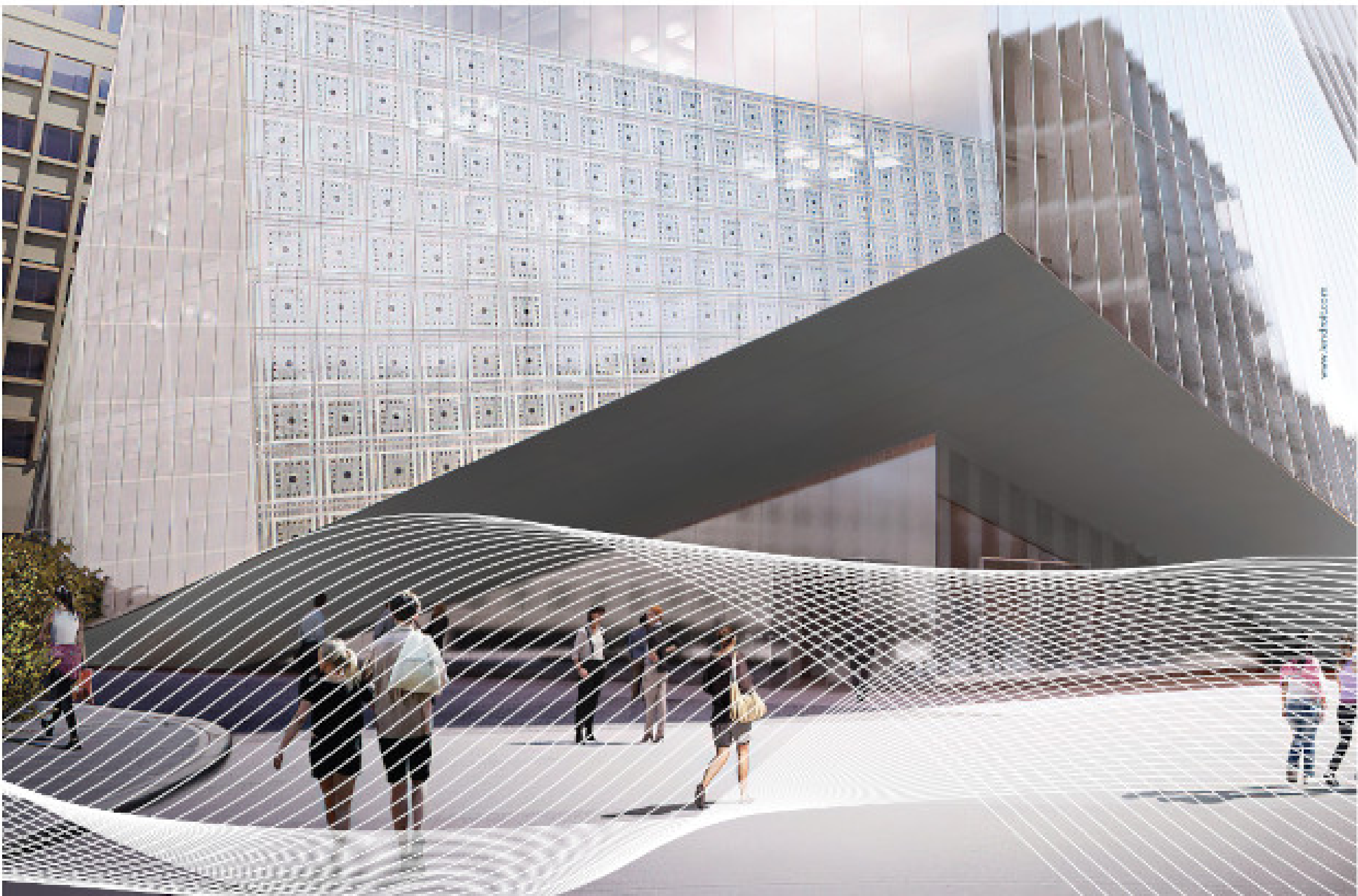
cfa
Paris 2025
27-30 avril
Congrès Français d'Acoustique

17e Congrès Français d'Acoustique
27-30 avril 2025, Paris

# 50 Years of Reverse Correlation: Replicating Ahumada et al.'s Pioneering Study

A. Le Bagousse [a], L. Varnet [a]

[a] Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France



In 1975, Ahumada, Marken & Sandusky addressed the fundamental question of tone-in-noise perception using an innovative psychoacoustic approach. They conducted a detection task involving a 500-Hz pure-tone target in white noise, and analyzed participants' trial-by-trial responses ('tone present' or 'tone absent') by correlating them with the spectrotemporal configuration of the noise presented on each trial. This approach enabled them to identify the acoustic cues essential for detection, demonstrating that listeners rely on the presence of energy at the target frequency, but also that the absence of energy in adjacent bands or immediately before the signal facilitates detection.

