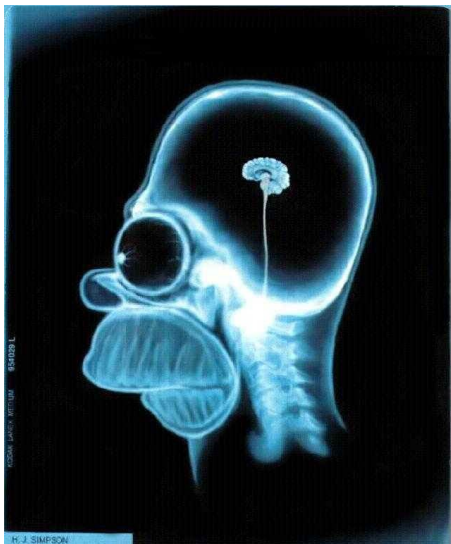
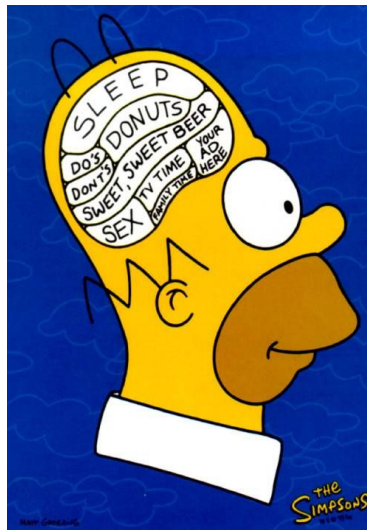


## MRI vs. fMRI

MRI (Magnetic Resonance Imaging) studies brain anatomy.



Functional MRI (fMRI) studies brain function.



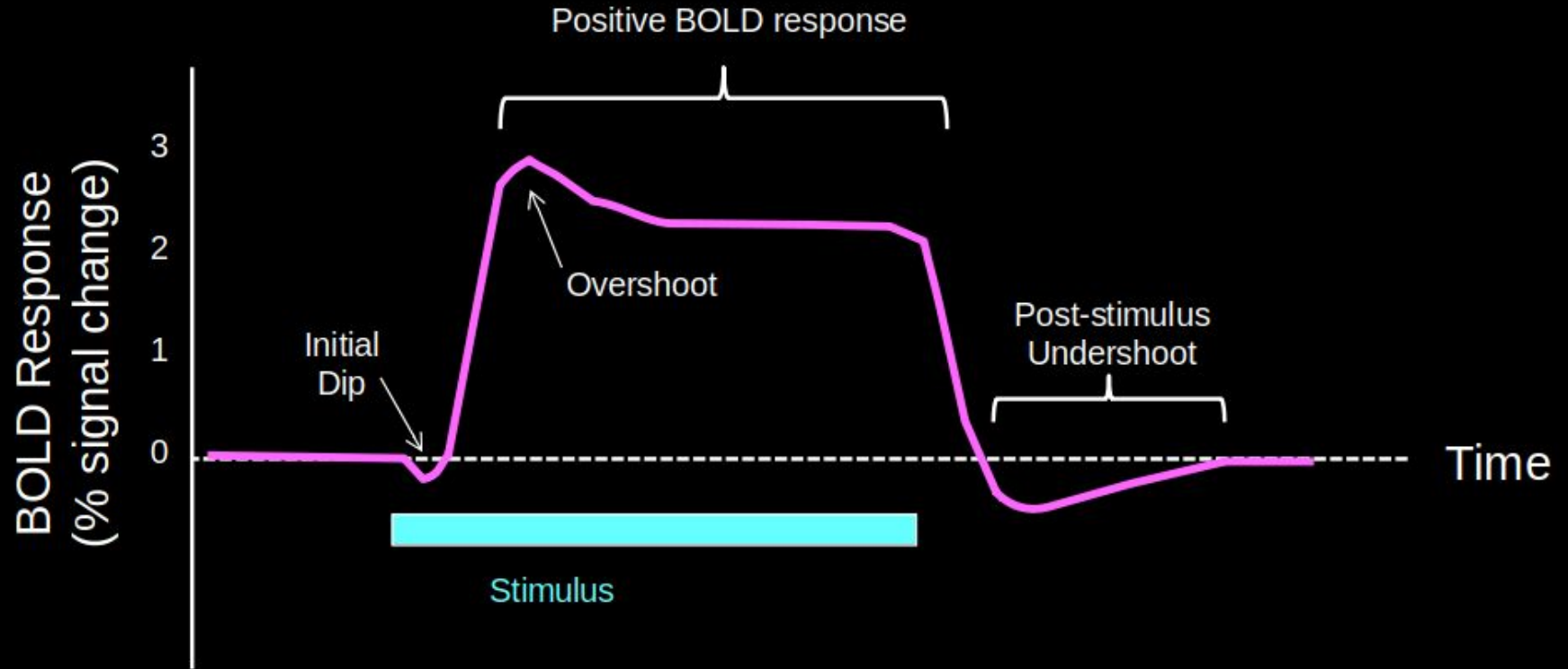
<https://lukas-snoek.com/NI-edu/index.html>

