Discriminative responses to auditory representations revealed by means of EEG frequency tagging of synthetic sounds

Martina Berto[1], Emiliano Ricciardi[1], Pietro Pietrini[1], Davide Bottari[1]

[1]Molecular Mind Lab, IMT School for Advanced Studies Lucca, Lucca, Italy

INTRODUCTION

In order to differentiate sound objects, the auditory system relies on two **modes of representation:** (i) processing of local temporal features and (ii) statistical averaging.

The processing of local features allows to discriminate sounds based on temporal and spectral fine-grained details; however, when the amount of information exceeds storage capacity, these local features cannot be retained and are summarized into a set of statistics (statistical averaging)[1]. Through the employment of a computational model[2,3] it is possible to synthesize sounds comprising the same set of statistics but different local features. These sounds will be difficult to discriminate, but as statistical averaging occurs over time, their perceived similarity will depend on sound duration.

→ couples of different short sounds are easy to discriminate, even when they contain similar statistics, as discrimination relies on the (i) **processing of local features.**

→ pairs of long sounds will be distinguishable if they comprise a different set of statistics, or perceptually very similar if they comprise the same statistics. This is independent from the precise local features they encompass, as discrimination of long sounds relies on (ii) **statistical averaging.**

Two main questions remain unsolved:

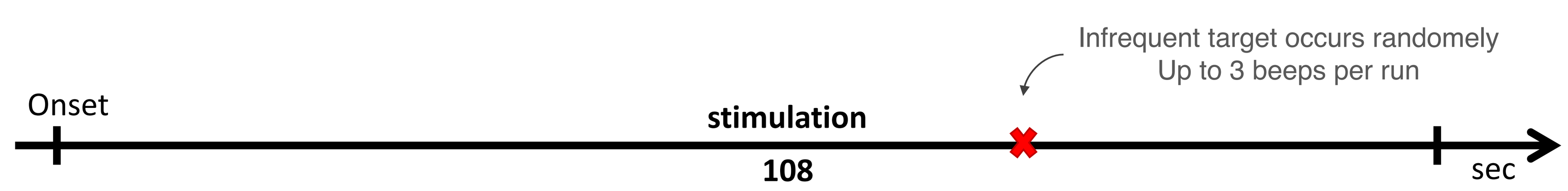
(1) Do discriminations based on these two modes occur **automatically**?

(2) Are these two modes **alternative** or **cumulative** processes and to which extent?

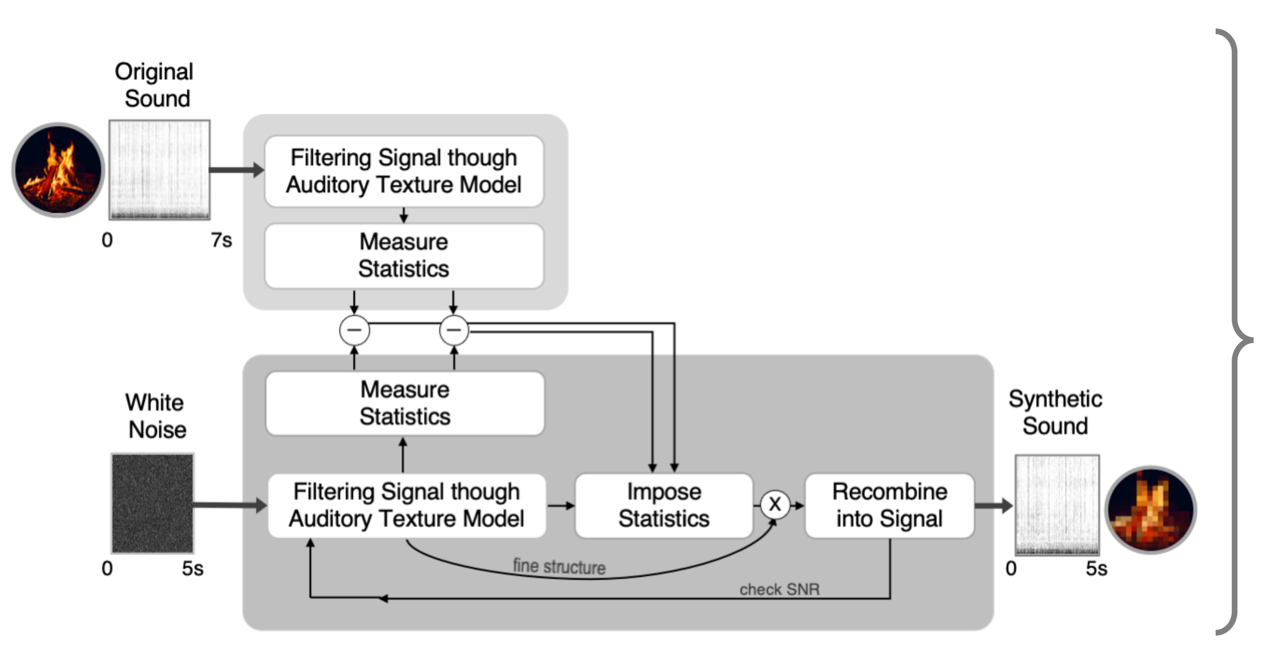
In order to address these issues, the study employed auditory EEG **frequency tagging** [4,5] to investigate neural correlates of automatic discriminative responses associated to the two modes; its combination with a **computational auditory model** allowed for total control of the statistics embedded in our experimental stimuli and enabled to interpret outcomes based on statistical variability between acoustic samples.

