

# НЕКОТОРЫЕ МАТЕМАТИЧЕСКИЕ ЗАДАЧИ ОБРАБОТКИ И АНАЛИЗА ДАННЫХ В НЕЙРОНАУКАХ

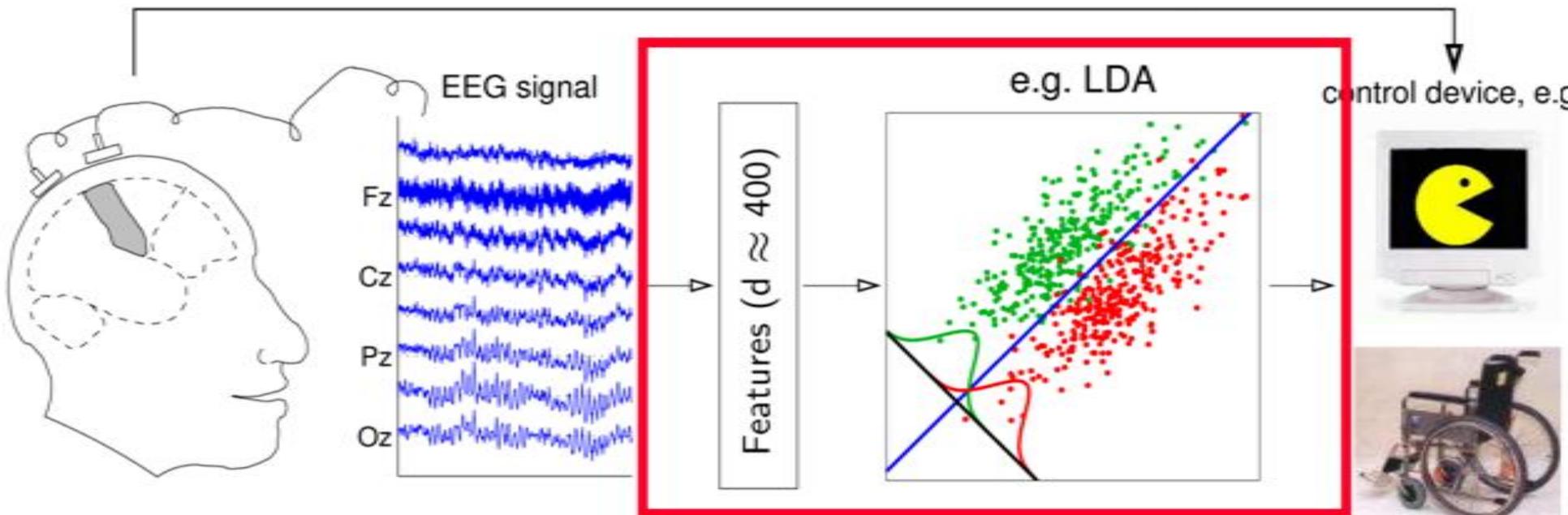
---

Михаил Беляев  
м.н.с. ИППИ РАН

# Интерфейсы мозг-компьютер

Ключевая задача ИМК - **интерпретировать сигналы мозга** и преобразовать их в управляющие команды некоторому устройству.

С математической точки зрения это **задача классификации**.



# Анализ данных в обработке ЭЭГ - пример

Многие факты о ЭЭГ получены с помощью **экспертного анализа данных**. Пример: изучение склонности к самоубийству при приеме антидепрессантов [1] (побочный эффект у ~10% людей).  
Экспертный анализ данных: падение мощности в тета диапазоне (4-8 Гц) в правой лобной доле у людей, которые проявят склонность к самоубийству в течении лечения.



## Математические методы классификации [2]:

- построение автоматических алгоритмов;
- возможность учесть вариабельность ЭЭГ и индивидуальные особенности;
- выявление более сложных зависимостей, чем может найти человек.

- 
1. Hunter, A. M. et al. (2010). Brain functional changes and worsening suicidal ideation and mood symptoms during antidepressant treatment. *Acta Psychiatrica Scandinavica*, 122(6)
  2. Müller, K. R. et al. (2008). Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *Journal of neuroscience methods*, 167(1)

# Задача классификации ЭЭГ

Задача классификации (на примере воображаемых движений)

1. Формирование обучающей выборки :

а. Регистрация ЭЭГ во время воображения движений по заданным инструкциям (какое движение необходимо вообразить в данный момент времени)

2. Обучение классификатора по данным:

записи ЭЭГ:

$$\{X_i \in R^{c \times t}\}_{i=1}^n$$

$c$  - число каналов ЭЭГ;

$t$  - длительность  
испытания;

$n$  - число испытаний.

метки класса:

$$\{y_i \in \{0, 1, 2\}\}_{i=1}^n$$



# Развитие систем регистрации ЭЭГ

В последние годы создан целый ряд портативных устройств для снятия ЭЭГ, в том числе

- с сухими сенсорами;
- беспроводных;
- с относительно низкой ценой (~300\$).

