Probing AM detection in noise with reverse correlation – a pilot study

Léo Varnet [1], *Christian Lorenzi* [1]

[1] Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, CNRS, Paris, France.
Contact: leo.varnet@ens.fr

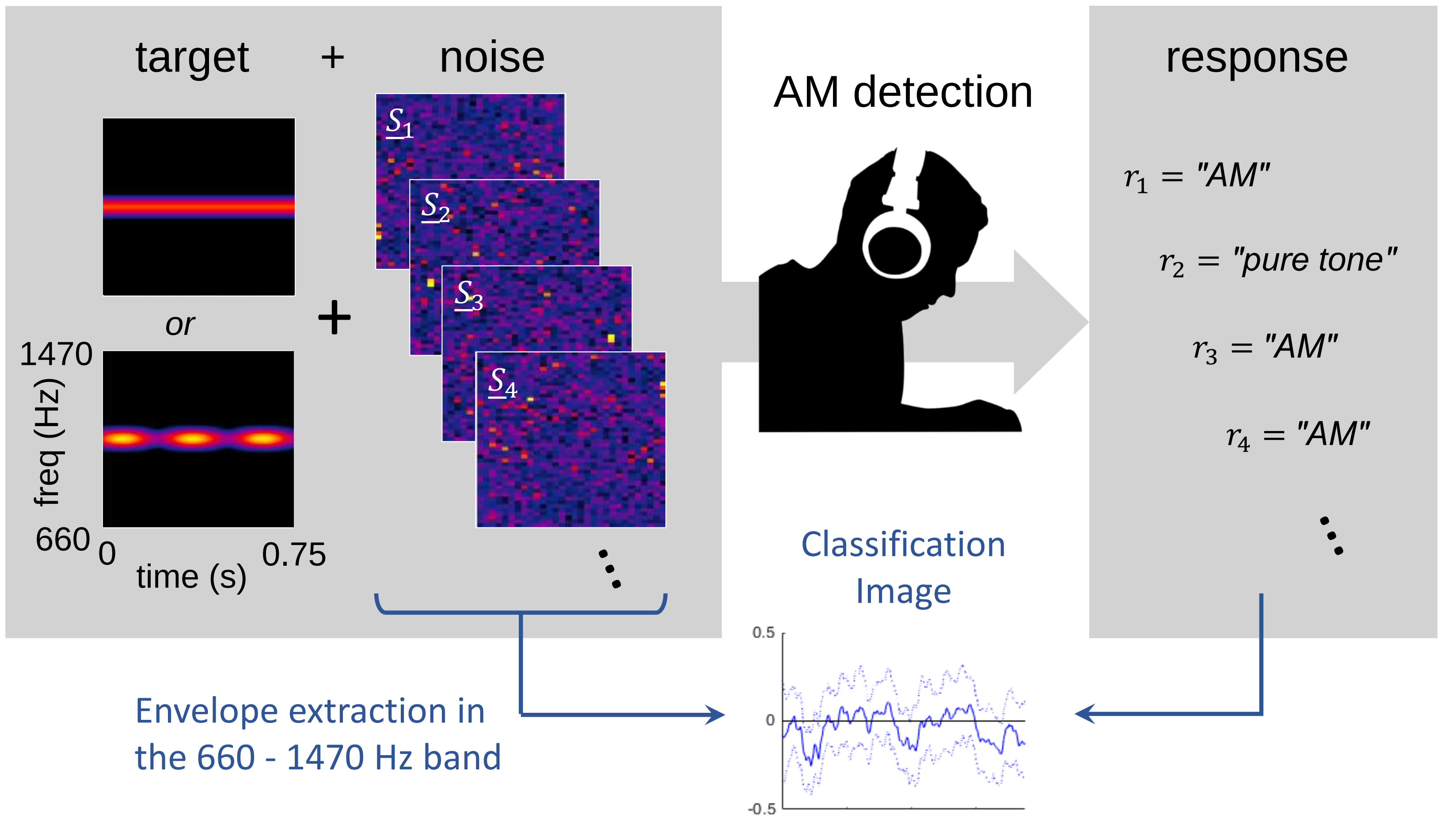
1. Introduction

- Evidence from psycholinguistic studies [1, 2] suggests that the impact of steady noise on envelope perception has two main components:

1. A **masking effect**: at low SNR, crucial envelope cues in the signal become masked, compromising recognition accuracy.
2. A **confounding effect**: intrinsic envelope modulations arising from the filtering of noise into critical bands can be confused with useful modulations in the signal.

Whereas masking leads to random answers in a phoneme identification task (i.e. stochastic errors), introduction of conflicting envelope cues results in predictable confusions (i.e. systematic errors).