Beyond these findings, the article marks an important milestone in psychophysical research as it introduced for the first time the "psychophysical reverse correlation" method. This technique, which consists in analyzing the influence of noise on perception on a trial-by-trial basis, has since gained considerable momentum in psychoacoustics. The increasing interest in the approach is reflected in the recent development of fastACI, an open-source Matlab toolbox for designing, conducting and analysing reverse correlation experiments with any type of acoustic signal.

Here, we present a replication of the original study by Ahumada et al. as an illustration of the possibilities of fastACI. The experiment was reimplemented within the toolbox, and data were collected from nine participants. These subjects completed the task under conditions as close as possible to those of the original experiment, except from the SNR that was adjusted to match the performance levels reported by Ahumada et al. Four of them then repeated the task using the SNR from the original study. The results are broadly consistent with the original findings. However, the comparison between the two SNR conditions provide valuable insights into the type of mental templates that are used in such detection-in-noise tasks.

# 1 Introduction

Auditory reverse correlation (revcorr) is a psychophysical paradigm that allows to identify the acoustic features in the test stimuli that participants effectively used as perceptual cues during a listening experiment, with only minimal prior assumptions on the cues being sought. This method relies on two key components: (1) the introduction of random fluctuations into the stimulus (such as adding background noise) and (2) the trial-by-trial analysis of the relationship between the specific noise samples and the corresponding participant responses [7, 8]. More precisely, the revcorr approach is based on a detailed analysis of the confusions made by participants. By examining which noise features systematically interfere with their decision, this technique enables researchers to infer the acoustic cues that listeners rely on in the stimulus, and to visualize them in the stimulus space (usually a time-frequency space).

This method was applied for the first time by Ahumada et al. in a series of two experiments focusing on the ability to detect a pure tone in a white Gaussian noise masker [1, 2]. Their primary objective was to identify the acoustic features our auditory system relies on to detect the presence of the tone. At the time, significant research had been conducted using standard experimental protocols to investigate and model the role of energetic cues [5, 17]. In 1975, Ahumada Marken and Sandusky (AMS75, [2]) approached the same question by analyzing the relationship between the fine acoustic details of the noise and participants' trial-by-trial responses. They applied multiple regression to the spectrotemporal representation of the noise in each trial to estimate the contribution of different auditory features to the listener's decision regarding the presence or absence of the tone. Their results showed that, in this tone-in-noise detection task, the greatest weight is assigned to the signal frequency, with negative weights at frequencies above and below the signal frequency, and immediately before the signal. Figure 1 shows the individual results for the 6 participants engaged in the original study.

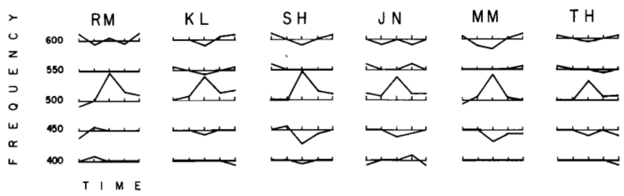



Figure 1: Results obtained for the 6 participants in the original study. Bold lines indicate the regression weights for different time-frequency bins (see Methods section). The central time-frequency bin corresponds to the position of the target tone. Adapted from [2].

Since this seminal study, the reverse correlation approach has become very popular in psychoacoustics. Recent applications include studies on loudness perception [16], modulation perception [18, 15], phoneme-in-noise perception [20, 19, 11, 3], word segmentation [13], and paralinguistic social judgements [14].

