# Memory for repeated auditory textures

Berfin Bastug, Vani Rajendran, Roberta Bianco, Trevor Agus, Maria Chait, Daniel Pressnitzer

# Memory for repeated auditory textures

Berfin Bastug[1,2], Vani Rajendran[1,3], Roberta Bianco[4,5], Trevor Agus[6], Maria Chait[5], & Daniel Pressnitzer[1]

1- Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France.
2- Ernst Strüngmann Institute for Neuroscience in Cooperation with Max Planck Society, Frankfurt am Main 60528, Germany
3- Institute of Neurobiology, National Autonomous University of Mexico, Campus Juriquilla, Querétaro, Qro 76230, México
4- Neuroscience of Perception & Action Lab, Italian Institute of Technology, Viale Regina Elena 291, 00161 Rome, Italy
5- UCL Ear Institute, 332 Gray's Inn Rd London WC1X 8EE, UK
6- School of Arts, English and Languages, Queen's University Belfast, University Road, Belfast BT7 1NN

**Declarations of interest:** none

**Corresponding authors:** Berfin Bastug (berfin.bastug@esi-frankfurt.de) & Daniel Pressnitzer (daniel.pressnitzer@ens.psl.eu)

## Highlights

- Individual exemplars of natural auditory textures can be memorized when repeated.
- For long exemplars, the corresponding memory traces are likely based on temporally-local features.
- Temporally-local features can co-exist with time-averaged summary statistics in the representations of natural textures.
- Auditory cognition can flexibly recruit multi-scale representations of complex sounds, shaped by task demands and fine-tuned through rapid plasticity

# 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 sound texture exemplars and used them in an implicit memory paradigm based on repetition, previously shown to induce rapid learning for artificial stochastic signals such as white noise. On the one hand, if the brain represented natural texture exemplars only as time-averaged summary statistics, then repetition detection and learning should all but vanish for longer exemplar durations. On the other hand, if the learning of repeated noise and repeated textures relied on similar processes, presumably involving temporally-local features, then results should be similar for noise and natural textures. Results were highly similar for artificial and natural sounds, showing an effect of duration on baseline repetition detection performance but little to no effect on learning afforded by repeated exposure. Thus, naturalistic texture exemplars were amenable to learning when repeated exposure was available. This finding is consistent with two interpretations: the existence of a special processing mode when acoustic repetition is involved, to which textures are not immune; or a convergence of the feature set versus summary statistics description of sound representations if a continuum of time scales is considered.

# 1. Introduction

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 et al., 2010). More generally, a pervasive role of memory in perception is at the core of current theories based on predictive coding, which assume that a model of the world has been somehow internalized through experience (Heilbron & Chait, 2018; Lange et al., 2018). 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 have been put forward: a representation based on temporally-local features (Agus & Pressnitzer, 2021), or a representation based on time-averaged summary statistics (McDermott et al., 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 et al., 2001). Subsequent experiments have confirmed that the features used to detect repetition in white noise generally seem to have a local time-frequency support (Kaernbach, 1993; Ringer et al., 2023). Recently, the technique has been extended to longer-term memory traces. When listeners were exposed to the same exemplar of white 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 et al., 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 identifier for a given noise exemplar (Agus et al., 2010). Neural correlates of the phenomenon were consistent with the feature set hypothesis, with the added proposal that feature sets could be at least partly idiosyncratic and thus unique to each listener/noise combination (Andrillon et al., 2015; Luo et al., 2013; Ringer et al., 2023). Finally, similar findings were obtained with stochastic sounds other than white noise, such as random melodies (Bianco et al., 2020, 2023), random rhythms (Kang et al., 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 et al., 2017). This all suggests that repetition automatically triggers the rapid formation of memory traces for many kinds of sounds.

