Cognition 268 (2026) 106350

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full Length Article

Memory for repeated auditory textures

Berfin Bastug [a,b,*], Vani G. Rajendran [a,c], Roberta Bianco [d,e,g], Trevor Agus [f], Maria Chait [e], Daniel Pressnitzer [a,*]

[a] Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France
[b] Ernst Strüngmann Institute for Neuroscience in Cooperation with Max Planck Society, Frankfurt am Main 60528, Germany
[c] Departamento de Neurociencia Cognitiva, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico
[d] Neuroscience of Perception & Action Lab, Italian Institute of Technology, Viale Regina Elena 291, 00161 Rome, Italy
[e] UCL Ear Institute, 332 Gray's Inn Rd, London WC1X 8EE, UK
[f] School of Arts, English and Languages, Queen's University Belfast, University Road, Belfast BT7 1NN, UK
[g] Department of Translational Research & New Technologies in Medicine and Surgery, University of Pisa, via San Zeno 31/33, 56123 Pisa, Italy

ARTICLE INFO

Keywords:
Auditory perception
Natural textures
Memory
Summary statistics
Sound representation

ABSTRACT

Even though memory plays a pervasive role in perception, the nature of the memory traces left by past sounds is still largely mysterious. Here, we probed the memory for natural auditory textures. For such stochastic sounds, two types of representations have been put forward: a representation based on sets of temporally local features, or a representation based on time-averaged summary statistics. We synthesized naturalistic texture exemplars and used them in an implicit memory paradigm based on repetition, previously shown to induce rapid learning for artificial sounds such as white noise. Results were similar for artificial and natural sounds, exhibiting a general trend for a decrease in repetition detection performance with increasing exemplar duration, although with some variation depending on texture type. This trend could be captured by a summary statistics model, but also by a new model based on the random sampling of temporally local features. Moreover, repeated exposure to a same natural texture or artificial noise exemplar systematically induced a performance gain, which was comparable across all sound types and exemplar durations. Thus, natural texture exemplars were amenable to learning when repeated exposure was available. The findings are consistent with two interpretations: the existence of a special processing mode when acoustic repetition is involved, to which natural textures are not immune, or a convergence of the local features versus summary statistics descriptions if a continuum of time scales is considered for auditory representations.

1. Introduction

1.1. Context and motivation

Auditory perception must combine the acoustic information reaching the ears at every moment in time with information from the past, stored in memory. This is obviously the case when rapidly recognizing sounds that have acquired meaning through exposure, such as for instance one's own ringtone (Roye, Schröger, Jacobsen, & Gruber, 2010). More generally, a pervasive role of memory in perception is at the core of theories based on Bayesian inference or predictive coding, as both approaches assume that a model of the world has been somehow internalized through experience (Heilbron & Chait, 2018; Kok, Mostert, & de Lange, 2017; Press, Kok, & Yon, 2020). The nature of the memory traces left by past sounds, however, is still largely mysterious. Here, we probe the memory for natural auditory textures. For such stochastic sounds, two types of representations can be hypothesized: a representation based on temporally local features (Agus, Thorpe, & Pressnitzer, 2010) and a representation based on time-averaged summary statistics (McDermott, Schemitsch, & Simoncelli, 2013).

The temporally local features hypothesis stems from a line of research that characterized the perception of repeated sounds. When hearing a repeated exemplar of white noise, listeners report the emergence of individual events, often described as "rasping" or "clanks" (Guttman & Julesz, 1963; Warren, Bashford, Cooley, & Brubaker, 2001). Subsequent experiments have confirmed that the features used to detect repetition in white noise generally seem to have a local time-frequency extent (Kaernbach, 1993; Ringer, Schröger, & Grimm, 2023). Recently,

* Corresponding authors.
E-mail addresses: berfin.bastug@esi-frankfurt.de (B. Bastug), daniel.pressnitzer@ens.psl.eu (D. Pressnitzer).

https://doi.org/10.1016/j.cognition.2025.106350
Received 8 April 2024; Received in revised form 14 October 2025; Accepted 17 October 2025
Available online 14 November 2025

the findings have been extended to longer-term memory traces. When listeners were exposed to the same exemplar of noise which reoccurred several times during an experimental block, behavioral evidence of a "memory for noise" lasting up to several weeks was observed (Agus et al., 2010; Viswanathan, Rémy, Bacon-Macé, & Thorpe, 2016). As the duration of the learnt noise exemplars extended to the multi-second range, it seemed unreasonable that listeners memorized the thousands of samples defining one particular noise exemplar. Rather, as was the case for the immediate repetition of noise, it was hypothesized that listeners stored a limited set of temporally local features, which could be used as a compact "watermark" for a given noise exemplar (Agus et al., 2010). Neural correlates of the phenomenon were consistent with the local features hypothesis, with the added proposal that feature sets could be idiosyncratic and thus unique to each listener/noise combination (Andrillon, Kouider, Agus, & Pressnitzer, 2015; Luo, Tian, Song, Zhou, & Poeppel, 2013; Ringer et al., 2023). Finally, similar findings were obtained with stochastic artificial sounds other than white noise, such as random melodies (Bianco et al., 2020; Bianco, Hall, Pearce, & Chait, 2023), random rhythms (Kang, Agus, & Pressnitzer, 2017), or tone clouds with a broad range of spectro-temporal complexities (Agus & Pressnitzer, 2021). Rapid plasticity was even evidenced with repeated exposure to noise exemplars during sleep (Andrillon, Pressnitzer, Léger, & Kouider, 2017). Overall, repetition seems to automatically trigger the rapid formation of memory traces for many kinds of sounds.

The summary statistics hypothesis stems from work on natural auditory textures (McDermott et al., 2013; McDermott & Simoncelli, 2011). Textures can be defined as sounds with stochastic but stationary characteristics. A *natural* texture is the sound emanating from an underlying stationary generative process in the environment, such as the sound of fire crackling, water flowing, or wind blowing. The first important finding of this line of research was that synthetic sounds matched to natural sounds in a few long-term statistics were readily identified as natural textures by listeners (Geffen, Gervain, Werker, & Magnasco, 2011; McDermott & Simoncelli, 2011). This showed that summary statistics were sufficient to recognize texture categories. Perhaps even more intriguingly, when asked to discriminate between two exemplars of the same texture (e.g., two instances of fire crackling), listeners' performances *decreased* as the exemplar durations increased. This seems counter-intuitive, as for many other tasks, longer durations usually result in improved discrimination performance (Teng, Tian, & Poeppel, 2016). However, such a finding could be understood if the discrimination was based on time-averaged summary statistics, and not on temporally local features that could be accrued as duration increased. To quote McDermott et al. (2013, abstract): *"These results indicate that once these sounds are of moderate length, the brain's representation is limited to time-averaged statistics, which, for different examples of the same texture, converge to the same values with increasing duration"*. Summary statistics for textures could be the auditory equivalent of "ensemble coding" for visual perception, which is an efficient way to capture the gist of natural images (Whitney & Leib, 2016).

Taken to an extreme, a consequence of the summary statistics hypothesis would be that different exemplars of the same natural texture cannot be memorized once they reach a moderate length, simply because they cannot be discriminated any more. Therefore, unlike for artificial sounds, repeated exposure to a natural texture exemplar may not induce a memory trace specific to that exemplar. Such a radical hypothesis is overly simplistic, however. In the original McDermott et al. (2013) study, texture exemplar discrimination did not fall to chance even at the longest duration tested. The authors thus left open the possibility that summary statistics may coexist with other types of representations. In their comment to the original study, Nelken and de Cheveigné (2013) strikingly summarized such a position by referring to "a skeleton of events on a bed of textures". Their point was that not every sound should be treated as a texture and thus summarized by statistics. For events, such as a bird call, local features would be preserved, leading to a dual representation of a sound scene. However, it is yet unknown if and how "events" may arise from natural textures themselves.

Another important point made by the texture literature is that the use of natural sounds could be critical to recruit ecologically relevant auditory processes. Such processes and their attending representations would not be called into action for artificial sounds (Theunissen & Elie, 2014). For instance, the auditory system may enter a "texture mode" when it recognizes a natural texture, and actively discard any temporally local features in favor of a more compact summary statistics representation (Nelken & de Cheveigné, 2013). Indeed, it makes much more sense to be able to recognize the physical cause of a texture (McDermott et al., 2013; McDermott & Simoncelli, 2011; Nelken & de Cheveigné, 2013) or even some of its characteristics, such as temperature for flowing water (Velasco, Jones, King, & Spence, 2013), than to recall the acoustic details of a given texture exemplar. Auditory cognition may thus be tuned to efficient representations of natural sounds and their statistical properties (Gervain & Geffen, 2019) in order to facilitate the categorization of the physical events making up our environment (Traer, Norman-Haignere, & McDermott, 2021).