Открывается возможность использования технологии в повседневной жизни.



# Будущее технологии ИМК

Выдержка из краткого резюме доклада

**The future in brain/neural-computer interaction:  
horizon 2020.** (ЕС, апрель 2015)

*“While the field has progressed rapidly from its infancy, a concerted effort supported by all stakeholders in the field is necessary to make further progress towards developing future BCI applications. That is, researchers will need to address the specific needs of potential end users together with companies to **produce BCI solutions that can be applied outside of laboratory settings.**”*

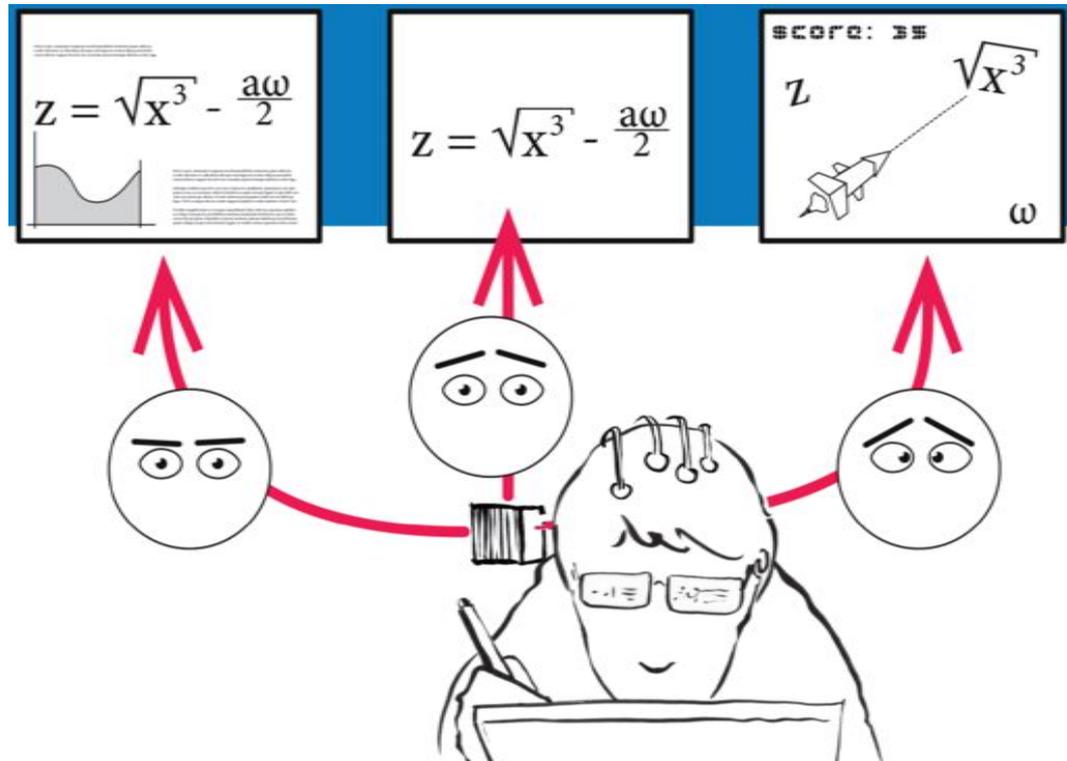
# Будущее технологии ИМК

Реабилитация постинсультных больных на дому: ЭЭГ + стимуляция + игровой подход [1]



# Будущее технологии ИМК

Адаптивная обучающая платформа, основанная на мониторинге ментального состояния ученика [1]



Прототипы подобных систем разрабатываются в DARPA [2]

[1] The future in brain/neural-computer interaction: horizon 2020 Roadmap

[2] Miranda, R. A et al. (2014). DARPA-funded efforts in the development of novel brain-computer interface technologies. *Journal of neuroscience methods*.

# Ускорение обучения

DARPA Accelerated Learning project:

1. Автоматическое определение “опасных” для обучения состояний и адаптивные обучающие материалы на основе распознавания ЭЭГ, ЭКГ, параметров дыхания:
  - a. неготовность к обучению (например, недостаток сна);
  - b. недостаточная вовлеченность.
2. Профилирование экспертов и оценка близости к профилю эксперта. Так, например, у профессиональных снайперов наблюдаются специфические паттерны ЭЭГ [1], которые можно сравнивать с ЭЭГ во время обучения.

Точность стрельбы новичков выросла после ускоренного обучения в **2.3** раза по сравнению с контрольной группой [2].

- 
1. Deeny, S. P. et al. (2003). Cortico-cortical communication and superior performance in skilled marksmen: an EEG coherence analysis. *Journal of Sport and Exercise Psychology*, 25(2), 188-204.
  2. Behneman, A. et al. (2012). Neurotechnology to accelerate learning: during marksmanship training. *Pulse, IEEE*, 3(1), 60-63.