The summary statistics hypothesis stems from work on auditory textures (McDermott et al., 2013; McDermott & Simoncelli, 2011). Such textures are natural sounds with stochastic but stationary characteristics, such as the sound of fire crackling, water flowing, or wind blowing. The first important finding was that synthetic sounds matched in certain long-term statistics to natural sounds, as extracted from an auditory model, were readily identified as natural sound textures by listeners (Geffen et al., 2011; McDermott & Simoncelli, 2011). This showed that summary statistics were sufficient to recognize textures. 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 et al., 2016). However, such a seemingly paradoxical 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"*. Similar results were observed with random melodies, whereby listeners could estimate the mean frequency of a melody without precise reports of the frequencies or temporal position of individual tones (Piazza et al., 2012, but see Bianco et al., 2020). Thus, 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).

At face value, the logical consequence of the summary statistics hypothesis is that different exemplars of the same texture cannot be memorized, simply because they cannot be discriminated in the first place. Thus, unlike other stochastic sounds or even tone clouds with similar spectro-temporal statistics, repeated exposure to a texture exemplar should not trigger any memory trace specific to that exemplar. As suggested by Nelken & de Cheveigné (2013), this stark difference could be because the auditory system may enter a "texture mode" when it recognizes one. Indeed, it makes much more sense to recognize the physical cause of a texture

(McDermott et al., 2013; McDermott & Simoncelli, 2011) or even some of its characteristics, such as temperature for flowing water (Velasco et al., 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) to facilitate the categorization of the physical events making up our environment (Traer et al., 2021). However, an alternative to this "textures are special" idea is that "repetition is special", in that repetition may trigger powerful plasticity phenomena able to overcome or complement the summary statistics code. This possibility is experimentally tested here. We synthesized naturalistic sound textures using the original McDermott & Simoncelli algorithm (2011) and used them in the "memory for noise" paradigm of Agus et al. (2010), based on repetition.

## 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 repeat 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), reoccurs 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 & 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 Figure 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. For the RN/RTx and RefRN/RefRTx conditions, this corresponded to exemplar durations of 125 ms, 250 ms, 500 ms, 1000 ms, and 2000 ms, respectively. Testing different duration was critical as the summary statistics hypothesis predicts that the representation of longer textures will converge to the same statistics within a texture category, thus making the task harder.

### *2.2. Participants*

72 individuals (13 female), 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. 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.

### *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. Unlike for training, 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 et al., 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 $\alpha$-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'*, so 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 $\eta^2$, notated $\eta^2_G$, 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).

