# Auditory Reverse Correlation on a Phoneme-Discrimination Task: Assessing the Effect of Different Types of Background Noise

Alejandro Osses Vecchi & Léo Varnet

Laboratoire des Systèmes Perceptifs (LSP), École Normale Supérieure (ENS), PSL University, Paris, France.
*contact: alejandro.osses@ens.psl.eu, leo.varnet@ens.psl.eu

**5-9 February 2022**
45th Annual MidWinter Meeting
Virtual Conference

## 1. Introduction

- **Reverse correlation** is a behavioral method that can be used to explore which properties of a sound are used as perceptually-relevant cues. In our case, the method assesses which acoustic cues of the **phonemes** /b/ and /d/ relate to correct or incorrect **listener responses** in a phoneme-in-noise task. Based on the tested noises, a fine-grained time-frequency map of acoustic cues can be obtained: The **auditory classification image (ACI**, Varnet et al. 2013, 2015**)**. We used three noise types and compared the robustness of the ACIs with respect to **noise statistics**.
- In this poster, we focus on the effect of the **statistics** of the background **noises** with respect to the efficiency and robustness of the method. We used three different noise types that have a flat long-term spectrum, but differ in the amount of temporal envelope fluctuations: **White noise (WN)**, white noise low-pass filtered in the modulation power spectrum **(MPS)**, and bump noise **(BP)**.

## 2. Methods: Experiments

- **Stimuli:** two natural male speech productions of **/aba/** and **/ada/** were taken from the OLLO database (Meyer et al. 2010, speaker S43M). The sounds were pre-processed to have equal duration and same acoustic energy in the first syllable. The speech level was adjusted according to the trial SNR, using one of three noise types.
- **Background noises**: 4000 realizations of each noise –**WN, MPS**, and **BP**– were used. The noises were presented at 65 dB SPL. The total level of the noisy trials was varied (roved) by ± 2.5 dB.
- **Participants:** two normal-hearing listeners.
- **Task:**
Each participant performed **4000 phoneme categorizations** for each noise, indicating whether the **last syllable was / ba/ or /da/.** The SNR was continuously adapted using a one-up one-down weighted rule to correct response rate of 70.7%.
In the the fastACI toolbox (Osses & Varnet, 2021a):

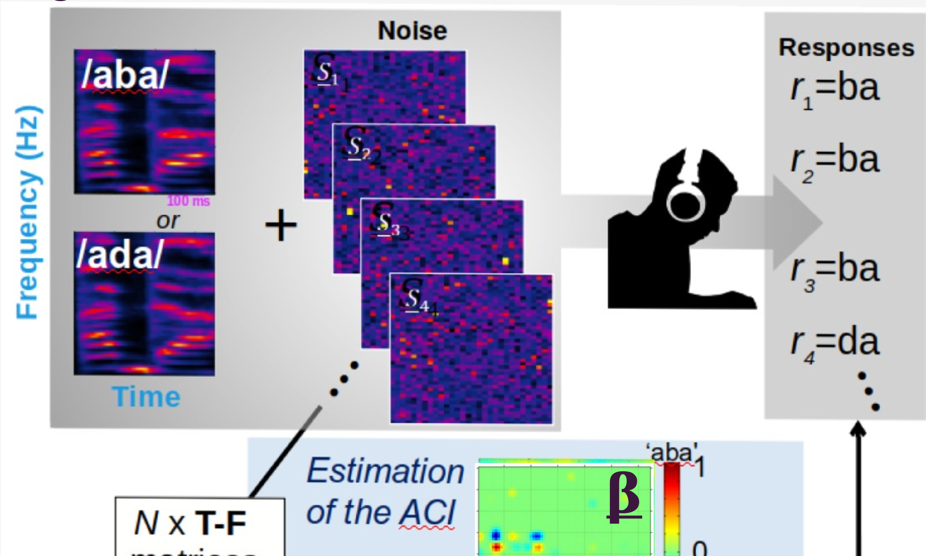



**Fig. 1. Revcorr method**

```
experiment = 'speechACI_Logatome-abda-S43M';
fastACI_experiment(experiment,Subject_ID,noise_type);
```