# Математические задачи анализа данных

Для создания эффективного доступного в повседневной деятельности ИМК необходимо разработать **устойчивые и надежные** алгоритмы классификации ЭЭГ, решив следующие проблемы:

- построения устойчивых оценок;
- устранения артефактов в ЭЭГ (движения, моргания, мимика лица, ...);
- нестационарности ЭЭГ;
- существенного различия ЭЭГ между разными пользователями.

# Решение задачи классификации ЭЭГ

## Распространенные методы решения:

- байесовский подход (все наблюдения одного класса порождены многомерным нормальным распределением)
- Common Spatial Pattern (выделение признаков) + Линейный дискриминантный анализ

## Common Spatial Pattern. Основная идея (бинарная задача):

- средняя матрица ковариации для каждого класса

$$\Sigma_k = \frac{1}{\#\{y_i = k\}} \sum_{y_i=k} C_i, \quad C_i = \frac{1}{t-1} X_i X_i^T, \quad k = \{0, 1\}$$

- выбор наиболее информативных комбинаций каналов

$$w^* = \operatorname{argmax}_w \frac{w \Sigma_0 w^T}{w (\Sigma_0 + \Sigma_1) w^T}$$

- в качестве признаков используются

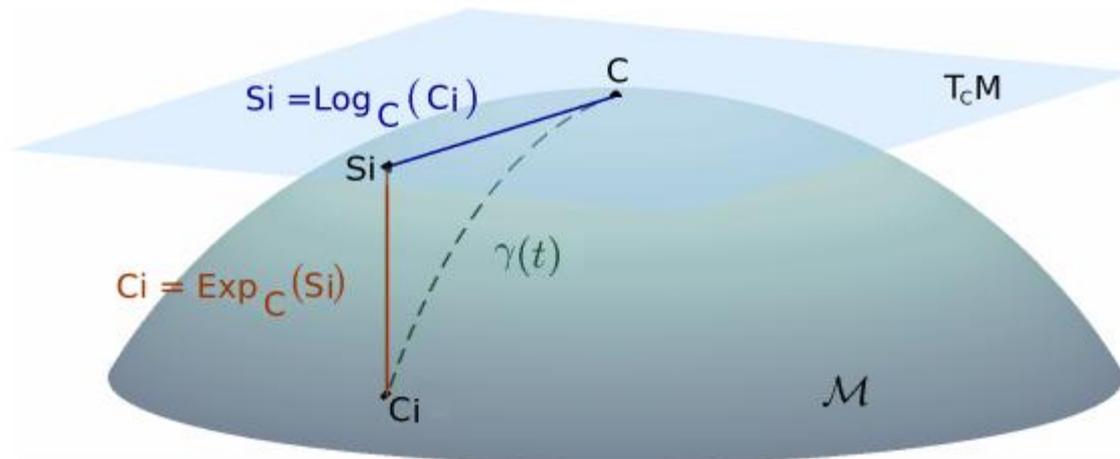
$$z_i = \log(\operatorname{var}(w^* X_i))$$

# Новые подходы к классификации ЭЭГ

- В рамках байесовского подхода мы моделируем все данные с помощью одной ковариационной матрицы
- **Что если** рассмотреть ковариационную матрицу для каждого эксперимента отдельно и **классифицировать матрицы?**
- **Проблема:** методы классификации разработаны для работы с векторами
- Пространство всех симметричных положительно определенных матриц размера  $s \times s$  образует дифференцируемое многообразие  $M$  в пространстве всех матриц размера  $s \times s \Rightarrow$  **риманова геометрия!**
- CSP можно рассматривать как некоторое приближение риманова расстояния [1].

# Новые подходы к классификации ЭЭГ

- Строим **касательное пространство** к  $M$ .
- Проецируем ковариационные матрицы на касательное пространство.
- С проекциями можно работать как с обычными векторами.
- **Построение классификации по проекциям [1].**



