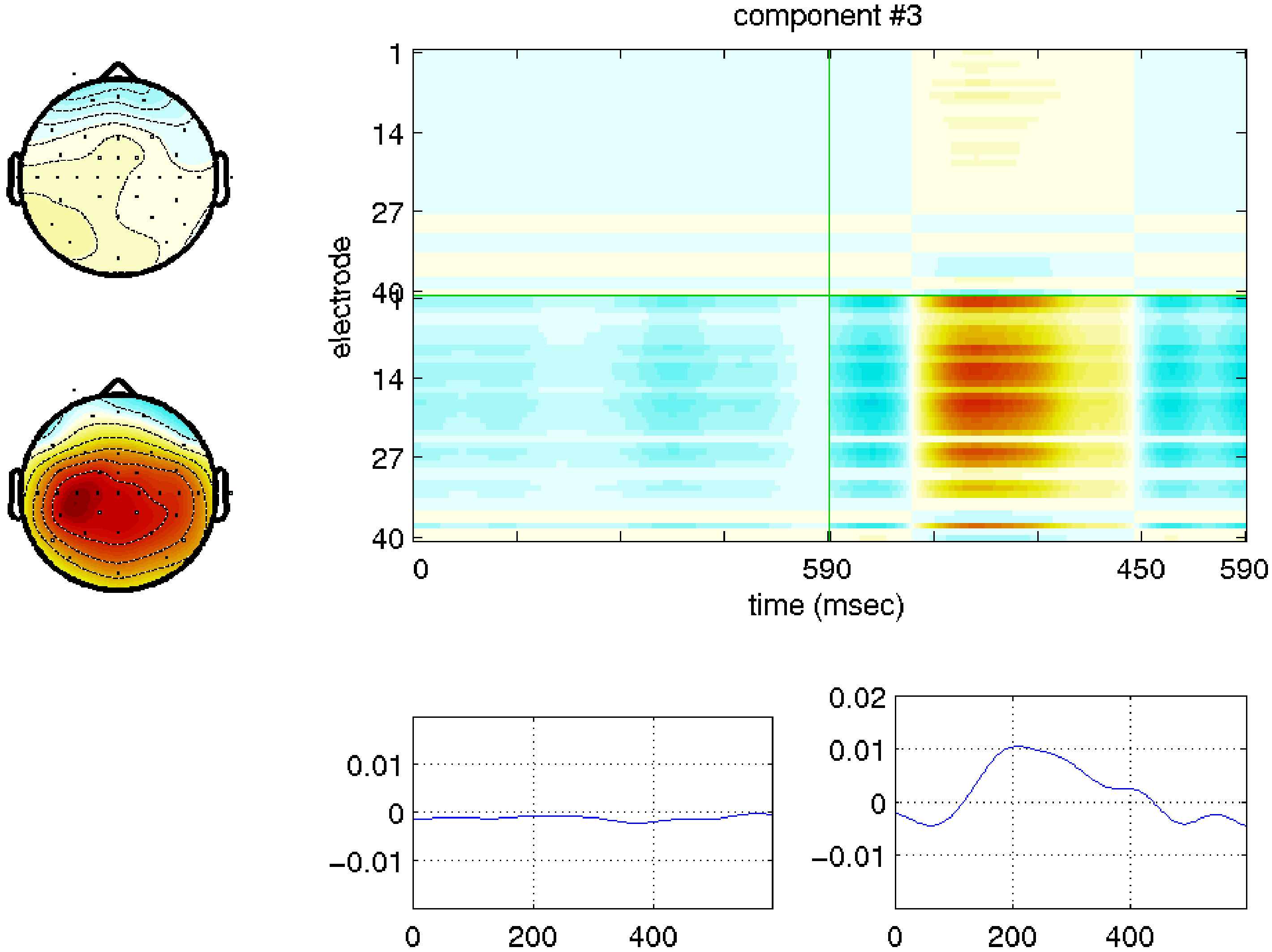
Example decomposition (auditory)

Auditory EEG data:



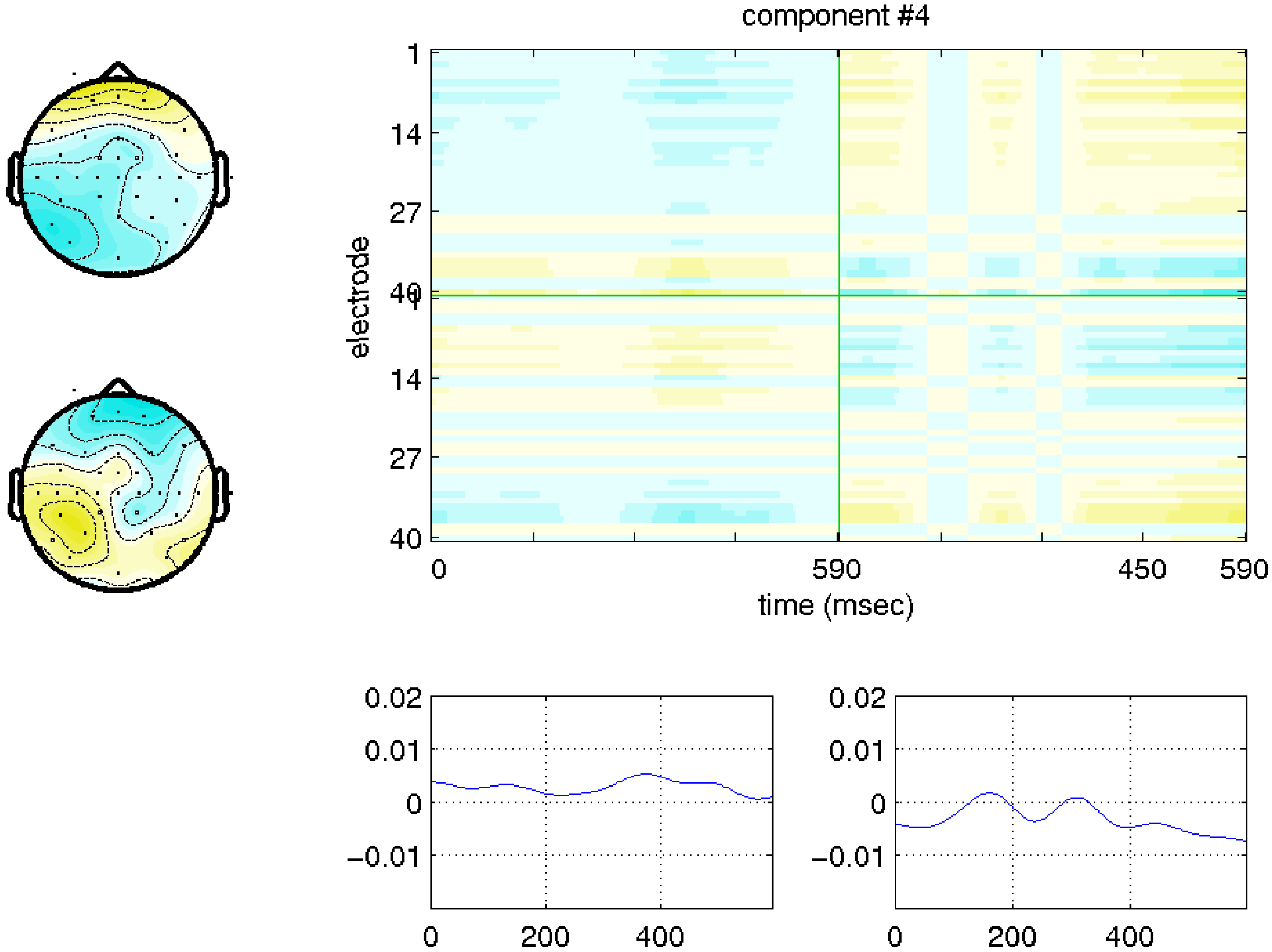
Example decomposition (auditory)