- **Data analysis:** The probability of "ga" answer was linked via a **Generalized Linear Model** (GLM) to a time-frequency representation in an ERB space of the presented stimulus. As we used in a previous study (Osses & Varnet 2021b), the GLM was fitted by **penalized likelihood maximization** with **sparseness prior**, a tradeoff between fitting the data well and obtaining a smooth ACI.
- The ACI (β, **Fig. 1**) shows **how the presence of energy at each time-frequency point influences the decision** (i.e. which parts of the stimulus serve as cues for categorization). Positive clusters of weights correspond to regions favoring response "ba", whereas negative clusters correspond to "da" regions.

## 3. Methods: Noise characterization

**All sounds are sampled at 16 kHz and have a duration of 0.85 ms.**

**Fig. 2A**
- **Power spectrum**: 512-point FFT, 90% overlap, 32-ms analysis windows

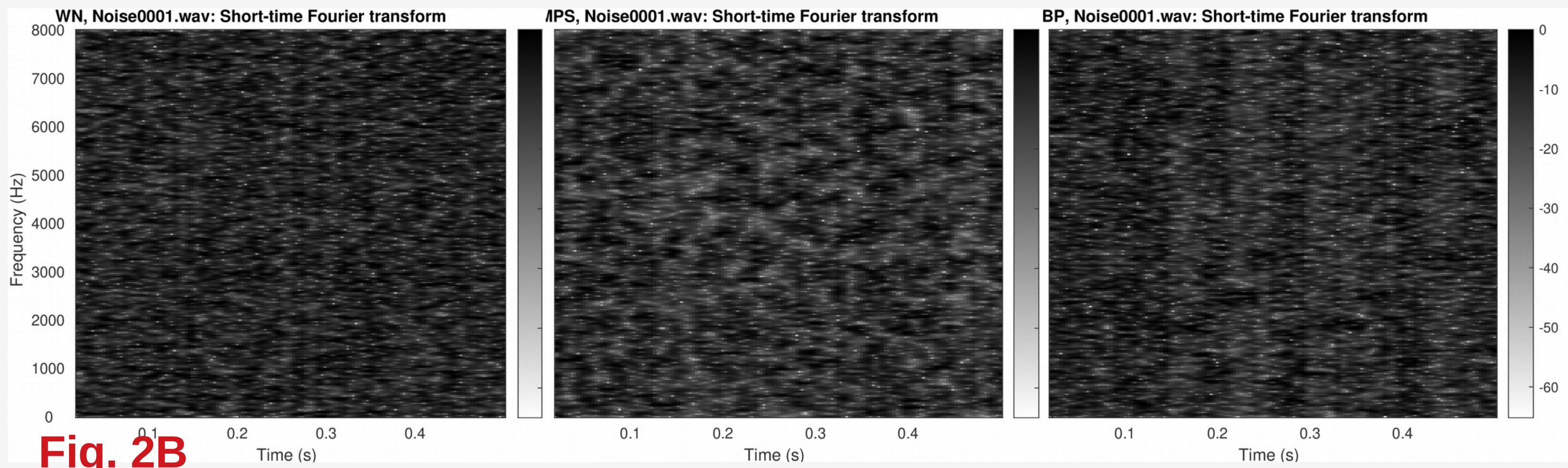
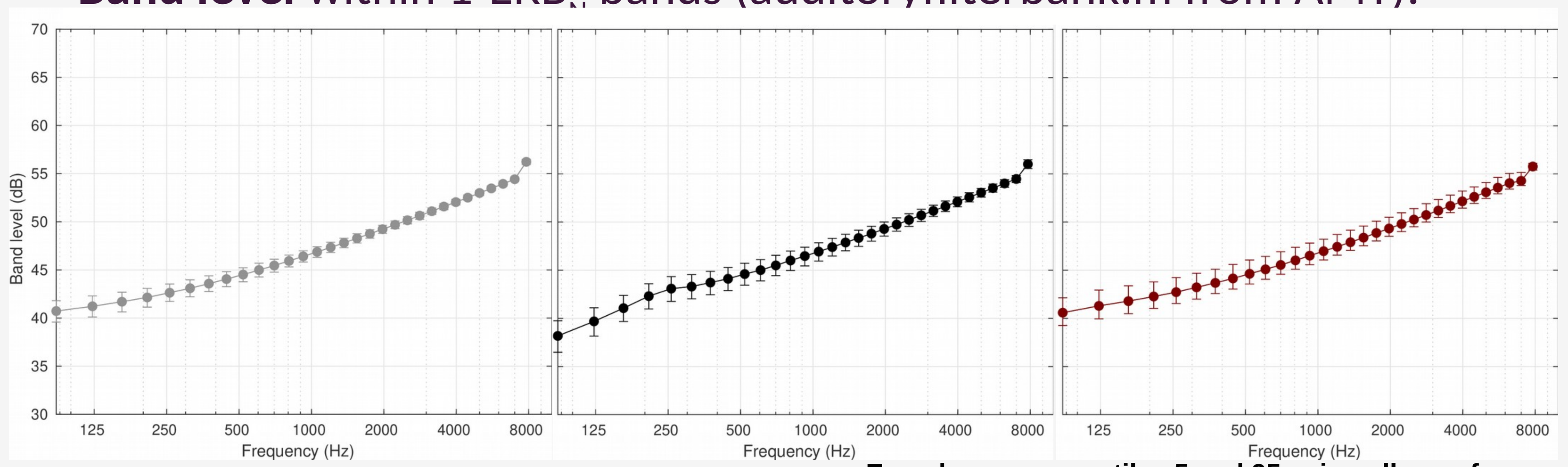
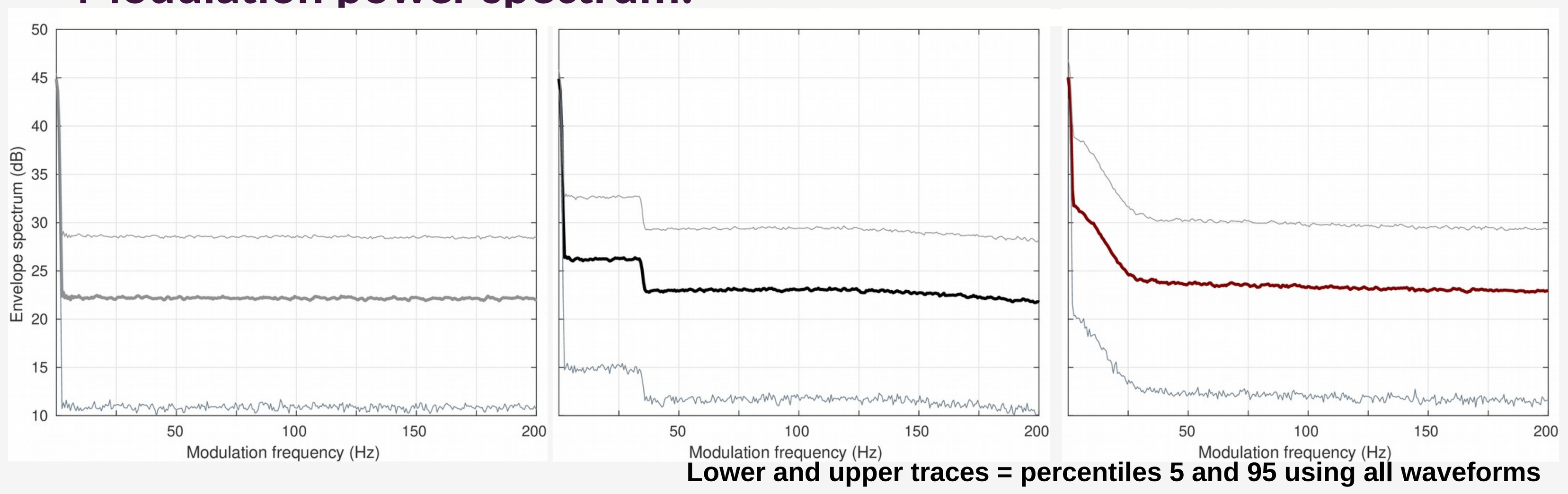