[https://andysbrainbook.readthedocs.io/en/latest/fMRI\\_Short\\_Course/fMRI\\_Intro.html](https://andysbrainbook.readthedocs.io/en/latest/fMRI_Short_Course/fMRI_Intro.html)

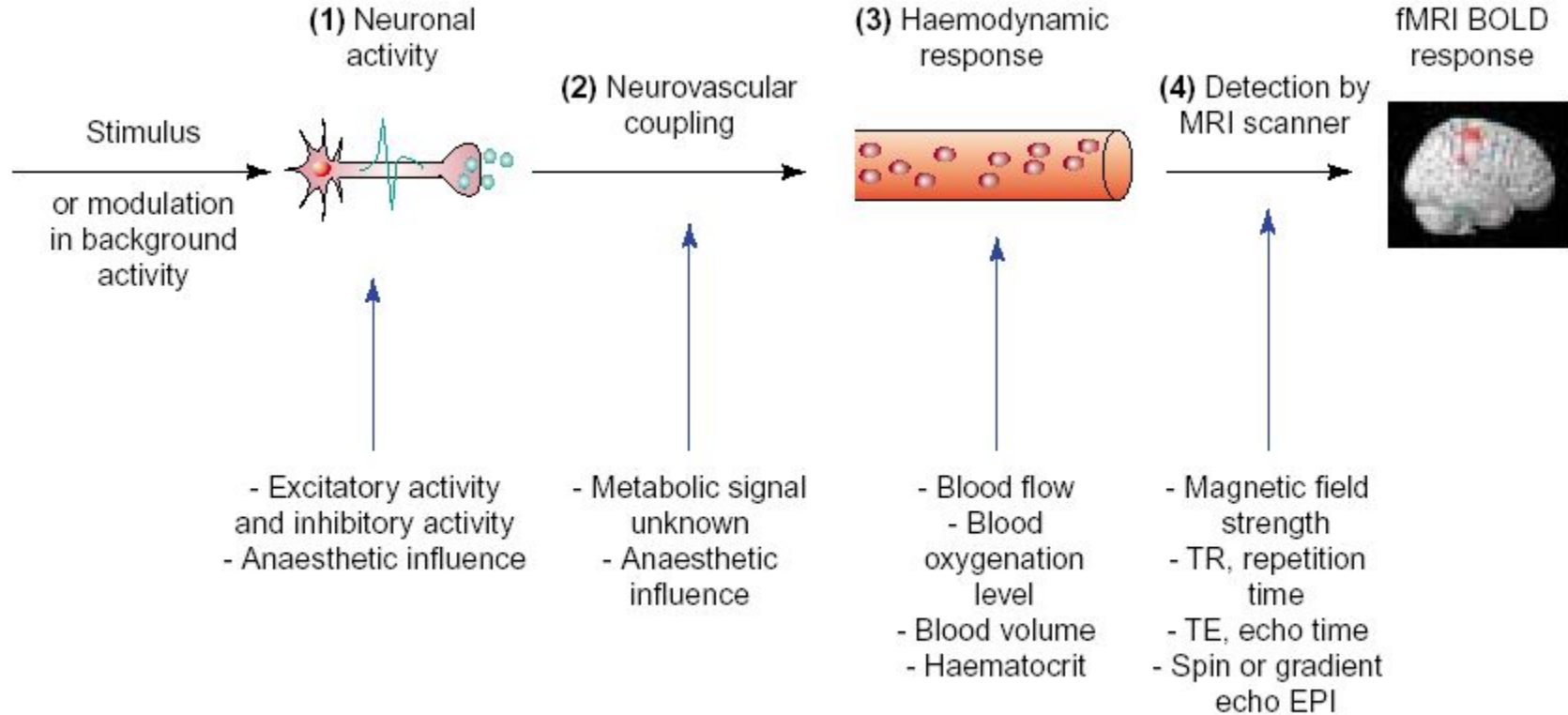
<https://fmriprep.org/en/stable/#>

[https://nipy.org/nibabel/coordinate\\_systems.html#introducing-someone](https://nipy.org/nibabel/coordinate_systems.html#introducing-someone)