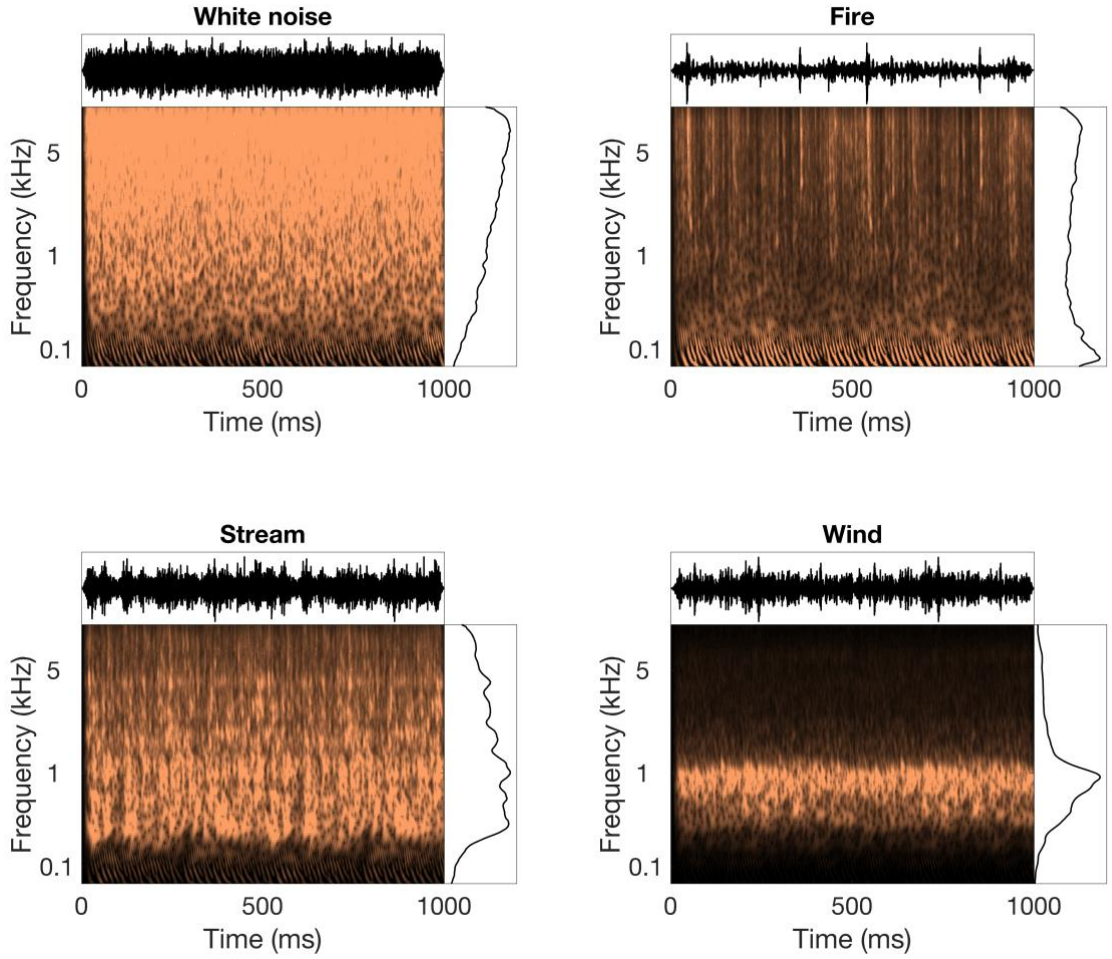


**Figure 1: Acoustic characteristics of white noise and natural textures**. Illustrative examples of cochleagrams for the four different sound categories. The time-frequency cochleagrams use brighter colors to represent energy within simulated auditory filters. In all cases, repeated trials are shown, 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.

*2.7. Acoustic analyses*

Illustration of the stimuli are provided in Figure 1. Stimuli were passed into an auditory model as described in Agus et al. (2012), which consisted of a broadband pre-emphasis bandpass filter (0.4-8.5 kHz), a gammatone auditory filterbank, half-wave rectification, square root compression, and low-pass filtering at 100 Hz. Such a time-frequency representation, termed a "cochleagram", is intended to roughly mimic the information available after peripheral auditory processing.

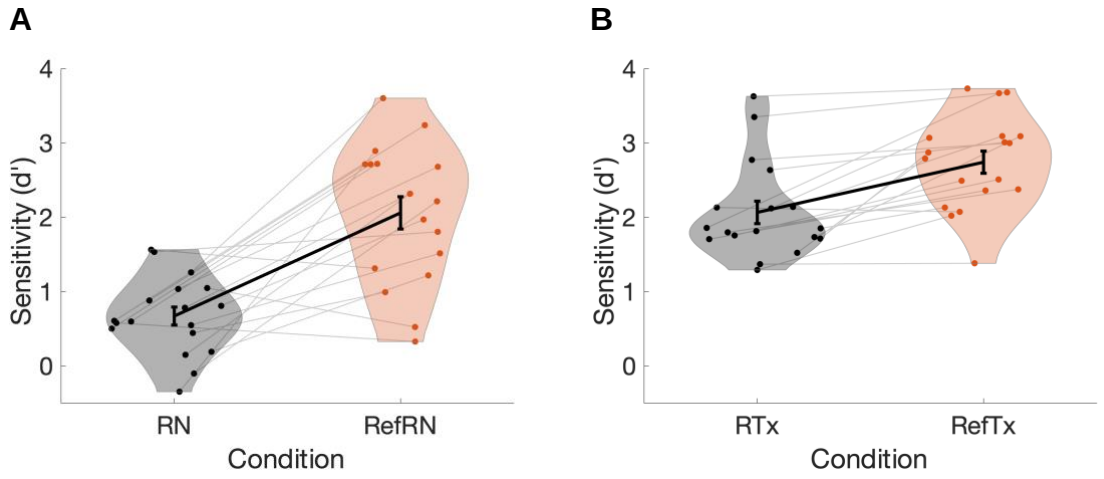


**Figure 2: Performance for 500-ms long exemplars**. **A). White noise.** Repetition detection performance is shown, expressed as the *d'* sensitivity index of signal detection theory. For the Repeated Noise condition (RN), the noise exemplar was novel in each trial. For the Reference Repeated Noise condition (RefRN), the same noise exemplar re-occured 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).