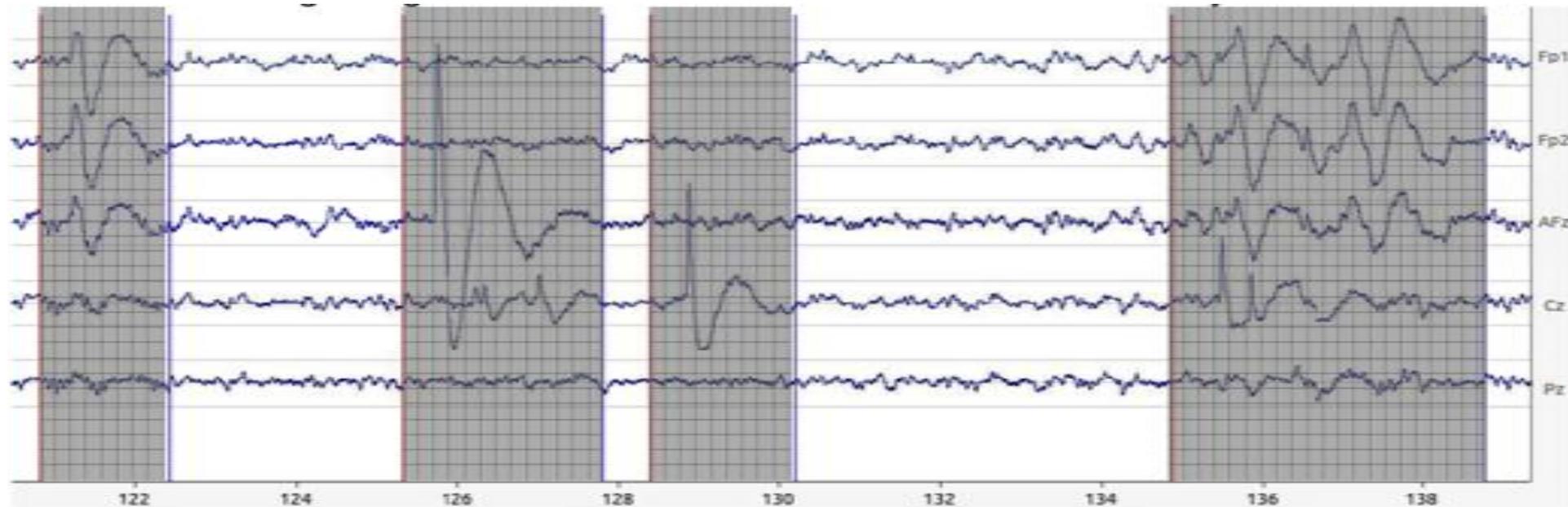
- 
1. Barachant, A. et al. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112, 172-178.

# Снижение размерности

- **Ковариационные матрицы** в конкретном эксперименте **лежат на многообразии меньшей размерности** (чем  $c * (c + 1) / 2$ )
- **Линейные методы** снижения размерности:
  - Метод главных компонент - 1901 год [1]
  - Аналог для симметричных положительно-определенных матриц - 2004 год [2]
- **Нелинейные методы** снижения размерности
  - LTSA, Grassmann-Stiefel Eigenmaps, ... [3]
  - Необходимо **обобщение на случай римановой геометрии**

- 
1. *Pearson K.*, On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, (1901) 2, 559—572
  2. Fletcher, P. T., & Joshi, S. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis* (pp. 87-98).
  3. Bernstein, A., & Kuleshov, A. (2014). Low-Dimensional Data Representation in Data Analysis. In *Artificial Neural Networks in Pattern Recognition* (pp. 47-58).

# Удаление артефактов ЭЭГ



Лабораторные условия:

- человек, как правило, неподвижен;
- инструкции: смотреть прямо, не моргать.

Использование в повседневной жизни:

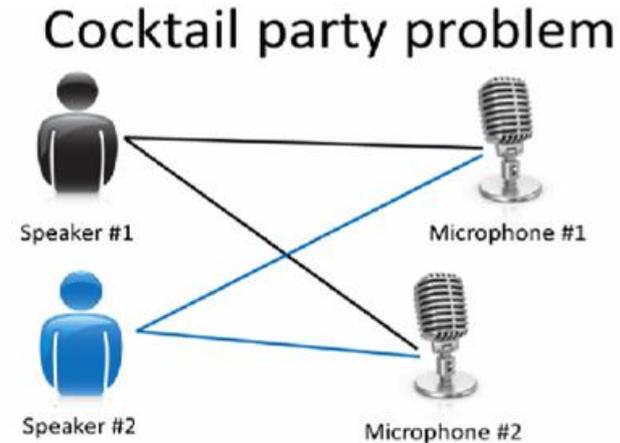
- человек находится в движении, используется мимика;
- моргания, взгляд в сторону;
- необходим анализ в режиме реального времени.

# Методы очистки от артефактов

Существующие алгоритмы основаны на методе независимых компонент (ICA).

Ограничения ICA

- Если в присутствии более одного гауссовского источника, то ICA не может однозначно идентифицировать источники. Решение - метод Non-Gaussian Components Analysis [1]
- модельные ограничения в контексте обработки ЭЭГ
  - предположение о независимости источников.
  - предположение о пространственной стационарности.Развиваются методы, снимающие эти модельные ограничения [2]