Auditory EEG data:



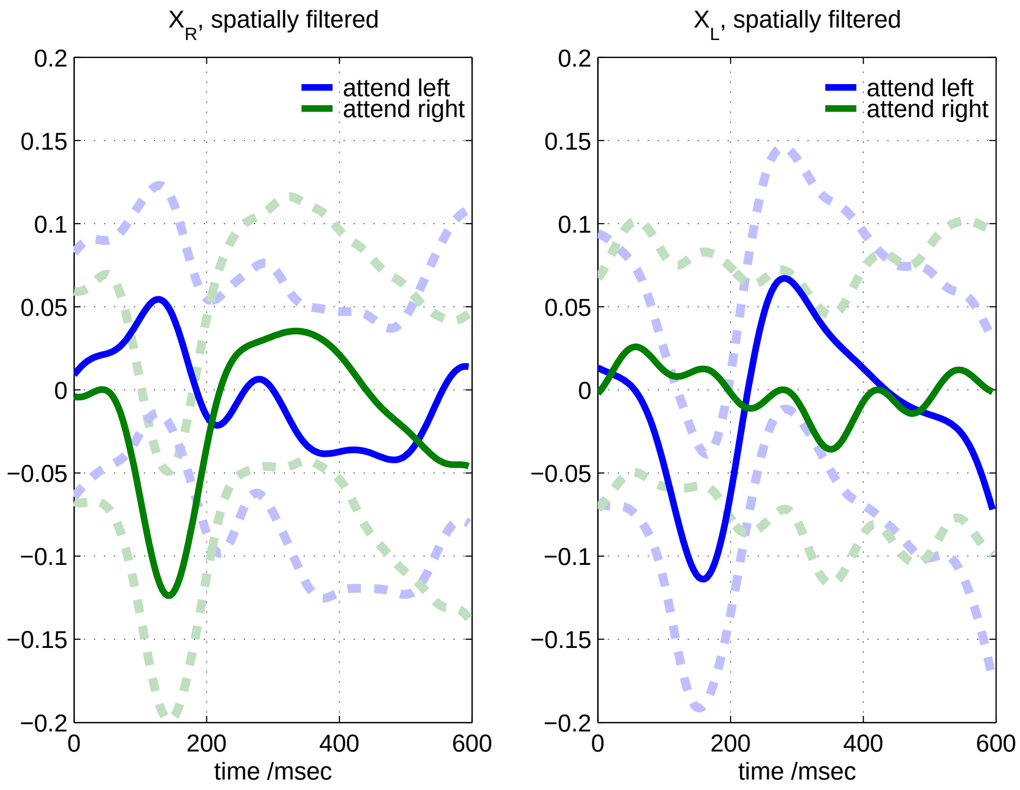
Example decomposition (auditory)

Auditory EEG data:

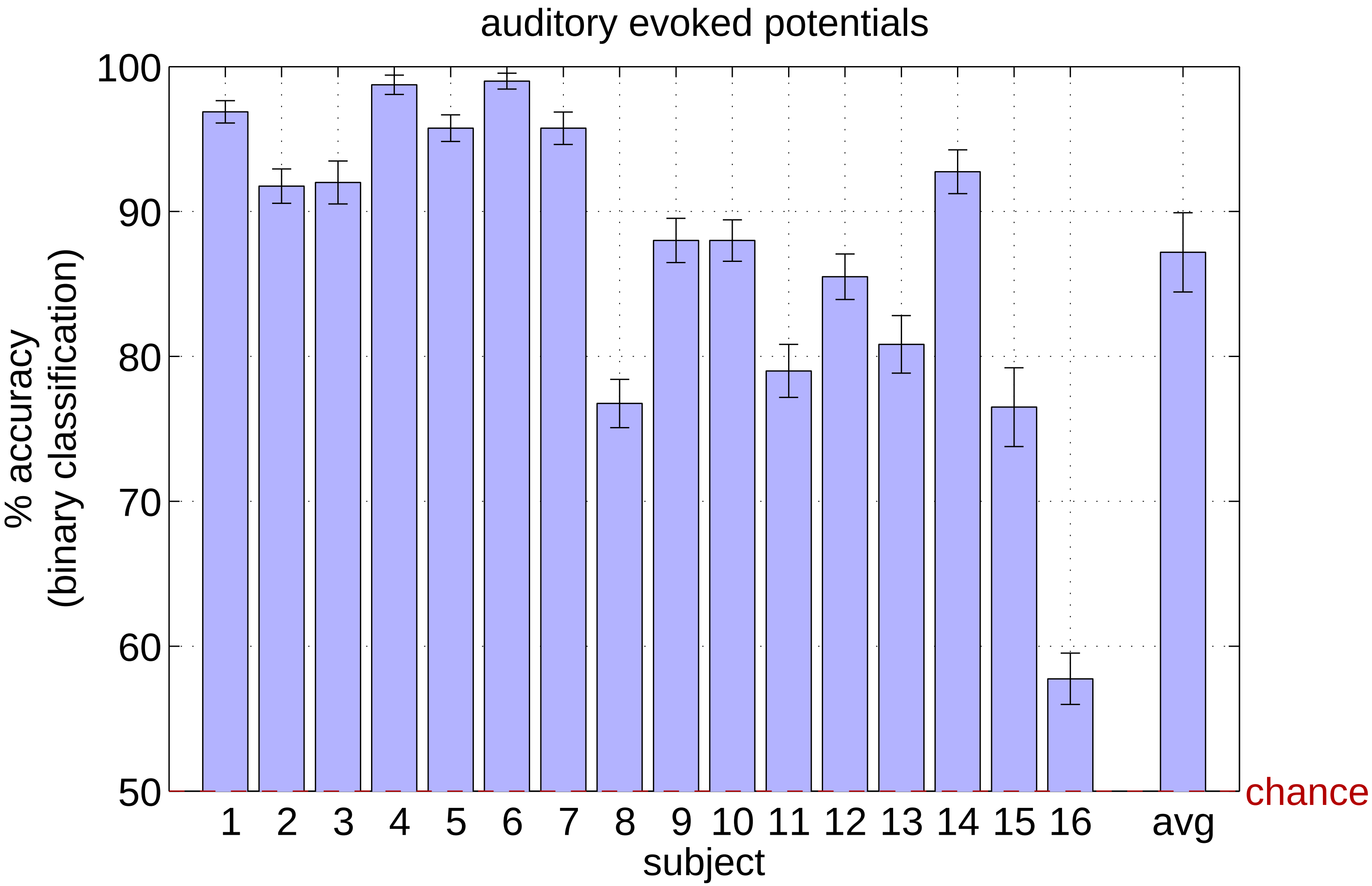


Example decomposition (auditory)

Auditory EEG data:

