Decoding auditory attention in polyphonic music based on EEG: a new dataset and a preliminary study

Giorgia Cantisani, Slim Essid, Gaël Richard
LTCI, Télécom Paris, Institut polytechnique de Paris, France

AESoP Symposium 2019, Leuven, Belgium

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

Table of contents

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**.
- previous works: [Treder et al., 2014].

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**.
- previous works: [Treder et al., 2014].

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**.
- previous works: [Treder et al., 2014].

Data & Methods

Dataset

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

Each attended instrument was previously heard in solo, as part of a *training phase*.

Features

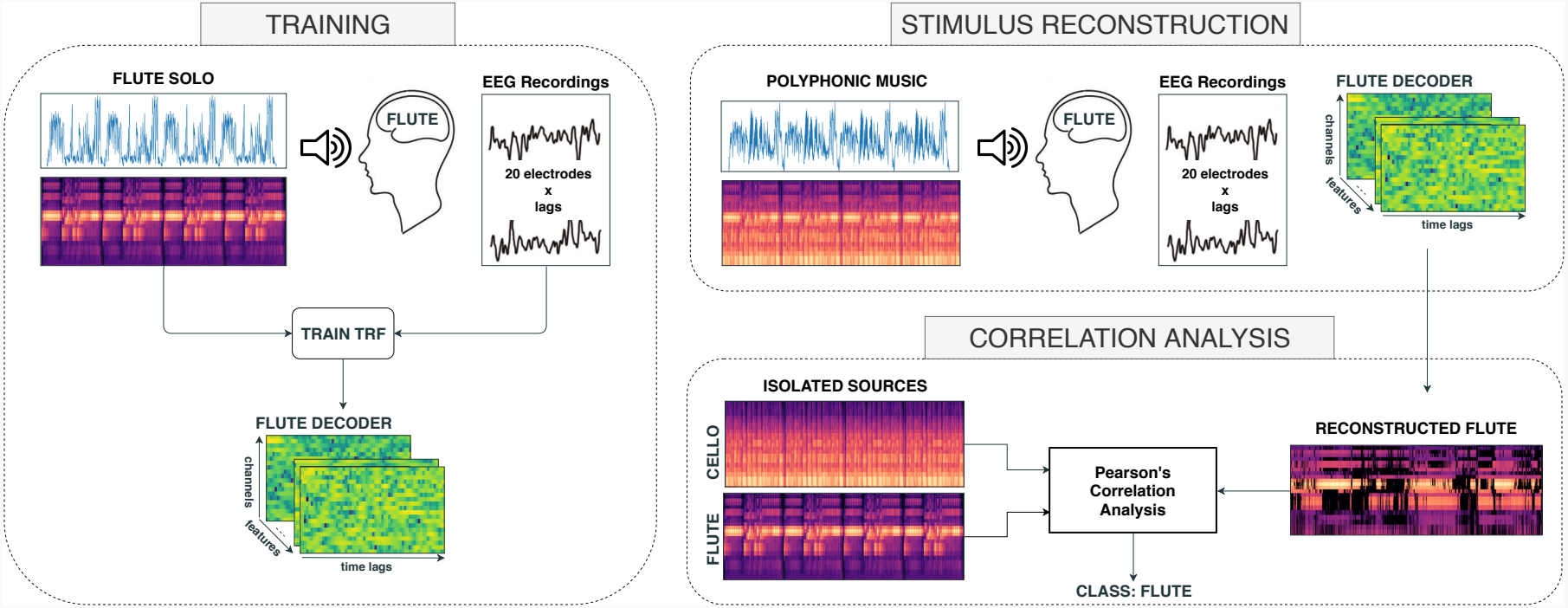
- **well-synchronized** musical stimuli and EEG responses;
- **real music compositions**;