**Fig. 2B**
- **Band level** within 1-ERB, bands (auditoryfilterbank.m from AMT):



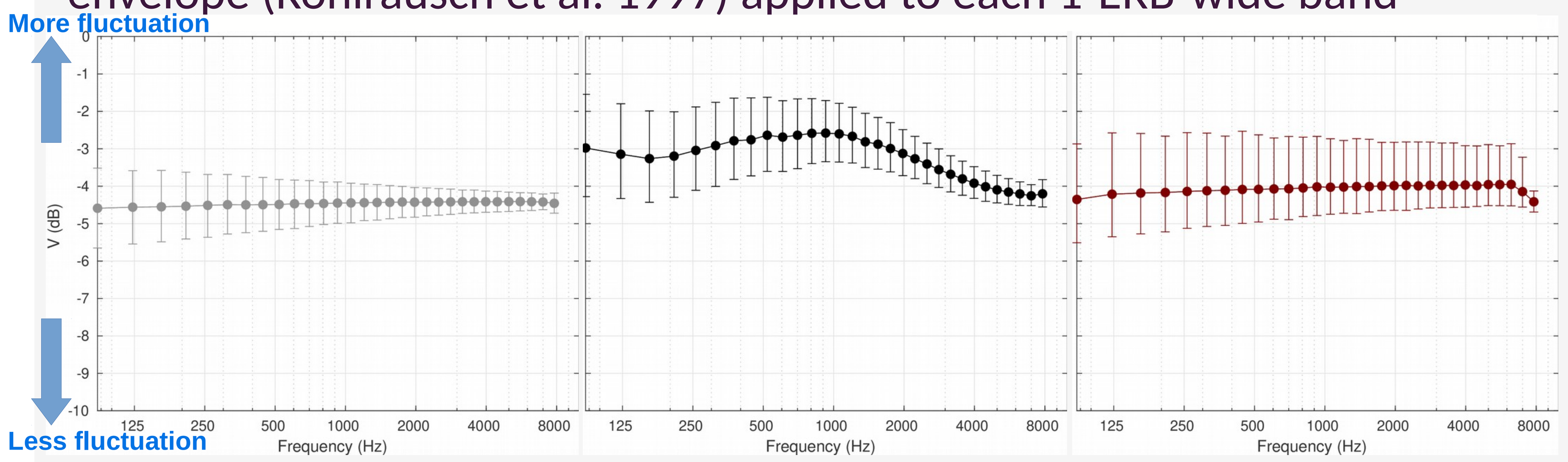

Error bars = percentiles 5 and 95 using all waveforms

**Fig. 2C**
- **Modulation power spectrum:**




Lower and upper traces = percentiles 5 and 95 using all waveforms

**Fig. 2D**
- **V metric**: Ratio between the standard deviation and the mean envelope (Kohlrausch et al. 1997) applied to each 1-ERB-wide band




Error bars = percentiles 5 and 95 using all waveforms

**References**

**Meyer et al. (2010)**. Human phoneme recognition depending on speech-intrinsic variability. J. Acoust. Soc. Am., 128, 3126–3141.
**Osses & Varnet (2021a)**. fastACI toolbox: the MATLAB toolbox for investigating auditory perception using reverse correlation (v1.0). doi:10.5281/zenodo.5500139
**Osses & Varnet (2021b)**. Consonant-in-noise discrimination using an auditory model with different speech-based decision devices. DAGA, 298–301.
**Varnet et al. (2013)**. Using auditory classification images for the identification of fine acoustic cues used in speech perception. Front. Hum. Neurosci., 7, 1–1
**Varnet et al. (2015)**. A psychophysical imaging method evidencing auditory cue extraction during speech perception. PLoS One, 10(3), 1–23