Our team has recently developed a Matlab toolbox designed to simplify the implementation and analysis of revcorr experiments. The fastACI toolbox [12], available at https://zenodo.org/records/5500138, provides a set of functions that makes it easy to create a revcorr experiment by specifying the stimuli used and the experimental parameters, as well as to conduct the experiment with human participants or auditory models, and to analyze the collecte data. This article aims to demonstrate a practical application of the toolbox by replicating the seminal study of AMS75 while also providing new insights into this foundational work. The experiment was reimplemented within the toolbox and data were collected from nine participants. These subjects completed the task under conditions as close as possible to those of the original experiment except from the SNR that was adjusted to match their performances with those reported by AMS75. Four of them then repeated the task using the SNR from the original study.

As the deviations from the original protocol were only minor, we expected to replicate the main result of the original study (Figure 1). Namely, the presence of noise energy in the region of the target tone should correlate positively with the perception of a tone. A similar positive correlation is expected in the segments following the tone, in the same frequency band. Conversely, the correlation should be negative in higher and lower frequency bands during the signal interval.

# 2 Methods

The experiment is available within the toolbox under the name `replication_ahumada1975`. Similar to the original study, 3,200 stimuli were presented consisting in 500-msec Gaussian white noise generated with a sampling frequency of 10,000 Hz. Half of these stimuli also included a 100-msec 500-Hz tone, added to the masker noise with a fixed signal-to-noise ratio ($10 \log_{10}(E_s/N_0) = 11.8$ dB). Tone-present and tone-absent stimuli were presented in a random order, through headphones, at a 65 dB sound-pressure level. Given the difficulty of the experiment, probe trials with a more favorable SNR (31 dB higher, although note that this parameter was not specified in the original article) were presented every tenth stimulus.

As described in the Results section, our participants performed substancially worse than those in the original study when tested at the same SNR. One possible explanation is that the experiment carried by AMS75 may have involved an extensive, though unreported, training phase. To address this issue, we included a second condition with an adjusted SNR to better match the performance levels reported by Ahumada et al. The SNR in this second condition was set to $10 \log_{10}(E_s/N_0) = 18$ dB. Nine participants performed the 18-dB condition and four of them also completed the 11.8 dB condition.

For the purpose of the demonstration, we chose to deviate from the original experiment in three ways. Firstly, participants were instructed to provide a yes/no response (tone present vs. tone absent) instead of a 4-point Likert scale judgment. The use of binary responses was preferred because it aligns with modern reverse correlation studies and the type of statistical models that are now used to analyze revcorr data. Secondly, the initial study included only a single block of 400 stimuli repeated eight times in random order. As this methodological choice was likely guided by computational constraints of the time, we decided instead to generate 3,200 independent stimuli, divided into 8 random blocks of 400 trials. Finally, stimuli were presented at a 65 dB instead of 85 dB sound-pressure level, as the latter was deemed too loud by the participants.

In line with the original study, the time-frequency representation of the noise was obtained by computing the energy values within a 5-by-5 matrix of 25 spectrotemporal regions, each spanning 50-Hz by 100-ms. This processing

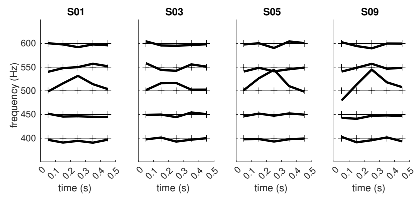



Figure 2: Results obtained for the 4 participants in the 11.8-dB condition (SNR matched with AMS75).

of the sounds is carried by an experiment-specific function `replication_ahumada1975_dataload` available in the toolbox. Perceptual weight matrices were obtained by correlating these coarse spectrograms with the behavioral responses (type of analysis: `correlation`). This is mathematically equivalent (up to a scaling factor) to the linear model used by AMS75, assuming the noise energy values are independent – a reasonable approximation given the size of the spectrotemporal regions.

The experiment can be easily run by installing the fastACI toolbox and using the command `fastACI_experiment`. All analyses and figures from the present article can be reproduced using the function `publ_lebagousse2025_figs`.

# 3 Results

The two experimental conditions yielded very different performance levels. In the first condition, where the SNR was similar to that of AMS75, the average sensitivity was $d' = 0.485$ (average correct response rate = 59%), much lower than the sensitivity reported in the original study (average $d' = 1.947$). Therefore, we introduced a second condition with a more favorable SNR, which led to a substantial increase in sensitivity (average $d' = 1.99$, correct response rate = 83%) closely matching the results of the original study.