Stimuli variants

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

[1] 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

Goal & Pipeline

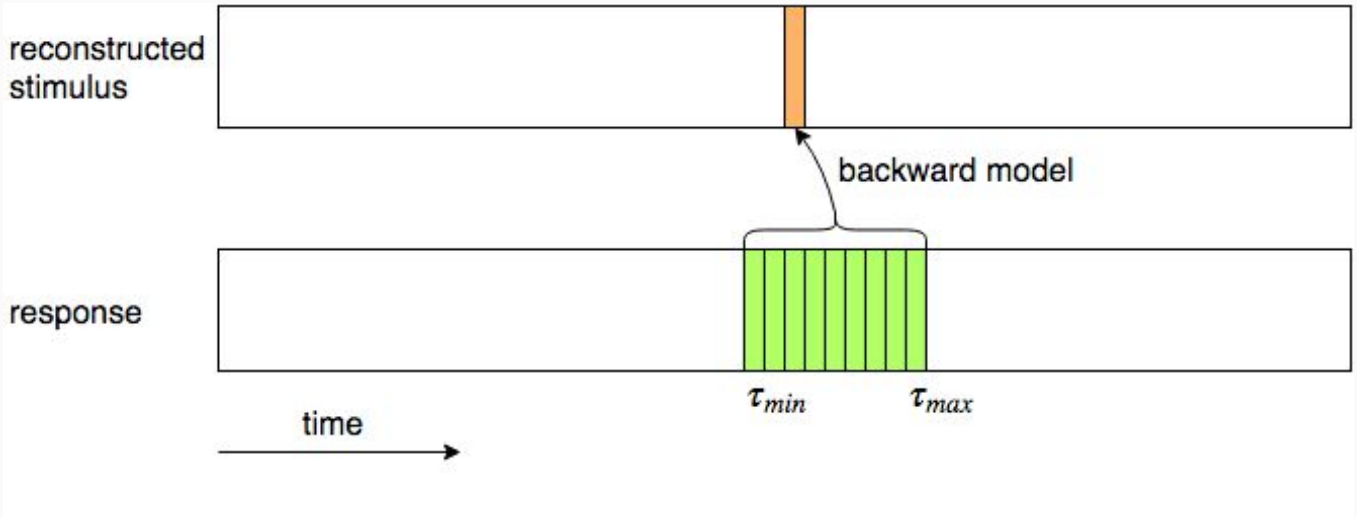


For each of the 8 subjects train on 14 solos, test on 40 duets and 24 trios.

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: $\sum_t \sum_f [s(t,f) - \hat{s}(t,f)]^2$

Results & Discussion

Research questions

- *Are we* **tracking attention** *or a general music entertainment*?
- *Are we tracking the* **target instrument**?
- *Which is the most suitable* **audio descriptor** *for such a task*?
- *How much* **variants in the stimuli** *influence the performances*?