METHODS

- **Participants:** 14 normal-hearing individuals (mean age = 26.64 years; std = 2.37);
- **Task:** Detect an infrequent target sound (beep at a random rate);

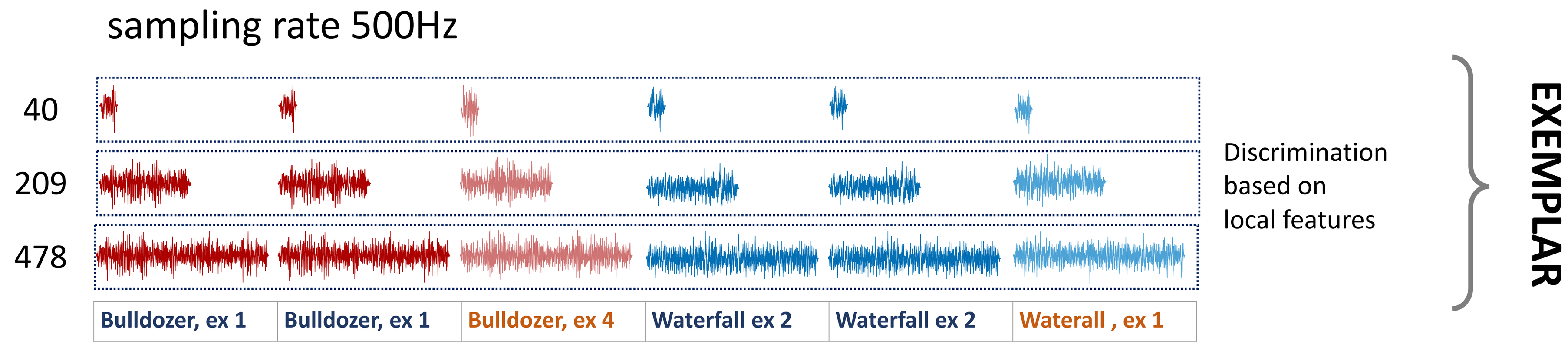


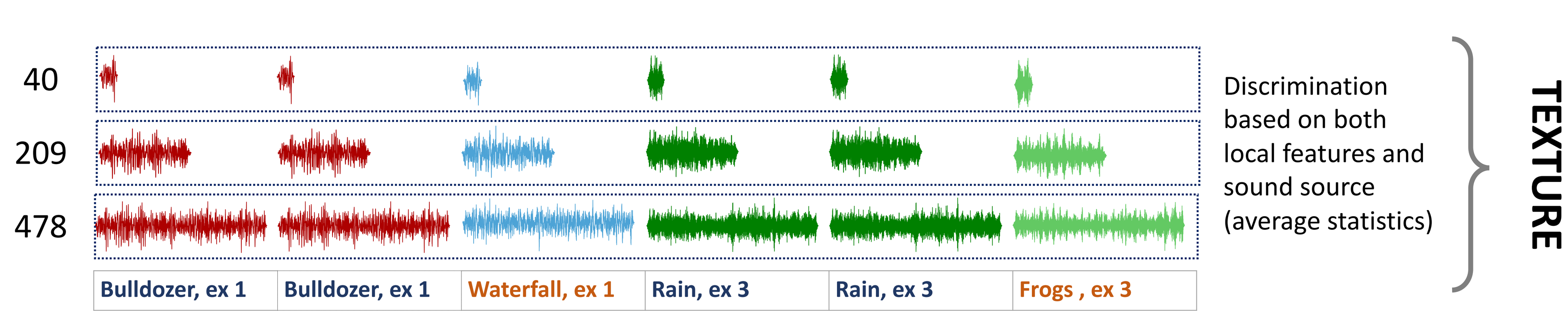
- **Auditory Frequency Tagging** combined with electroencephalography (EEG);
- **Stimuli:** Sound Textures synthesized through Auditory Texture Model[2] by imposing a set of auditory statistics (env mean, env var, env skewness, env cross-correlation, mod power, C1, C2).