# 3. Results

## *3.1. Validation of the online testing procedure*

So far, all studies using variants of the "memory for noise" paradigm but one (Dauer et al., 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.

Figure 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.

Figure 2B shows the first set of results using natural textures, again for the 0.5-s duration, with performance for all three texture categories. For these relatively short-duration 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 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*

The main prediction of the summary statistics hypothesis is that texture exemplars should become more similar as their duration increases. Therefore, if performance on the memory for noise task relies on summary statistics, it should markedly decrease as texture duration increases.

With the durations tested here, this is not what was observed. Figure 3 shows the results for white noise and textures, as a function of exemplar duration (remember that 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 performance gain observed for RefRT was approximately constant across the whole range of durations. For textures, performance in the RTx condition was always good (average *d'* above 1), but had a band-pass shape with a peak at 250 ms. Importantly, again, the performance advantage for RefTx was observed throughout the whole range of durations.

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 ($F(1,17) = 44.81$, $p < 0.001$, $\eta^2_G = 0.32$) and Duration ($F(4,68) = 11.14$, $p < 0.001$, $\eta^2_G = 0.18$) were observed, without any interaction between the two factors ($F(4,68) = 1.42$, $p = 0.24$, $\eta^2_G = 0.03$). Similar findings were obtained with textures, with significant effects of Condition ($F(1,17) = 470.4$, $p < 0.001$, $\eta^2_G = 0.47$) and Duration ($F(2.53,46.0.6) = 14.28$, $p < 0.001$, $\eta^2_G = 0.23$). For textures, there was a significant interaction between Condition and Duration, with a small effect size ($F(3.28,55.68) = 5.00$, $p = 0.003$, $\eta^2_G =$

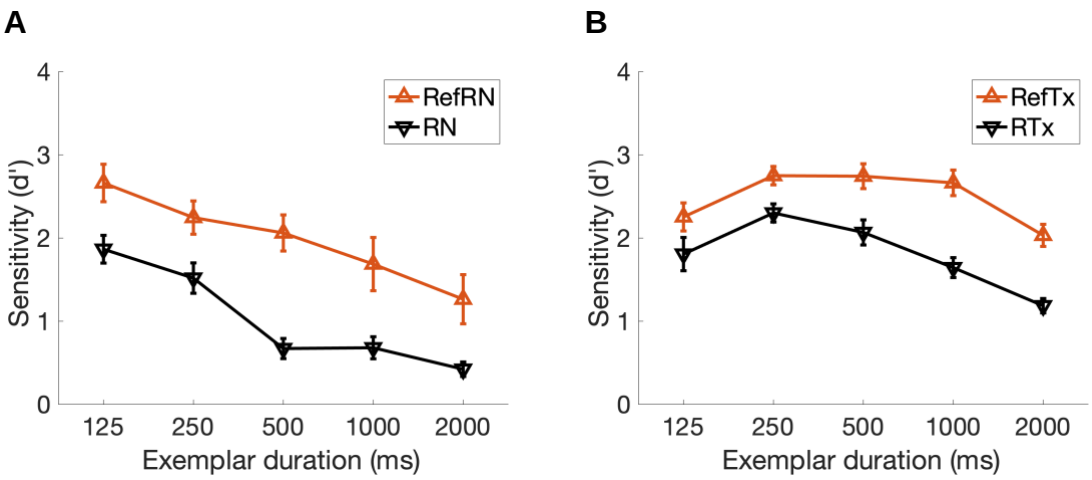


**Figure 3: Effect of exemplar duration**. **A). White noise.** Mean performance across participants is shown for the RN and RefRN conditions, as a function of exemplar duration. Error bars represent standard error about the mean. **B). Natural textures.** Performance averaged for the natural textures.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