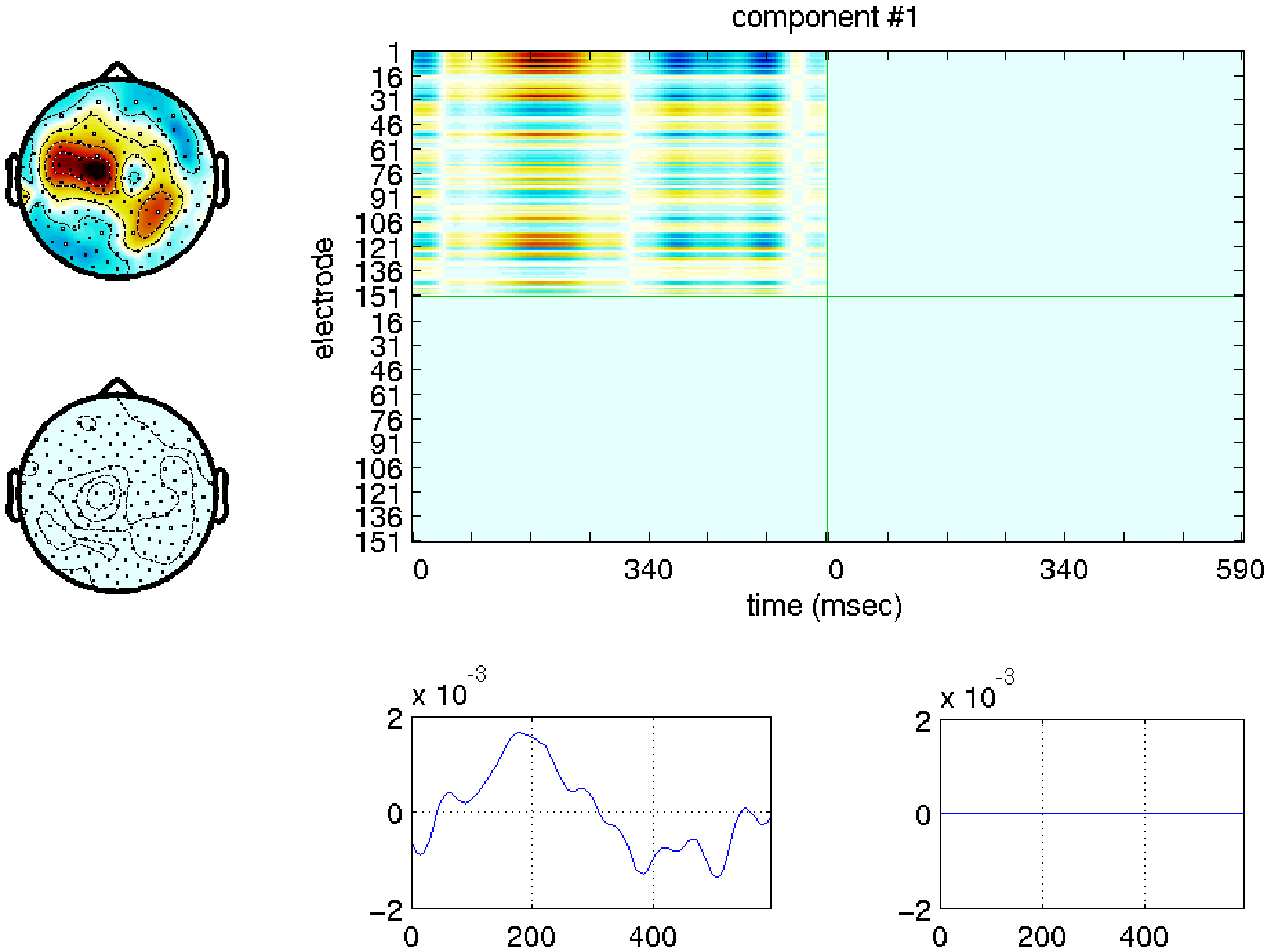


Classification performance (auditory)



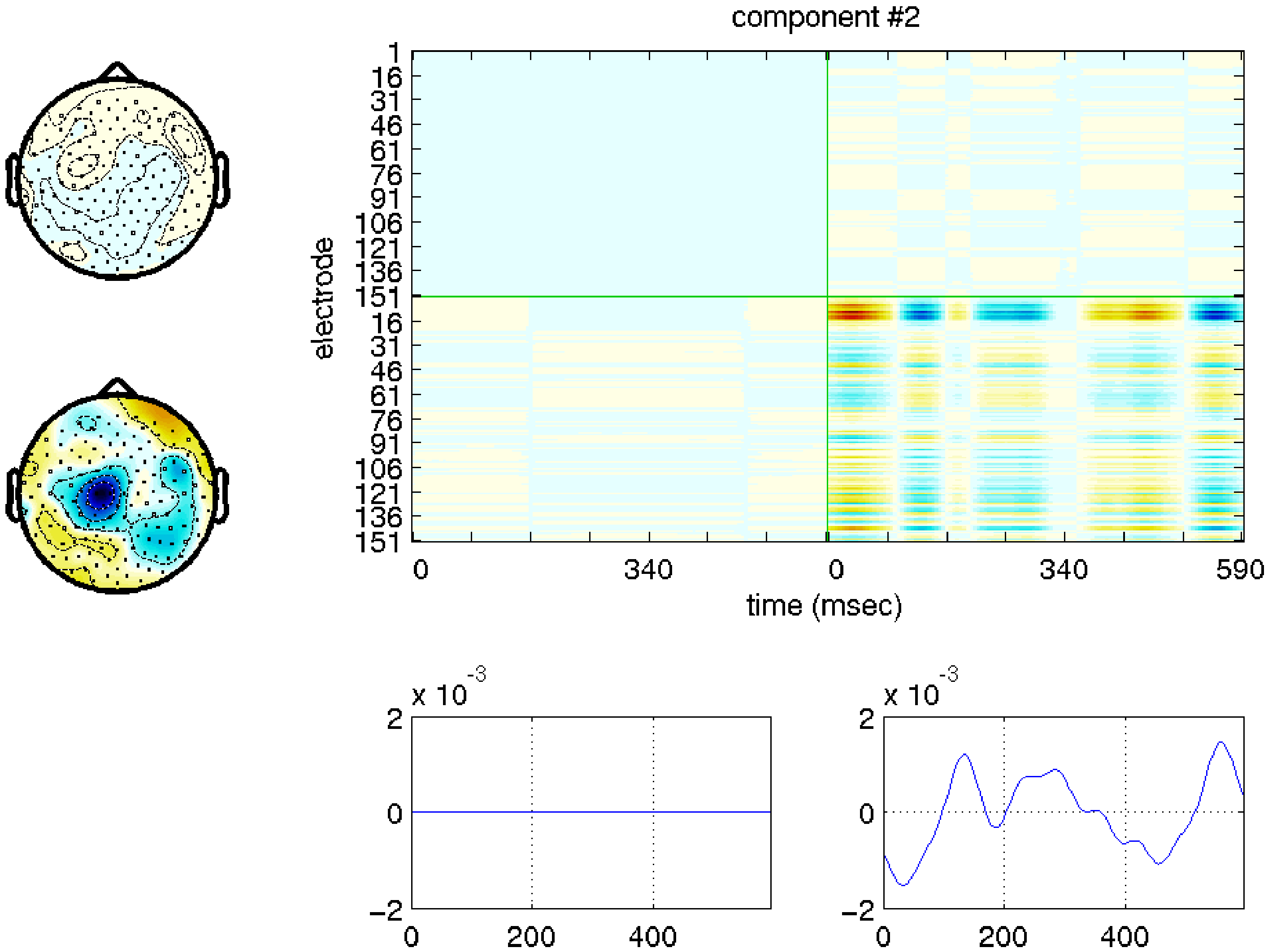
Example decomposition (tactile)

Subset of tactile MEG data (left little finger versus right little finger):