The sounds that have been used in the memory for noise paradigm, such as white noise or tone clouds, can be seen as artificial textures. However, natural textures have not been used yet in such a paradigm. Here, we synthesized naturalistic sound textures using the original McDermott and Simoncelli (2011) and used them in the repetition-based "memory for noise" paradigm of Agus et al. (2010), alongside one condition using white noise for comparison.

1.2. Previous results and specific predictions

The memory for noise paradigm introduced by Agus et al. (2010) required participants to discriminate between trials containing fully random noise and trials made of abutting repetitions of the same noise exemplar. This repetition-detection task is possible for a range of exemplar durations, from tens of milliseconds to several seconds (Guttman & Julesz, 1963; Kaernbach, 2004; Warren et al., 2001). When a single repetition is presented, performance decreases as exemplar duration increases, reaching chance for exemplar durations of about 6 s (Kaernbach, 2004, their Fig. 2). Agus et al. (2010) introduced an additional experimental condition: some noise exemplars re-occurred over many trials in their experiments. Perhaps surprisingly, a constant "memory gain" was observed for re-occurring exemplars, irrespective of duration for exemplars up to 2 s long (the longest duration tested in Agus et al., 2010, their Fig. 4B). This was interpreted as signaling a memory trace based on a few, temporally local features at all durations.

From the now abundant literature about natural textures, the most relevant results for the present study are that of Experiment 2 of McDermott et al., (2013, their Fig. 2b). They introduced a texture exemplar discrimination task, where participants heard three different sounds in each trial. All three sounds were from the same natural texture category (e.g. Fire). Two of them were acoustically identical, whereas the third one was a different exemplar from the same texture category. Participants had to indicate the odd one out. Exemplar durations ranged from 40 ms to 2500 ms. Performance decreased with duration, from about 90 % correct at 40 ms down to about 75 % correct at 2500 ms (the longest duration they tested). This was interpreted by the convergence of summary statistics towards their expected value for the texture category as exemplar duration increased.

From both sets of results, a decrease in performance with duration is therefore expected for a repetition-detection task, for natural textures and noise. For repetition-detection with re-occurring exemplars, a constant memory gain is expected for artificial noise. For natural textures, there are two possible predictions. Either the representation of repeated natural textures includes temporally local features, and then they should display constant memory gain for all durations just like noise. Or, because of a specific "texture mode" that actively discards temporally local features, their representation is exclusively based on summary statistics. As longer-term memory for summary statistics has not yet been investigated, the memory gain in this case is essentially unknown.

2. Methods

2.1. General procedure

All experiments were performed online, as data collection took place during the pandemic. The method was otherwise identical to the original "memory for noise" study (Agus et al., 2010). Briefly, in such a paradigm, each trial consists of a single sound: either noise (N), or repeated noise (RN), that is, noise for which the first half is identical to the second half. The repetition is seamless, with no acoustic cue nor silent interruption between halves. The participant's task is to report whether the trial contained a repetition or not. For some trials, the RN is randomly drawn anew, so participants only hear each RN stimulus once. Such a condition taps into short-term memory processes and provides a baseline repetition-detection performance, which may depend on various stimulus parameters such as duration (Kaernbach, 2004; Warren et al., 2001). However, without informing the participants, a third condition is introduced: one RN exemplar, called the reference RN (RefRN), re-occurs over different trials throughout an experimental block. An improved performance for RefRN trials compared to RN trials is interpreted as learning of the RefRN exemplar.

The stochastic stimuli used here were white noise, replicating previous studies, but we also introduced natural textures. Three texture categories were chosen: fire crackling (Fire), water running down a stream (Stream), and wind blowing (Wind). Texture trials (Tx) were all different and generated as in McDermott and Simoncelli (2011). Repeated textures (RTx) and reference repeated texture (RefRTx) trials were obtained by cross-fading two copies of a same texture exemplar. Illustrations of the stimuli are shown in Fig. 1. As can be seen, even though the choice of texture was largely arbitrary, they all differed in their spectro-temporal characteristics, which in turn differed from white noise.

Different trial durations were tested on different experimental blocks: 250 ms, 500 ms, 1000 ms, 2000 ms, and 4000 ms. This corresponded to exemplar durations of 125 ms, 250 ms, 500 ms, 1000 ms, and 2000 ms, respectively, matching exactly Agus et al. (2010) for noise and approximately McDermott et al. (2013) for natural textures.

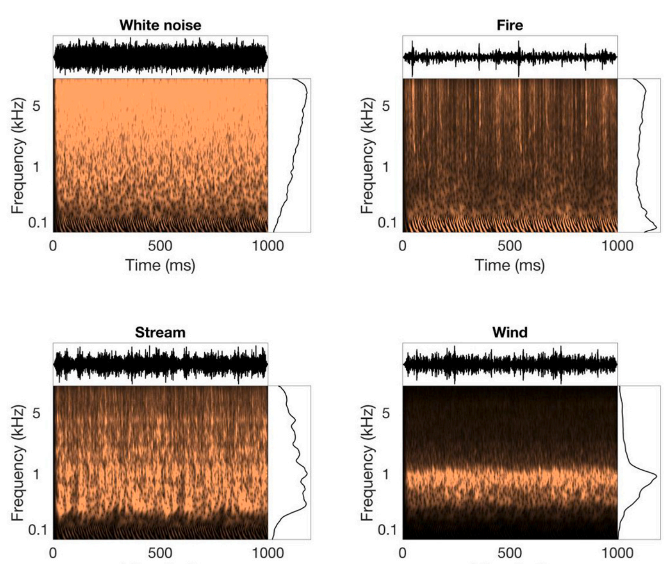

Fig. 1. Acoustic characteristics of white noise and natural textures. Illustrative examples of cochleagrams for the four different sound categories. The time-frequency cochleagrams use brighter colors to represent energy within simulated auditory filters (see Methods). In all cases, repeated trials are shown, so the first 500-ms are identical to the last 500-ms. Note that there is no acoustic discontinuity at the repetition onset. The temporal waveforms (top inset of each panel) and spectral average (right inset of each panel) are also provided.

2.2. Participants

72 individuals (13 female, 59 male), aged between 18 and 38 ($M = 29.7$ $SD = 1.66$), with self-reported normal hearing participated in the online experiment in return for monetary compensation. This corresponded to 18 participants per texture, similar to the number of participants in previous comparable in-lab studies. The sample size was preregistered (see below). Participants were recruited through Prolific (Oxford, UK). Before the experiment, all participants provided informed consent. At the end of the experiment, an online debriefing text was presented. The UCL Research Ethics Committee approved the protocol (#1490/009).

2.3. Stimuli

Because generating naturalistic textures can be computationally intensive, they were synthesized offline and stored as sound files, which were loaded to the participant's browser during the experiment. The synthesis algorithm was the one from McDermott and Simoncelli (2011) as available online (http://mcdermottlab.mit.edu/downloads.html). In total, 9 sound files with a different random seed, each 392 s long, were synthesized for each of the three categories of natural sound textures (Fire, Stream, Wind). For symmetry, 9 sound files, also 392 s long, were generated for white noise. Each of the 36 unique sound files (9 random seeds x 4 sound types) was used twice, but always for different participants. To generate a trial, short exemplars of the desired duration were cut sequentially (no overlap between exemplars) from the 392 s-long sounds. For repeated trials (RN/RTx and RefRN/RefRTx), the same exemplar was collated twice, with a 10-ms crossfade. For non-repeated trials (N/Tx), two different exemplars were collated, with the same crossfade technique. All sounds were presented as uncompressed *.wav* files.

2.4. Procedure

The experiment was conducted using the Gorilla platform (Cambridge, UK). Before starting the experiment, several checks were run to ensure data quality, including browser checks and headphone checks (Milne et al., 2021). Individuals who failed any of these checks were rejected from participating, so all participants are assumed to have been wearing headphones. Participants were then presented with an information sheet and gave their informed consent.

Each participant completed five experimental blocks, each of which corresponded to a different exemplar duration, all for the same sound type. Each participant was thus only tested on one sound type (e.g. White noise or Fire). Each block was preceded by a brief familiarization phase with feedback. The participant first heard a sound with 10 repetitions of a given exemplar, to illustrate the cues to repetition at the duration of the block. This familiarization sound could be played up to three times. Then, four training trials were provided. In the training trials, the stimulus either consisted of an exemplar repeated 10 times or of 10 different exemplars. Participants were instructed to report whether they heard a repetition. Immediate feedback was given. Further training trials followed, with gradually increasing difficulty. Those training trials contained 4 repetitions (10 trials), 3 repetitions (12 trials), and finally 2 repetitions as in the main experiment (20 trials). Training trials were always 50 % RN/RTx and 50 % N/Tx (no RefRN/RefRTx). The training session was immediately followed by the experimental block at the same duration. In the experimental block, participants did not receive immediate feedback but did see their cumulative accuracy (percent correct) at the end of each block. Each block consisted of 40 N/Tx trials, 20 RN/RTx trials, and 20 RefRN/RefRTx trials, with those conditions presented in a pseudorandom order (Ref stimuli were never presented on successive trials).