0.09). Overall, the ANOVAs confirm that, while duration affected performance, the advantage provided by repeated exposure to a same exemplar was about constant for all 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 summary statistics hypothesis, namely the possible advantage of RefTx over RTx, reflecting learning of the texture exemplar. 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. 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 details 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

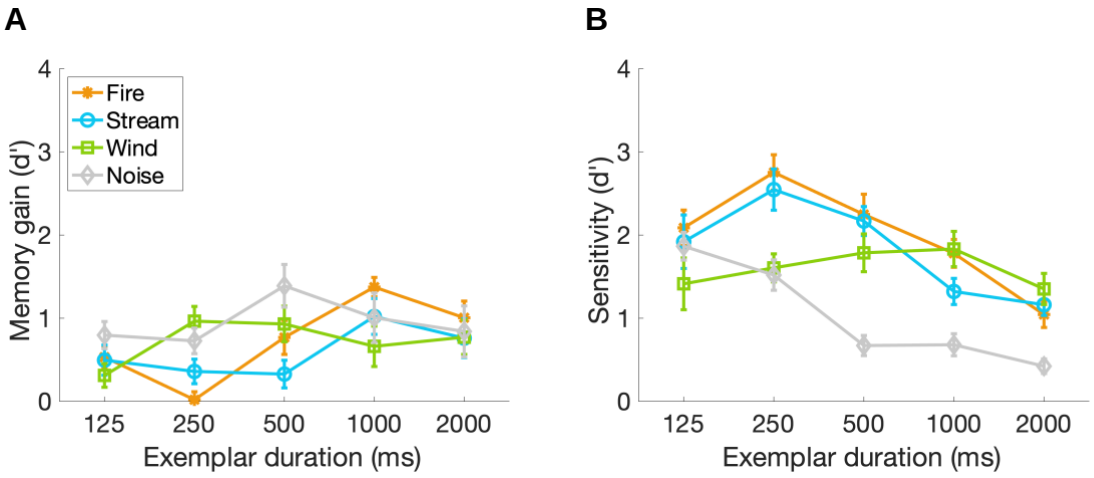


**Figure 4: Effect of sound categories**. **A).** Learning was assessed by computing the sensitivity increase due to repeated exposure. The memory gain was defined as performance for the trials where sound exemplars re-occurred across 20 trials (RefRN/RefTx) minus performance where sound exemplars were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

Condition * Texture * Duration, were also significant ($p$ = 0.003 or less), with relatively small effect sizes ($\eta^2$ = 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 a "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 Figure 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, which would both be predictions of the summary statistics for natural textures hypothesis.

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, although with a small effect size ($F(4,272) = 5.04$, $p < 0.001$, $\eta^2_G = 0.05$), suggesting that the memory gain was in fact larger for longer durations. Crucially, no effect of Sound category was found ($F(3,68) = 2.21$, $p = 0.10$, $\eta^2_G = 0.02$). The

Duration * Sound category was significant, with a medium effect size ($F(12,272) = 2.57$, $p = 0.003$, $\eta^2_G = 0.08$).

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

A mixed-design ANOVA was performed on the performance in the RN/RTx condition only, with factors "Duration" (5 levels, [125, 250, 500, 1000, 2000] ms), and "Sound category" (4 levels, White noise, Fire, Stream, Wind). Like for the memory gain analysis, an effect of Duration was observed ($F(4,272) = 24.14$, $p < 0.001$, $\eta^2_G = 0.19$). Unlike for the memory gain analysis, this time an effect of Sound category was observed ($F(3,68) = 17.68$, $p < 0.001$, $\eta^2_G = 0.21$). The Duration * Sound category was also significant ($F(12,272) = 3.88$, $p < 0.001$, $\eta^2_G = 0.10$).

A post-hoc comparison of all data points for Figure 4B was performed (Duration * Sound category interaction). Briefly, possibly because of large number of comparisons, the robustness of the pattern of performance seen in Figure 4B could not be 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 differences across sound categories, 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. However, all sound categories produced about the same amount of memory gain for all durations tested. Importantly, there was no difference between noise and natural textures for the memory gain. In particular, no clear advantage of noise, which produced the amount of memory gain expected from previous studies, was found over natural textures, which produced sizeable memory gains even at the longest durations tested.