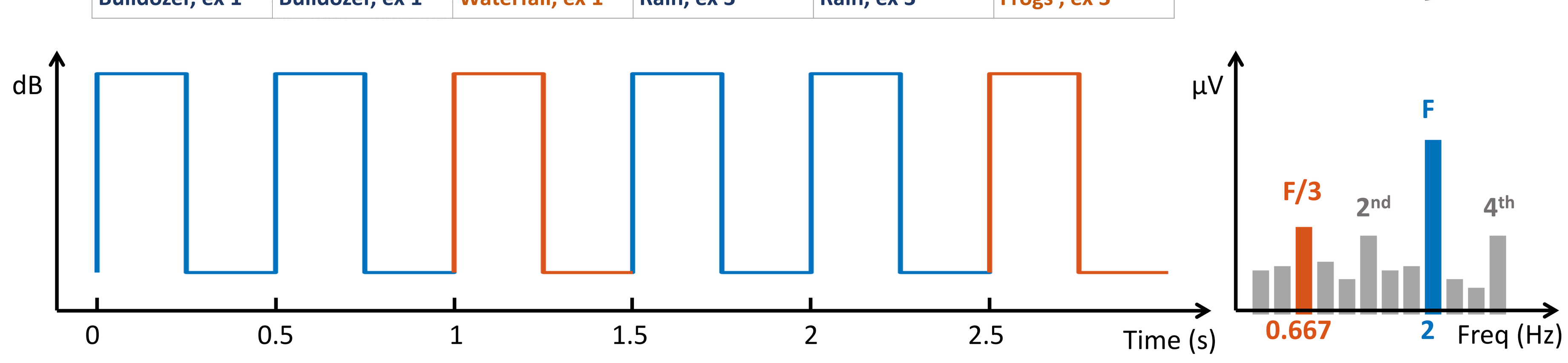


This procedure is repeated four times, resulting in four different **synthetic sound exemplars** for each **original Sound Texture**, all comprising the same long-term average statistics but different local details

- **Two experiments** (based on behavioral paradigm by McDermott et al., 2013)[1]: (1) Exemplar Discrimination (EX), (2) Texture Discrimination (TT).
- **Conditions for each experiment:** Only periodic; three durations: 40, 209, 478ms; one sound every 500ms.
- **EEG recordings**: 64-channel EGI system, referenced to central electrode E65, sampling rate 500Hz