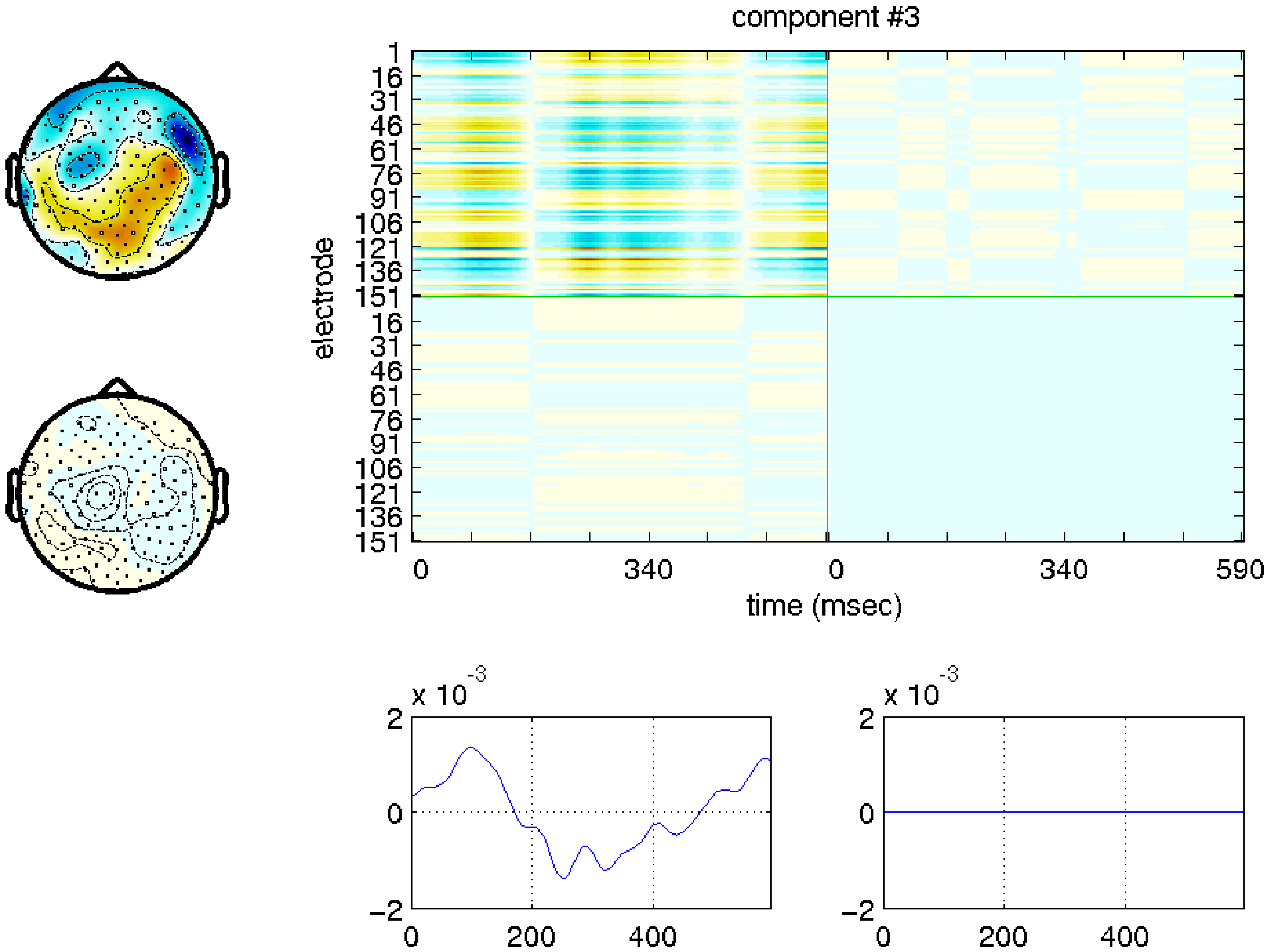
Example decomposition (tactile)

Subset of tactile MEG data (left little finger versus right little finger):



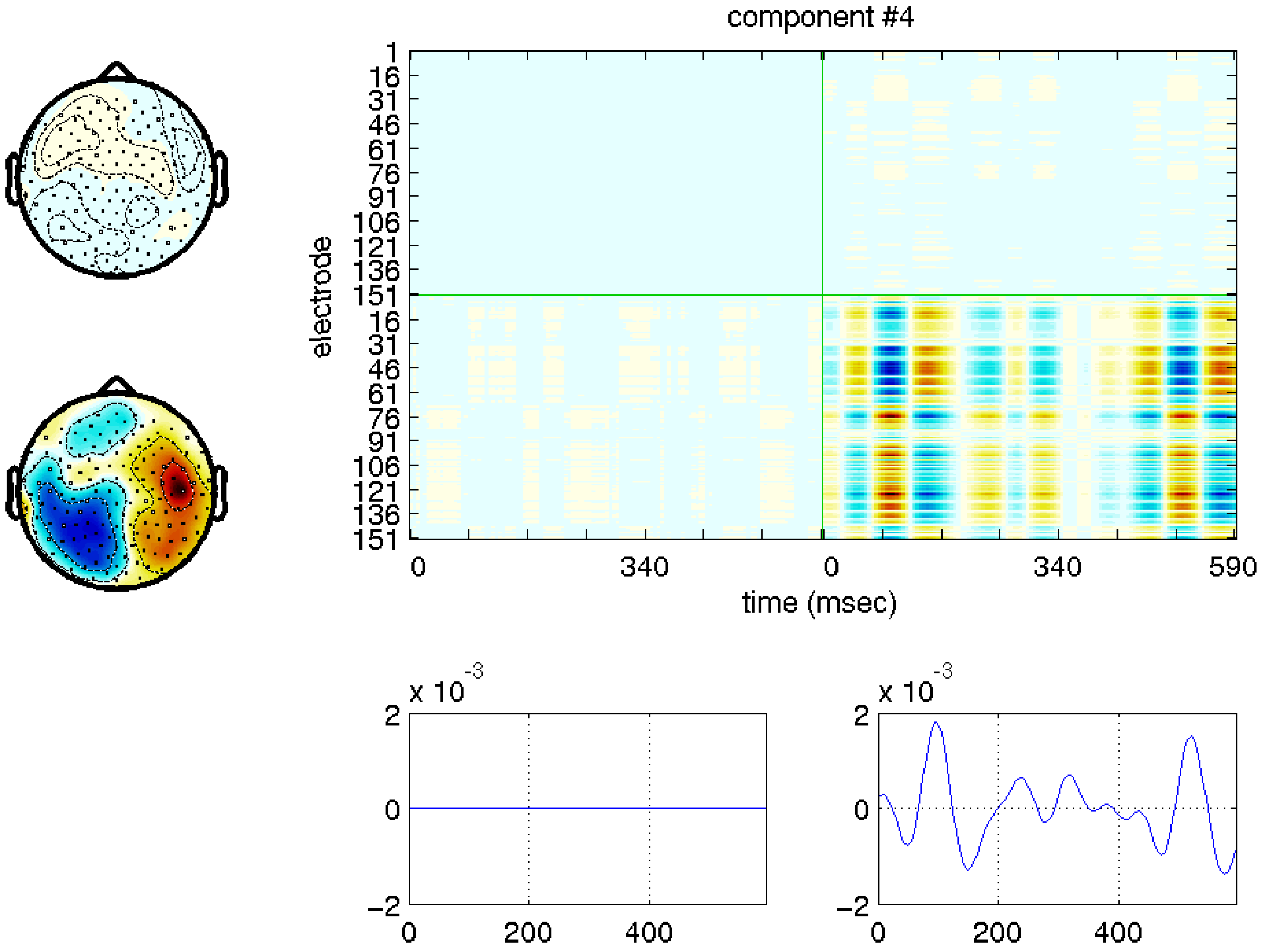
Example decomposition (tactile)

Subset of tactile MEG data (left little finger versus right little finger):