# 4. Discussion

Combining ideas and techniques from two different lines of research, we applied a "memory for noise" paradigm to natural auditory textures. Two different predictions were made based on the putative representation of natural textures. If natural textures were only represented as time-averaged summary statistics, then repetition-detection and learning should be impossible for longer exemplar durations, because all texture exemplars should be represented by virtually identical summary statistics. However, if the rapid plasticity mechanisms induced by repeated exposure were able to complement the summary statistics representation by temporally-local features, unique to each exemplar, then results should be similar for white noise and natural textures.

The experimental findings strongly supported the second hypothesis. Results replicated previous findings for white noise, showing an effect of duration on baseline repetition detection performance but little to no effect on learning afforded by repeated exposure. Incidentally, this confirmed that the presumably subtle acoustic features required to memorize different noise exemplars were available in an online setting (Dauer et al., 2022). For natural textures, results were highly similar. Importantly, in spite of differences across sound categories in baseline immediate repetition-detection, a similar amount of learning with repeated exposure was observed for all sound categories tested.

### *4.1. Texture discrimination versus texture repetition detection*

It is useful at this point to provide more details about the experimental data that led to the original formulation of the summary statistics hypothesis (McDermott et al., 2013). The crucial finding was from McDermott et al.'s Experiment 2. Participants heard three different sounds in each trial. All 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. This decrease showed that texture exemplars were perceived as more and more similar with duration, and this was interpreted by remarking that summary statistics converged towards their mean value for the texture category as duration increased.

Here, we used the exact same texture generation algorithm and tested exemplar durations over a similar range. A difference between the two studies, however, is that in the original texture discrimination paradigm, there was no immediate repetition of texture exemplars (McDermott et al., 2013). All sounds within a trial were surrounded by silent gaps. Here, within-trial repetitions were seamless. Could this account for the different outcomes, as repetition regularity (Rajendran et al., 2016) and predictability (Dauer et al., 2022) have been shown to modulate noise learning? This is a possibility, but with caveats. First, because of the trial structure of the texture discrimination task, there was always a direct repeat of the texture exemplars (the odd-one-out was either the first or the last sound of a trial). Second, the exact value of the inter-stimulus interval made no difference to the findings of McDermott et al. (2013). Finally, noise learning can occur even with non-contiguous repetition (Andrillon et al., 2015; Bianco et al., 2020; Kaernbach, 2004; Ringer et al., 2023), although in these experiments the gaps were not silent. Noise learning has also been observed without immediate repetitions and intervening silences, albeit less reliably than when there are immediate repetitions (Agus & Pressnitzer, 2013).

Perhaps a more critical difference is that texture exemplars were never presented more than twice in the original texture discrimination paradigm, as new exemplars were used on each trial. Here, the reference texture exemplars re-occurred over 20 trials, randomly interspersed within an experimental block. We suggest that this is the main difference between the two paradigms: the amount of exposure to a given texture exemplar.

### *4.2. Repetition is special?*

Given the similarity between the learning patterns for noise and textures, as well as similarities with previous results with tone clouds of comparable spectro-temporal complexity (Agus & Pressnitzer, 2021), it appears that the memory processes induced by repetition generalized to natural textures.

There are several speculative arguments suggesting that repetition should be special for the auditory modality. From an ecological point of view, it is not obviously possible to actively search “back in time” for additional information once a sound has ended, unlike for visual search (Demany et al., 2010; Garnier-Allain et al., 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 predictive coding framework, events that repeat in the past may reasonably be attributed to a higher-than-baseline probability of repeating in the future, so repeated sounds should be expected to alter neural processing in such a framework (Baldeweg, 2006). Perhaps relatedly, repetitions have been shown to actually change the very perceptual qualities of sound, such as in the speech-to-song illusion (Deutsch et al., 2011), which, interestingly, also generalizes to natural textures (Rowland et al., 2019).