Participants were incentivized through bonus payments that would be calculated based on their overall accuracy at the end of the experiment (Bianco, Mills, de Kerangal, Rosen, & Chait, 2021). Finally, participants were also informed that their data would be rejected if they scored less than 60 % accuracy on the task to discourage participants from guessing at random.

2.5. Statistical analyses

We used the *d'* sensitivity index of signal detection theory to estimate performance. Hits were defined as "repeated" responses for RN/RTx and RefRN/RefRTx trials. False alarms were defined as "repeated" responses for N/Tx trials. When the proportion of hits or false alarms reached 0 or 1 for a given participant and measurement, a correction corresponding to plus or minus half-a-trial was applied (Macmillan & Creelman, 2004).

Analyses of variance (ANOVAs) and *t*-tests were used as inferential tests, with an α-level of 0.05. Reporting convention follows the APA guidelines, 7th edition, so all *p*-values less than 0.001 are reported as $p < 0.001$. The main analysis was a mixed-design ANOVA, fully reported in Table A1. Because the false alarm rate was shared between RN/RTx and RefRN/RefRTx condition, which could have introduced correlations across measurements, the ANOVAs and the paired *t*-tests contrasting conditions were performed on the *z*-transformed hit rates used to compute *d'* (Agus & Pressnitzer, 2021). We further checked that performing the same analyses on *d'*, corresponding to *z*-transformed hit rates minus *z*-transformed false alarms, led to strictly identical conclusions. A Greenhouse-Geisser correction was applied when Mauchly's test indicated a violation of the sphericity assumption ($p < 0.05$). Further partial ANOVAs and t-tests were run to help interpret the main analysis, as it included several factors and revealed second- and third-order interactions. The partial ANOVAs used repeated-measures or mixed-design ANOVAs as appropriate. When performing partial analyses where all natural textures were considered together, and because different participants ran different texture blocks, a "participant" was defined as the average of individual results in the three natural texture blocks, in order of enrolment. All ANOVAs were run using JASP (JASPTeam, 2024). Effect sizes are reported as generalized η^2, notated η^2_G, as recommended for mixed designs (Lakens, 2013, p. 6).

2.6. Preregistration and data availability

The study was preregistered (ResearchBox #2762). There were minor deviations to the preregistration: the noise sound category was added for comparison; the hit-rate over time analyses were omitted due to the large number of conditions; the ANOVAs were run on *z*(hits) and not *d'* as justified above; partial ANOVAs were added to interpret the outcome of the full analysis. The main characteristics of the design (sample size, exclusion criteria, duration conditions, test procedure, main analyses) exactly followed the preregistration. The full dataset is available online (ResearchBox #2762).

2.7. Summary statistics model

Summary statistics were computed for the exact stimulus set that was presented to participants. The model of McDermott et al. (2013) was used, downloaded from the author's website and using the default parameters corresponding to the published study. To avoid repetition artifacts, only the first half of our stimuli was fed to the model. The variance of statistics was estimated across exemplars, as in McDermott et al. (2013). In our case, for each texture type and duration, we had 540 available exemplars (40 N trials, 20 RN trials, from 9 different batches). We arbitrarily grouped the exemplars in groups of 10, computed the statistics' variance for each group of 10, and derived the median and interquartile range of the statistics' variance across the 54 groups.

2.8. Local feature sampling

A novel feature sampling simulation was also introduced. Its implementation is described alongside the simulation description in Section 6. The simulation code is available online (ResearchBox #2762).

2.9. Acoustic analyses

To obtain the illustrations of the stimuli provided in Fig. 1., stimuli were passed into an auditory model as described in Agus, Suied, Thorpe, and Pressnitzer (2012). The model consisted of a broadband pre-emphasis bandpass filter (0.4–8.5 kHz), a gammatone auditory filterbank, half-wave rectification, square root compression, and lowpass filtering at 100 Hz. Such a time-frequency representation, termed a "cochleagram", is intended to roughly mimic the information available after peripheral auditory processing.

3. Results

3.1. Validation of the online testing procedure

So far, all studies using variants of the memory for noise paradigm but two (Dauer, Henry, & Herrmann, 2022; Ringer, Schröger, & Grimm, 2022) were performed under highly controlled laboratory conditions. It was thus unclear whether the findings, presumably dependent on subtle acoustic cues, would be robust enough to translate to online testing.

Fig. 2A shows the average results for the 0.5-s duration, for white noise, as this is the duration condition that was most extensively tested in previous investigations. Performance is expressed as the sensitivity index *d'* of signal detection theory. Baseline performance for the within-trial repetition detection task RN, for which the repeated noise exemplar was novel on each trial, was modest but still above chance ($M = 0.67$; *t*-test against the chance value of $d' = 0$: $t(17) = 5.54$, $p < 0.001$). Importantly, in the RefRN condition for which the same noise exemplar re-occurred throughout a block, performance improved ($M = 2.06$; paired *t*-test against RN performance $t(17) = 5.40$, $p < 0.001$). This pattern of results replicates in-lab findings using white noise (Agus et al., 2010; Agus & Pressnitzer, 2013), validating the online procedure.

Fig. 2B shows the first set of results for natural textures, again for the 0.5-s duration, with performance averaged for all three texture categories. For these relatively short-duration natural texture exemplars, performance was generally higher than for noise. In the baseline RTx condition, performance was well above chance ($M = 2.07$; *t*-test against the chance value of 0: $t(17) = 13.67$, $p < 0.001$). In the RefRTx condition, performance further improved ($M = 2.74$; paired *t*-test against RTx

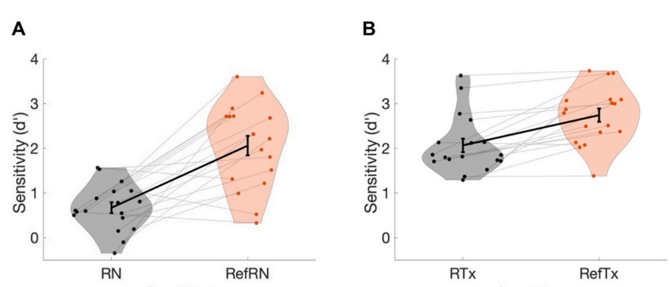

Fig. 2. Performance for 500-ms long exemplars. A). White noise. Repetition detection performance is shown, expressed as the *d'* sensitivity index of signal detection theory. For the Repeated Noise condition (RN), the noise exemplar was novel in each trial. For the Reference Repeated Noise condition (RefRN), the same noise exemplar re-occurred on 20 trials randomly interspersed in the experiment. Dots represent individual participants, connected by thin lines across conditions. Mean performance and standard error about the mean are shown as thick lines. B). Natural textures. Performance averaged for the three natural textures. Same as A).

performance $t(17) = 5.64$, $p < 0.001$). This shows that within-trial repetition detection and across-trial learning are possible with relatively short texture exemplars.

3.2. Repetition detection and rapid learning for short and long texture durations

Behavioral performance was measured for repeated (RN/RTx) stimuli and for repeated and re-occurring stimuli (RefRN/RefTx). Repetition detection was expected to decrease with duration and a memory gain was hypothesized for re-occurring stimuli, at least for noise. Fig. 3 shows the results for noise and textures as a function of exemplar duration (trial duration itself was twice as long). For noise, there was a steady decrease in performance in the RN condition from short to long durations, but, importantly, the memory gain observed for RefRN was approximately constant across the whole range of durations. For textures, performance in the RTx condition was always good (average *d'* above 1), but, contrary to the expectation, had a band-pass shape with a peak at 250 ms. Importantly again, a memory gain for RefTx was observed throughout the whole range of durations, a novel finding.

These observations were formally tested by two separate repeated-measures ANOVAs, one for noise and one for textures, with factors "Condition" (2 levels, RN and RefRN for noise or RTx and RefRTx for textures) and "Duration" (5 levels, [125, 250, 500, 1000, 2000] ms). For noise, significant effects of Condition ($F(1,17) = 44.81$, $p < 0.001$, $\eta_G^2 = 0.32$) and Duration ($F(4,68) = 11.14$, $p < 0.001$, $\eta_G^2 = 0.18$) were observed, without any interaction between the two factors ($F(4,68) = 1.42$, $p = 0.24$, $\eta_G^2 = 0.03$). Similar findings were obtained with textures, with significant effects of Condition ($F(1,17) = 470.4$, $p < 0.001$, $\eta_G^2 = 0.47$) and Duration ($F(2.53,46.0.6) = 14.28$, $p < 0.001$, $\eta_G^2 = 0.23$). For textures, there was a significant interaction between Condition and Duration, with a small effect size ($F(3.28,55.68) = 5.00$, $p = 0.003$, $\eta_G^2 = 0.09$). Overall, the ANOVAs confirmed that, while duration affected performance, the advantage provided by repeated exposure to a same exemplar was at least as large for longer durations compared to shorter durations, for both noise (RefRN vs RN) and textures (RefTx vs RTx).