## 3. Experimental results

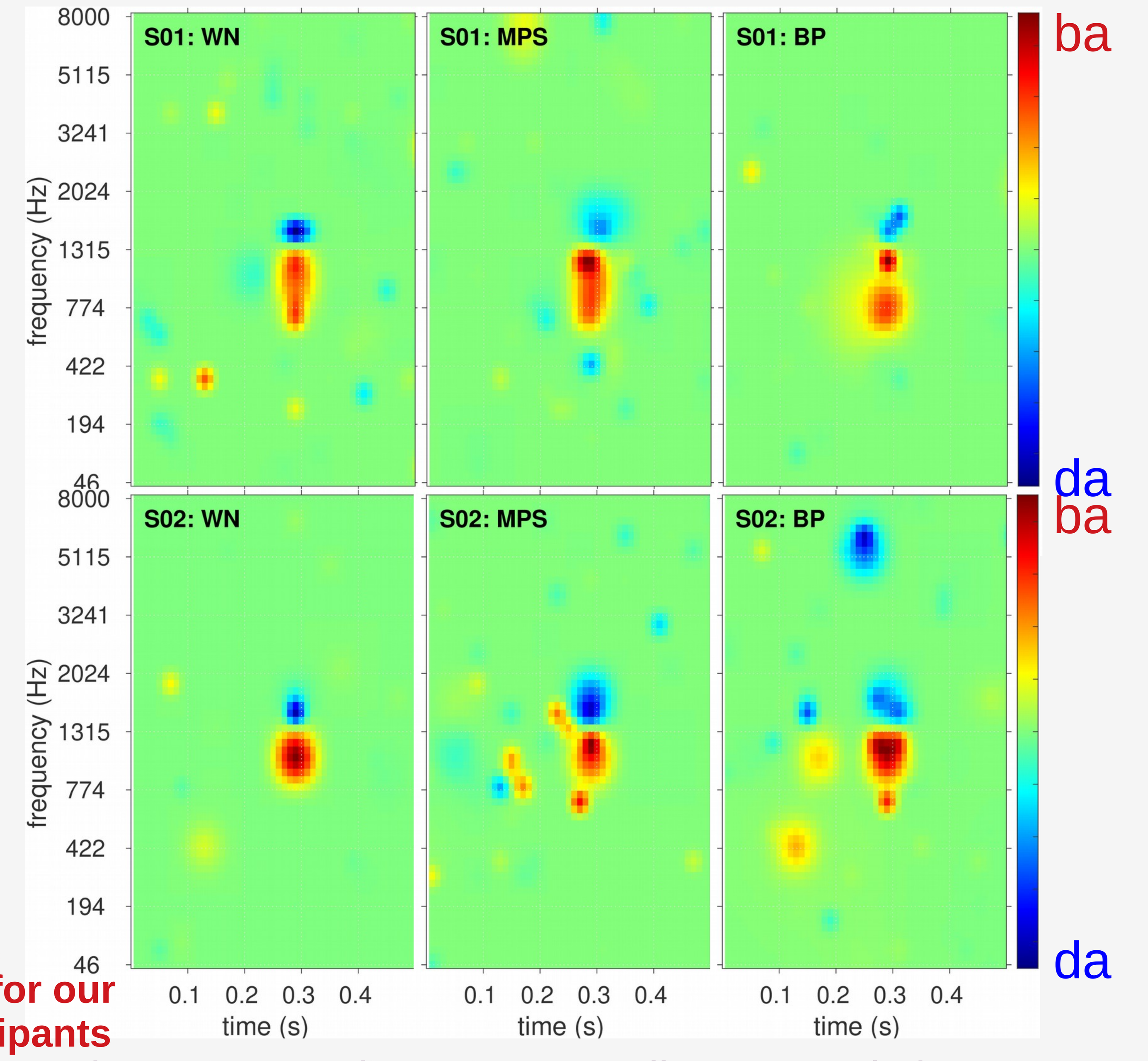



**Fig. 3. ACIs for our participants**

- In line with previous studies using **WN**, all **ACIs** revealed a primary cue around the time onsets of the second syllable and the frequency region of the F1 and F2 formants.