Thus, one interpretation of the present findings is that repetition "overcame" the default processing mode for textures. If the detailed acoustic information could be initially preserved for at least the duration of the exemplar, 2000ms in our case, then an immediate repeat could be detected and the "repetition mode" recruited. This could then result in complementing the well-established summary statistics representation of natural textures, used in texture class recognition, with temporally-local features, useful to memorize unique exemplars. Such a mechanism could be related to auditory scene analysis processes, creating foreground events emerging from the background texture when repetition is involved (McDermott et al., 2011). The possibility of different processing modes for natural textures was floated in the original studies (McDermott et al., 2013; Nelken & de Cheveigné, 2013). The present data provides experimental support for the idea.

### *4.3. Time scales of representation*

Another possible interpretation relies on a loose interpretation of the feature set versus summary statistics dichotomy. The core difference between the two possible 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 could be based on multiple or even flexible temporal time scales.

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 (Albouy et al., 2020; Joris et al., 2004; Norman-Haignere et al., 2022) findings. The details of the underlying 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), 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 for features/statistics may coexist in auditory representation of complex sounds.

Given such findings, the boundary between a relatively long-duration feature and a relatively short-duration statistics becomes blurry. 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. To come back to the present findings, when presented with repeating sounds, the auditory systems could trigger plasticity mechanisms that require a compact but distinctive representation of the sound exemplar being repeated. This representation could be viewed as a feature set, or as a collection of short summary statistics. Note that such an interpretation also applies to the original texture discrimination findings, where performance decreased with duration but, importantly, still remained above chance even for the longest textures. In this task, a single repeat of a texture exemplar was available, so it may have been able to trigger the memory for distinctive features/statistics in at least in a subset of trials.

## 5. 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 textures are not immune; or a convergence of the feature set versus summary statistics description of sound representations, if a continuum of time scales is considered.

To conclude, 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 useful to generate sparse representations of meaningful sounds (Gervain & Geffen, 2019; Wang et al., 2020). We speculate that, more generally, it is the sign of the auditory system adapting its very processing to the statistical peculiarities of its environment, even before the subsequent inferential processes required to recognize natural sound categories.

## Appendix A. Supplemental statistical reporting.

**Table A1.** Results of the main mixed-design ANOVA on the full dataset.

**Within Subjects Effects**

| Cases | Sphericity Correction | Sum of Squares | df | Mean Square | F | p | $\eta^2_G$ |
|---|---|---|---|---|---|---|---|
| Condition | None | 102.212 | 1.000 | 102.212 | 228.180 | < .001 | 0.260 |
| Condition ✻ Category | None | 2.968 | 3.000 | 0.989 | 2.209 | 0.095 | 0.010 |
| Residuals | None | 30.460 | 68.000 | 0.448 | | | |
| Duration | None | 31.668 [a] | 4.000 [a] | 7.917 [a] | 21.986 [a] | < .001 [a] | 0.098 |
| | Greenhouse-Geisser | 31.668 | 3.442 | 9.201 | 21.986 | < .001 | 0.098 |
| Duration ✻ Category | None | 16.000 [a] | 12.000 [a] | 1.333 [a] | 3.703 [a] | < .001 [a] | 0.052 |
| | Greenhouse-Geisser | 16.000 | 10.325 | 1.550 | 3.703 | < .001 | 0.052 |
| Residuals | None | 97.944 | 272.000 | 0.360 | | | |
| | Greenhouse-Geisser | 97.944 | 234.032 | 0.419 | | | |
| Condition ✻ Duration | None | 6.868 | 4.000 | 1.717 | 5.043 | < .001 | 0.023 |
| | Greenhouse-Geisser | 6.868 | 3.612 | 1.901 | 5.043 | 0.001 | 0.023 |
| Condition ✻ Duration ✻ Category | None | 10.565 | 12.000 | 0.880 | 2.586 | 0.003 | 0.035 |
| | Greenhouse-Geisser | 10.565 | 10.837 | 0.975 | 2.586 | 0.004 | 0.035 |
| Residuals | None | 92.599 | 272.000 | 0.340 | | | |
| | Greenhouse-Geisser | 92.599 | 245.634 | 0.377 | | | |

*Note.* Sphericity corrections not available for factors with 2 levels.

