# Auditory memory for random time patterns

HiJee Kang, Trevor R. Agus, and Daniel Pressnitzer

Citation: The Journal of the Acoustical Society of America **142**, 2219 (2017); doi: 10.1121/1.5007730
View online: http://dx.doi.org/10.1121/1.5007730
View Table of Contents: http://asa.scitation.org/toc/jas/142/4
Published by the Acoustical Society of America

# Auditory memory for random time patterns

HiJee Kang,[a] Trevor R. Agus,[b] and Daniel Pressnitzer
*Laboratoire des Systèmes Perceptifs, Département d'études cognitives, École Normale Supérieure, PSL Research University, Centre National de la Recherche Scientifique, 29 Rue d'Ulm, 75005 Paris, France*

(Received 25 May 2017; revised 20 September 2017; accepted 29 September 2017; published online 20 October 2017)

The acquisition of auditory memory for temporal patterns was investigated. The temporal patterns were random sequences of irregularly spaced clicks. Participants performed a task previously used to study auditory memory for noise [Agus, Thorpe, and Pressnitzer (2010). Neuron **66**, 610–618]. The memory for temporal patterns displayed strong similarities with the memory for noise: temporal patterns were learnt rapidly, in an unsupervised manner, and could be distinguished from statistically matched patterns after learning. There was, however, a qualitative difference from the memory for noise. For temporal patterns, no memory transfer was observed after time reversals, showing that both the time intervals and their order were represented in memory. Remarkably, learning was observed over a broad range of time scales, which encompassed rhythm-like and buzz-like temporal patterns. Temporal patterns present specific challenges to the neural mechanisms of plasticity, because the information to be learnt is distributed over time. Nevertheless, the present data show that the acquisition of novel auditory memories can be as efficient for temporal patterns as for sounds containing additional spectral and spectro-temporal cues, such as noise. This suggests that the rapid formation of memory traces may be a general by-product of repeated auditory exposure. © *2017 Acoustical Society of America.*

https://doi.org/10.1121/1.5007730

[JJL] Pages: 2219–2232

## I. INTRODUCTION

Listening often requires matching current sensory evidence with stored memory traces, such as when associating a series of artificial tones to one's new phone (Roye *et al.*, 2010), or recognizing natural sounds like voices and musical instruments (Agus *et al.*, 2012; Norman-Haignere *et al.*, 2015). In fact, as auditory information develops over time, most aspects of auditory perception are likely to be shaped by memory processes on different time scales (Demany and Semal, 2007). How such memories are acquired through experience is not well understood. Here, we investigated auditory memory for temporal patterns, using random sequences of irregularly spaced clicks, over a broad range of time scales regarding the time-intervals forming the patterns.

In previous studies probing the acquisition of novel auditory memories, another type of random stimulus has been extensively used—white noise, obtained by drawing successive audio samples from a Gaussian distribution (Agus *et al.*, 2010; Andrillon *et al.*, 2015; Goossens *et al.*, 2008; Guttman and Julesz, 1963; Kaernbach, 2004; Luo *et al.*, 2013; Rajendran *et al.*, 2016; Viswanathan *et al.*, 2016; Warren *et al.*, 2001). Studies of the memory for noise differed from each other in many experimental details, but, essentially, an exemplar of noise selected at random was presented to listeners more than once during an experiment. Putative memory traces were then probed by tracking psychophysical or physiological measures for the re-occurring noise exemplar, as exposure increased, either within a trial or across trials. These studies have demonstrated remarkable features of auditory memory. Immediate auditory memory for noise, often probed by asking participants to detect ongoing repetitions of a "frozen" noise exemplar, can extend up to noise durations of several seconds (Guttman and Julesz, 1963; Kaernbach, 2004; Warren *et al.*, 2001). Longer-term memory traces can be measured by presenting the same exemplar over different trials or experimental blocks, and in this case memory be established within tens of seconds and persist for minutes or even weeks (Agus *et al.*, 2010; Viswanathan, 2016; Viswanathan *et al.*, 2016). Real-time physiological correlates of the memory traces can be observed in electroencephalography or magnetoencephalography, with a response timing and topography suggesting an origin within auditory cortex (Andrillon *et al.*, 2015; Luo *et al.*, 2013). Finally, these physiological correlates can be formed incidentally, even with a task diverting attention away from the noises to be remembered (Andrillon *et al.*, 2015), or during REM sleep (Andrillon *et al.*, 2017).

A puzzling question has been raised by most of these studies: what had been learnt by listeners when they displayed evidence of a memory for noise? Answering this question has implications for whether the memory for noise is representative of the memory for sounds in general. It is implausible that listeners learnt the exact time series of the thousands of audio samples comprising the noise waveform, first because of the sheer amount of information involved, and second because audio waveforms are transformed as

---

[a] Electronic mail: hijee.kang@ens.fr
[b] Also at: Sonic Arts Research Centre, School of Arts, English and Languages, Queen's University Belfast, University Road, Belfast, BT7 1NN, UK.

they ascend the auditory system so the central nervous system does not have access to the audio samples themselves. Rather, a memory for noise should rely on a subset of auditory cues emerging after processing along the auditory pathways. Such cues can belong to two broad classes: spectral cues, like the random distribution of frequencies over a short time frame, or temporal cues, like the occurrence of salient "events" at random times during the stimulus. Even though the opposition between spectral and temporal cues is somewhat arbitrary at the audio signal level, being essentially imposed by the time frame of analysis, the distinction is meaningful and important from an auditory perspective—and even more so from an auditory *memory* perspective. To simplify, spectral cues may be encoded by the place pattern of spiking rates in frequency-tuned populations of auditory neurons, whereas temporal cues may be encoded by the timing between spikes within a neuron or across neurons in a population. Even though the mapping between acoustic cues and their underlying neural representation is not well-established (Cariani and Delgutte, 1996; Lim *et al.*, 2016; Shamma and Lorenzi, 2013), both place (Rothschild *et al.*, 2010) and temporal (Lu *et al.*, 2001; Petkov and Bendor, 2016) cues are abundant along the auditory pathways, from the auditory periphery up to at least the auditory cortex. However, the two types of cue require qualitatively different mechanisms when learning and neural plasticity are considered. Spectral cues in a place code are available as simultaneously active neurons, so the classic "fire together, wire together" principle of neural plasticity can apply (Klampfl and Maass, 2013; Masquelier *et al.*, 2008). For temporal cues, however, such a mechanism does not apply directly because what needs to be learnt develops over time. So, neurons would need somehow to encode past and present information simultaneously to learn temporal patterns (Karmarkar and Buonomano, 2007; Lim *et al.*, 2016).