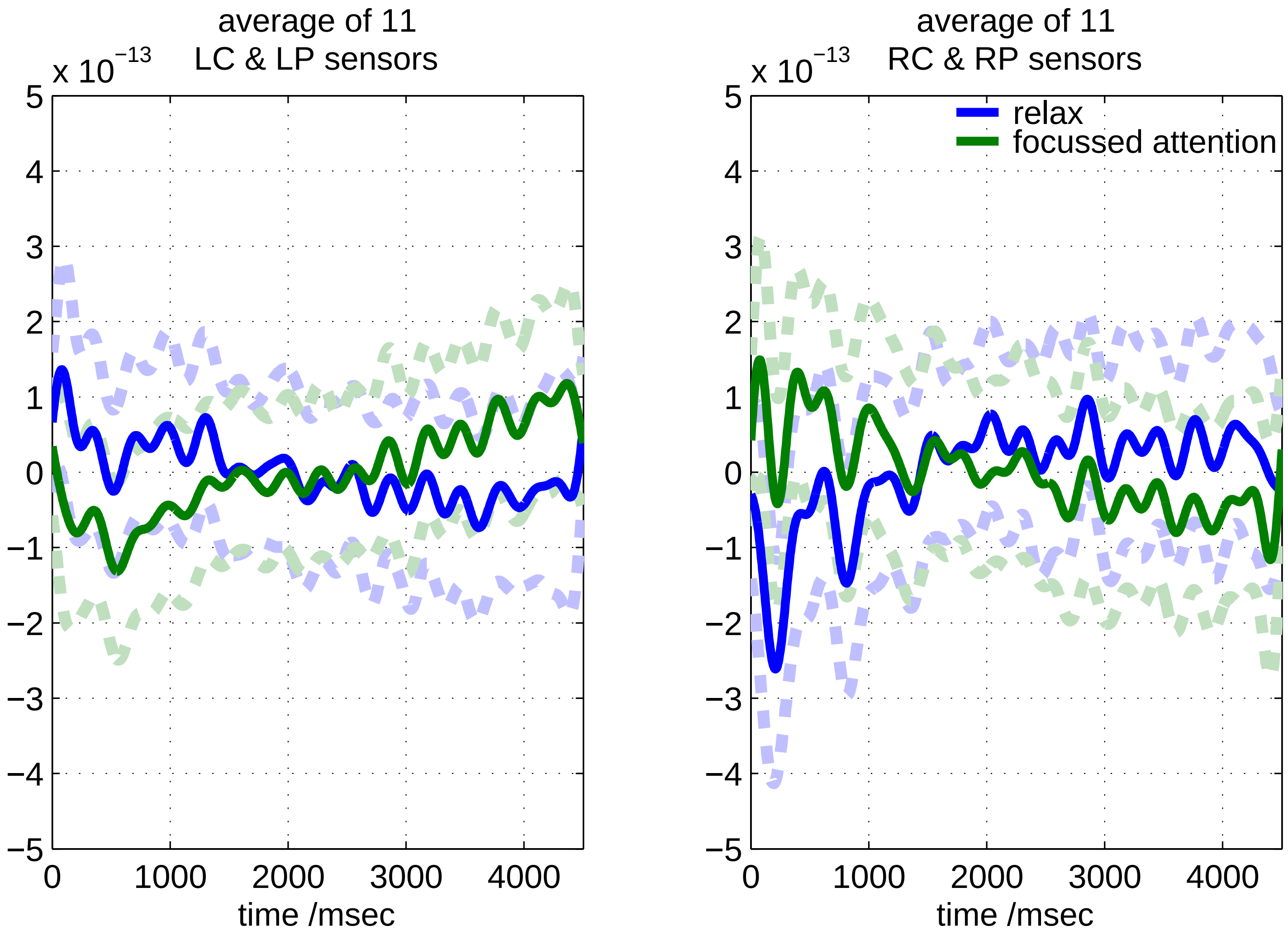


Example decomposition (tactile)

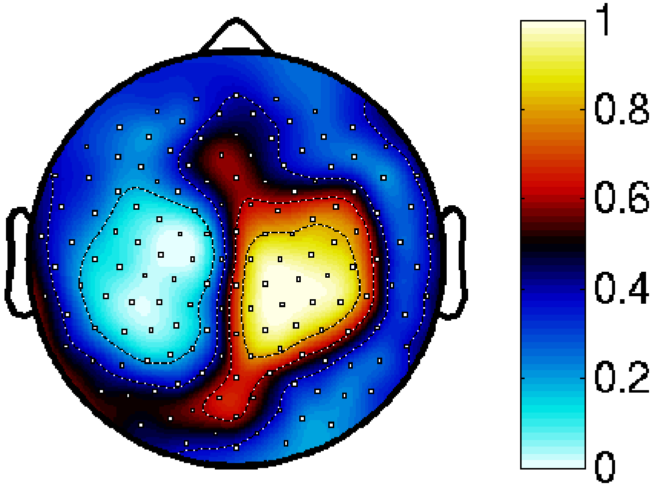
Subset of tactile MEG data (left little finger versus right little finger):



