

# MAD-EEG

*an EEG dataset for decoding auditory attention to a target instrument in polyphonic music*

---

**Giorgia Cantisani, Gabriel Trégoat, Slim ESSID, Gaël Richard**

LTCI, Télécom Paris, Institut polytechnique de Paris, France

*Speech, Music and Mind 2019, TU Wien*



MIPFRONTIERS PROJECT

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 765068.

# Table of contents

1. Introduction
2. MAD-EEG
3. Data Validation
4. Conclusions

# Introduction

---

# Auditory Attention Decoding

**Auditory attention decoding** aims at determining, from the brain's activity, which sound source a subject is "focusing on" while listening to a complex auditory scene.

## State of the Art

- Previous works to decode the **attended speaker** in **multi-speaker environments** [Mesgarani and Chang, 2012, O'sullivan et al., 2014, Crosse et al., 2016];
- the **neural activity** tracks dynamic changes in the speech stimulus;
- a feature representation (*amplitude envelope, magnitude spectrogram, ...*) of the attended speech stimulus is reconstructed from the **unaveraged EEG** response.

*How can we transpose this problem to music?*

# Auditory Attention Decoding

**Auditory attention decoding** aims at determining, from the brain's activity, which sound source a subject is "focusing on" while listening to a complex auditory scene.

## State of the Art

- Previous works to decode the **attended speaker** in **multi-speaker environments** [Mesgarani and Chang, 2012, O'sullivan et al., 2014, Crosse et al., 2016];
- the **neural activity** tracks dynamic changes in the speech stimulus;
- a feature representation (*amplitude envelope, magnitude spectrogram, ...*) of the attended speech stimulus is reconstructed from the **unaveraged EEG** response.

*How can we transpose this problem to music?*

# Auditory Attention Decoding

**Auditory attention decoding** aims at determining, from the brain's activity, which sound source a subject is "focusing on" while listening to a complex auditory scene.

## State of the Art

- Previous works to decode the **attended speaker** in **multi-speaker environments** [Mesgarani and Chang, 2012, O'sullivan et al., 2014, Crosse et al., 2016];
- the **neural activity** tracks dynamic changes in the speech stimulus;
- a feature representation (*amplitude envelope, magnitude spectrogram, ...*) of the attended speech stimulus is reconstructed from the **unaveraged EEG** response.

*How can we transpose this problem to music?*

# Auditory Attention Decoding

## Study Case:

Detect and characterize auditory attention to **an instrument in polyphonic music**.

## Useful for

improving **MIR** tasks such as *music transcription*, *score following*, *source separation*...

## However...

- musical stimuli are **complex**;
- musical stimuli are intrinsically **different from speech**;
- there is a **lack of data**.
- **no previous works** (except [Treder et al., 2014] using averaging techniques).

# Auditory Attention Decoding

## Study Case:

Detect and characterize auditory attention to **an instrument in polyphonic music**.

## Useful for

improving **MIR tasks** such as *music transcription*, *score following*, *source separation*...

## However...

- musical stimuli are **complex**;
- musical stimuli are **intrinsically different from speech**;
- there is a **lack of data**.
- **no previous works** (except [Treder et al., 2014] using averaging techniques).



# Auditory Attention Decoding

## Study Case:

Detect and characterize auditory attention to **an instrument in polyphonic music**.

## Useful for

improving **MIR tasks** such as *music transcription*, *score following*, *source separation*...

## However...

- musical stimuli are **complex**;
- musical stimuli are intrinsically **different from speech**;
- there is a **lack of data**.
- **no previous works** (except [Treder et al., 2014] using averaging techniques).

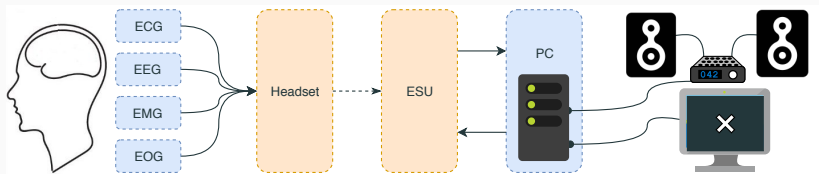
# MAD-EEG

---

**MAD-EEG**<sup>1</sup>: 20-channel surface electroencephalographic (EEG) signals recorded from 8 subjects while they were attending to a particular instrument in polyphonic music.

## Features

- **well-synchronized** musical stimuli and EEG responses;
- additional physiological signals: **EOG, EMG, ECG**;
- self-assessment of **attention, stress and fatigue**.



<sup>1</sup>G. Cantisani, G. Trégoat, S. Essid, G. Richard, "MAD-EEG: an EEG dataset for decoding auditory attention to a target instrument in polyphonic music", SMM19 workshop, 2019

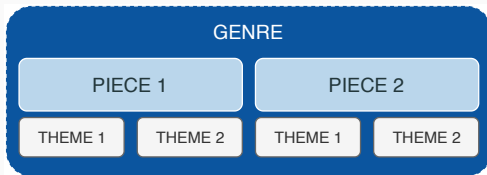
# Stimuli

## Stimuli

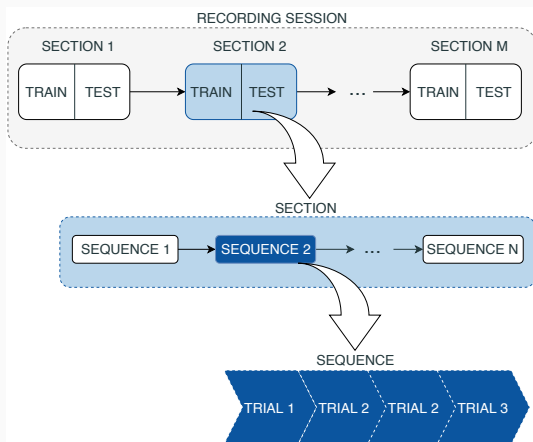
Real music compositions.