For noise, the evidence available so far points strongly towards a memory representation based on brief spectral cues. In their original report, Guttmann and Julesz (1963) showed that listeners could detect repeating copies of a noise segment even for periodicities outside of the pitch range. They went on to describe the perceptual cues used to detect repetitions with noise: a series of brief events, such as "rasps" and "clangs," which led them to the following statement: "*It appears to us that the basis of pitchless periodicity detection must lie in the detection of short-term power-spectrum recurrence*" (Guttman and Julesz, 1963). This conjecture, based on introspection, was confirmed experimentally by Kaernbach (1993), who used a technique based on participants tapping to a repeated noise exemplar. The replacement of various portions of the memorized noise exemplar with fresh noise showed that repetition-detection was based on short segments, lasting less than 100 ms. Further experiments using behavioral reverse-correlation suggested that, while some of those short features could be described as spectro-temporal, most were spectral edges (Kaernbach, 1999, 2000, 2017). More recently, when studying longer-term memory for noise, Agus *et al.* (2010) showed that the memory representation was robust to the time reversal of the noise segment. Time-reversal would leave short-term power spectra unaffected, but would disrupt most cues based on timing. Viswanathan *et al.* (2016) reported similar conclusions using time shuffling instead of time reversal. Finally, Andrillon *et al.* (2015) observed sizeable EEG event-related potentials after learning. As event-related potentials require a precisely time-locked neural response, this finding was accounted for by positing neural responses locked to brief spectral features within the noises after learning.

Considering all of these pieces of indirect but converging evidence, together with the well-established availability of neural mechanisms to learn spectral cues, it is legitimate to ask whether the remarkable characteristics of auditory memory observed with noise are restricted to sounds containing spectral cues. Here we aimed to answer this question directly by using random time patterns devoid of spectral cues.

The perception of auditory time patterns has been studied extensively, mostly using discrimination experiments. The simplest possible temporal pattern is a single time interval of variable duration. The discrimination between two such intervals yields thresholds corresponding to a Weber fraction, of about 10% for intervals in the range of about 200 ms to 2 s—although this range may vary with experimental details (Grondin, 2010; Teki and Griffiths, 2014). When the time interval to be discriminated was embedded within longer, isochronic patterns, or relatively simple patterns containing only two different interval durations, the Weber law held (Hirsh *et al.*, 1990). However, when the embedding patterns were more complex, other listening strategies appeared. Using a discrimination task between two fully random temporal patterns, Sorkin (1990) showed that behavioral results could be accounted for by a model based on the cross-correlation between time intervals, each interval being corrupted by additive noise (which would be inconsistent with a Weber fraction). When the embedding pseudo-random patterns were manipulated to induce a musical meter of varying strength (Povel and Essens, 1985), the detailed structure of the sequence influenced discrimination performance, demonstrating that not all intervals had the same perceptual weight (Ross and Houtsma, 1994; Teki and Griffiths, 2014). Finally, in yet another departure from the Weber law, when the interval to be discriminated was preceded by another, random time interval, this systematically distorted the perceived duration of the target interval (Karmarkar and Buonomano, 2007).

A comprehensive theory accounting for all of these findings about temporal pattern discrimination is still lacking, but, as outlined by Ross and Houtsma (1994), it may be useful to consider at least three types of strategy to discriminate complex time patterns: an analytic strategy, focusing on each interval independently (Hirsh *et al.*, 1990); a holistic strategy, considering the whole pattern (Sorkin, 1990); and an emphasis on a small subset of pattern motifs (sometimes termed "runs" as they occur in time), which may stand out because of meter (Povel and Essens, 1985), perceptual organization (Royer and Garner, 1966), or informational content (Pollack, 1968a). Similar strategies could be relevant for the memory representation of time patterns, but this has not been investigated yet.

In the present set of experiments, we used irregularly spaced click trains in a memory paradigm similar to what has been used previously for noise (Agus *et al.,* 2010). The intervals between clicks were randomly derived from a uniform (experiment 1) or Poisson distribution (experiment 2), to evaluate the influence of interval statistics on performance. A minimum gap between clicks was always imposed, as well as high-pass filtering, to prevent any overlap in the internal representation of successive clicks and thus restrict auditory cues to purely timing cues (Methods, see also Fig. 1). One condition consisted of a random series of clicks, lasting 2 s in total, which we termed *C* (Fig. 1, top). A second condition, termed repeated clicks or *RC*, consisted of a random series of clicks lasting 1 s, immediately repeated for another 1 s (Fig. 1, bottom). The task assigned to participants for all experiments was to report whether a trial contained a repetition or not. There was a third condition, not mentioned to participants during the instruction phase: a Reference RC, or *RefRC*, which was an RC that re-occurred over several trials throughout an experimental block. As we showed previously, longer-term memory can be probed by comparing performance on RCs, heard in one and only one trial, to performance on RefRCs, heard identically over several trials (Agus *et al.,* 2010; Agus and Pressnitzer, 2013). Improved performance for RefRCs compared to RCs can be taken as an indication of longer-term memory traces. In experiment 3, we tested if time-reversal would affect performance after learning, as a memory representation based on local cues such as extreme values of the inter-click intervals could be resilient to time-reversal. Finally, we used Sorkin's cross-correlation model (Sorkin, 1990) to see if it could help interpret some characteristics of repetition-detection (RC performance) for our click trains.

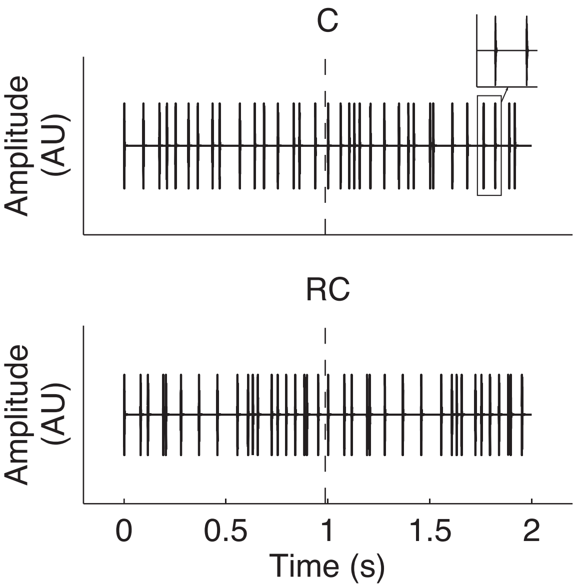

FIG. 1. Illustration of the random auditory time patterns used in all experiments. Stimuli were 2 kHz high-pass filtered click trains, combined with low-pass noise, with a minimum time gap of 10 ms imposed between clicks. The condition illustrated corresponds to a uniform distribution between intervals and an 18-Hz average click rate, with the low-pass noise omitted for clarity. The output of a single auditory filter (gammatone) centered at 4 kHz is represented. The vertical dashed lines indicate the midpoint of the stimulus. On the top panel, a random pattern is shown (condition C, see text). On the bottom panel, the first and second halves of the stimulus are the same random pattern (conditions RC or RefRC). The inset on the top panel illustrates that the minimum gap between clicks ensured that there was no overlap between clicks, even after auditory filtering.