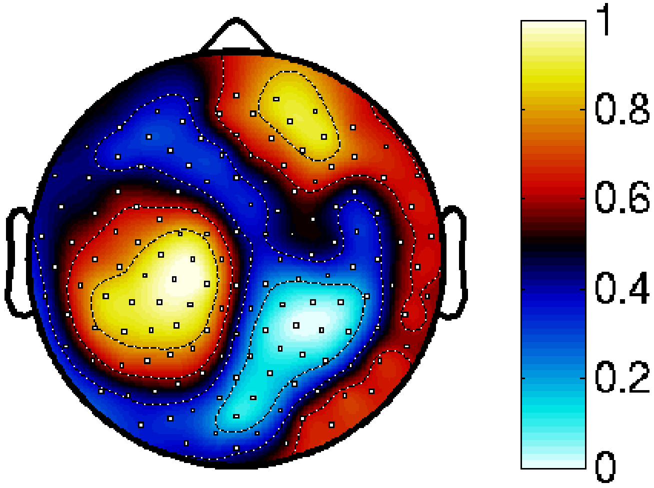
Slow waves indicating attention



Slow waves indicating attention

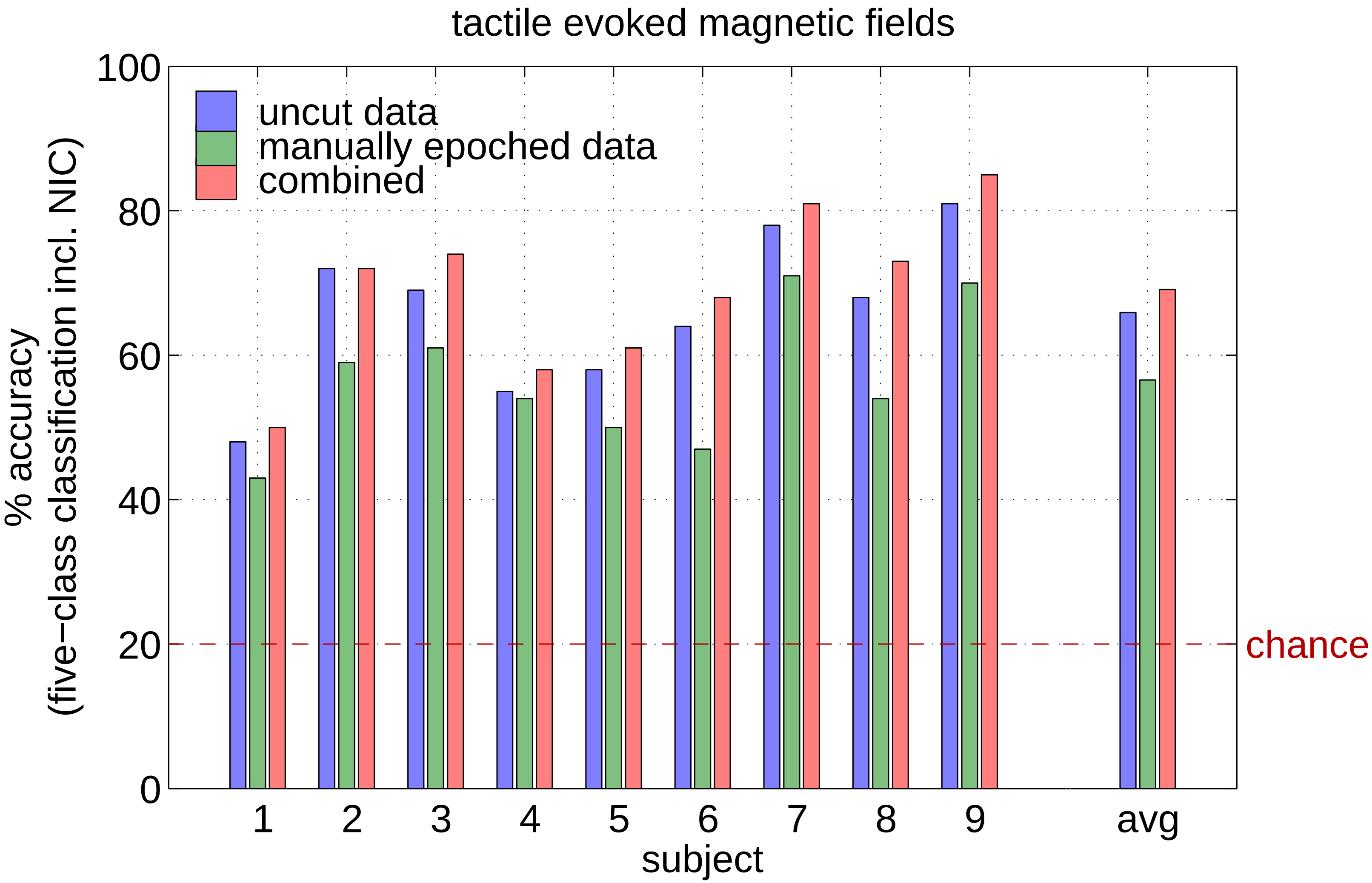


AUC scores at t=0.2 sec

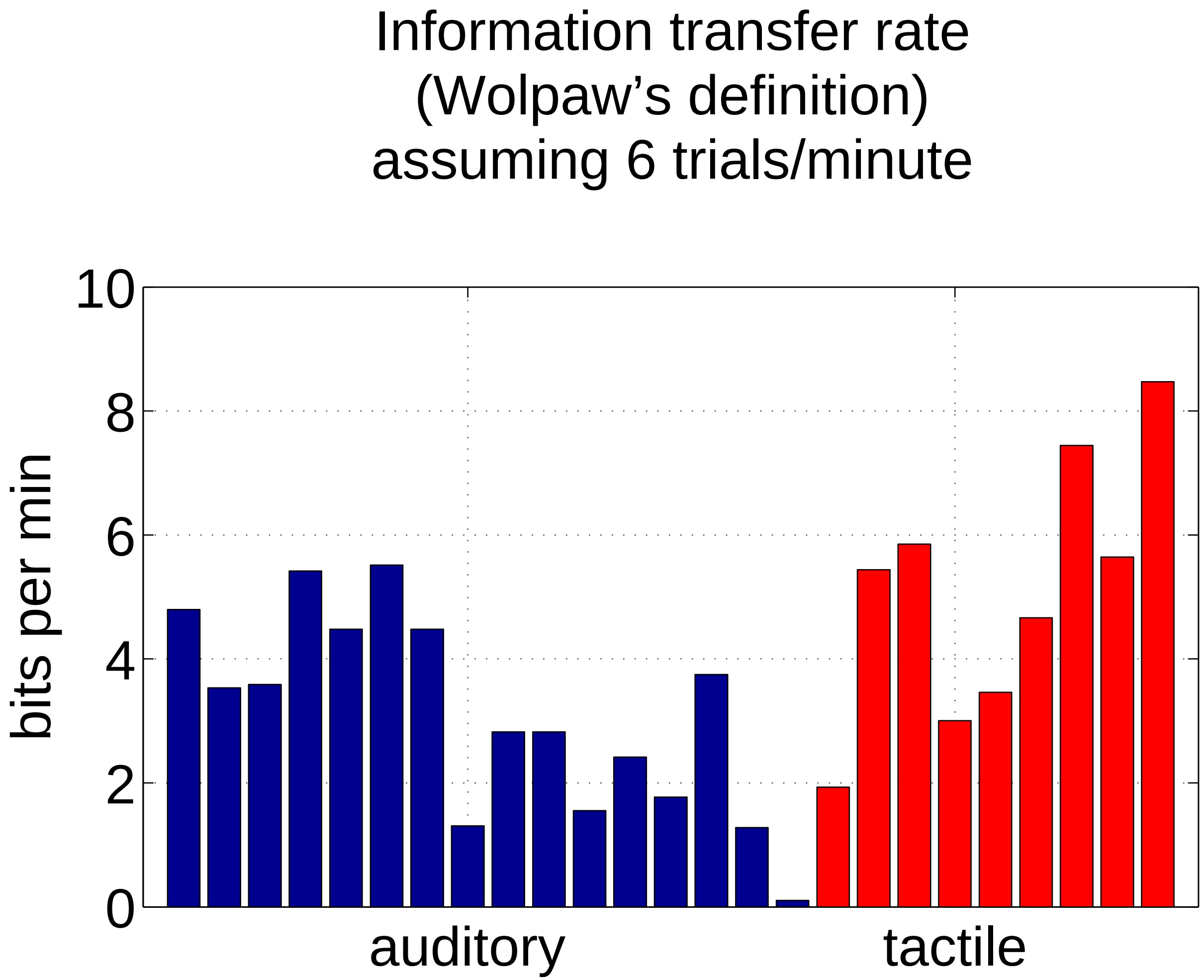


AUC scores at t=4 sec

Classification performance



Bit rates


Information transfer rate
(Wolpaw's definition)
assuming 6 trials/minute
bits per min
10
8
6
4
2
0
auditory
tactile

Conclusions

- It is possible to classify attention modulation of evoked responses to auditory and tactile stimuli, from 4.5–second signal segments.
- The stimuli are frequent (not an "oddball" paradigm).
- The classifier tends to rely more heavily on early (100–200 msec) components, although P300-like components are also useful.
- In addition, slow waves may help us to distinguish the control/no control problem.