## Variants

- number and type of instruments in the mixture;
- melody/rhythmical pattern that is played;
- music genre;
- spatial rendering.



# Recording protocol

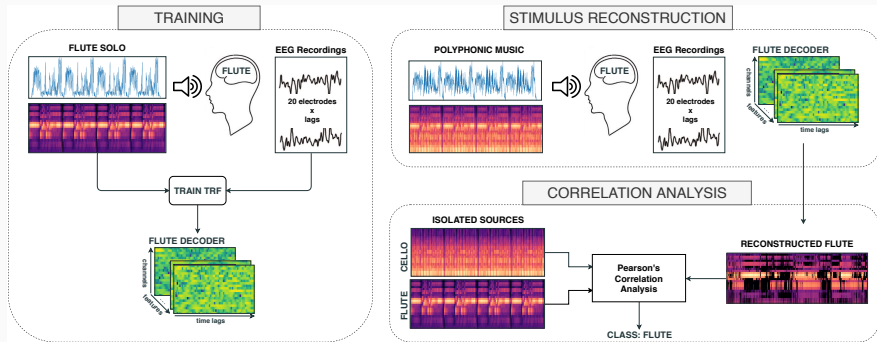


- long-enough stimuli sequences to study **single-trial** techniques;
- stimuli repetitions to study **averaging** techniques;

# Data Validation

---

# Goal & Pipeline

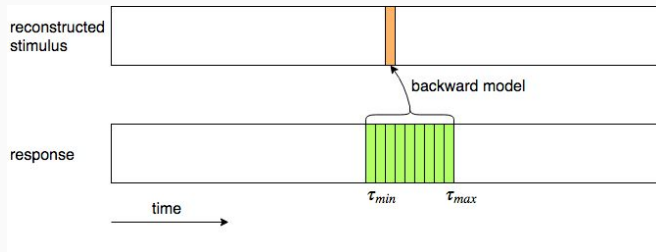


*For each of the 8 subjects train on 14 solos, test on 40 duets.*

# Stimulus Reconstruction

A stimulus representation  $\hat{s}$  is estimated from multi-channel neural data  $r$  through a model  $g$  which behaves like a *multi-channel Wiener filter*:

$$\hat{s}(t, f) = \sum_n \sum_{\tau} g(\tau, f, n) r(t - \tau, n)$$

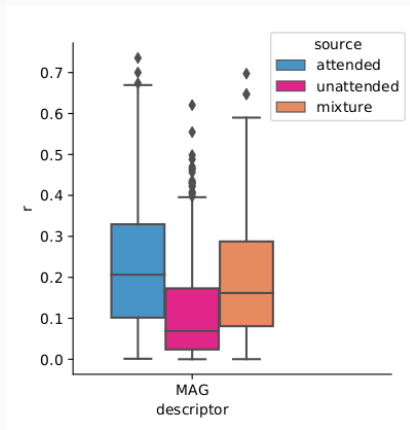


The filter is learned by solving a **linear regression** problem:  $\sum_t \sum_f [s(t, f) - \hat{s}(t, f)]^2$



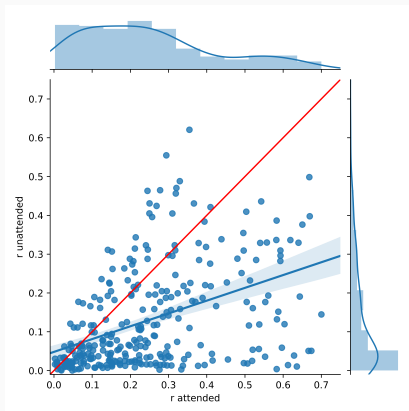
# Research questions

- Are we tracking attention or a general music entertainment?
- Are we tracking the target instrument?



- ATTENDED-MIXTURE: the two distributions are significantly different ( $p < 10^{-14}$ , Wilcoxon test)
- ATTENDED-UNATTENDED: the two distributions are significantly different ( $p < 10^{-27}$ , Wilcoxon test)

# Research questions



- Correlation ranges are comparable with the ones reported for speech by [Mesgarani and Chang, 2012].
- Accuracy rate of 78% ( $p < 10^{-15}$ , randomization test over 10000 repetitions).
- F1 score of 74% ( $p < 10^{-}$ , randomization test over 10000 repetitions).

# Conclusions

---

## Outcomes

- the neural activity is correlated with **musically relevant features of the attended source**.
- even this simple linear model is capable **tracking attention** and not a general entertainment to the music;

## Future works:

- extend the number of **recordings**;
- add more **variants** in the stimuli;
- include more **behavioural data**.

Data soon available on our lab's web page

<https://www.tsi.telecom-paristech.fr/aa0/en/2019/07/19/mad-eeg/>

**Thank you for the attention!**

---



Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016).

**The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli.**

*Frontiers in human neuroscience*, 10:604.



Mesgarani, N. and Chang, E. F. (2012).

**Selective cortical representation of attended speaker in multi-talker speech perception.**

*Nature*, 485(7397):233.



O'sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor, E. C. (2014).

**Attentional selection in a cocktail party environment can be decoded from single-trial eeg.**

*Cerebral Cortex*, 25(7):1697–1706.



Treder, M. S., Purwins, H., Miklody, D., Sturm, I., and Blankertz, B. (2014).

**Decoding auditory attention to instruments in polyphonic music using single-trial eeg classification.**

*Journal of neural engineering*, 11(2):026009.