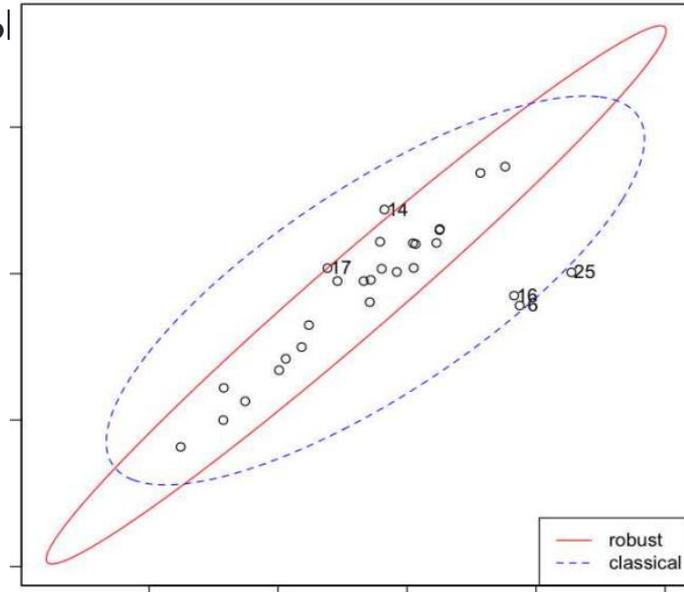


1. Blanchard, G., Spokoiny, V., et al. (2006). In search of non-Gaussian components of a high-dimensional distribution. *The Journal of Machine Learning Research*, 7, 247-282.
2. Makeig, S. et al (2012). Evolving signal processing for brain-computer interfaces. *Proceedings of the IEEE*, 100(Special Centennial Issue), 1567-1584.

# Построение устойчивых классификаторов

- В данных могут содержаться выбросы
  - неправильно размеченные данные
  - не до конца почищенные артефакты
- Многие классические методы неробастны (чувствительны к выбросам)!
- Пример: CSP основан на усреднении ковариационных матриц:

$$\Sigma_k = \frac{1}{\#\{y_i = k\}} \sum_{y_i=k} C_i, \quad C_i = \frac{1}{t-1} X_i X_i^T, \quad k = \{0, 1\}$$



# Построение устойчивых классификаторов

$$D_{\beta}(p(x) \parallel q(x)) = \int \left[ \frac{1}{\beta} \{p^{\beta}(x) - q^{\beta}(x)\} p(x) - \frac{1}{\beta + 1} \{p^{\beta+1}(x) - q^{\beta+1}(x)\} \right] dx,$$

Распределения:

$p(x)$  - эмпирическое

$q(x)$  - параметрическое

Выбирая  $q(x)$  как распределение Уишарта, получим итеративный способ робастной оценки средней ковариационной матрицы:

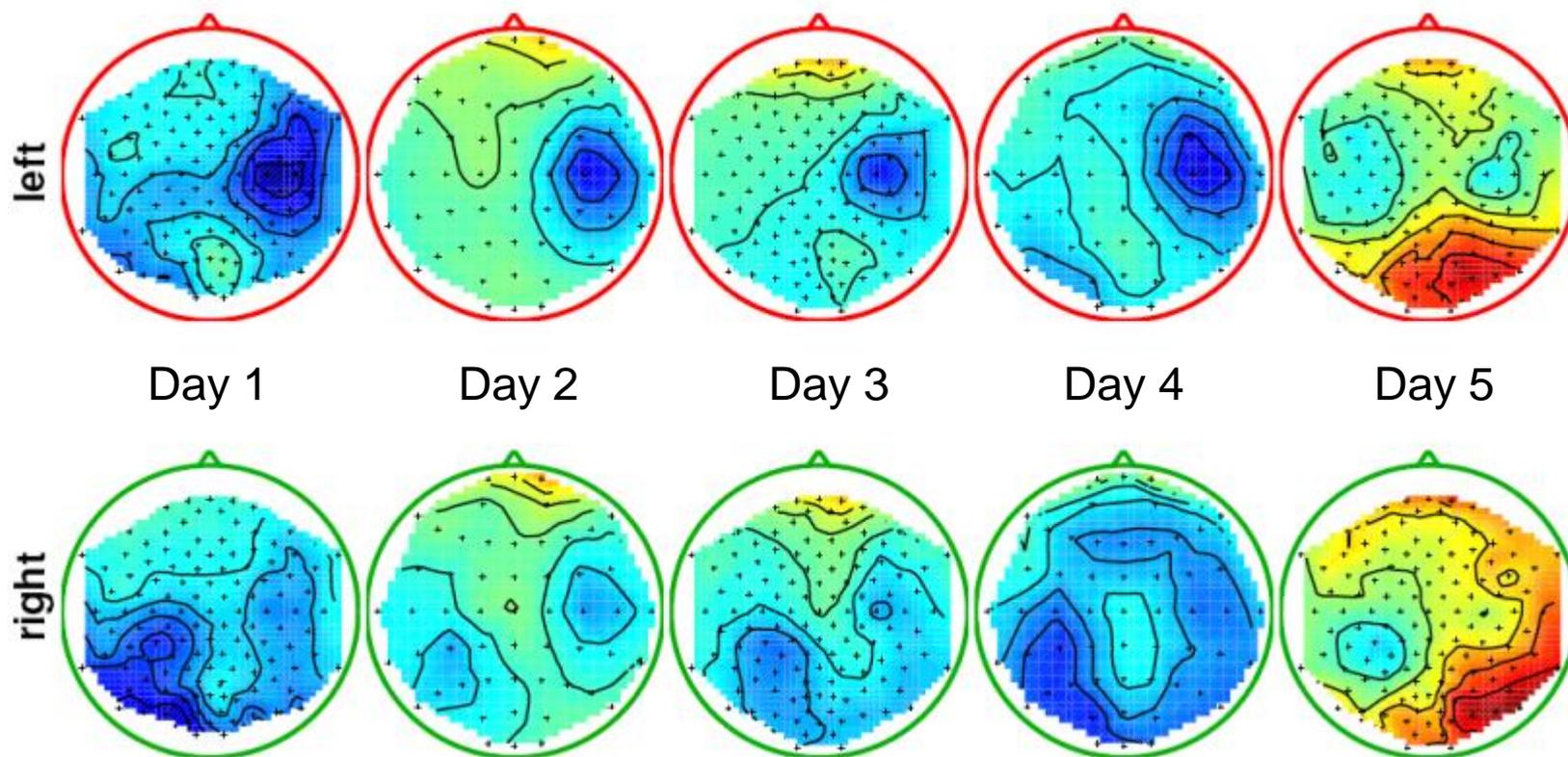
$$\Sigma^{(k+1)} = \frac{\sum_{i=1}^n \psi_{\beta}(S_i; \Sigma^{(k)}, \nu) S_i}{\nu \sum_{i=1}^n \psi_{\beta}(S_i; \Sigma^{(k)}, \nu) - \gamma |\Sigma^{(k)}|^{\frac{(\nu-C-1)\beta}{2}}} \leftarrow \text{нормировка}$$