Unlike previous studies of auditory temporal patterns, we probed both the immediate memory processes needed for discrimination, reflected by the performance on RC trials, and longer-term memory processes over the course of minutes reflected by a possible advantage of RefRC trials over RC trials. The experimental design precisely controlled for the amount of exposure to RefRC, as participants would not have heard the exact same random time pattern before, so we could characterize the time-course of memory acquisition for temporal patterns. Finally, we could use the same psychophysical task to test a broad range of time scales (inter-click intervals), covering sequences sounding either like a succession of isolated clicks or sequences sounding like continuous buzzing sounds.

## II. EXPERIMENT 1: UNIFORMLY DISTRIBUTED TIME PATTERNS

### A. Method

#### *1. Participants*

Fourteen participants were tested (age in year $M = 25$, $SD = 4$; 6 female). All had normal hearing, as established by an audiogram administered before the experiment [20 dB hearing level (HL) or less for all tested frequencies of 125, 250, 500, 1000, 2000, 4000, and 800 Hz]. Participants gave informed consent and were paid 10 euros per hour. The study was approved by the ethical committee of U. Paris Descartes, France (CERES, IRB: 20154000001072).

#### *2. Stimuli*

Click trains were generated in the digital domain at a sampling rate of 44.1 kHz. Each click was a rectangular pulse of two samples (about 50 $\mu$s). Inter-click intervals (ICIs) were drawn from a uniform distribution, whose lower boundary was maintained constant at 10 ms for all experiments. The upper boundary of the ICI distribution was an experimental parameter. Five logarithmically spaced values were tested: 25, 50, 100, 200, and 400 ms. When taking account of the upper and lower boundaries, the five experimental conditions corresponded to average click rates of 57, 33, 18, 10, and 5 Hz. For consistency with experiment 2, average click rate is used to identify conditions.

Auditory filter ringing may introduce overlap in the excitation caused by successive clicks, and thus spectral cues that can be potentially be used to discriminate stochastic click trains (Pollack, 1968b). To prevent the appearance of any such cues, all click trains were high-pass filtered using an 8th order Butterworth filter with a 2 kHz filter cutoff. The cut-off value was chosen as the 20th potential "harmonic" of the minimum ICI of 10 ms.

The nominal duration of each click train was 2 s. To form the first half of the stimulus, a first click was set at time 0 and successive ICIs were randomly drawn from a uniform distribution until the nominal duration of 1 s was surpassed, and then the stimulus was truncated to 1 s minus the minimum gap of 10 ms. In the C condition, the procedure was repeated for the second half of the stimulus, with a fresh random draw so as to obtain a fully random click train lasting 2 s (Fig. 1, top). In the RC condition, the first half of the stimulus was repeated identically to obtain a repeated click train lasting 2 s (Fig. 1, bottom). For C and RC, the random ICIs were generated anew for each trial. For RefRC, the same generation procedure as RC was adopted, but, additionally, ICIs were fixed for all trials within an experimental block. Note that the truncation at 1 s caused a slight difference in the long-term distribution of the last ICI of the first half of the stimulus compared to all other ICIs. However, such a difference was only apparent on the long-term average of ICIs (each individual draw was compatible with the target distribution) and, in any case, it was present for all conditions.

Low-pass pink noise was finally added to the stimuli, to mask any possible distortion and to obtain stimuli lasting exactly the same duration. The noise pedestal was generated in the spectral domain with components between 50 Hz and 2 kHz. It started 200 ms before the click train and finished 200 ms after its nominal ending, including 50 ms raised cosine fade in and fade out. The root-mean square (rms) level of the pink noise was set at $-20$ dB relative to the click train rms. The full stimuli were presented at an overall level of 65 dB sound pressure level (SPL), A-weighted. As overall level was maintained constant, this means that peak level varied across (but not within) average rate conditions.

### *3. Apparatus and procedure*

Participants were tested individually in a double-walled sound-insulated booth (Industrial Acoustics). Stimuli were played diotically through an RME Fireface UC soundcard, at a sampling rate of 44.1 kHz and 16-bit resolution. They were presented over Sennheiser HD 600 headphones. Participants were instructed to report within-trial repetitions. They responded through a computer keyboard in a self-paced manner.

Prior to data collection, a training period was included. Training was intended to help participants discover the types of perceptual cues elicited by repeated click trains. We modeled this training period on what had been previously used for repeated noise experiments (Agus *et al.*, 2014). Training started with ten repeats of 1-s long random click trains, a task in which informal listening indicated that repetition detection should be easy. The click trains were generated as Cs and RCs in the main experiment, with an average click rate of 12 Hz. This rate was chosen to be approximately in the middle of the experimental parameter range, but different from any subsequent experimental parameter. Participants received visual feedback on the repetition detection task during training. The first training block presented ten trials, half of them C and the other half RC. Then, training blocks were run with 4, then 3, then 2 repeats only, for 40 trials in each training block. Note that there were no RefRCs in any of these training blocks.

When data collection began, no feedback was further provided. The instructions to participants remained the same. In particular, at no time were they instructed about the possible re-occurrence of a sound (RefRC) during the block, nor about our aim to test memory. In each experimental block, the experimental parameter (average click rate) was fixed and there were 40 C, 20 RC, and 20 RefRC trials in pseudo-random order. The only constraint on order was that RefRCs could not appear on successive trials (Agus *et al.*, 2010). Participants completed four blocks for each of the five click-rate conditions. Blocks for a given parameter were run in succession, but the order of parameters was random and counterbalanced across subjects. Training and testing was completed over two experimental sessions, each lasting about 2 h and run on consecutive days.

### *4. Data analysis*

The pattern of "yes" and "no" responses was first converted to hit rate (proportion of "yes" responses for RC or RefRC) and false alarms ("yes" response for C). The sensitivity index $d'$ from signal detection theory was then computed for each participant and RC or RefRC conditions (sharing false alarms for a given experimental block). Statistical testing was performed on $d'$ with repeated-measures analyses of variance (ANOVAs), with trial type (RefRC or RC) and average click rate (5, 10, 18, 33, 57 Hz) as within-subject factors and participants as random factors. We verified that the data conformed to the assumptions required for ANOVA and applied Greenhouse-Geisser corrections when necessary. Along with the $F$ statistics and $p$-values, with a significant level set at $p < 0.05$, we also report effect size with the generalized eta-squared measure, $\eta^2_g$ (Lakens, 2013). *Post hoc* $t$-tests were conducted, corrected for multiple comparisons by using the Holm-Bonferroni method as necessary.

When analyzing the time course of performance, responses were converted to hit rate and aggregated in trial bins over all blocks and participants. The first bin represents the first time participants heard, e.g., a RefRC during a block, but not necessarily the first trial of the block and not necessarily the same trial number for all participants. These average values were then fitted with either a flat line corresponding to the average performance (1 parameter) or exponential functions (three parameters) using a least-squared method. The goodness of fit was estimated by penalizing models with extra parameters (Motulsky and Christopoulos, 2004). The choice of model then led us to conclude whether the performance of RC and RefRC evolved over time, perhaps because of learning, or stayed statistically constant over the duration of the blocks (Agus *et al.*, 2010).