Reconstruction performances

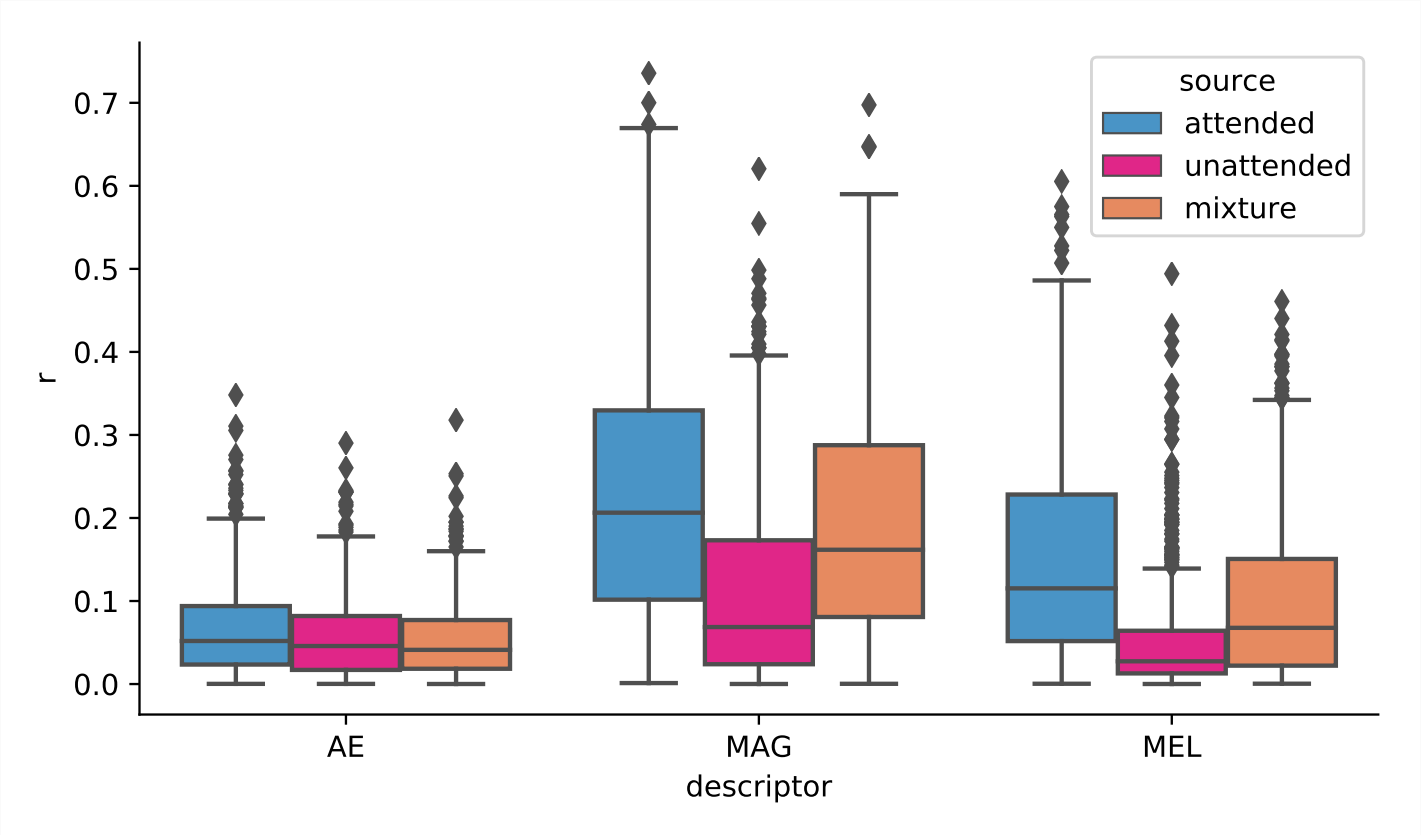


Figure 1: Pearson's correlation coefficients of the reconstructed stimulus with the attended source (blue), the unattended one (pink) and the mixture (orange) for the three audio descriptors.