RESULTS

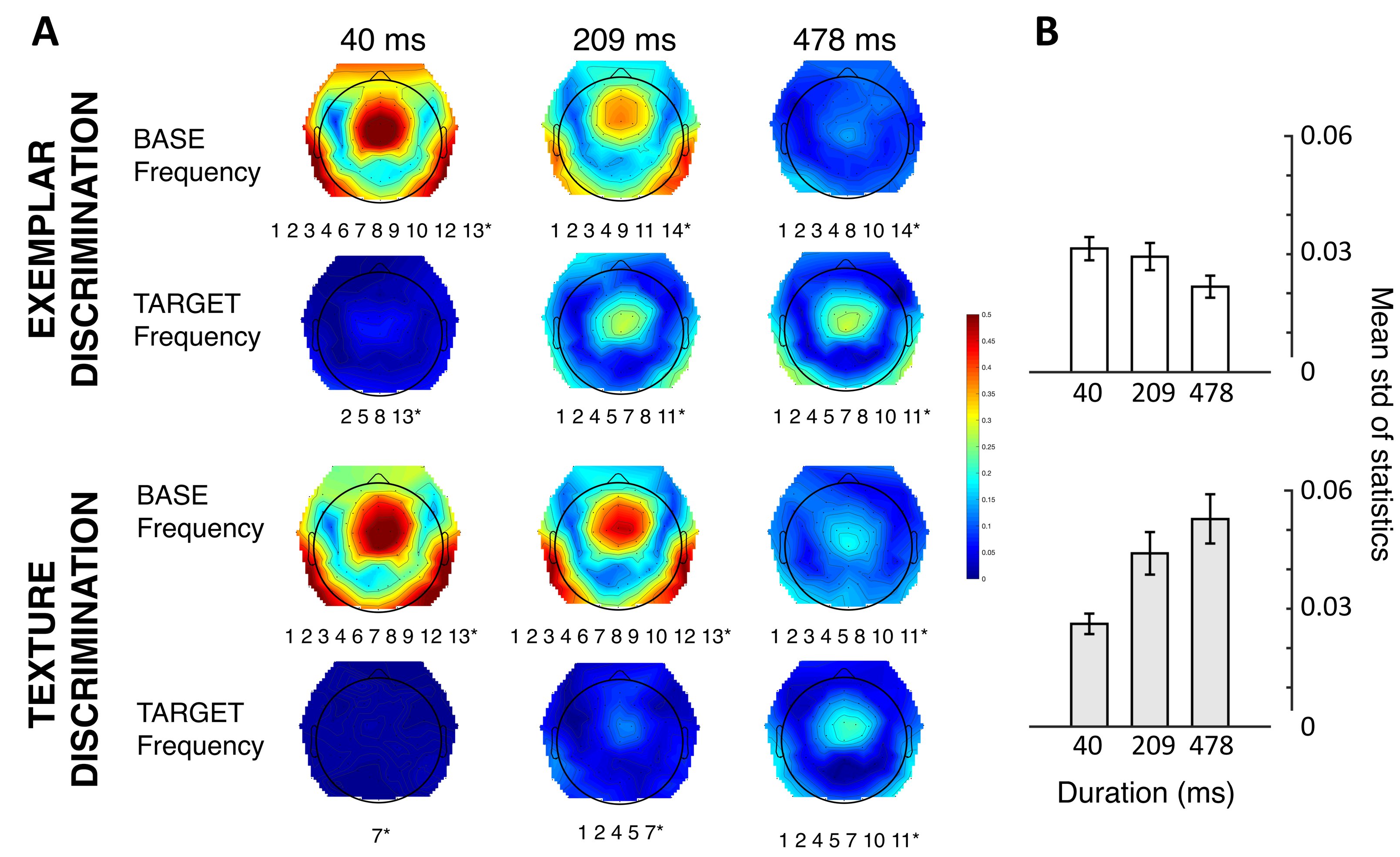


Figure 1.

(A) Topographical maps of grand average whole response of baseline subtracted data at base frequency (2Hz) and significant harmonics and at target frequency (0.667Hz) and significant harmonics, for both experiments (Exemplar and Texture) at each duration (40, 209, 478ms). **Sig Harmonics (z-score > 1.64) summed up to compute whole response;* No difference across experiment was observed at base frequency.

At target Frequency response increased with duration in both conditions. Overall, response to target frequency was always larger for Exemplar Discrimination as compared to Texture ($p < 0.001$). Note that stimuli durations were identical across experiments. Thus, differences were driven by the statistical properties of stimuli used in each experiment.

(B) Average statistics variability between standard and oddball sounds in both experiments. In Exemplar (top), as duration increases statistical variability tended toward zero as long-term average statistics converged; in Texture (bottom), as duration increased, statistical variability also increased, as the two sounds pertained to two different sound-objects.