## B. Results

The top panel of Fig. 2 displays the average performance across participants as the sensitivity index $d'$, for RC and RefRC trials, and over the five values of click rates tested. A first feature of the results is that performance generally decreased with increasing click rate, from $d'$ values around 2 at the lowest rate of 5 Hz, to $d'$ values below 1 at

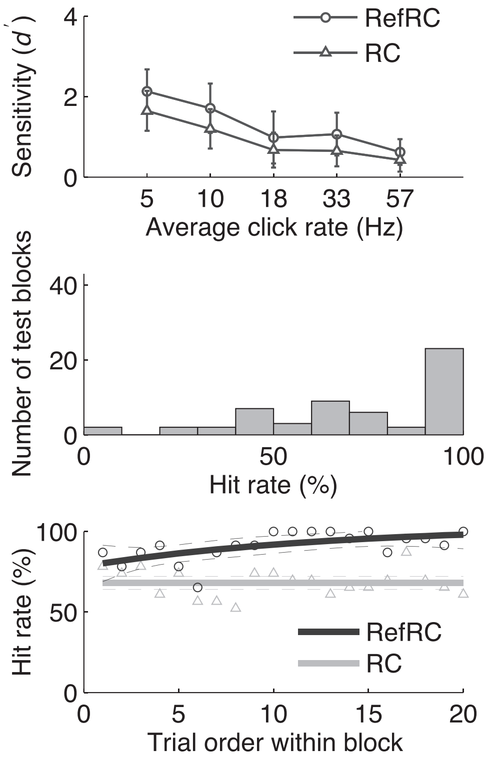

FIG. 2. Results for experiment 1, Uniform distribution. Top panel: The sensitivity index $d'$ is shown for each stimulus type and average click rate conditions, averaged across listeners. Error bars are 95% confidence intervals about the mean. Middle panel: Histogram of the proportion of hit rates for the last ten trials of each block in the RefRC condition, over all blocks for all listeners. Here, only the condition with 18 Hz average rate is shown, but all conditions displayed similar results. Lower panel: The time course of hit-rate for all blocks with 90% or more hit-rate over the last ten trials of a block. Data points are average values over blocks and listeners for each trial order. The solid lines are best-fit exponentials for each stimulus condition, with dashed lines indicating 95% confidence interval. Again, only the 18 Hz condition is illustrated.

the fastest rate of 57 Hz. Stimulus duration was kept constant, so increasing rates also corresponded to more clicks and thus more time intervals in the patterns. A second feature is that performance was generally higher for RefRC than RC, with an advantage of about 0.5 $d'$ units for all rates.

A repeated-measures ANOVA on $d'$ (see Sec. II A) confirmed that both trial type [$F(1, 13) = 41.56$, $p < 0.001$, $\eta^2_g = 0.05$] and click rate [$F(4, 52) = 25.29$, $p < 0.001$, $\eta^2_g = 0.27$] had a reliable effect on performance. No interaction reached our significance criterion (all $p > 0.05$). *Post hoc* comparisons confirmed that the advantage of RefRC over RC was present for all rates [5 Hz: $t(13) = -3.10$, $p < 0.01$, 10 Hz: $t(13) = -3.86$, $p < 0.01$, 18 Hz: $t(13) = -2.22$, $p < 0.05$, 33 Hz: $t(13) = -3.45$, $p < 0.01$, 57 Hz: $t(13) = -2.84$, $p < 0.01$; reported $p$-values are uncorrected but all $p < 0.05$ after Holm-Bonferroni correction].

A superior performance for RefRC over RC could indicate learning of RefRC thanks to increasing exposure to the reference click train as the block progressed. But it could also simply be the chance result of the particular set of randomly chosen RefRCs being easier to detect from the outset. To test for this possibility, we examined the changes in performance over trials throughout the block. Previous experiments with white noise showed that not all blocks produced learning (Agus *et al.*, 2010). Moreover, RefRC performance was associated to higher variance than RC in Fig. 2, which would also suggest an additional source of variability for this condition. Therefore, we first tested for the presence of "good" blocks and "bad" blocks in the present data. "Good" blocks were defined as blocks for which performance was high for RefRCs at the end of the block. The middle panel of Fig. 2 shows the histogram of hit rates for RefRCs over the last ten presentations of RefRC during the blocks, illustrated here for the 18 Hz rate condition. A one-sample Kolmogorov-Smirnov test confirmed that hit rates did not follow a normal distribution [$D(56) = 0.55$, $p < 0.001$]. This suggests that two types of blocks contributed to the histogram: "bad" blocks, with hits around 50%, and "good" blocks, with hits above 90%. We then selected those "good" blocks (41% of blocks in the 18 Hz rate condition) and computed performance for RC and RefRC for each trial position within a block. The result is shown in the lower panel of Fig. 2. Performance was not constant for RefRC: from the beginning to the end of the block, performance increased. In contrast, for RC there was a no change in performance.

The same qualitative description of the data was verified in all other conditions, and tested statistically. We first tested for the presence of good and bad blocks: all distributions deviated from normal [one-sample Kolmogorov-Smirnov test, 5 Hz: $D(56) = 0.95$; 10 Hz: $D(56) = 0.55$; 18 Hz: $D(56) = 0.55$; 33 Hz: $D(56) = 0.58$; 57 Hz: $D(56) = 1$, all $p < 0.001$]. When applying the criterion of selecting blocks with hit rates above 90% for the last ten presentations of RefRC, the proportion of good blocks was 64%, 55%, 41%, 39%, and 23% for rate conditions of 5, 10, 18, 33, and 57 Hz, respectively. We next tested whether performance changed over the course of an experimental block, for RC and RefRC trials, by fitting exponential functions testing whether this explained the data better than a flat line at the mean hit rate (see Sec. II A). This test confirmed that performance increased over the course of a block for RefRC, for all click rate conditions except 10 Hz [5 Hz: $F(2, 17) = 12.70$, $p < 0.001$, 10 Hz: $F(2, 17) = 2.18$, $p = 0.143$, 18 Hz: $F(2, 17) = 4.73$, $p < 0.05$, 33 Hz: $F(2, 17) = 11.85$, $p < 0.001$, 57 Hz: $F(2, 17) = 8.19$, $p < 0.01$]. The same test suggested that performance for RC did not change over the course of a block for any of the click rates (all $p > 0.05$). RefRCs regarded as "bad blocks" also showed no performance change of the course of a block in any of the click rate conditions (all $p > 0.05$).

### C. Interim discussion

The results obtained so far are similar to what had been observed using white noise in the same experimental paradigm (Agus *et al.*, 2010). Across-trial re-occurrence of a stochastic time pattern, defined by a series of clicks with random ICIs, led to improved performance on a within-trial repetition detection task.