- Traditional objective measures of speech-in-noise intelligibility rely on an estimate of the masking effect [e.g. STI, 3]. However, the concept of **signal-to-noise ratio in the envelope power domain** [$\mathrm{SNR}_{\mathrm{env}}$, 4, 5], quantifying the strength of speech modulations relative to a floor of spurious modulations, was found to yield better predictions.
- **Reverse correlation** techniques (aka **Classification Image**) are particularly suitable for exploring the systematic effect of noise on perception. They have been used in tone detection [6] and phoneme categorization [7] tasks. However, the only attempt to derive a Classification Image in an AM-detection task yielded mixed results [8].
- The present project aims at exploring the systematic effect of noise on AM detection. For this purpose, we used the Classification Image method to relate the modulation content of the noise (intrinsic fluctuations) with the response of a listener.



2. Methods

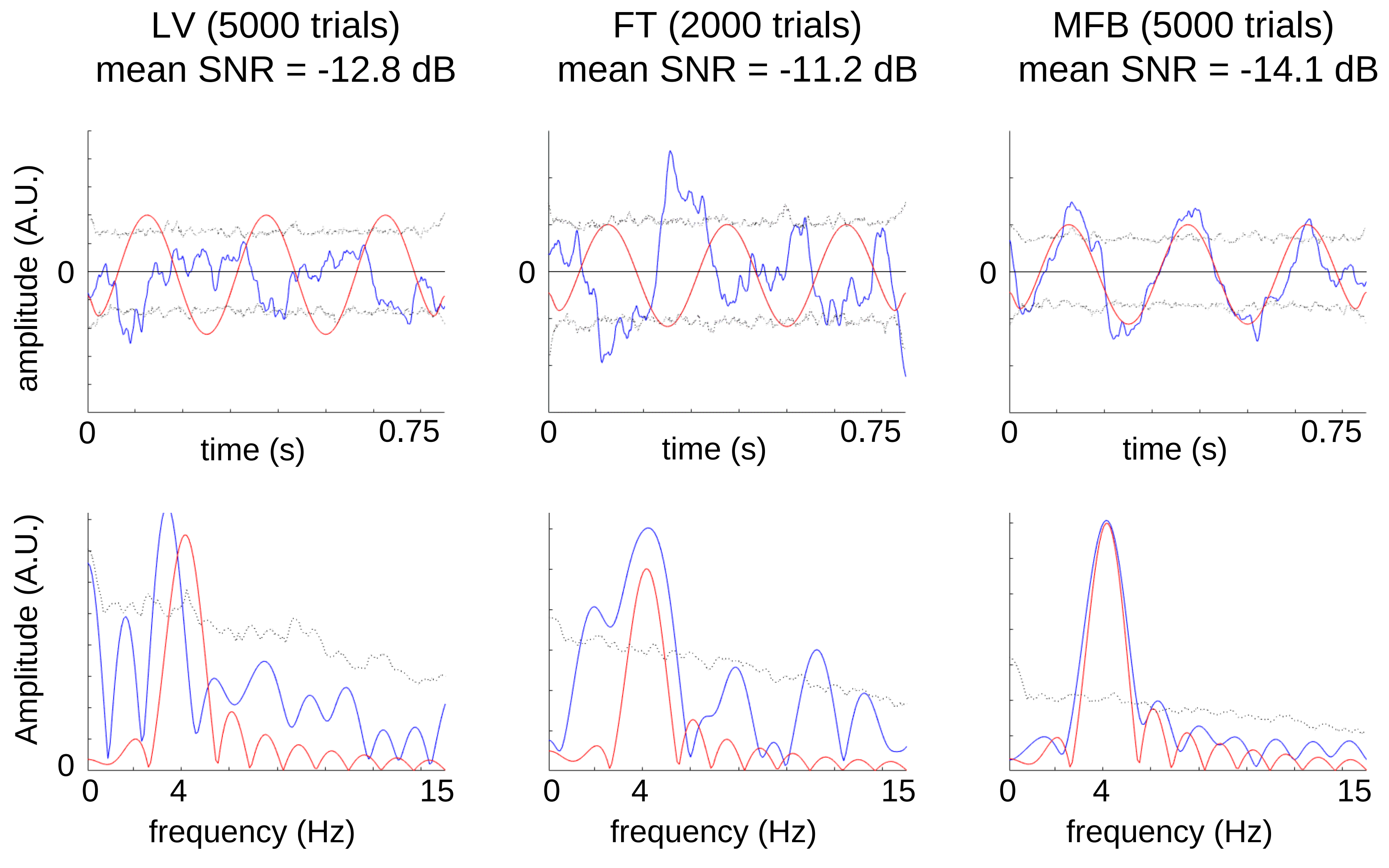
- Target signals: **Pure tone** vs. **4-Hz AM tone**. Duration = 0.75 s; carrier frequency = 1000 Hz; fixed AM phase.
- **Task:** 1-interval yes/no task with white noise masker.
- Modulation depth determined in a preliminary staircase experiment to yield ~75% correct answers at -10 dB SNR.
- SNR adapted continuously during the main experiment to ensure a correct response rate of 75%.
- **Participants**: LV (author, 5000 trials); FT (naive, 2000 trials). We also generated simulated data using the Modulation Filterbank Model [MFB, 9] on the same task (5000 trials).
- **Data analysis**: the envelope of each noise stimulus in the 660 - 1470 Hz band (+/- 3 ERBn around carrier frequency) is extracted. Then, envelopes are averaged conditional on the response of the participant. Finally, a **Classification Image** (estimate of the psychophysical kernel) is obtained as the difference between the two mean envelopes.
- An **ideal template** is also computed, for comparison purpose, by subtracting the envelopes of AM tone and pure tone.