The resulting correlation matrices for the 4 participants engaged in the SNR-matched condition and the 9 participants engaged in the performance-matched condition are shown in Figure 2 and 3, respectively. These matrices reflect the influence of the noise in each spectrotemporal region on the final decision ("tone present" vs. "tone absent"). Therefore, they provide insights into the auditory mechanisms involved in tone-in-noise detection.

A consistent feature across all derived matrices is the positive correlation coefficient observed in the central region corresponding to the spectrotemporal location of the target tone. This positive weight indicates that the presence of energy in the same region of the tone increases the likelihood that the listener perceives a tone. Complementary analyses conducted separately for tone-present or tone-absent trials confirm that noise influence perception in both scenarios: in tone-present trials, it enhances tone detectability, resulting in more Hit responses, while in tone-absent trials, it causes listener to mishear a tone, resulting in a larger number of false

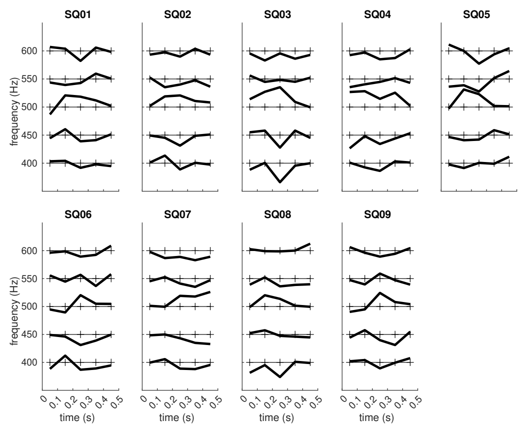



Figure 3: Results obtained for the 9 participants in the 18-dB condition (performance matched with AMS75).

alarms. Although this result was expected, it also provides additional confirmation that the participants are actively engaged in the task.

For most participants, the positive correlation was not limited to the central region but extended to the temporal segments preceding and/or following the target tone. This reflects temporal uncertainty regarding the precise location of the target tone: listeners may confuse the noise energy at 500 Hz in the second or fourth segments as the presence of a target tone in the third segment. The extent of temporal uncertainty, however, appears to vary between participants, and it was overall larger in target-absent trials.

Finally, the region immediately below the target frequency (i.e., around 450 Hz) exhibited a consistent negative correlation with the response. This suggests that high levels of noise energy in this range may mask information in the target region, and/or that the absence of energy in this range makes it more likely for the listener to perceive it as a tone. For most participants, the region immediately above the target frequency (550 Hz), but also the regions at 600 Hz and 400 Hz also displayed negative correlations, although weaker. The trial-specific analysis revealed that this influence of adjacent frequencies was in fact limited to target-present trials.

# 4 Discussion

All participants successfully completed both conditions with a rate of correct detection higher than chance (50%), although several found the task very difficult. Correlation matrices were derived individually for each set of data collected (Figures 2 and 3).

## 4.1 Comparison with AMS75

A major difference between the AMS75 data and our replication is the overall performance level in the task. In the original study, listeners achieved an average sensitivity level of $d' = 1.947$, substantially higher than the $d' = 0.485$ observed in our study under comparable experimental conditions. We do not believe that this discrepancy can be attributed to the minor deviations from the original experimental protocol. It is more likely that some experimental details were omitted in the AMS75 paper, such as the inclusion of an extensive training phase. For instance, previous work by Ahumada involved training sessions of up to 2,000 trials, which could have significantly enhanced participant performance.

The pattern of correlations in conditions 11.8 dB and 18 dB generally confirms AMS75's observation that the presence of noise energy on the target interval, but also in the following intervals, biases the listener towards perceiving a tone. Furthermore, the listeners assigned negative weights to energy surrounding the target frequency, consistent with AMS75 findings.

Although the overall pattern of correlation weights was consistent between the original study and the replication, some differences were also observed. First, temporal uncertainty appeared to be larger in both the 11.8-dB and 18-dB condition, with a strong primacy effect (weights in the segments 0.1 s and 0.2 s can even be larger than the target segment in some cases, in particular in target-absent trials). This contrasts with AMS75, which reported a weak negative influence of noise energy in segments preceding the target. Similar primacy effects, however, have been observed in related psychoacoustic tasks, such as loudness judgments [16, 9, 10] and amplitude modulation detection [18, 6]. This anticipation effect was attributed to different mechanisms, including a cross-correlation decision process with a limited time lag [18], or an evidence integration process [9]. However, it is important to note that these two models accounted for part, but not all, of the observed phenomenon.