This was not observed for all experimental blocks, but only on between 23% and 64% of all blocks, depending on the baseline difficulty of the within-trial repetition detection task. These numbers bracket what was observed with naive listeners and 0.5 s-long white noise, where the proportion of good blocks was about a third. It is possible that the actual click train characteristics systematically differed between good and bad blocks, even though their generative parameters were the same. To test for this possibility, we compared between good and bad blocks the average ICIs, the standard deviation of ICIs, and the presence of exceptionally long or short ICIs. The analysis is detailed in the Appendix. Overall, we did not find any systematic difference between good and bad blocks for these acoustic characteristics. As each RefRC was unique to a block and a participant, it is not possible to ascertain whether good blocks were caused by intrinsically easier RefRC patterns, with acoustic characteristics remaining to be elucidated, or if good and bad blocks rather resulted from an idiosyncratic interaction between pattern and participant.

The rapid time-course of learning was also similar with Agus *et al.* (2010), as 90% performance or better was reached after only ten presentations of the RefRC sequence. Looking at the time course of the good blocks, we could rule out chance selection of easier stimuli for RefRC, but rather attribute the RefRC advantage to true learning. Thus, stochastic time patterns can rapidly form memory traces with increasing exposure.

The observed effect of click rate, a parameter that could not be tested with white noise, was perhaps in part predictable: slower rates, associated with clicks that could easily be heard out individually, produced better performance compared to faster rates. However, the advantage of longer-term learning in the case of RefRC was always of about 0.5 of a $d'$ unit, with no interaction between this RefRC advantage and click rate. This suggests that even though faster rates led to poorer baseline performance and also less good blocks, the benefit of longer-term learning for RefRC remained about the same.

It is possible that the other aspects of the ICI patterns besides mean rate could have impacted performance. For instance, as we kept stimulus duration constant, faster rates led to more clicks and thus to more time intervals to be learnt. Even though it is intuitively appealing that more intervals should induce poorer performance, this has not always been observed in discrimination tasks. When the variance of ICIs is low, for instance, performance actually improves with number of intervals (Pollack, 1968a). We therefore repeated the experiment with a different random distribution of ICIs to test for the influence of additional ICI statistics besides the average click rate.

## III. EXPERIMENT 2: POISSON-DISTRIBUTED TIME PATTERNS

### A. Method

#### 1. Participants

Participants were 14 normal-hearing listeners, who had not participated in experiment 1 (age in year $M = 24$, $SD = 3$; 11 female). They were recruited as in experiment 1, and their audiogram was confirmed as normal using the same procedure.

#### 2. Stimuli

Stimuli were click trains generated as for experiment 1, with a single difference: the ICIs were now drawn from a Poisson distribution, instead of a uniform distribution. We imposed a refractory time of 10 ms on the Poisson distribution, and adjusted its rate parameter to have the same average click rate conditions as for experiment 1. To obtain average click rates of 5, 10, 18, 33, and 57 Hz, we used Poisson rate parameters of 5, 10, 22, 50, and 133 Hz, respectively. Stimuli were then high-pass filtered at 2 kHz and noise was added, exactly as described for experiment 1. We again used conditions C, RC, and RefRC.

#### 3. Apparatus, procedure, and data analysis

Apparatus, procedure, and data analysis were identical to experiment 1.

### B. Results

The top panel of Fig. 3 displays the average performance across participants, for RC and RefRC trials, and for the five values of click rates tested. Results were broadly similar to those of experiment 1: performance decreased with average click rates, and RefRC produced generally better performance than RC. However, the overall performance was higher here than for experiment 1, with $d'$ values between 2 and 3 for the slowest rates, and $d'$ values above 1 for the fastest rates. The difference between RefRC and RC

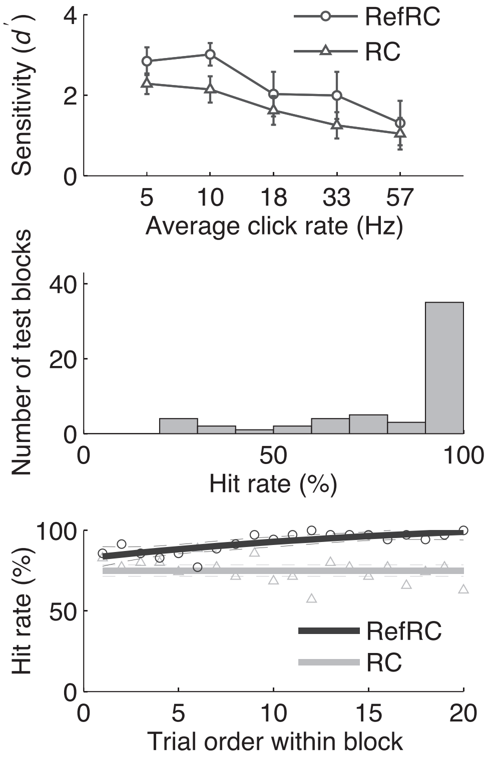

FIG. 3. Results for experiment 2, Poisson distribution. Same format as Fig. 2.

was in the same range or larger as what was observed for experiment 1, between 0.5 $d'$ units (5 Hz, 18 Hz, 57 Hz) and 1 $d'$ unit (10 Hz, 33 Hz).

A repeated measures ANOVA was used to test those observations (see Sec. III A, degrees of freedom corrected as appropriate). Both average click rate [$F(2.46,\ 31.93) = 25.25$, $p < 0.001$, $\eta^2_g = 0.38$] and trial type [$F(1,\ 13) = 82.40$, $p < 0.001$, $\eta^2_g = 0.14$] had a reliable effect on performance. An interaction with small effect size was found between average click rate and trial type [$F(4,\ 52) = 3.02$, $p < 0.05$, $\eta^2_g = 0.02$]. *Post hoc* $t$-test confirmed that RefRCs produced higher performance than RCs for all average click rates [5 Hz: $t(13) = -2.42$, $p < 0.05$, 10 Hz: $t(13) = -4.20$, $p < 0.01$, 18 Hz: $t(13) = -2.59$, $p < 0.05$, 33 Hz: $t(13) = -6.92$, $p < 0.001$, $t(13) = -4.39$, $p < 0.001$; all $p < 0.05$ after the Holm-Bonferroni correction]. The interaction may thus have been caused by an especially large performance advantage of RefRC over RC for 10 and 33 Hz. Given the small effect size of the interaction, we did not explore it any further.