$$\psi_{\beta}(S; \Sigma, \nu) = |S|^{\frac{(\nu-C-1)\beta}{2}} \exp \left\{ -\text{tr} \left( \frac{\beta}{2} \Sigma^{-1} S \right) \right\} \leftarrow \text{адаптивный вес, у выбросов } \sim 0$$

- 
1. Samek, W., & Kawanabe, M. (2014). Robust common spatial patterns by minimum divergence covariance estimator. In *Acoustics, Speech and Signal Processing (ICASSP)*

# Нестационарность ЭЭГ

Вариабельность усредненных (140 испытаний) паттернов ЭЭГ у одного человека, но в разные дни (воображение движения рукой)



# Учет нестационарности ЭЭГ

## Principal Component Analysis (PCA)

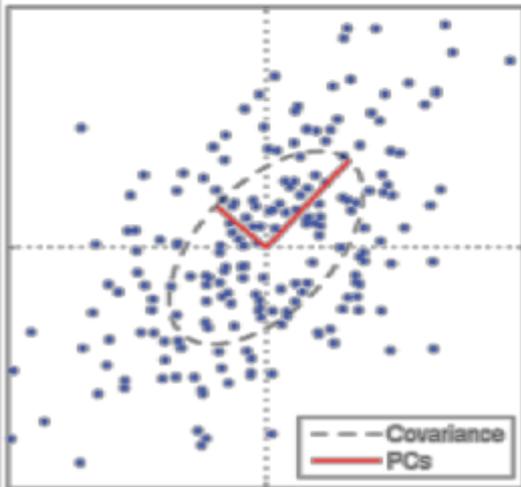
uncorrelated sources

orthogonal mixing

$$X = A \begin{bmatrix} S^{(1)} \\ \vdots \\ S^{(d)} \end{bmatrix}$$

max. variance

min. variance

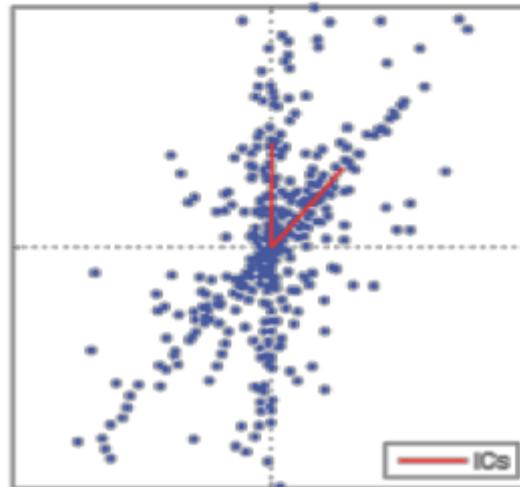


## Independent Component Analysis (ICA)

independent sources

arbitrary mixing

$$X = A \begin{bmatrix} S^{(1)} \\ \vdots \\ S^{(d)} \end{bmatrix}$$



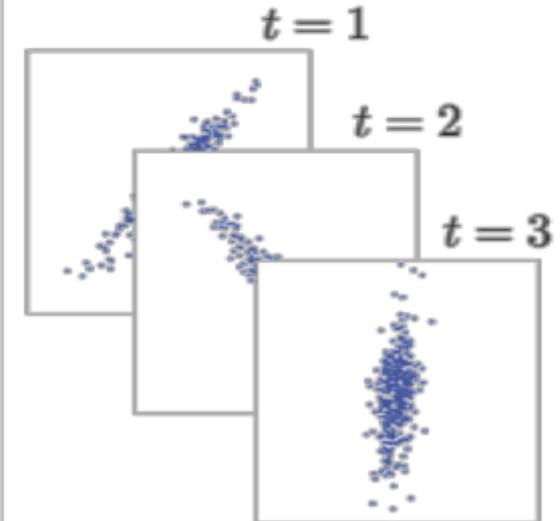
## Stationary Subspace Analysis (SSA)