The findings of [6] on a similar task (detecting an intensity increment in noise) may also help explain the apparent increase in the integration window. After collecting an average of 5,500 trials per subject, the researchers compared the kernels from the beginning and end of the experiment, revealing a change in the modulation tuning over the course of the experiment, from a low-pass filter at the beginning of the experiment to a band-pass filter at the end of the experiment. It is possible that the 3,200 trials in our replication capture an early stage of this adaptation, whereas the data from Ahumada et al., collected after extensive training, reflect a more advanced phase. If this interpretation holds, the broader modulation filter observed in our study could, in theory, account for the larger integration window.

Interestingly, the correlation matrices obtained in the 11.8-dB condition appear to be “closer to optimality” compared to AMS75, as they place less weight on the non-target frequency bands. Only in the 18-dB condition did participants exhibit off-frequency weights comparable to that observed in the original experiment. These findings suggest that the 18-dB condition is closer perceptually to AMS75, both in terms of sensitivity and listening strategy.

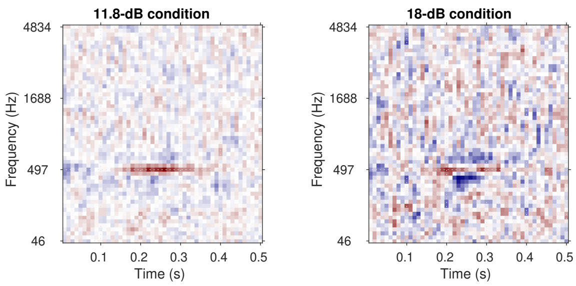



Figure 4: Mean ACIs in the 11.8-dB and 18 dB condition. The weights are normalized to the maximum absolute value.

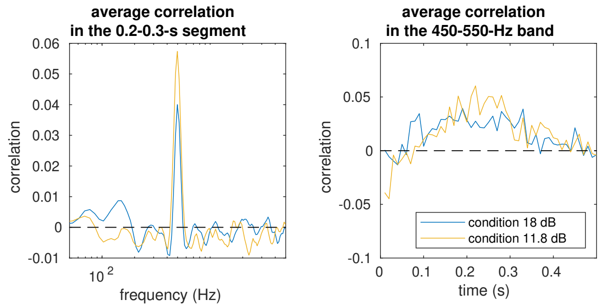



Figure 5: Marginalized target-absent ACIs along the temporal and spectral dimensions. The spectral correlation profile (left) corresponds to the 0.2-0.3 s segment only, while the temporal correlation profile (right) corresponds to the 450-550 Hz frequency band.

## 4.2 Temporal and spectral resolutions

In this section, we present a complementary analysis to further determine the extent of the temporal and spectral windows used for the detection of the target tone. High-resolution correlation matrices, commonly referred to as Auditory Classification Images (ACIs), were computed using the same dataset described previously. The processing pipeline remained identical as described in the Methods section, except that the coarse 5-by-5 spectrogram was replaced by a high-resolution 110-by-50 gammatone-based spectrogram. This allows for a more detailed representation of the temporal and spectral features relevant to tone detection. The resulting ACIs, averaged across participants for each condition, are shown in Figure 4.

Subsequently, we restricted the analysis to target-absent trials (i.e. noise-only trials), for two reasons. First, target-absent ACIs are known to provide a more accurate estimate of the underlying template, because they are less affected by non-linearities in auditory processing. Second, these trials are identical in both conditions, therefore allowing for direct comparison.

Target-absent ACIs were derived for all participants. Figure 5 presents the resulting correlations, averaged within the 0.2-0.3 s time window (spectral correlation profile) or within the 450–550 Hz frequency band (temporal correlation profile).

As expected, the temporal and spectral correlation profiles obtained for the two conditions are very similar overall. The main differences include an increased sensitivity to the spectro-temporal location of the target for the 11.8 dB condition, and a stronger negative effect of the presence of energy at the onset of the stimulus.

The spectral profile in the 0.2-0.3 s segment provides an indirect estimate of the human auditory system's frequency selectivity in the task. The marginalized ACI exhibits positive correlations around the target frequency from 430 Hz to 578 Hz. This is consistent with the width of critical bands in this region (Equivalent Rectangular Bandwidth at 500 Hz ~ 75 Hz).