The middle panel of Fig. 3 illustrates performance for RefRCs for different blocks in the 18 Hz average click rate condition. As for experiment 1, the distribution of hit rate over the last ten trials for RefRC was not normally distributed. This was confirmed by one-sample Kolmogorov-Smirnov tests for all rate conditions [5 Hz: $D(56) = 0.64$, $p < 0.001$, 10 Hz: $D(56) = 0.73$, $p < 0.001$, 18 Hz: $D(56) = 0.58$, $p < 0.001$, 33 Hz: $D(56) = 0.57$, $p < 0.001$, 57 Hz: $D(56) = 0.50$, $p < 0.001$]. When applying the criterion of hit rate above 90% for the last ten trials, we observed 79%, 91%, 63%, 57%, and 38% of good blocks for rate conditions of 5, 10, 18, 33, and 57 Hz. The lower panel of Fig. 3 illustrates performance for RC and RefRC during the course of good blocks. As for experiment 1, performance increased for RefRCs while there was a small trend for a decrease for RC. Statistical analyses of the exponential fits showed that performance increased for RefRC during a block [5 Hz: $F(2,\ 17) = 20.80$, $p < 0.001$, 10 Hz: $F(2,\ 17) = 14.25$, $p < 0.001$, 18 Hz: $F(2,\ 17) = 11.21$, $p < 0.001$, 33 Hz: $F(2,\ 17) = 7.95$, $p < 0.01$, 57 Hz: $F(2,\ 17) = 13.64$, $p < 0.001$]. For RC, performance decreased for 10 Hz [$F(2,\ 17) = 4.99$, $p < 0.05$] but was constant for all other rates (all $p > 0.05$).

### C. Interim discussion

Experiments 1 and 2 used the same psychophysical paradigm and average stimulus characteristics, but with uniformly distributed (experiment 1) or Poisson-distributed (experiment 2) stochastic time intervals. The average performance and proportion of good blocks differed between the two experiments, with a general advantage to the Poisson distribution. We thus decided to compare formally the results of these two experiments.

Figure 4 shows the individual data obtained by all participants of experiments 1 and 2, averaged across all blocks (good and bad) and averaged over all average click rates for each participant. A first observation, consistent with all analyses presented so far, is that in all cases but one (out of 28 participants), performance for RefRC was better than performance for RC.

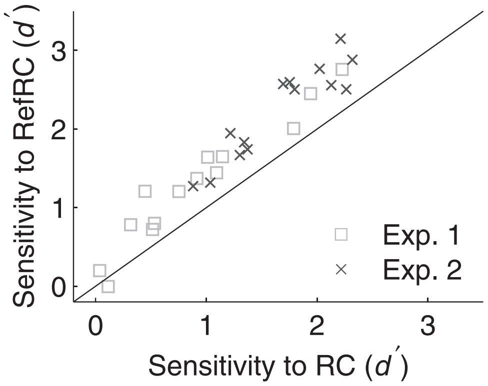

FIG. 4. Comparison between experiments 1 and 2. The sensitivity index $d'$ is plotted for individual participants (squares: experiment 1, crosses: experiment 2) for RC and RefRC conditions. In all but one cases, the points are above the diagonal, indicating a performance advantage for RefRC.

It also appears that, in spite of individual variability, performance was generally better for experiment 2. This was tested statistically with a mixed-design ANOVA on $d'$ averaged across rates, with factors of experiment (1 or 2) and trial type (RC or RefRC). Both experiment [$F(1,26) = 12.37$, $p < 0.01$, $\eta^2_g = 0.09$] and trial type [$F(1,26) = 121.42$, $p < 0.001$, $\eta^2_g = 0.23$] had sizeable effects. A small interaction was also found [$F(1,26) = 4.52$, $p < 0.05$, $\eta^2_g = 0.004$].

Both experiments therefore clearly showed an advantage of RefRC over RC, but performance was generally higher for Poisson-distributed intervals compared to uniformly distributed intervals. The means of both distributions were matched across experiments, but distributions still differed in terms of higher order statistics. In particular, the Poisson distribution had higher variance. Using 1000 simulated click trains per condition, we estimated that the variance between successive time intervals was between 18% and 70% higher for Poisson-distributed stimuli, depending on average rate. As suggested before, higher variance may facilitate the discrimination between temporal patterns (Pollack, 1968a; Sorkin, 1990) and thus explain the difference between experiments 1 and 2. We evaluate this interpretation quantitatively in the modeling section below.

Another difference between experiments 1 and 2 could have been in the local distribution of clicks. It is possible that salient “motifs” or “runs” were used by participants to perform the task (Pollack, 1968a; Ross and Houtsma, 1994). For instance, a chance succession of very short or very long ICIs could have been sufficiently salient to support the repetition detection task, even if the whole time pattern had not been learnt. Such runs may have occurred more often for Poisson distributions because of its higher variance. These runs are still temporal cues, but highly local.

To test for the importance of local cues on the results presented so far, we quantified the possibility of transferring the learning of one sequence to its time-reversed counterpart. If the whole time pattern was learnt, time reversal should dramatically impair performance. In contrast, if only local runs were learnt, such as distinctively long or short ICIs, or repetitions of identical ICIs, memory should transfer almost perfectly to time-reversed sequences. The same reasoning was followed by Agus *et al.* (2010) for white noise. They showed an almost-perfect memory transfer after time reversal for noise stimuli, suggesting local spectral cues for the memory representation of noise. In experiment 3, we tested whether the same would hold for click trains.

## IV. EXPERIMENT 3: MEMORY TRANSFER AFTER TIME REVERSAL

### A. Method

#### 1. Participants

Twelve participants who had previously participated in psychophysical experiments in our laboratory were invited back as "experienced" participants (age in year $M = 25$, $SD = 3$; 5 female). We chose experienced participants because the experiment tested memory transfer, so for efficiency we wanted to ensure that as many blocks as possible would display initial learning. Participants' audiograms were again verified as normal. All other details are as in experiments 1 and 2.

#### 2. Stimuli

The stimulus generation method of experiment 2 was used, with Poisson-distributed ICIs. Here we only tested the average rate conditions of 5, 18, and 57 Hz. In addition to the C, RC, and RefRC trial types, a new "reversed-RefRC" trial type was introduced. This was simply achieved by reversing the order of ICIs of the RefRC trials (effectively, reversing the time axis of the stimulus before high-pass filtering was applied).

We used two minor variants of the stimulus generation procedure. In the method of experiment 2, there was always a click at the beginning of the stimulus. As a consequence, all reversed-RefRCs would have displayed a click at the very end of the sequence, and thus a slightly different time-course relative to the noise pedestal compared to all other trial types. In a first variant of the stimuli, we shifted the reversed-RefRCs in time so that they would include a click at time 0 s like all other conditions. In a second variant, we omitted the first click of the sequence for all trial types before reversal (C, RC, RefRC). This led to more variability overall in the start and end time of the stimuli relative to the noise pedestal, as well as a small change in average rate compared to experiment 2, but here all conditions received the same treatment. Participants were randomly assigned to one or the other variant (six participants for each).

#### 3. Apparatus and procedure

The procedure was the same as experiment 2, with minor differences. We used BeyerDynamics DT 770 Pro headphones. Also, as participants were experienced, no training session was provided.

The first half of each block contained 30 C, 15 RC, and 15 RefRC trials in pseudo-random order. The second half of the block seamlessly followed the first half without any interruption, and introduced the reversed-RefRCs. It contained another 30 C, 15 RC, and 15 reversed-RefRC trials. Four blocks were run in succession for each average rate. The order of presentation of the rates was random and counterbalanced across participants. Each experimental session lasted about 2 h.