arbitrary mixing

$$X_t = A \begin{bmatrix} S_t^s \\ S_t^n \end{bmatrix}$$

stationary sources

non-stationary sources



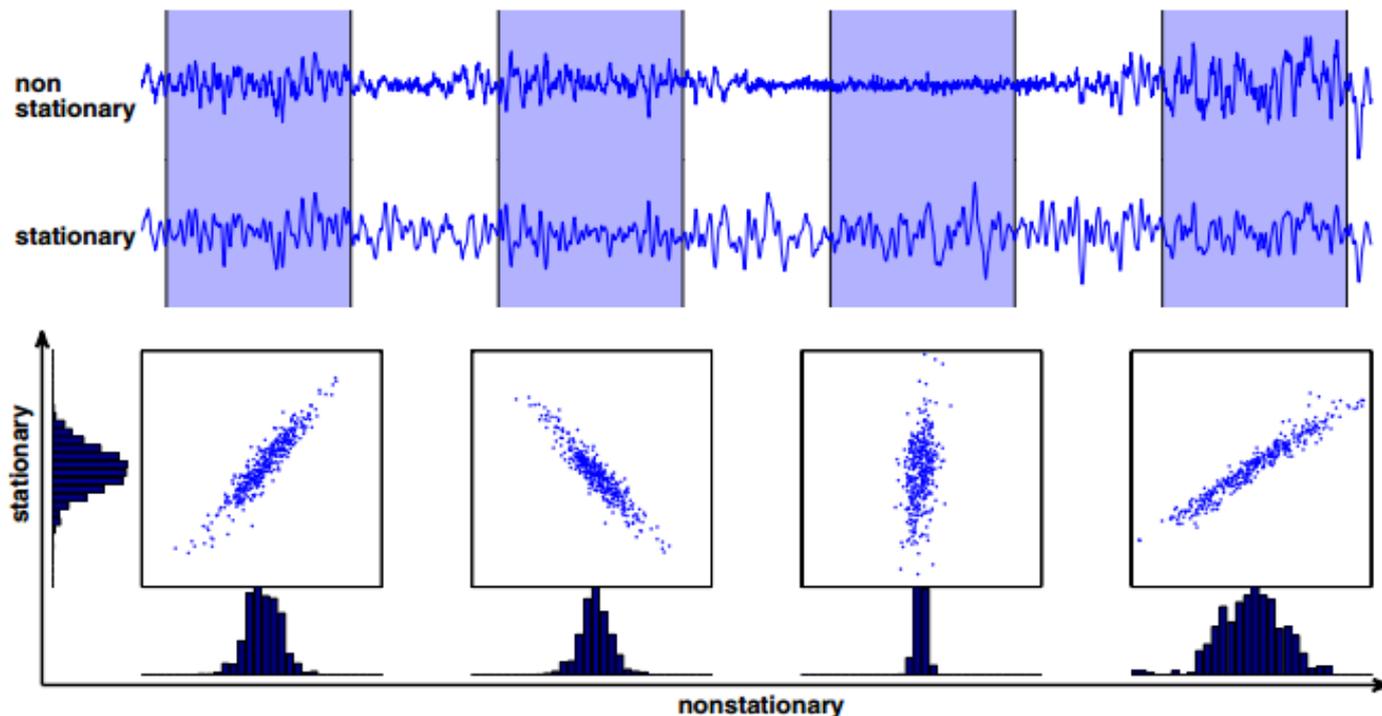
von Bünau, P. et al (2009). Finding stationary subspaces in multivariate time series. *Physical Review Letters*, 103(21), 214101.

# Нахождение стационарного подпространства

1. Введем меру нестационарности на основе расстояния Кульбака-Лейбера между средней ков. матрицей и ков. матрицей каждого испытания