Finally, for textures, post hoc tests were run to compare each data point with all others, using a conservative Bonferroni correction (45 comparisons). We only report the crucial tests for the novel texture condition, namely, the possible memory gain for RefTx over RTx, reflecting learning of individual texture exemplars. At all tested durations, including the longest ones, a significant effect of RefTx was observed over RTx (all Bonferroni-corrected $p < 0.001$, except for 125 ms for which $p = 0.003$ and 250 ms for which $p = 0.002$).

3.3. Detailed effect of sound categories

To investigate in further detail whether the different texture categories influenced performance, a mixed-design ANOVA was run with factors "Condition" (2 levels: RN/RTx and RefRN/RefRTx), "Duration" (5 levels: [125, 250, 500, 1000, 2000] ms), and "Sound category" (4 levels: White noise, Fire, Stream, Wind). This main analysis is reported in full detail in the Appendix, Table A1. To summarize, all main effects were significant ($p < 0.001$). The two-way interactions Condition * Duration and Texture * Duration, as well as the three-way interaction Condition * Texture * Duration, were also significant ($p = 0.003$ or less), with relatively small effect sizes ($\eta_G^2 = 0.05$ or less).

To help interpret the effect of sound category in relation to our question of interest, the memory for natural textures, we transformed the data to directly estimate the amount of learning that was afforded by repeated exposure to a sound exemplar. To this effect, we computed the memory gain, defined as performance for the trials where the sound exemplar re-occurred throughout a block (RefRN/RefTx) minus performance where sound exemplars were novel on each trial (RN/RTx). The memory gain obtained for the different sound categories and durations is displayed in Fig. 4A. Even though the results were somewhat noisy, there was no trend for a smaller gain at longer texture durations, nor for a systematic advantage of white noise over natural textures.

A mixed-design ANOVA was performed on the memory gain with factors "Duration" (5 levels: [125, 250, 500, 1000, 2000] ms), and "Sound category" (4 levels: White noise, Fire, Stream, Wind). An effect of Duration was observed ($F(4,272) = 5.04$, $p < 0.001$, $\eta_G^2 = 0.05$), suggesting that the memory gain was, perhaps paradoxically, slightly *larger* for longer durations. However, due to the small effect size, we did not investigate the effect further with post hoc tests, and conservatively interpret this finding as a largely constant memory gain over all tested durations. Crucially, no effect of Sound category was found ($F(3,68) = 2.21$, $p = 0.10$, $\eta_G^2 = 0.02$). The Duration * Sound category was significant, with a medium effect size ($F(12,272) = 2.57$, $p = 0.003$, $\eta_G^2 = 0.08$). The interaction suggests that the small increase in memory gain with duration was mostly due to two natural textures, Fire and Stream.

Finally, we examined the baseline performance for individual sound categories in the repetition detection task, the conditions RN/RTx. The results are shown in Fig. 4B. Breaking-up the data into sound categories revealed that the peak at 250 ms observed in the mean texture data (Fig. 3B) was due to the Fire and Stream textures. The Wind texture instead displayed a peak at 1000 ms.

A mixed-design ANOVA was performed on the performance in the RN/RTx condition, with factors "Duration" (5 levels, [125, 250, 500, 1000, 2000] ms), and "Sound category" (4 levels, White noise, Fire, Stream, Wind). As for the memory gain analysis, an effect of Duration was observed ($F(4,272) = 24.14$, $p < 0.001$, $\eta_G^2 = 0.19$). Unlike for the memory gain analysis, this time an effect of Sound category was

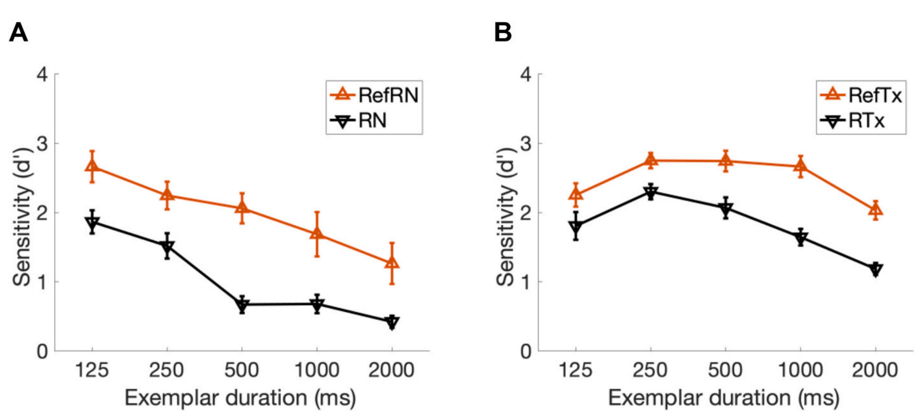

Fig. 3. Effect of exemplar duration. A). White noise. Mean performance across participants is shown for the RN and RefRN conditions, as a function of exemplar duration. Error bars represent standard error about the mean. B). Natural textures. Performance averaged for the natural textures. Same as A.).

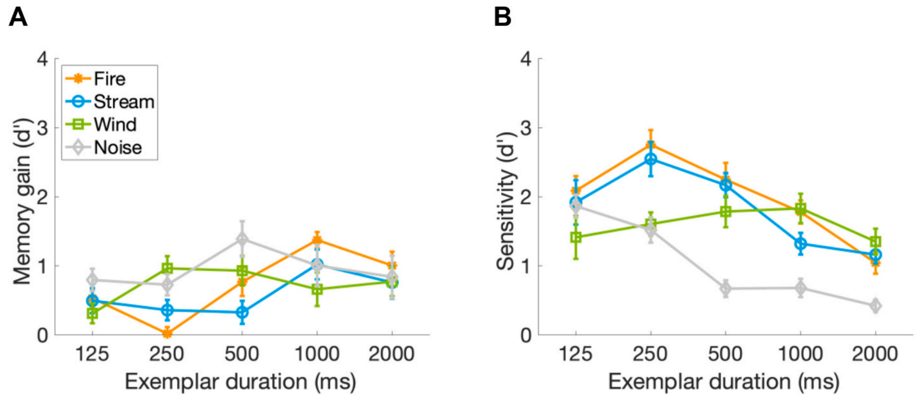

Fig. 4. Effect of sound categories. A). Learning was assessed by computing the sensitivity increase due to repeated exposure. A "Memory gain" was defined as performance for the trials where sound exemplars re-occurred across 20 trials (RefRN/RefTx) minus performance where sound exemplars were novel on each trial (RN/RTx) B). Performance for the within-trial repetition detection task (RN/RTx).

observed, with a large effect size ($F(3,68) = 17.68$, $p < 0.001$, $\eta_G^2 = 0.21$). The Duration * Sound category was also significant ($F(12,272) = 3.88$, $p < 0.001$, $\eta_G^2 = 0.10$).

A post hoc comparison of all data points for Fig. 4B was performed (Duration * Sound category interaction). Briefly, likely because of large number of comparisons, the robustness of the pattern of performance seen in Fig. 4B could not be formally confirmed. In particular, the apparent peaks in performance at 250 ms for Fire and Stream and at 1000 ms for Wind were not significantly different from their neighbors (Table A2). So, whereas there were significant differences across sound categories with large effect size, pinpointing them to specific durations and textures would require further experimental data.

In summary, these analyses show that the texture type can influence baseline repetition-detection performance, in terms of overall performance but also in terms of performance change with exemplar durations. For re-occurring exemplars, all sound categories exhibited a memory gain. Importantly, there was no clear advantage of noise, which produced the amount of memory gain expected from previous studies, over natural textures. Natural textures produced sizeable memory gains, even at the longest durations tested.

4. Signal model based on summary statistics

A major insight of the summary statistics approach is that, for stochastic sounds with an underlying stationary generative process, features derived from auditory models will converge to their expected mean as duration increases (McDermott et al., 2013; McDermott & Simoncelli, 2011). This convergence allows for a compact