Reconstruction performances - variants

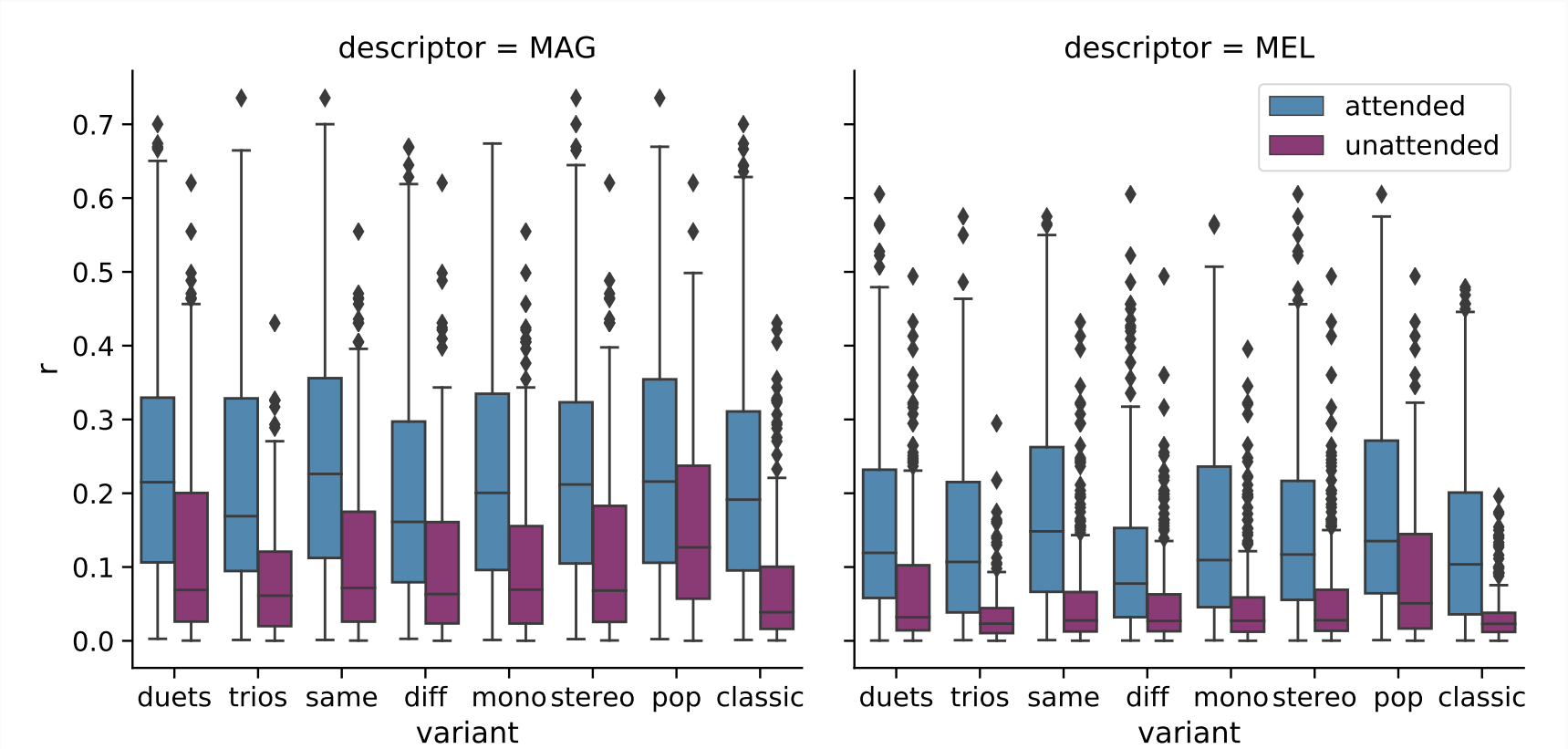


Figure 2: $r_{attended}$ and $r_{unattended}$ scores for the all the stimuli variants. Only MAG and MEL descriptors are considered. $r_{attended}$ and $r_{unattended}$ distributions are significantly different for all the variants ($p < 0.001$, Wilcoxon test).

Decoding performances

	F1 score (%)								
	all	ensemble		melody/rhythm		rendering		genre	
		duets	trios	same	diff	mono	stereo	pop	classic
AE	51 *	58 *	37 n.s.	48 n.s.	53 *	53 *	48 n.s.	54 *	48 n.s.
MAG	72 **	74 **	66 **	76 **	**65** **	73 **	**72** **	**64** **	79 **
MEL	**73** **	**79** **	**73** **	**79** **	60 **	**74** **	71 **	60 **	**83** **

Table 1: F1 scores for different subsets of the test set corresponding to variants in the stimuli and different audio descriptors. "**" denotes high ($p < 0.001$), "*" good ($p < 0.01$), and "n.s." no ($p > 0.05$) statistical significance of the results.