3. Results

- Classification Images for participants LV and FT reveal a **systematic effect of noise** on AM detection: specific intrinsic noise envelope patterns significantly bias the listener towards one response or the other.
- These patterns show a **strong 4-Hz component** (target rate), as confirmed by a Fourier transform.

References:

[1] Dubbelboer & Houtgast (2007). *A detailed study on the effects of noise on speech intelligibility*. J. Acoust. Soc. Am. 122, 2865–2871.
[2] Drullman (1995). *Temporal envelope and fine structure cues for speech intelligibility*. J. Acoust. Soc. Am. 97, 585–592.
[3] Houtgast & Steeneken (2002). *Past, present and future of the Speech Transmission Index*. TNO reports.
[4] Dubbelboer & Houtgast (2008). *The concept of signal-to-noise ratio in the modulation domain and speech intelligibility*. J. Acoust. Soc. Am. 124, 3937–3946.
[5] Jørgensen & Dau (2011). *Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing*. J. Acoust. Soc. Am. 130, 1475–1487.
[6] Ahumada, Marken & Sandusky (1975). *Time and frequency analyses of auditory signal detection*. J. Acoust. Soc. Am. 57, 385–390.
[7] Varnet, Knoblauch, Meunier & Hoen (2013). *Using auditory classification images for the identification of fine acoustic cues used in speech perception*. Front. Hum. Neurosci. 7, 865.
[8] Ardoint, Mamassian & Lorenzi (2007). *Internal representation of amplitude modulation revealed by reverse correlation*. in proceedings of the 30th ARO midwinter meeting, Denver, Colorado, USA.
[9] King, Varnet & Lorenzi (2019). *Accounting for masking of frequency modulation by amplitude modulation with the modulation filter-bank concept*. J. Acoust. Soc. Am. 145, 2277–2293.
[10] Jørgensen, Ewert & Dau (2013). *A multi-resolution envelope-power based model for speech intelligibility*. J. Acoust. Soc. Am. 134, 436–446.



Blue lines: estimated psychophysical kernels for human (LV, FT) and simulated (MFB) participants, in the envelope (top) and Fourier (bottom) domains. Red lines: ideal template for the task. Dotted lines: 99% confidence interval under null hypothesis (permutation test).

- The **MFB model successfully recovers the ideal template** for this task. However, **human psychophysical kernels are in phase opposition with the ideal template**. Psychophysical weights are positive in the valleys of the AM target and negative in the peaks.
- This suggest that, unlike the MFB model, real participants represent the input signal in terms of **short-term $\mathrm{SNR}_{\mathrm{env}}$** [10]. Spurious envelope fluctuations are not confused with target envelope fluctuations, as it is the case for the MFB. Rather, dips in the intrinsic noise envelope are interpreted as modulations of the pure tone.

4. Conclusions

1. In this AM detection task, **noise acts primarily as a confounder** (rather than a masker): errors are not randomly distributed but directly predictable from the modulation content of the noise.

2. Whereas the MFB model (as most AM detection models) makes no distinction between useful modulations coming from the target and uninformative modulations coming from the interferer, participants are able to set apart the two sources of modulation.

Coming soon: more participants, 4-Hz and 16-Hz targets, FM detection task...

Acknowledgements: Léo Varnet and Christian Lorenzi are supported by the EUR Frontiers in Cognition grant ANR-17-EURE-0017.