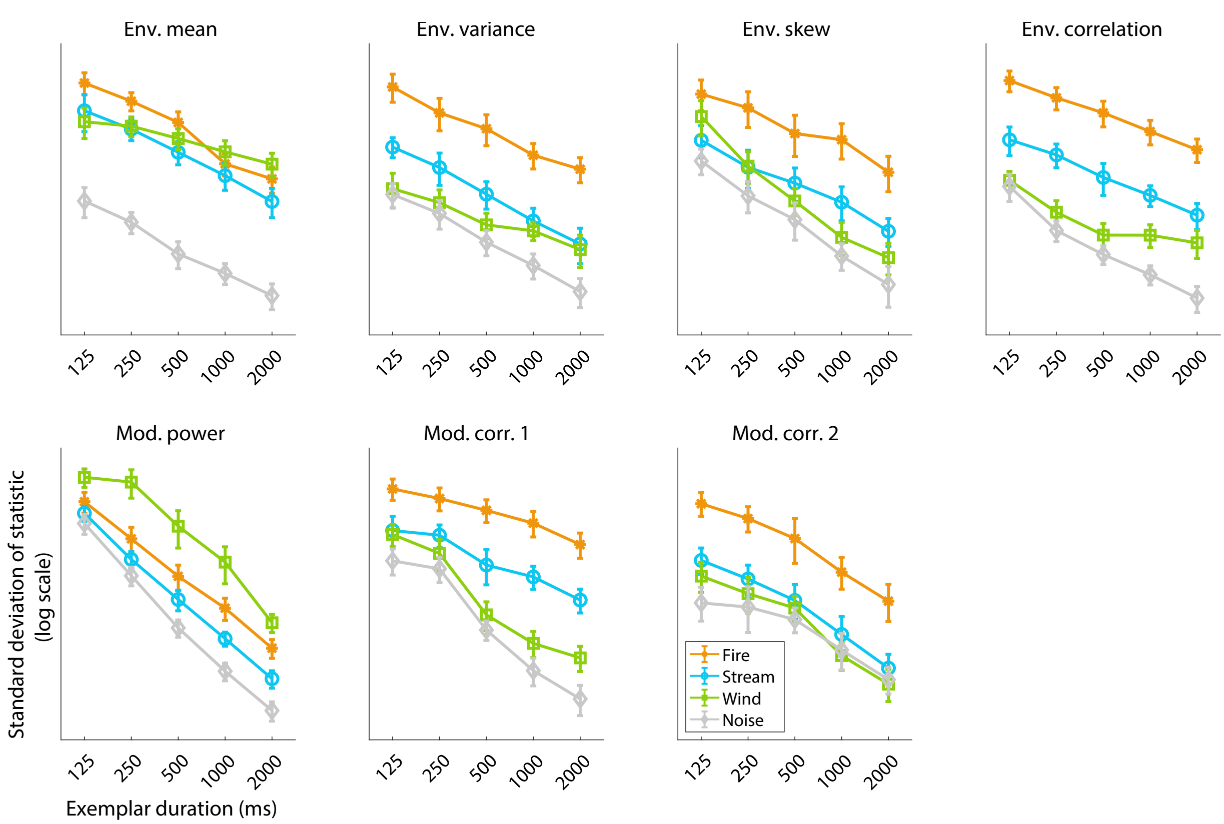

Fig. 5. Variance of summary statistics. The variance of seven summary statistics was evaluated over the full stimulus set used in our experiments. Median values are shown on a logarithmic scale, with error bars corresponding to the interquartile range (summary statistics and display format as in McDermott et al., 2013, Fig. 1c).

representation of natural auditory textures. The corresponding reduction in the variance of statistics has been proposed as the explanation for the otherwise counterintuitive finding that discrimination performance between two exemplars of the same texture type decreases with longer durations. Our repetition-detection task can be construed as a kind of exemplar discrimination task, so could the same model apply?

We computed the variance of summary statistics over all exemplars of our stimulus set, using the model of McDermott et al. (2013). Results are displayed in Fig. 5. The reduction in statistics' variance with duration was confirmed. The variability of our variance estimates was similar to previous reports. This confirms that our stimulus generation strategy was appropriate, even though we only matched the statistics of the synthesized stimuli to those of target natural textures over 392 s (see Methods) with no additional constraint for the shorter exemplars actually used in the experiments.

A qualitative comparison of the variance patterns can be made with the results of our repetition-detection task (Fig. 4B). The summary statistics model captured the most important trend of decreasing performance with duration, as expected. The model was also mostly successful in predicting the better performance observed for the Fire and Stream texture types at short durations, reflected by higher statistics' variance. However, the model did not predict the more subtle differences observed across texture types. In particular, the non-monotonic pattern observed for some natural textures categories was not reflected in any summary statistics variance pattern. The ordering of texture types in terms of statistics variance was also dependent on the statistics being considered, and the convergence in performance for Fire, Stream and Wind at longer durations was not captured.

Overall, the core prediction of reduced statistics' variance with longer exemplar duration was verified in our stimulus set. It is possible that changing model parameters could lead to a better fit with the finer details our data. However, the aim here was to document the results of the standard summary statistics model, so no such fit was attempted.

5. Simulation based on local features sampling

We now turn to a different idea: in addition to or instead of summary statistics, listeners may use a small set of temporally local features to perform the repetition-detection task. Such a hypothesis has been put forward many times to interpret otherwise puzzling aspects of repetition-detection for stochastic signals, such as the distinctive features subjectively heard in repeating noise (Guttman & Julesz, 1963; Warren et al., 2001), the ability to tap consistently to repeating noise (Kaernbach, 1993; Ringer et al., 2023), or the constant memory gain observed across sound durations (Agus et al., 2010; Agus & Pressnitzer, 2021; Andrillon et al., 2015; Kang et al., 2017). Moreover, idiosyncratic patterns of performance suggest that not all listeners use the same features for a given stochastic signal (Agus et al., 2010; Andrillon et al., 2015; Kaernbach, 1993). To the best of our knowledge, no attempt has been made to derive quantitative predictions from such a proposal.

We do so here by simulating the performance expected if listeners based their perceptual decisions on a small number of features *randomly sampled* from the sound. Such an idea is directly inspired by the classic "Stimulus Sampling Theory" (Estes, 1950). As an early attempt to mathematically derive quantitative behavioral characteristics of learning, the stimulus sampling theory has led to numerous variants, for both memory and perceptual tasks (see Kent, Guest, Adelman, & Lamberts, 2014, for a review). More recently, stochastic sampling of discrete stimulus features has been put forward as a unifying framework for computational and neural models of visual working memory (Schneegans, Taylor, & Bays, 2020).

Our implementation was kept as minimal as possible, to highlight the generic consequences of stimulus sampling. A texture exemplar was abstractly represented as a list of discrete features: a list of integers randomly drawn from a dictionnary of size *M,* the first parameter of the simulations. The number of features representing an examplar was decided by the second parameter of the simulation, *R*, the number of features per second. Random sampling of the features list was implemented by Poisson thinning of the full list, as in Schneegans et al. (2020). Reporting a repetition was finally decided by checking for multiple appearances of the same feature in the sampled list.

Fig. 6A further illustrates how the simulation was applied to our repetition-detection task. First, for each texture examplar, a list was drawn of (*R**duration) random integers, uniformly distributed between 1 and *M*. To simulate repeated trials, a texture examplar *A* was drawn and the trial was constructed as the catenated list of features *AA*. For non-repeated trials, another texture examplar *B* was drawn and the trial was constructed as the catenated list of features *AB*. Random sampling of the full feature list was performed by Poisson thinning with rate λ and equal weight for each feature (Schneegans et al., 2020). The simulated decision was "Repeated" if there were repeated features in the sampled list.

The simulation's formalism can be interpreted as follows. The size of the dictionary *M* and the feature rate *R* are characteristics of a texture type: some textures may display a broader variety of features than others, and features may (on average) be longer or shorter for different texture types. The sampling process is the core idea we wished to probe: not all features will be available for the perceptual decision, presumably because of constraints on memory processes, but also because stochastic stimuli such as noise or natural textures sound initially "featureless". Note that our model does not simulate an online comparison of each and every incoming feature to features stored in memory. Instead, the sampling occurs on the full stimulus representation, in order to present only a subset of features to the perceptual decision process. Finally, the decision rule is a based on a simple criterion that counts the number of repeated features in the sampled feature list.

This minimal implementation was sufficient to predict decreasing performance with texture duration. Intuitively, hits happen in the simulation when the random sampling process picks up twice the same feature in the *AA* list. This is more likely if the *AA* list is short. False alarms happen when the same feature appears by chance in the *AB* list. This is more likely if the *AB* list is long when *M* is large enough. Thus, in general, hits will decrease and false alarm will increase with duration, predicting declining performance. However, the simulation parameters can modulate the qualitative prediction: if *M* is low, there are fewer different features to sample from for a given texture type, so both hits and false alarms should increase. If *R* is low, the feature list will be short compared to λ at short durations, again impacting hits and false alarms. We therefore turned to numerical simulations to further probe the simulations' behavior.

Grid searches on the texture-dependent parameters *M* and *R* were performed to find the best fits to the repetition-detection results of Fig. 4B. For simplification, and after pilot simulations showed the robustness of the predictions for different choices, all parameters related to the sampling and decision process were kept fixed: a Poisson rate parameter of $\lambda = 8$ was used throughout, corresponding to an expected rate of 8 sampled features per stimulus. The criterion for deciding on a repeat was set at 1, meaning at least a single repeated feature after sampling was enough to report a repeat. Results of the best fits (minimization of quadratic error) are shown in Fig. B1. This figure confirms that decreasing performance can be predicted purely through sampling, without any temporal averaging operation.

To improve the fit to our data, and in particular to capture the non-monotonic trends observed for some texture types, we introduced a modification to the sampling process. The Poisson rate parameter λ was limited to the total number of features available in a stimulus, meaning that the expected number of sampled features was either 8 or the total number of stimulus features, whichever was smallest. The best fits are shown in Fig. 6B. The distribution of the goodness of fits over *M* and *R* are shown in Fig. 6C. With this modification, satisfactory fits were obtained for all texture types. For Noise, simulation performance smoothly decreased with duration, which is the default behavior of Fig. B1.