#### 4. Data analysis

To test for the possible influence of the two possible variants in the stimulus generation procedure, the sensitivity index $d'$ from signal detection theory was used to evaluate performance for RefRC and RC on the first half of the blocks. We then performed a repeated measures ANOVA on $d'$ with factors of stimulus type, rate condition, and stimulus generation procedure variant. To test for memory transfer, we estimated performance at various time points during the block by computing $d'$ for subsets of trials. We first selected "good" blocks to ensure that the initial RefRCs had been learnt, using the criterion of 90% or more hit rate for the last five trials over which this stimulus was presented (which, for five trials, amounts to 100% hit rate). Then, for such good blocks, we computed $d'$ for the RefRCs for the first five trials (i.e., before learning) and last five trials (i.e., after learning) of the first half of the block, estimating false alarms from the corresponding C trials. The same analysis was then carried out for reversed-RefRCs, for the second half of the selected blocks (i.e., when reversed-RefRCs replaced the original RefRCs). A perfect memory transfer from RefRC to reversed-RefRC would be reflected the same performance for the last five trials of RefRC and the first five trials of reversed-RefRC, before any re-learning could occur. To test for this, we used paired $t$-tests comparing $d'$ on the last five trials for RefRC and first five trials for reversed-RefRC, for each rate condition. On some rates, one subject did not show any good blocks and was excluded from this particular analysis, hence the reduced degrees of freedom. All differences reported as statistically significant are at $p < 0.05$ after Holm-Bonferroni correction.

### B. Results

We first confirmed that there was no effect of the stimulus generation variant on general task performance [$F(1,8) = 3.31$, $p > 0.05$] nor any interaction with other factors (all $p > 0.05$). All further analyses thus used aggregated data for the 12 participants, irrespective of stimulus generation variant.

A one-sample Kolmogorov-Smirnov test confirmed that hit rate for the last five trials of each block did not follow a normal distribution [5 Hz: $D(48) = 0.6423$, $p < 0.001$, 18 Hz: $D(48) = 0.5721$, $p < 0.001$, 57 Hz: $D(48) = 0.5929$, $p < 0.001$]. We thus selected "good" blocks with hit rate of 100% for the last five trials, which led to selecting 67%, 70%, and 52% of blocks for 5, 18, and 57 Hz, respectively. For these blocks, the RefRC has been learnt, so we could meaningfully test for memory transfer to the reversed-RefRC.

Figure 5 displays the sensitivity to RefRCs and reversed-RefRCs over the course of a block. The important feature is what happened immediately after the introduction of time reversal (last white bar compared to first black bar). In all cases, performance dropped for reversed-RefRC, down to levels similar of those observed with initial exposure to a novel RefRC (first white bar). Paired *t*-test confirmed reliable performance drops after time reversal for rates of 18 Hz [$t(1,10) = 3.37$, $p < 0.001$] and 57 Hz [$t(1,10) = 6.96$, $p < 0.001$]. The 5 Hz rate condition did not meet our significance criterion [5 Hz: $t(1,11) = 1.87$, $p = 0.11$].

### C. Interim discussion

The pattern of results showed a drop in performance when a RefRC was replaced by its time-reversed version, with performance reverting to levels observed with novel RefRCs. The one condition that failed the significance test, 5 Hz average rate, did nevertheless exhibit the same pattern. In addition, this condition displayed equal initial performance for reversed-RefRC and RefRC, so the effect of time reversal was basically as large as it could have been. The lack of a statistically significant difference can thus be attributed to a ceiling effect, as $d'$ values were high (above 4) at this slow rate with our group of experienced listeners.

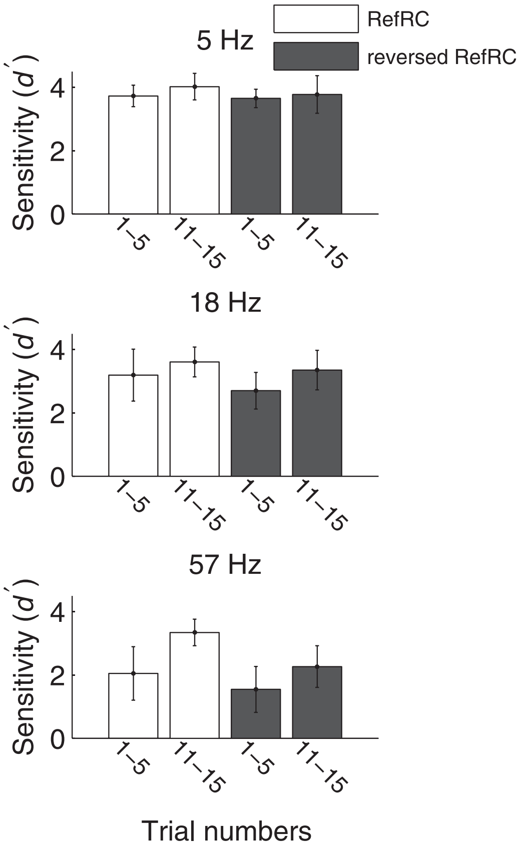

FIG. 5. Results for experiment 3, the effect of time-reversal. For each average click rate condition tested, performance is plotted at the beginning of the block (trials 1–5, white), just before the introduction of time reversals (trials 11–15, white), just after the introduction of time reversal (trials 1–5, black) and at the end of the block (trials 11–15, black). The effect of time reversal can be assessed by comparing the last white bar to the first black bar. Only blocks with 90% or more hit rate for the original RefRC are included. Results are averaged across listeners. Error bars are 95% confidence intervals about the mean.

Performance for the reversed-RefRCs followed a pattern compatible with the learning of a novel stimulus, with an increase in performance between the beginning and end of the second half of the block. The end performance for reversed RefRCs was generally not as high as that observed for the RefRC presented on the first half of the block, but that is likely because "good" blocks were selected based on performance for the initial RefRC only.

Overall, the data are consistent with little or no memory transfer from a learnt temporal pattern to its time-reversed version, which has to be re-learnt from scratch. This lack of memory transfer suggests that participants learnt extended temporal features, and not local runs of salient ICIs. This finding is perhaps especially intriguing for the high rate condition, 57 Hz. Such a high average rate is above the lower limit of pitch for isochronous click trains (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). Accordingly, our stochastic sequences generally sounded like buzzes at this rate, with individual clicks difficult to hear out. However, even in this case, the memory representation appeared to retain an extended series of time intervals and their order of appearance.

## V. CROSS-CORRELATION MODEL

### A. Rationale and model description

We adapted a model of sequence discrimination (Sorkin, 1990) to interpret the effect of ICI distributions on the repetition-detection task. The behavioral task that was assigned to participants was to decide whether each trial contained a repetition or not. For modeling purposes, this can be construed as a discrimination task between the two halves, A and B, of the click train presented on any given trial. If A were judged sufficiently similar to B, the participant would respond "Yes" to the repetition detection task. For repeating trials, A and B were identical, so errors (misses) would be caused by internal noise. For non-repeating trials, A and B were different, so errors (false alarms) would occur if noisy representations of A and B were by chance as similar as noisy representations of identical stimuli.