# BOLD Time Course Blood Oxygenation Level-Dependent Signal



## Stimulus to BOLD



# Машинное обучение для работы с видео

## Общие идеи для работы с видео

- 3D свёртки по времени и пространству (минусы: много параметров, вычислительно дорого)
- Факторизация 3D свёрток на 2D для пространства и 1D для времени
- Two-Stream подход
- Inflated 3D CNN – использование предобученных на картинках 2D свёрток для начального приближения
- Использование разных типов attention

## Задачи, на которые обратить внимание

- Video-to-Video
- Video Understanding (текстовое описание видео)

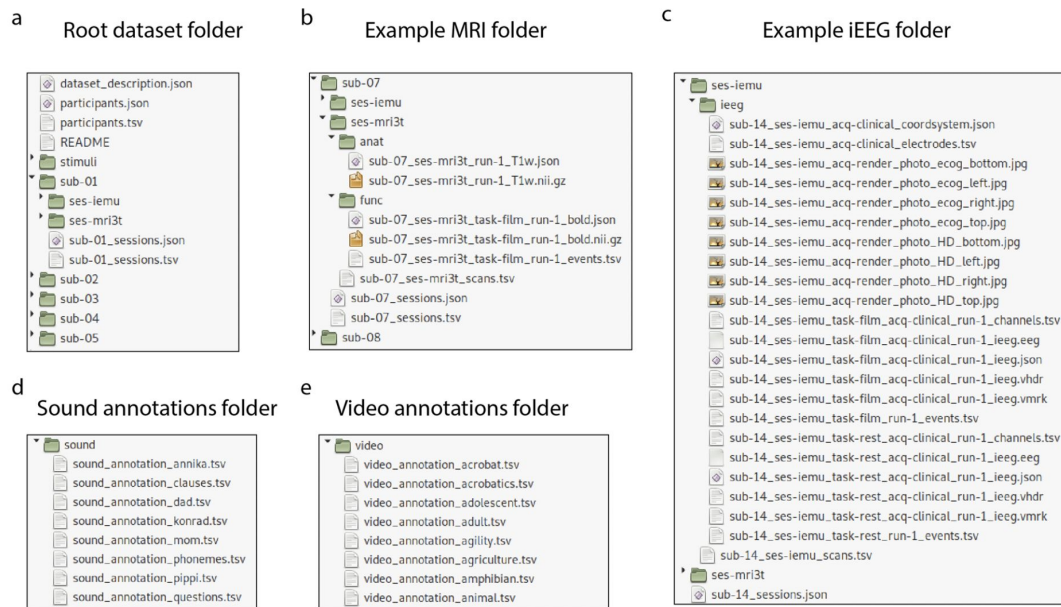
## Статьи

- Серия обзоров методов работы с видео  
(<https://towardsdatascience.com/deep-learning-on-video-part-one-the-early-days-8a3632ed47d4>)
- Inflated 3D CNN (<https://arxiv.org/pdf/1705.07750.pdf>)
- Трансформеры для видео (<https://arxiv.org/pdf/2201.04288v4.pdf>, <https://arxiv.org/pdf/2102.05095v4.pdf>) — на следующую неделю

## Обзор статьи

## Некоторые характеристики датасета

- 30 участников с данными fMRI
- 13 чередующихся эпизодов по 30 секунд (речь / музыка)
- Всё переведено на голландский
- аудио и видеодорожка аннотированы (время начала и конца появления каждого объекта)
- Снимки fMRI в формате BIDS
- Папка видео → 135 tsv файлов (для 129 объектов, и 6 персонажей)



**Fig. 2** Overview of data records. (a–e) Structure of folders and files in the dataset with example (f) MRI and iEEG folders from one subject.

# Speech Prediction in Silent Videos using Variational Autoencoders

## Обзор статьи

- Решается задача озвучивания видео с применением вариационного автоэнкодера
- Во время обучения на вход подается последовательность кадров и звуковой сигнал
- Сначала с помощью энкодеров получают эмбединги независимо для каждого аудио/видео фрейма
- Далее применяют LSTM и полносвязный слой для получения среднего и дисперсии вариационных распределений  $q$
- Оптимизируют ELBO (evidence lower bound)

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \sum_{t=1}^N \mathbb{E}_{q_{\phi_a}(z|a_t)} [\lambda \log p_{\theta_a}(a_t|z)] - \beta KL[q_{\phi_a}(z|a_t) || q_{\phi_f}(z|f_t)]$$

## Комментарий

- Чтобы решать нашу задачу, нужно будет заменить первичную обработку звукового сигнала (перевод в спектрограмму и audio encoder), а также последующее восстановление сигнала из спектрограммы (audio decoder, Griffin-Lim reconstruction) на соответствующие компоненты для обработки fMRI
- В исходном виде данную модель придется обучать отдельно для каждого участника эксперимента