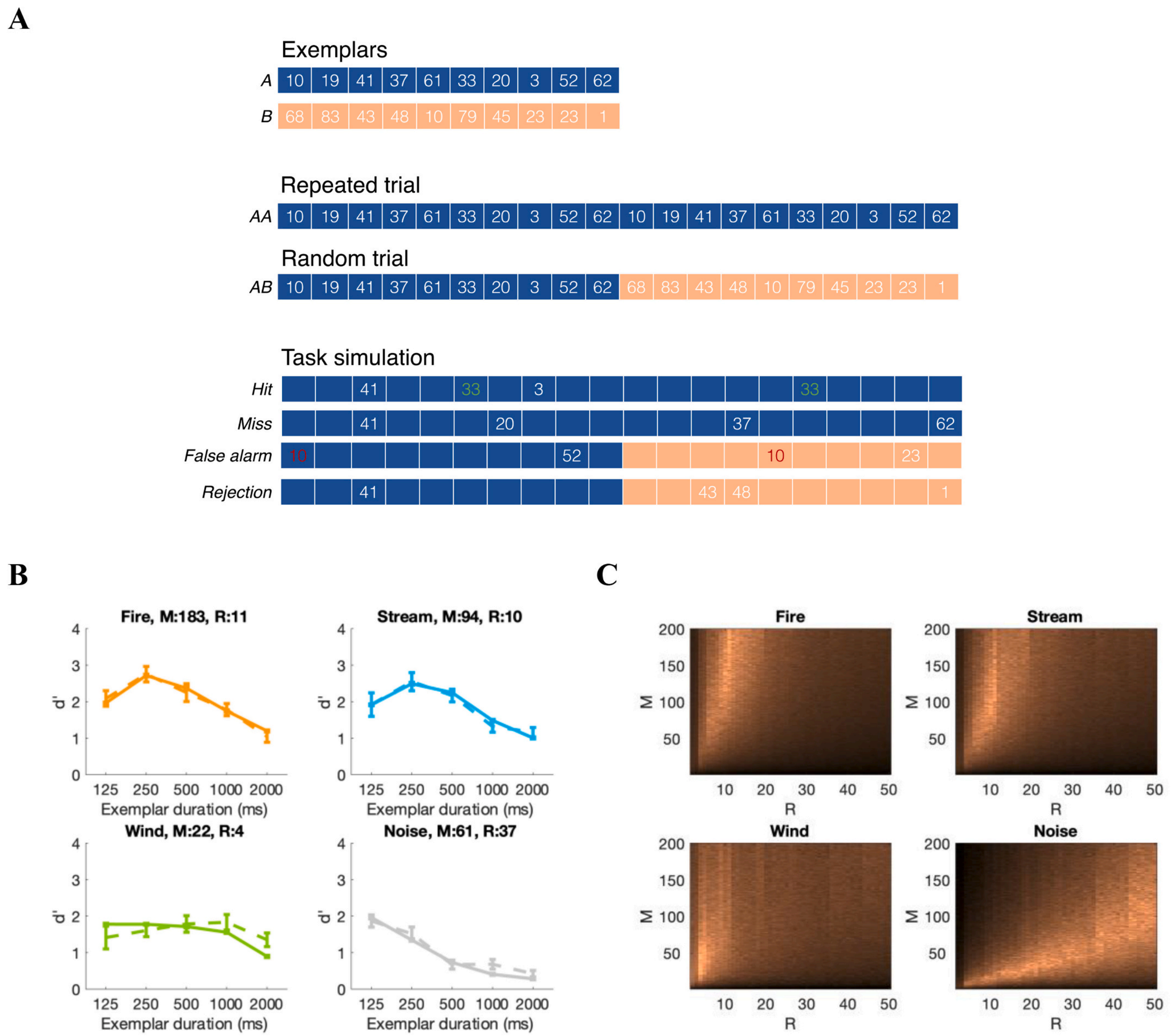

Fig. 6. Local features sampling. A. Each texture exemplar is represented by a list of features, random numbers from 1 to M. The parameter M represents the number of different features available for a given texture type. The number of features per exemplar depends on a second parameter, R, the rate of features per second. Here, two exemplars, A and B, are shown, with parameters [M = 100; R = 10] and duration = 1 s. Trials for the experiments are lists of features, constructed from texture exemplars. In repeated trials, the same exemplar is repeated (*AA*). In non-repeated trials, two different exemplars are collated (*AB*). To reach a perceptual decision, features are randomly sampled from the trial's list of features. If a same feature appears more than once in the sampled features, then a "Repeated" answer is provided. From this strategy, all possible experimental outcomes can be simulated. B. For each texture type, the d' values simulated (solid lines) are compared to the behavioral results (dashed lines, reproduced from Fig. 4B). The panel title indicates the parameters of the simulation. C. Distribution of goodness of fits (shown as -log (err^2)) over the parameter space investigated in the grid search.

Perhaps as a result, good fits were obtained over a broad parameter range, with the best fit observed for $M = 61$ and $R = 37$. The non-monotonic pattern of performance for Fire and Stream was simulated thanks to a lower feature rate ($R = 11$ or 10) and a large number of features ($M = 183$ or 84). Finally, the relatively flat performance for Wind was obtained for a very low feature rate ($R = 4$) and a small number of possible features ($M = 22$). The numerical values of M and R at the best fit should not be taken too literally, though, as there were many parameter combinations that produced approximately equally good fits (Fig. 6C). Nevertheless, the regions where good fits were observed seemed to match intuition, as Fire and Stream sound sparser than Noise (lower R), but with more possible distinct features than Noise (higher M). Wind, which in our stimulus set covered the narrowest frequency band, was fitted with both low R and low M.

6. General discussion

Combining ideas and techniques from two different lines of research, we applied a memory for noise paradigm to natural auditory textures and artificial white noise. Exemplars from different texture types were used in a repetition detection task, with parametrically varied exemplar duration. Results were similar for natural and artificial sounds: repetition detection performance was accurate for short durations and poorer for long duration, even though a non-monotonic pattern of results was observed for natural textures. An increase in performance for exemplars that re-occurred throughout an experimental block was observed in all cases, which we interpreted as a memory gain. The memory gain was at least as large for long durations than for short durations. Computational modeling showed that two a priori distinct representations, summary statistics and local feature sampling, could be used to predict the general decrease in performance with duration in the repetition detection task. After briefly comparing the behavioral results to the literature, the data and models will be discussed in the light of the long-standing issue of auditory representation for perception and memory.

6.1. Texture repetition-detection and texture exemplar discrimination

The results for noise replicate previous findings in repetition-detection tasks in a laboratory setting. Agus et al. (2010) observed a mean performance ranging from about $d' = 1.5$ to about $d' = 0.5$ for exemplar durations from 125 ms to 2000 ms. This is similar, or perhaps even slightly poorer, than what was observed here in an online setting (Fig. 3A, RN). Moreover, re-occurring exemplars exhibited a memory gain of about 1 d'-unit in both sets of findings (Fig. 3A, RefRN). This further validates the use of online testing with "memory for noise" paradigms (Dauer et al., 2022; Ringer et al., 2022).

The present results for natural textures may also be compared to the texture discrimination task of McDermott et al. (2013, Experiment 2). We used the exact same texture generation algorithm and tested exemplar durations over a similar range. A difference between the two studies is that, in the texture exemplar discrimination task, all exemplars within a trial were surrounded by silent gaps (McDermott et al., 2013). Here, in contrast, within-trial repetitions were seamless. Nevertheless, there was also a direct repetition in the texture examplar discrimination task (the odd-one-out was either the first or the last sound of a trial) and the exact value of the silent gaps made no difference to performance. Considering that noise learning can occur even with non-contiguous repetition (Andrillon et al., 2015; Bianco et al., 2020; Kaernbach, 2004; Ringer et al., 2023), a direct comparison between our findings for natural textures and the texture discrimination task of McDermott et al. (2013) seems warranted.

We interpolated the percentage correct values provided in Fig. 2b of McDermott et al. to the durations tested here, and converted the results to d'. This led to interpolated performance of $d' = [1.9, 1.5, 1.1, 0.9, 0.7]$ for durations of [125, 250, 500, 1000, 2000] ms. This is broadly consistent with what was observed in the present repetition-detection task (Fig. 3B, RTx). The additional finding of the present study concerns natural texture exemplars that re-occurred over 20 trials (Fig. 3B, RefRTx). In this case, increased exposure led to improved performance for all tested durations.

6.2. Summary statistics model and local feature sampling simulation

Both the summary statistics model and the feature sampling simulation predicted the main trend of the behavioral results: a general decrease in performance as exemplar duration increased. Summary statistics predicted this trend by a reduced variance of statistics as duration increased. Feature sampling predicted this trend by the probabilistic effect of selecting a fixed number of local features from increasingly long sets of possible features as duration increased. That these two independent frameworks led to similar predictions is a novel insight, suggesting that either one, or both, could be at play for repetition-detection and exemplar discrimination with natural textures.