Reconstruction performances

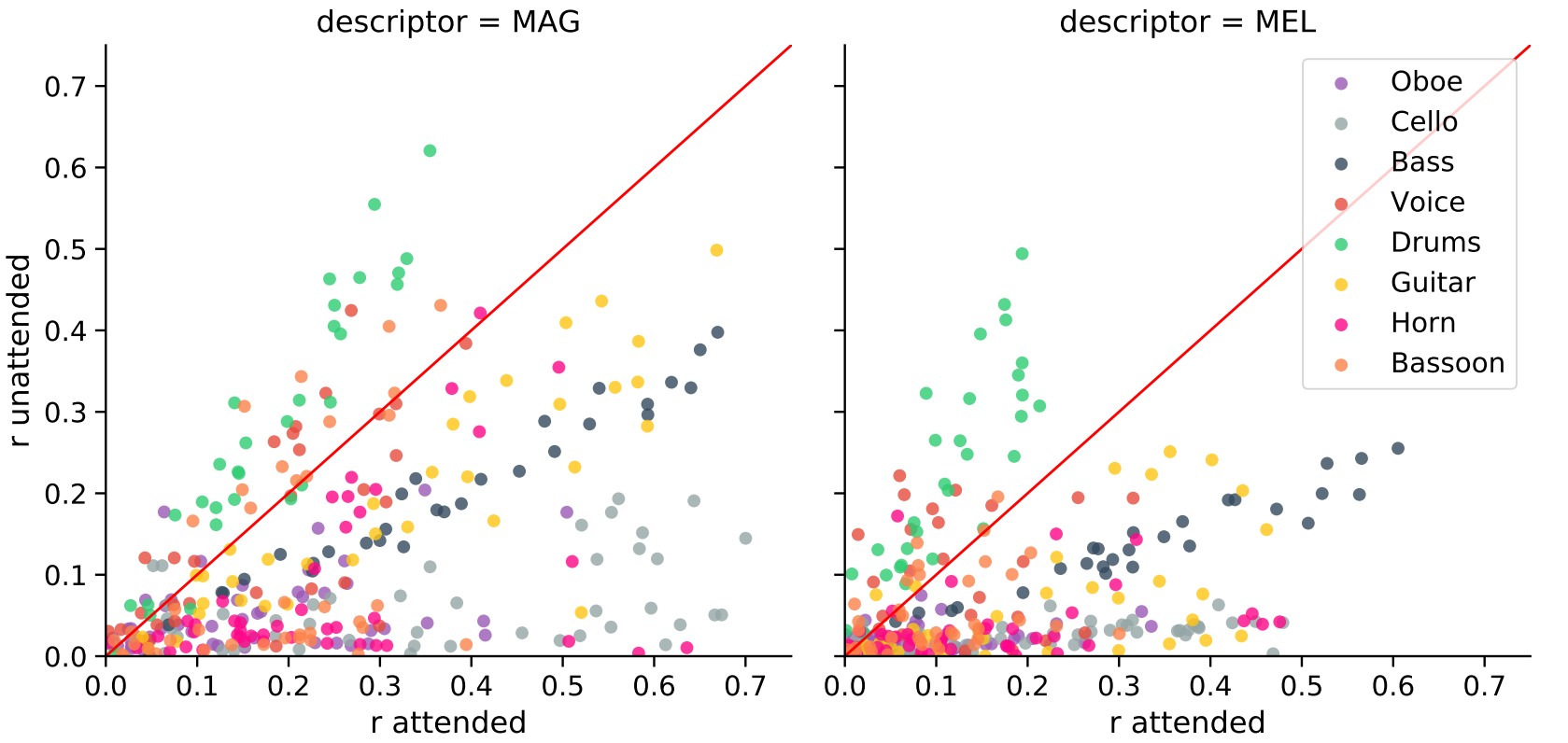


Figure 3: $r_{attended}$ against $r_{unattended}$ (only duets) for the MAG and MEL descriptors. The instruments are marked with different colors.

Conclusions

Conclusions

Take-home

- the model is **tracking attention** and not a general entertainment to the music;
- the neural activity tracks **musically relevant features of the attended source**;
- MAD-EEG: a **new dataset** to study the problem (soon available online!).

Limitations

- **limited generalization capability**;
- the model is tracking mostly the **pitch/harmonic contour** of the attended source;
- the **more instruments** in the mixture, the **more difficult** is the attention task.

Contributions

- G. Cantisani, S. Essid, G. Richard, *"EEG-based decoding of auditory attention to a target instrument in polyphonic music"*, WASPAA, 2019
- 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, 2019

Data soon available on our lab's web page
`https://www.tsi.telecom-paristech.fr/aao/en/2019/07/19/mad-eeg/`

Thank you for the attention!

References i


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.

References ii

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.