The temporal profile in the 450-550 Hz band reveals that noise energy present as early as 200 ms before the tone onset influences participants' responses. As mentioned in the previous section such a large temporal window is at odds with the findings of AMS75, suggesting that the participants in the present study may have used a different decision process. Although it is not clear what mechanism accounts best for the sloppiness of the auditory system in the modulation domain, it can be argued that the marginal ACI provides a direct insight into the tone-detection process.

## 4.3 Comparison between the 11.8-dB and 18-dB conditions

Although very similar, the measured profiles for the 11.8-dB and 18-dB conditions exhibit several differences: an enhanced sensitivity to the spectro-temporal location of the target for the 11.8-dB condition and an inhibitory effect at the beginning of the stimulus (i.e. noise energy in this region decreases the probabilty that the participant makes a false alarm). It is important to note that these differences stem solely from changes in the template used by the subjects between the two conditions, as the acoustic characteristics of the trials remain identical (non-target stimuli consisting of noise alone). This indicates that the same sounds are not processed in the same way between the two conditions, i.e., that listeners included the difference in SNR into their internal template of the target.

These results challenge the standard modeling of detection tasks in noise as a comparison with a noise-free template (constructed as a noise-free representations of the target [18], or as the difference between noisy representations of the target and non-target signals [4]). Such approaches would not account for the observed difference, as both conditions would correspond to the same templates. Only models using a template that includes not only the target but also the noise could accound for the result.

These observations therefore provide valuable insights into the decision-making device, which has been relatively understudied. A similar analysis based on non-target trials conducted in a previous study also allowed us to conclude that the template was not updated on a trial-by-trial basis when the SNR varied throughout the experiment [18]. Taken together, these results imply that, in detection-in-noise tasks, the auditory system relies on a noisy template rather than an ideal, noise-free template. However, it appears unable to update this representation in real-time based on variations in noise level.

# 5 Conclusion

In summary, our study successfully replicated the findings of AMS75, confirming that listeners engaged in a tone-in-noise detection task are affected not only by the presence of noise at the spectrotemporal location of the target but also in adjacent time-frequency regions. More broadly, this study showcases the practical utility of the fastACI toolbox. Providing an efficient and user-friendly framework for designing, conducting, and analyzing revcorr experiments, fastACI simplifies the process of generating new studies and replicating previous findings. It also helps setting up fully computationally reproducible workflows, as demonstrated by the `publ_lebagousse2025_figs` script for the present study. Our results also provide new insights into the decision stage of the auditory system, suggesting that the internal templates used in detection-in-noise tasks incorporate both the target and the noise, rather than a clean version of the target. As interest in reverse correlation approaches continues to grow across the psychoacoustics and psycholinguistics communities, we hope that fastACI will be a valuable resource for researchers seeking to explore various aspects of auditory processing.

## Data availability

All data from this study is available within a Zenodo repository: https://zenodo.org/records/14972392. The experiment can be reproduced and all analyses can be replicated using the fastACI MATLAB Toolbox, available at https://zenodo.org/records/5500138.


## Acknowledgements

This study was supported by the French National Research agency through the ANR grants "fastACI" (Grant No. ANR-20-CE28-0004), and "FrontCog" (Grant No. ANR-17-EURE-0017).


## References


[1] Albert J Ahumada Jr and John Lovell. "Stimulus Features in Signal Detection". In: *The Journal of the Acoustical Society of America* 49.6B (1971), pp. 1751–1756.

[2] Albert J Ahumada Jr, R Marken, and A Sandusky. "Time and frequency analyses of auditory signal detection". eng. In: *The Journal of the Acoustical Society of America* 57.2 (Feb. 1975), pp. 385–390. ISSN: 0001-4966.

[3] Géraldine Carranante et al. "Mapping the spectrotemporal regions influencing perception of French stop consonants in noise". en. In: *Scientific Reports* 14.1 (Nov. 2024). Publisher: Nature Publishing Group, p. 27183. ISSN: 2045-2322. DOI: 10.1038/s41598-024-77634-w.