Outright, it is important to stress that a quantitative comparison of the two approaches is unwarranted on the basis of the present results only. The summary statistics model is a complete signal processing pipeline, with parameters that were not adjusted to our dataset. Also, we made no attempt to convert the model's output (variance of a set of statistics) to a single d' measure. In contrast, the feature sampling simulation was purely conceptual. By design, it provided a d' output so its parameters could be fitted to the data. Therefore, we do not wish to imply that one model "outperformed" the other. Instead, our point is that both approaches can predict the general decrease in performance with duration – which was expected for summary statistics, but perhaps less obvious for local features sampling. Moreover, the two underlying representations need not be exclusive to each other (see Subsection 6.4).

As the application of local features sampling to auditory repetition-detection is novel, we focus the rest of this subsection on its strengths and shortcomings. First, for the shortcomings, unrealistic assumptions were made in our simulations. To cite a few, all possible features for a texture category were equally likely to appear for a given texture exemplar; all features for a texture type had the same duration; all features were equally likely to be sampled. These assumptions were made to keep the simulation as close as possible to its core principles and thus clarify the intuitions derived from stimulus sampling. For natural textures, it is instead almost certain that some features will be more likely to appear than others across exemplars, because of the physical properties of the generative processes. Moreover, some features may exhibit greater perceptually saliency than others and thus become more likely to be sampled, contrary to the equiprobable sampling enforced here. Future signal models implementing stimulus sampling for auditory stimuli would need to address these shortcomings. In parallel, behavioral experiments could try to characterize the local features that arise from repeated natural textures, for instance by using a reverse correlation approach (Kaernbach, 1993).

The local features sampling approach does open new perspectives, however. Perhaps most importantly, it is congruent with introspection when listening to repeated noises or textures: one clearly hears individual "events" (Guttman & Julesz, 1963; Warren et al., 2001). The simulations show that such sparse events could form the basis of repetition-detection performance. Moreover, the formalism of the simulations suggests potential cognitive mechanisms for the auditory repetition-detection task. Speculatively, the stimulus sampling process could represent the transfer of a limited number of distinctive features from implicit auditory sensory memory stores (Cowan, 1984; Kaernbach, 2004; Nees, 2016) to working memory (Schneegans et al., 2020). To further account for the memory gain observed when a same exemplar reoccurs many times, one could hypothesize that reoccurring features are transferred from working memory to long-term memory, with memorized features more likely to be sampled in future presentations. Such a change in likelihood can naturally be introduced in the Poisson sampling framework, through a change of feature weights (Schneegans et al., 2020). Finally, there could be further possible refinements to the sampling mechanism, for instance by introducing temporal decays on the memory traces of features (Harrison, Bianco, Chait, & Pearce, 2020; Kent et al., 2014). Those are beyond the scope of the present study but would be useful to introduce further plausible constraints to cognitive models of the repetition-detection task.

6.3. Is repetition special?

Given the similarity between results for artificial noise and natural textures, it appears that the memory processes induced by repetition and re-occurrence generalized to natural textures.

There are several speculative arguments suggesting that repetition should be special for the auditory modality. From an acoustic point of view, it is impossible to actively search "back in time" for additional information once a sound has ended, unlike for visual search (Demany, Semal, Cazalets, & Pressnitzer, 2010; Garnier-Allain, Pressnitzer, & Sergent, 2023). Thus, the auditory system may have evolved to be exquisitely tuned to repetitions, as they provide a unique opportunity to re-examine auditory cues with a deeper level of processing. Also, scene elements that reoccur likely indicate an agent in the environment that may be behaviorally significant. Finally, in a Bayesian or predictive coding framework, events that repeat in the past may reasonably be assigned a higher-than-baseline probability of repeating in the future, so repeated sounds should be expected to alter neural processing (Baldeweg, 2006). Perhaps relatedly, repetitions have also been shown to change the very perceptual qualities of sound, such as in the speech-to-song illusion (Deutsch, Henthorn, & Lapidis, 2011), which, interestingly, generalizes to natural textures (Rowland, Kasdan, & Poeppel, 2019).

Thus, one interpretation of the present findings is that repetition overcame a putative specialized processing mode for textures. If local perceptual features could be preserved for at least the duration of an exemplar, then an immediate repetition could be detected and a "repetition mode" recruited. As subsequent re-occurrences of the texture exemplar were presented, the preserved features could have been consolidated into memory. Such a repetition-induced representation could also be beneficial to auditory scene analysis, by creating foreground events emerging from the background texture when repetition is involved (McDermott, Wrobleski, & Oxenham, 2011). A feature-based representation could of course co-exist with a summary statistics representation, which would still be efficient for e.g. texture category recognition. Such a possibility of different processing modes for natural textures was floated in the original texture studies (McDermott et al., 2013; Nelken & de Cheveigné, 2013). The present data provide experimental support for the idea.

6.4. Time scales of representation

Another possible interpretation relies on a looser view of the dichotomy between local features versus summary statistics. Indeed, the core difference between the two types of representations is the time scale over which features versus statistics are estimated. Thus, there could be an overlap between the two notions if auditory representations were based on multiple or even flexible time scales of integration.

There is a large and growing body of evidence suggesting the auditory system represents sounds over different time scales, from behavioral (Divenyi, 2004; Teng et al., 2016) or neural findings (Asokan, Williamson, Hancock, & Polley, 2021; Joris, Schreiner, & Rees, 2004; Norman-Haignere et al., 2022). The details of the associated theories differ on important points, such as whether all time scales within the possible range are available simultaneously, whether they depend on the task and context (McWalter & McDermott, 2019 in the case of natural textures), or whether a limited number of fixed windows exist to encode fine and coarse details (Teng et al., 2016). However, all accounts suggest that short and long time-scales may coexist in the auditory representations of complex sounds.

As a result, the boundary between a relatively long-duration feature and a relatively short-duration statistic becomes blurry. Furthermore, the modeling results presented here show that representations based on either long-duration statistics or short-duration local features could provide qualitatively similar predictions, both in line with the behavioral data. Thus, instead of a dichotomy in kind, we suggest that the features versus statistics distinction may better be thought of in terms of a continuum over different time scales.

7. Conclusion

We have shown that naturalistic texture exemplars are amenable to learning when repeated exposure is available. In this respect, natural textures join the growing list of stochastic sounds that behave surprisingly similarly in a memory for noise paradigm. This main finding is consistent with two interpretations: the existence of a special processing mode when acoustic repetition is involved, to which natural textures are not immune, or a convergence of the feature set versus summary statistics representations, if a continuum of time scales is considered.

Whereas the computational appeal of summarizing a texture to its statistics is obvious, one may wonder what use there could be to store the detailed acoustic features of a given exemplar? It could be that such a finding is simply the by-product of powerful plasticity mechanisms triggered by repetition, which are otherwise useful to generate sparse representations of meaningful sounds (Wang et al., 2020). We speculate that, more generally, it is the sign of the auditory system adapting its internal representations to the statistical regularities of its environment.

CRediT authorship contribution statement

Berfin Bastug: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Vani G. Rajendran:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Roberta Bianco:** Writing – review & editing, Methodology, Conceptualization. **Trevor Agus:** Writing – review & editing, Formal analysis. **Maria Chait:** Writing – review & editing, Methodology, Conceptualization. **Daniel Pressnitzer:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization.

Funding

Roberta Bianco was funded by the European Union (MSCA, PHYLOMUSIC, 101064334). Maria Chait was partly supported by BBSRC project grant BB/P003745/1. Daniel Pressnitzer was partly supported by grants ANR-17-EURE-0017 and ANR-22-CE28-0023.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106350.