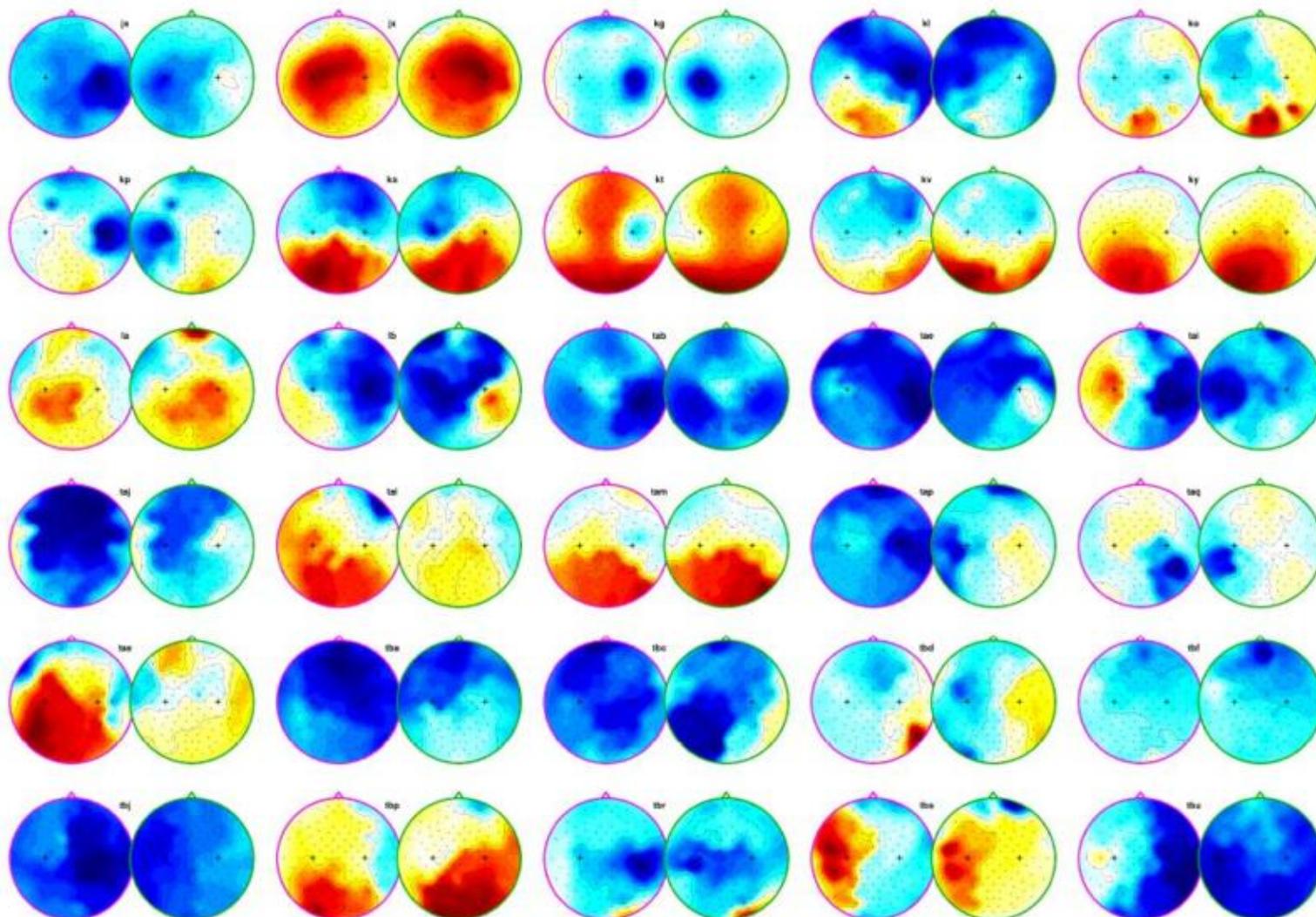
$$\operatorname{argmin}_W \sum_{i=1}^n D_{KL} \left[ N(W \mu_i, W C_i W^T), N(W \bar{\mu}_i, W \bar{C}_i W^T) \right]$$

2. Найдем такую матрицу  $W$ , которая позволит минимизировать введенную меру нестационарности



# Межсубъектная вариабельность

Каждая картинка - ЭЭГ одного человека (левая часть - изображение левой руки, правая - изображение правой)



from K.-R. Muller. Brain computer interfaces, MLSS 2012.

# Адаптация к новым пользователям

Две возможные постановки задачи

1. Для нового пользователя нет вообще никаких испытаний с известными метками класса
2. Есть некоторое количество испытаний, но их число мало.

Развиваемые подходы:

1. Ансамбли моделей моделей, построенных по  $S$  “обучающим” людям [1]
2. 1 + адаптация весов индивидуальных моделей по имеющимся данным
  - a. Даны ЭЭГ и метки - оценка весов через наилучшее приближение [2]
  - b. Даны только ЭЭГ - использование римановой геометрии (оценка схожести людей) - **калибровка “на лету”**

- 
1. A. Barrachant. Kaggle NER'15 contest winning solution.
  2. Fazli, S. et al. (2009). Subject-independent mental state classification in single trials. *Neural networks*, 22(9), 1305-1312.

# Заключение

1. Существует технологический и научный задел для выхода ИМК из лабораторных условий.
2. Для успешного использования в повседневной жизни необходимо повысить качество и устойчивость алгоритмов классификации ЭЭГ.
3. В мире (и, в частности, в ИППИ) ведется разработка новых математических методов анализа данных, призванных решить эту задачу.