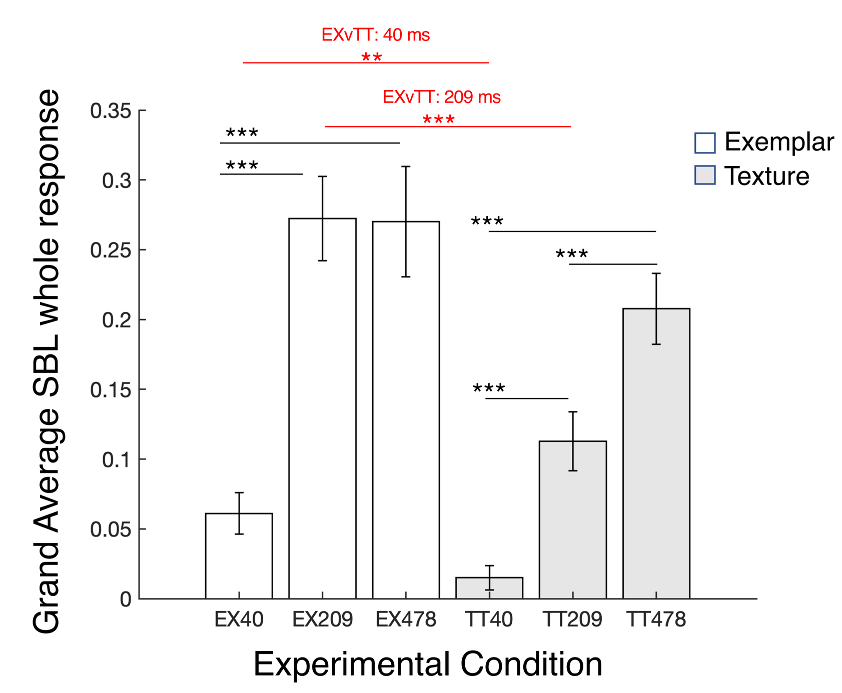


Figure 2.

Mean response at target frequency at central electrodes (Cz and FCz) for each condition and duration.

In both experiments, response significantly increased with duration. Response was bigger for Exemplar as compared to Texture (overall and for most durations).

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.5$

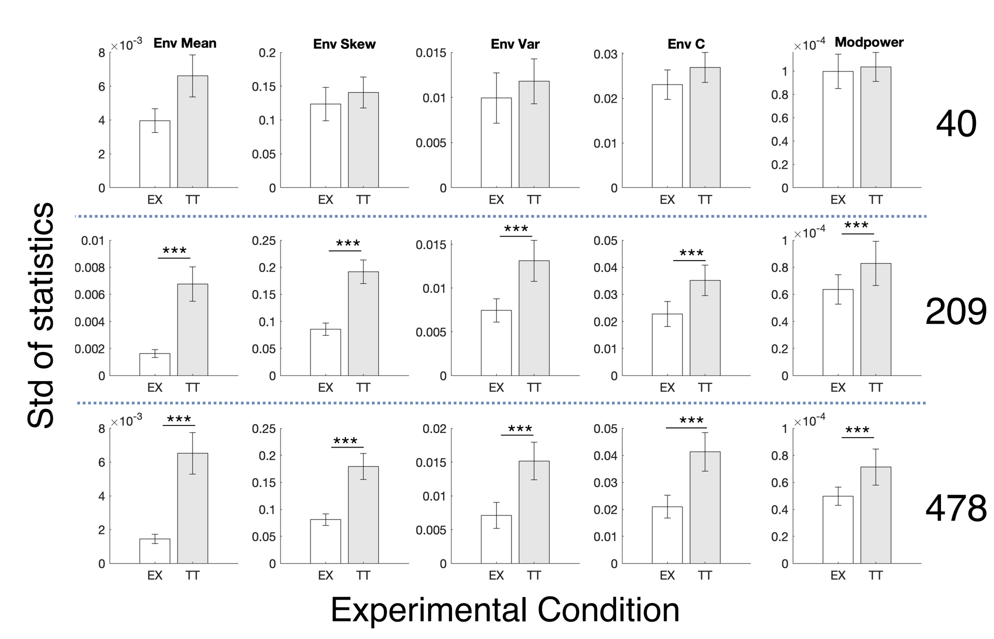


Figure 3.

Mean standard deviation of a set of five statistics, measured between couples of sounds (oddball and standard) employed in Exemplar (EX) or Texture (TT) Discrimination.

As duration increases, std decreases in Exemplar, but not in Texture, increasing the gap in statistical variability between the two experiments.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.5$

CONCLUSIONS

Automatic discriminative responses could be detected, and significantly diverged between the two experiments. In **Exemplar Discrimination**, response did not differ between intermediate (209ms) and long (478ms) durations. These findings suggest that local features are available for sounds lasting at least up to 478ms.

In **Texture discrimination**, response progressively increased following the increment in duration and statistics variability; however, overall response was smaller in Texture as compared to Exemplar Discrimination, suggesting that once statistical differences are detected, the response is dampened. We conclude that the two modes represent at least partially **alternative** processes which can be engaged **automatically** and according to sounds properties.

[1] McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature neuroscience.*

[2] McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron.*

[3] Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America.*

[4] Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of vision.*

[5] Nozaradan, S., Mouraux, A., & Cousineau, M. (2017). Frequency tagging to track the neural processing of contrast in fast, continuous sound sequences. *Journal of neurophysiology.*

Contact: martina.berto@imtlucca.it

momilab.imtlucca.it/