Data availability

Data shared on ResearchBox (anonymized) with the link available in the manuscript

References

Agus, T. R., & Pressnitzer, D. (2013). The detection of repetitions in noise before and after perceptual learning. *The Journal of the Acoustical Society of America, 134*(1), 464–473. https://doi.org/10.1121/1.4807641

Agus, T. R., & Pressnitzer, D. (2021). Repetition detection and rapid auditory learning for stochastic tone clouds. *The Journal of the Acoustical Society of America, 150*(3), 1735. https://doi.org/10.1121/10.0005935

Agus, T. R., Suied, C., Thorpe, S. J., & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *The Journal of the Acoustical Society of America, 131*(5), 4124–4133. https://doi.org/10.1121/1.3701865

Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron, 66*(4), 610–618. https://doi.org/10.1016/j.neuron.2010.04.014

Andrillon, T., Kouider, S., Agus, T., & Pressnitzer, D. (2015). Perceptual learning of acoustic noise generates memory-evoked potentials. *Current Biology: CB, 25*(21), 2823–2829. https://doi.org/10.1016/j.cub.2015.09.027

Andrillon, T., Pressnitzer, D., Léger, D., & Kouider, S. (2017). Formation and suppression of acoustic memories during human sleep. *Nature Communications, 8*(1), 1–15. https://doi.org/10.1038/s41467-017-00071-z

Asokan, M. M., Williamson, R. S., Hancock, K. E., & Polley, D. B. (2021). Inverted central auditory hierarchies for encoding local intervals and global temporal patterns. *Current Biology, 31*(8), 1762–1770.e4. https://doi.org/10.1016/j.cub.2021.01.076

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences, 10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Bianco, R., Hall, E. T. R., Pearce, M. T., & Chait, M. (2023). Implicit auditory memory in older listeners: From encoding to 6-month retention. *Current Research in Neurobiology, 5*, Article 100115. https://doi.org/10.1016/j.crneur.2023.100115

Bianco, R., Harrison, P. M., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *ELife, 9*, Article e56073. https://doi.org/10.7554/elife.56073

Bianco, R., Mills, G., de Kerangal, M., Rosen, S., & Chait, M. (2021). Reward enhances online participants' engagement with a demanding auditory task. *Trends in Hearing, 25*. https://doi.org/10.1177/23312165211025941, 23312165211025940.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96*(2), 341–370. https://doi.org/10.1037/0033-2909.96.2.341

Dauer, T., Henry, M. J., & Herrmann, B. (2022). Auditory perceptual learning depends on temporal regularity and certainty. *Journal of Experimental Psychology: Human Perception and Performance, 48*(7), 755–770. https://doi.org/10.1037/xhp0001016

Demany, L., Semal, C., Cazalets, J.-R., & Pressnitzer, D. (2010). Fundamental differences in change detection between vision and audition. *Experimental Brain Research, 203* (2), 261–270. https://doi.org/10.1007/s00221-010-2226-2

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America, 129*(4), 2245–2252. https://doi.org/10.1121/1.3562174

Divenyi, P. (2004). The times of Ira Hirsh: Multiple ranges of auditory temporal perception. *Seminars in Hearing, 25*(3), 229–239. https://doi.org/10.1055/s-2004-832857

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57*(2), 94–107. https://doi.org/10.1037/h0058559

Garnier-Allain, A., Pressnitzer, D., & Sergent, C. (2023). Retrospective cueing mediates flexible conscious access to past spoken words. *Journal of Experimental Psychology: Human Perception and Performance, 49*(7), 949–967. https://doi.org/10.1037/xhp0001132

Geffen, M. N., Gervain, J., Werker, J. F., & Magnasco, M. O. (2011). Auditory perception of self-similarity in water sounds. *Frontiers in Integrative Neuroscience, 5*, 15. https://doi.org/10.3389/fnint.2011.00015

Gervain, J., & Geffen, M. N. (2019). Efficient neural coding in auditory and speech perception. *Trends in Neurosciences, 42*(1), 56–65. https://doi.org/10.1016/j.tins.2018.09.004

Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *The Journal of the Acoustical Society of America, 35*(4), 610. https://doi.org/10.1121/1.1918551

Harrison, P. M. C., Bianco, R., Chait, M., & Pearce, M. T. (2020). PPM-decay: A computational model of auditory prediction with memory decay. *PLoS Computational Biology, 16*(11), Article e1008304. https://doi.org/10.1371/journal.pcbi.1008304

Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience, 389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

JASPTeam. (2024). *JASP version 0.18.3* (0.18.3) [Computer software] https://jasp-stats.org/.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84*(2), 541–577. https://doi.org/10.1152/physrev.00029.2003

Kaernbach, C. (1993). Temporal and spectral basis of the features perceived in repeated noise. *The Journal of the Acoustical Society of America, 94*(1), 91–97. http://pubmed.gov/8354764.

Kaernbach, C. (2004). The memory of noise. *Experimental Psychology, 51*(4), 240–248. https://doi.org/10.1027/1618-3169.51.4.240

Kang, H., Agus, T. R., & Pressnitzer, D. (2017). Auditory memory for random time patterns. *The Journal of the Acoustical Society of America, 142*(4), 2219–2232. https://doi.org/10.1121/1.5007730

Kent, C., Guest, D., Adelman, J. S., & Lamberts, K. (2014). Stochastic accumulation of feature information in perception and memory. *Frontiers in Psychology, 5*, 412. https://doi.org/10.3389/fpsyg.2014.00412

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America, 26*(39), 201705652–201705656. https://doi.org/10.1073/pnas.1705652114

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Luo, H., Tian, X., Song, K., Zhou, K., & Poeppel, D. (2013). Neural response phase tracks how listeners learn new acoustic representations. *Current Biology, 23*(11), 1–7. https://doi.org/10.1016/j.cub.2013.04.031

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory*. Psychology Press. http://books.google.com/books/content?id=2_V5AgAAQBAJ&printsec=frontcover&img=1&zoom=1&edge=curl&source=gbs_api.

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience, 16*(4), 493–498. https://doi.org/10.1038/nn.3347

McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron, 71*(5), 926–940. https://doi.org/10.1016/j.neuron.2011.06.032

McDermott, J. H., Wrobleski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences, 108*(3), 1188–1193. https://doi.org/10.1073/pnas.1004765108

McWalter, R., & McDermott, J. H. (2019). Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nature Communications, 10*(1), 5096. https://doi.org/10.1038/s41467-019-12893-0

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods, 53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology, 7*, 1892. https://doi.org/10.3389/fpsyg.2016.01892

Nelken, I., & de Cheveigné, A. (2013). An ear for statistics. *Nature Neuroscience, 16*(4), 381–382. https://doi.org/10.1038/nn.3360

Norman-Haignere, S. V., Long, L. K., Devinsky, O., Doyle, W., Irobunda, I., Merricks, E. M., … Mesgarani, N. (2022). Multiscale temporal integration organizes hierarchical computation in human auditory cortex. *Nature Human Behaviour, 6*(3), 455–469. https://doi.org/10.1038/s41562-021-01261-y

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences, 24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Ringer, H., Schröger, E., & Grimm, S. (2022). Perceptual learning and recognition of random acoustic patterns. *Auditory Perception & Cognition, 5*(3–4), 259–281. https://doi.org/10.1080/25742442.2022.2082827

Ringer, H., Schröger, E., & Grimm, S. (2023). Within- and between-subject consistency of perceptual segmentation in periodic noise: A combined behavioral tapping and EEG study. *Psychophysiology, 60*(2), Article e14174. https://doi.org/10.1111/psyp.14174

Rowland, J., Kasdan, A., & Poeppel, D. (2019). There is music in repetition: Looped segments of speech and nonspeech induce the perception of music in a time-dependent manner. *Psychonomic Bulletin & Review, 26*(2), 583–590. https://doi.org/10.3758/s13423-018-1527-5

Roye, A., Schröger, E., Jacobsen, T., & Gruber, T. (2010). Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(21), 7310–7313. https://doi.org/10.1523/jneurosci.1113-10.2010

Schneegans, S., Taylor, R., & Bays, P. M. (2020). Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences, 117*(34), 20959–20968. https://doi.org/10.1073/pnas.2004306117

Teng, X., Tian, X., & Poeppel, D. (2016). Testing multi-scale processing in the auditory system. *Scientific Reports, 6*(1), Article 34390. https://doi.org/10.1038/srep34390

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience, 15*(6), 355–366. https://doi.org/10.1038/nrn3731

Traer, J., Norman-Haignere, S. V., & McDermott, J. H. (2021). Causal inference in environmental sound recognition. *Cognition, 214*, Article 104627. https://doi.org/10.1016/j.cognition.2021.104627

Velasco, C., Jones, R., King, S., & Spence, C. (2013). The sound of temperature. *Journal of Sensory Studies, 28*(5), 335–345. https://doi.org/10.1111/joss.12052

Viswanathan, J., Rémy, F., Bacon-Macé, N., & Thorpe, S. J. (2016). Long term memory for noise: Evidence of robust encoding of very short temporal acoustic patterns. *Frontiers in Neuroscience, 10*, 610–611. https://doi.org/10.3389/fnins.2016.00490

Wang, M., Liao, X., Li, R., Liang, S., Ding, R., Li, J., Zhang, J., He, W., Liu, K., Pan, J., Zhao, Z., Li, T., Zhang, K., Li, X., Lyu, J., Zhou, Z., Varga, Z., Mi, Y., Zhou, Y., … Chen, X. (2020). Single-neuron representation of learned complex sounds in the auditory cortex. *Nature Communications, 11*(1), 4361. https://doi.org/10.1038/s41467-020-18142-z

Warren, R. M., Bashford, J. A., Cooley, J. M., & Brubaker, B. S. (2001). Detection of acoustic repetition for very long stochastic patterns. *Perception & Psychophysics, 63* (1), 175–182. http://pubmed.gov/11304013.

Whitney, D., & Leib, A. Y. (2016). Ensemble Perception. *Annual Review of Psychology, 69* (1), 1–25. https://doi.org/10.1146/annurev-psych-010416-044232