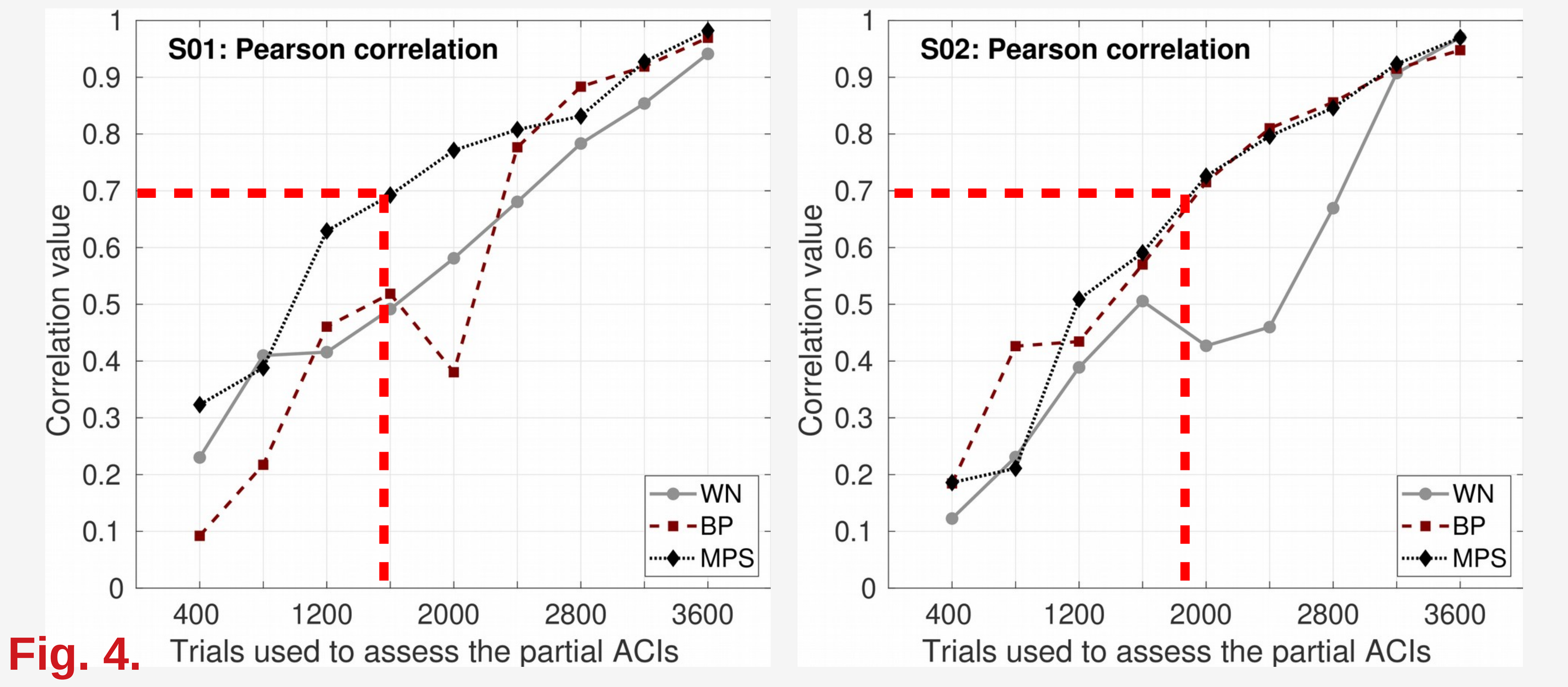



**Fig. 4. Correlation between "partial ACIs" and the ACI obtained with 4000 trials**

## 5. Summary

Our analyses suggest that the reverse correlation method applied to a consonant-in-noise discrimination task, the **ACI**, converges more quickly to a stable result when the background noises contain dominant components in the modulation frequency range between 0 and 40 Hz, which is the case for **MPS** and **BP** noises. The prominent envelope fluctuations in this range lead to more systematic confusion errors compared to white noise and, therefore, to higher prediction accuracy and more robust reverse correlation results.

**This research was supported by the French National Research Agency (Grant ANR-17-EURE-0017).**