[4] Torsten Dau, Dirk Püschel, and Armin Kohlrausch. "A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements". In: *The Journal of the Acoustical Society of America* 99.6 (June 1996), pp. 3623–3631. ISSN: 0001-4966.

[5] David Marvin Green and John A. Swets. *Signal Detection Theory and Psychophysics*. en. Peninsula Publishing, 1966. ISBN: 978-0-932146-23-6.

[6] Eva R. M. Joosten et al. "Dynamic Reweighting of Auditory Modulation Filters". In: *PLoS Computational Biology* 12.7 (July 2016). ISSN: 1553-734X.

[7] Richard F. Murray. "Classification images: A review". en. In: *Journal of Vision* 11.5 (Jan. 2011). ISSN: 1534-7362.

[8] Peter Neri. "Classification images as descriptive statistics". In: *Journal of Mathematical Psychology* 82 (Feb. 2018), pp. 26–37. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2017.10.004.

[9] Daniel Oberfeld, Alexander Fischenich, and Emmanuel Ponsot. "Non-uniform temporal weighting of intensity in audition and vision: The signature of an evidence integration process?" en. In: *Acta Acustica* 8 (2024). Publisher: EDP Sciences, p. 57. ISSN: 2681-4617. DOI: 10.1051/aacus/2024061.

[10] Daniel Oberfeld and Tina Plank. "The temporal weighting of loudness: effects of the level profile". eng. In: *Attention, Perception & Psychophysics* 73.1 (Jan. 2011), pp. 189–208. ISSN: 1943-393X. DOI: 10.3758/s13414-010-0011-8.

[11] Alejandro Osses and Léo Varnet. "A microscopic investigation of the effect of random envelope fluctuations on phoneme-in-noise perception". In: *Journal of the Acoustical Society of America* 155.2 (Feb. 2024). Publisher: Acoustical Society of America, pp. 1469–1485. DOI: 10.1101/2022.12.27.522040.

[12] Alejandro Osses and Léo Varnet. *fastACI toolbox: the MATLAB toolbox for investigating auditory perception using reverse correlation (v1.0)*. Oct. 2021. DOI: 10.5281/zenodo.5500139.

[13] Alejandro Osses et al. "Prosodic cues to word boundaries in a segmentation task assessed using reverse correlation". In: *JASA Express Letters* 3.9 (Sept. 2023), p. 095205. ISSN: 2691-1191. DOI: 10.1121/10.0021022.

[14] Emmanuel Ponsot et al. "Cracking the social code of speech prosody using reverse correlation". en. In: *Proceedings of the National Academy of Sciences* 115.15 (Apr. 2018). Publisher: National Academy of Sciences Section: Biological Sciences, pp. 3972–3977. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1716090115.

[15] Emmanuel Ponsot et al. "Mechanisms of Spectrotemporal Modulation Detection for Normal- and Hearing-Impaired Listeners". eng. In: *Trends in Hearing* 25 (2021), p. 2331216520978029. ISSN: 2331-2165.

[16] Emmanuel Ponsot et al. "Temporal loudness weights for sounds with increasing and decreasing intensity profiles". In: *Journal of the Acoustical Society of America* 134.4 (Oct. 2013). Publisher: Acoustical Society of America, EL321–6. DOI: 10.1121/1.4819184.

[17] Chalmers W. Sherwin et al. "Detection of Signals in Noise: A Comparison between the Human Detector and an Electronic Detector". In: *The Journal of the Acoustical Society of America* 28.4 (July 1956), pp. 617–622. ISSN: 0001-4966. DOI: 10.1121/1.1908424.

[18] Léo Varnet and Christian Lorenzi. "Probing temporal modulation detection in white noise using intrinsic envelope fluctuations: A reverse-correlation study". In: *The Journal of the Acoustical Society of America* 151.2 (Feb. 2022). Publisher: Acoustical Society of America, pp. 1353–1366. ISSN: 0001-4966. DOI: 10.1121/10.0009629.

[19] Léo Varnet et al. "A Psychophysical Imaging Method Evidencing Auditory Cue Extraction during Speech Perception: A Group Analysis of Auditory Classification Images". In: *PLoS ONE* 10.3 (Mar. 2015), e0118009.

[20] Léo Varnet et al. "Using auditory classification images for the identification of fine acoustic cues used in speech perception." In: *Frontiers in Human Neuroscience* 7: (2013), p. 865.