*Note.* Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

**Between Subjects Effects**

| Cases | Sum of Squares | df | Mean Square | F | p | $\eta^2_G$ |
|---|---|---|---|---|---|---|
| Category | 45.467 | 3 | 15.156 | 14.834 | < .001 | 0.135 |
| Residuals | 69.477 | 68 | 1.022 | | | |

*Note.* Type III Sum of Squares

**Assumption Checks**

**Test of Sphericity**

| | Mauchly's W | Approx. $X^2$ | df | p-value | Greenhouse-Geisser ε | Huynh-Feldt ε | Lower Bound ε |
|---|---|---|---|---|---|---|---|
| Duration | 0.761 | 18.169 | 9 | 0.033 | 0.860 | 0.912 | 0.250 |
| Condition ✻ Duration | 0.796 | 15.187 | 9 | 0.086 | 0.903 | 0.960 | 0.250 |

**Table A2.** Selected post-hoc comparisons for Figure 4B, showing the comparisons of peak values to their neighbors for the natural textures. Due to the high number of comparisons, the lenient Holm correction was applied.

**Post Hoc Comparisons - Texture ✻ Duration**

| | | Mean Difference | SE | t | $p_{holm}$ |
|---|---|---|---|---|---|
| Fire, 125 | Fire, 250 | -0.477 | 0.187 | -2.552 | 1.000 |
| Stream, 125 | Stream, 250 | -0.161 | 0.187 | -0.861 | 1.000 |
| Fire, 250 | Fire, 500 | 0.572 | 0.187 | 3.062 | 0.281 |
| Stream, 250 | Stream, 500 | 0.214 | 0.187 | 1.148 | 1.000 |
| Wind, 500 | Wind, 1000 | -0.026 | 0.187 | -0.139 | 1.000 |
| Wind, 1000 | Wind, 2000 | 0.190 | 0.187 | 1.019 | 1.000 |

*Note.* P-value adjusted for comparing a family of 190

## CReDIT authorship

Berfin Bastug: Conceptualization; Formal Analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft; Writing – review & editing.

Vani Rajendran: Conceptualization; Investigation; Methodology; Supervision; Writing – original draft; Writing – review & editing.

Roberta Bianco: Conceptualization; Methodology; Writing – review & editing.

Trevor Agus: Formal Analysis; Writing – review & editing.

Maria Chait: Conceptualization; Methodology; Writing – review & editing

Daniel Pressnitzer: Conceptualization; Formal Analysis; Methodology; Supervision; Visualization; Writing – original draft; Writing – review & editing.

## 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- 0023-01.

## 6. 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), 4124–10. 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

Albouy, P., Benjamin, L., Morillon, B. & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*(6481), 1043–1047. https://doi.org/10.1126/science.aaz3468

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

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*, 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*, e56073. https://doi.org/10.7554/elife.56073

Bianco, R., Mills, G., Kerangal, M. de, Rosen, S. & Chait, M. (2021). Reward Enhances Online Participants' Engagement With a Demanding Auditory Task. *Trends in Hearing*, *25*, 23312165211025940. https://doi.org/10.1177/23312165211025941

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

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), 610–610. https://doi.org/10.1121/1.1918551

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

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

Lange, F. P. de, Heilbron, M. & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

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

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., Feldstein, N. A., McKhann, G. M., Schevon, C. A., Flinker, A. & 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

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A. & Whitney, D. (2012). Humans Use Summary Statistics to Perceive Auditory Sequences. *Psychological Science*, *24*(8), 1389–1397. https://doi.org/10.1177/0956797612473759

Rajendran, V. G., Harper, N. S., Abdel-Latif, K. H. A. & Schnupp, J. W. H. (2016). Rhythm Facilitates the Detection of Repeating Sound Patterns. *Frontiers in Neuroscience*, *10*(60), 464–467. https://doi.org/10.3389/fnins.2016.00009

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), 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

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

Traer, J., Norman-Haignere, S. V. & McDermott, J. H. (2021). Causal inference in environmental sound recognition. *Cognition*, *214*, 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

597