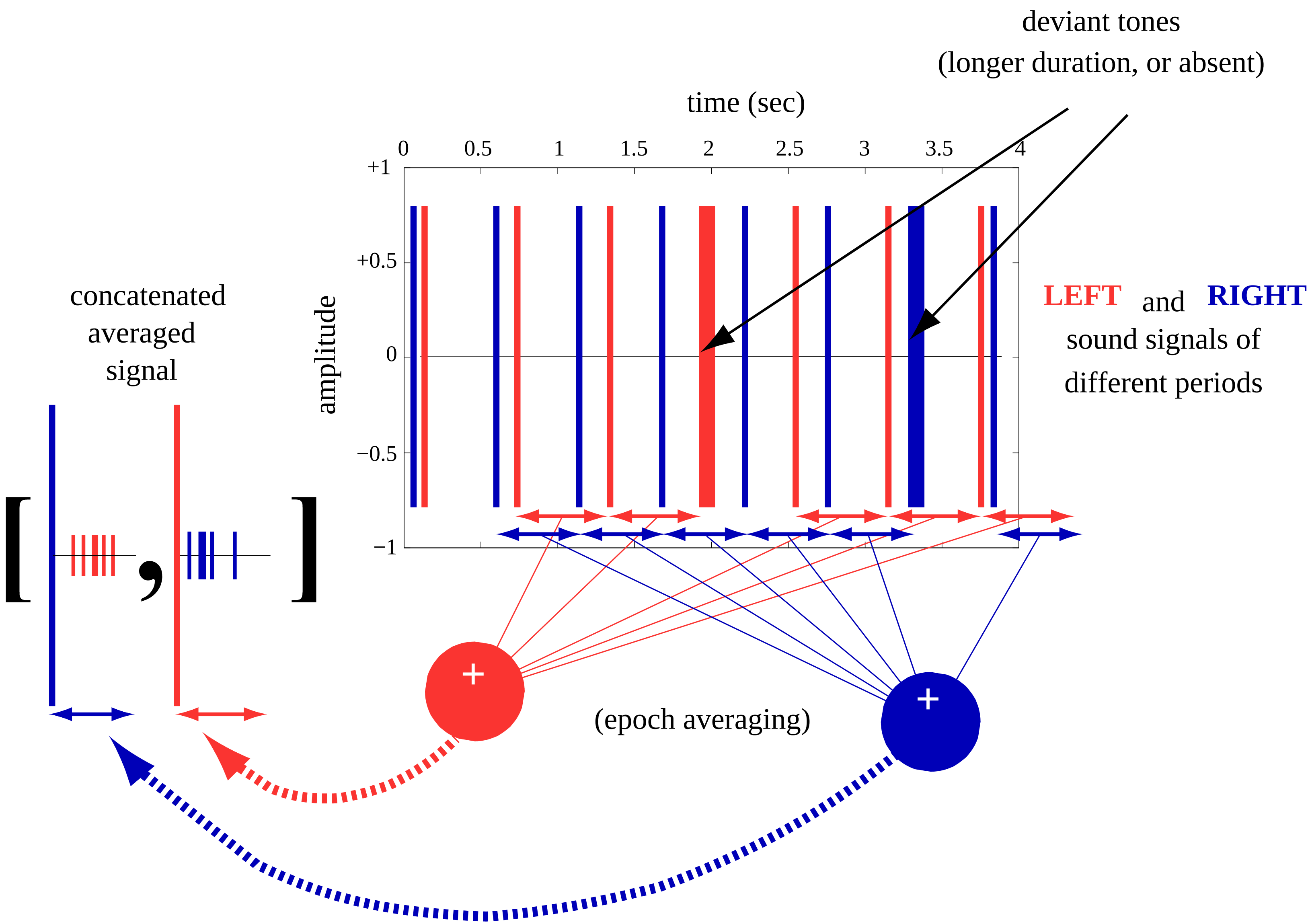
Conclusions

- It is possible to classify attention modulation of evoked responses to auditory and tactile stimuli, from 4.5–second signal segments.
- The stimuli are frequent (not an "oddball" paradigm).
- The classifier tends to rely more heavily on early (100–200 msec) components, although P300-like components are also useful.
- In addition, slow waves may help us to distinguish the control/no control problem.

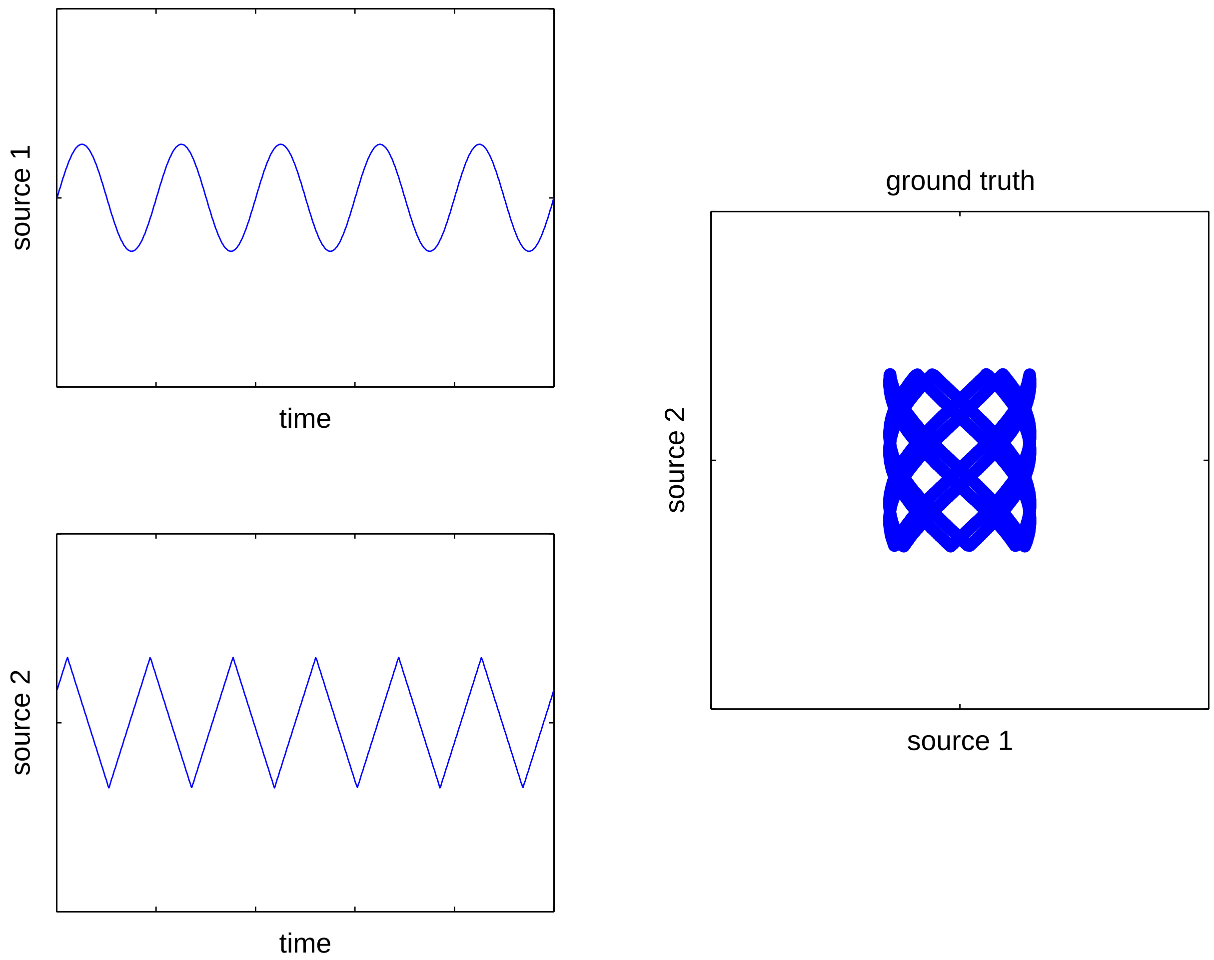
- How well will this work as an online BCI?
- How much can the stimuli be speeded up?
- How high can the number of classes go, in the tactile experiment?
- How will speed and number of classes interact?

Thank you for your attention.

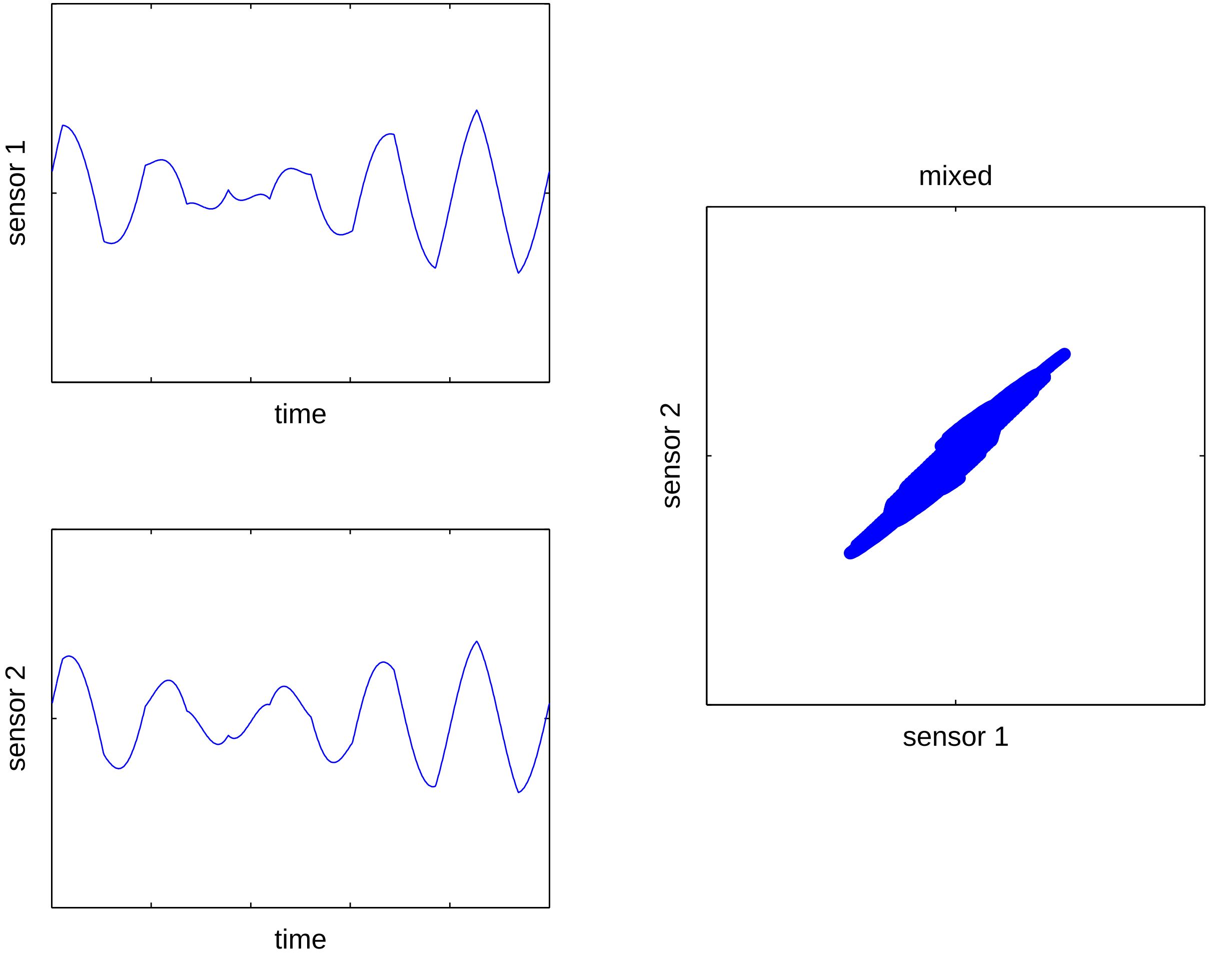
Auditory stimulus design



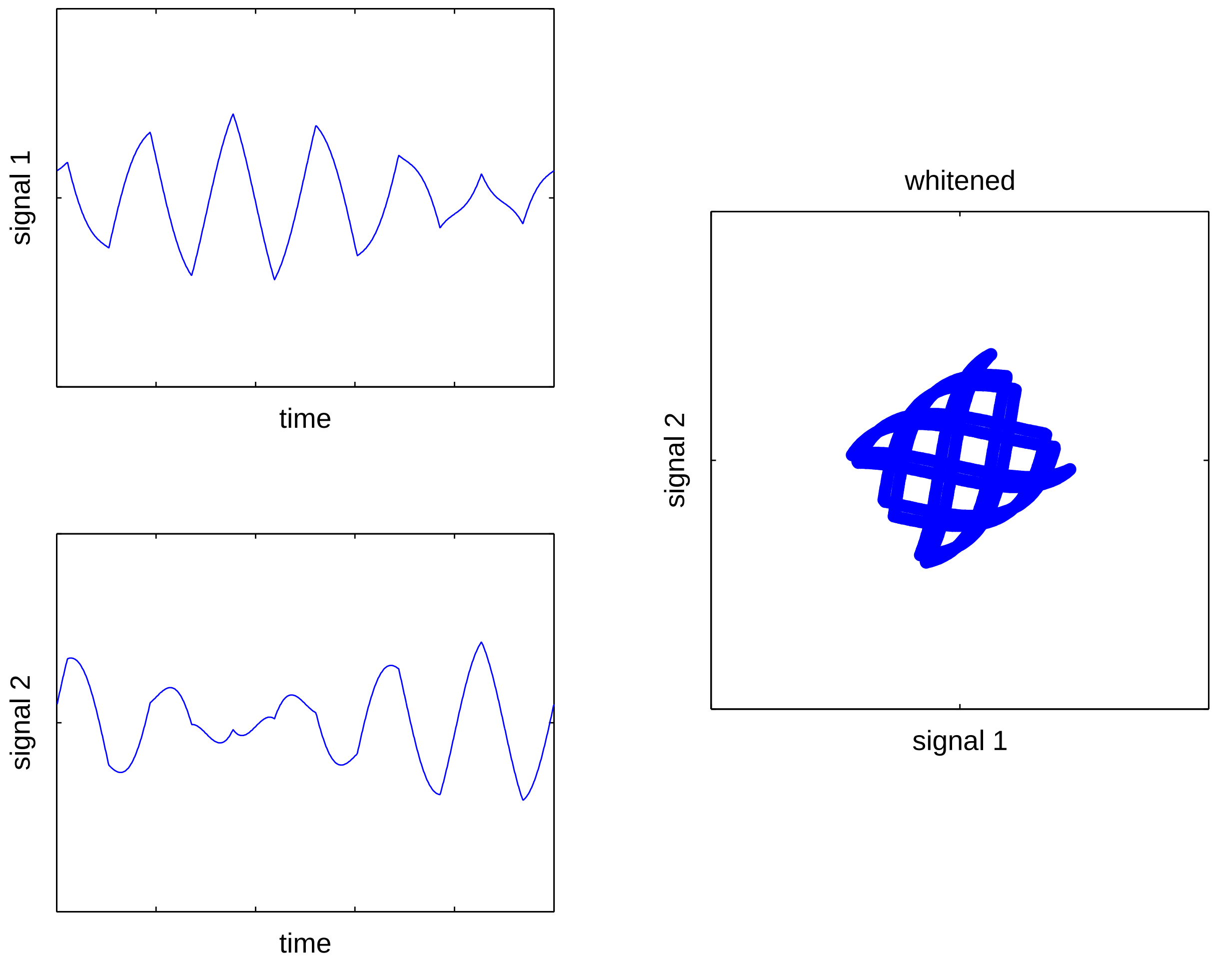
Whitening and rotation



Whitening and rotation



Whitening and rotation



Whitening and rotation