This formalization of the task can be implemented using cross-correlation as a similarity measure between series of time intervals, corrupted by internal noise (Sorkin, 1990). The first half of the stimulus can be notated as $\mathrm{A} = \{\mathrm{t}_{\mathrm{A}1}, \mathrm{t}_{\mathrm{A}2}, \mathrm{t}_{\mathrm{A}3}, \ldots, \mathrm{t}_{\mathrm{Am}}\} + \mathtt{N}(0,\sigma^2)$, with $\mathrm{t}_{\mathrm{Ai}}$ the time interval corresponding to the *i*th ICI and $\mathtt{N}(0,\sigma)$ an additive noise term, following a normal distribution with mean 0 and variance $\sigma^2$, added independently to each $\mathrm{t}_{\mathrm{Ai}}$. Similarly, the second half of the stimulus can be expressed as $\mathrm{B} = \{\mathrm{t}_{\mathrm{B}1}, \mathrm{t}_{\mathrm{B}2}, \mathrm{t}_{\mathrm{B}3}, \ldots, \mathrm{t}_{\mathrm{Bn}}\} + \mathtt{N}(0,\sigma)$. To estimate the similarity between A and B, a cross-correlation was computed between the series $\{\mathrm{t}_{\mathrm{A}1}, \mathrm{t}_{\mathrm{A}2}, \mathrm{t}_{\mathrm{A}3}, \ldots, \mathrm{t}_{\mathrm{Ak}}\} + \mathtt{N}(0,\sigma^2)$ and $\{\mathrm{t}_{\mathrm{B}1}, \mathrm{t}_{\mathrm{B}2}, \mathrm{t}_{\mathrm{B}3}, \ldots, \mathrm{t}_{\mathrm{Bk}}\} + \mathtt{N}(0,\sigma^2)$, with $\mathrm{k} = \min(\mathrm{m},\mathrm{n})$.

To compute $d'$ values simulating our behavioral dataset, we ran the correlation model on ICIs derived from the experimental stimuli generation procedure, for the RC and C conditions. For each simulated trial, we selected a "Yes" response if the cross-correlation between A and B was above a fixed criterion, *Crit*. If the cross-correlation was below the criterion, a "No" response was selected. This provided a proportion of hits and false alarms for each trial type from which we could compute $d'$.

### B. Results

After an initial exploration of the parameter range, the two free model parameters $\sigma$ and *Crit* were systematically varied and the resulting goodness of fit of the model evaluated, using the squared error between behavioral data and model predictions. We computed $d'$ values on 100 simulated C and RC trials, for all rate conditions and uniform or Poisson distributions, and for all tested $\sigma$ and *Crit* values. The search space for $\sigma$ was from 0 to 100 ms, in 5 ms steps, while for *Crit* it was from $-0.9$ to 0.9, in 0.1 steps.

A broad range of parameters led to a satisfactory fit to the data (Fig. 6, top). The best fit was obtained for $\sigma = 50$ ms and $Crit = -0.2$. This estimate is different from the value of $\sigma = 15$ ms suggested in the original model (Sorkin, 1990). However, we verified that running the model with $\sigma = 15$ ms produced the same trends in the predictions, so the interpretations outlined below would hold with either value of $\sigma$.

For the fitted model parameters, 10 000 simulated trials were computed to estimate $d'$. The model predictions are represented in Fig. 6, bottom, along the corresponding behavioral data. The model correctly captured the two main trends in the data: first, lower click rates led to better performance; second, Poisson distribution led to better performance. However, there were still discrepancies between the model and data. In particular, performance decreased in the model for the lowest rates, whereas it kept increasing in the data. Also, performance was generally underestimated for the highest rates, especially for the Poisson distribution.

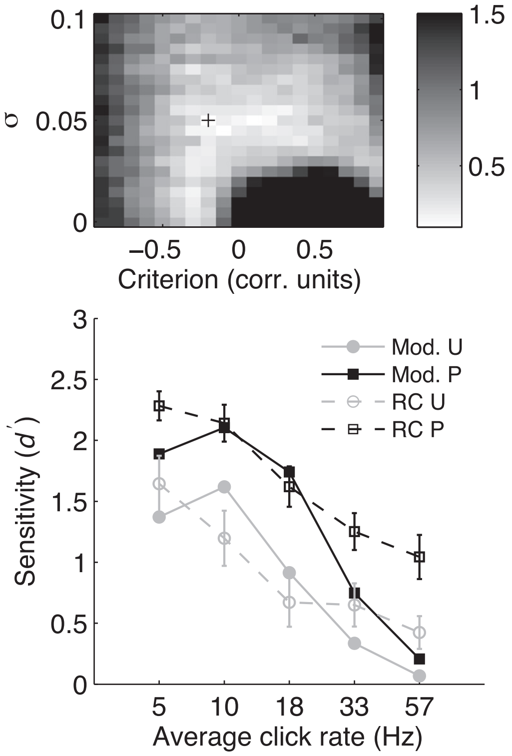

FIG. 6. Predictions of a correlation model. Top panel: mean quadratic error relative to the behavioral data. The black cross indicates the best fit. Lower panel: the data for the RC condition are replotted from experiments 1 (U) and 2 (P), along with the model predictions. See text for details.

### C. Interim discussion

A model of temporal pattern discrimination based on cross-correlation (Sorkin, 1990) was able to capture two important trends of the within-trial repetition-detection performance observed in our dataset: the better performance for long ICIs, and the better performance for Poisson-distributed compared to uniformly distributed ICIs. We interpret the model behavior by considering the ratio between stimulus-related variance and internal-noise variance. Stimulus-related variance was beneficial to model performance, as it made patterns more distinguishable from each other for the cross-correlation operation. In contrast, internal noise reduced performance by corrupting the internal representation of the patterns. The important characteristic of the model here is that internal noise had constant variance. As interval durations decreased for high click rates, their representation in the model was proportionally more affected by internal noise, and the model performance decreased. As the intervals had higher stimulus-related variance for the Poisson distribution compared to the uniform distribution, their representation in the model was more robust to internal noise, and the model performance was higher for the Poisson distribution compared to the uniform distribution.

The main aim of the model was to gain an intuition about the effect of ICI distributions, so we kept its structure as simple as possible, but this implied some important limitations. In particular, this "toy" model had perfect memory. The only limiting factor for performance was the additive noise added to each interval, irrespective of interval duration or total number of intervals. An extension of the original model has been proposed including both additive and multiplicative noise (Sorkin, 1990). We tried including this multiplicative noise parameter, but it only provided an overall gain for the model predictions and did not change the qualitative differences between model and data. Another simplification is that the model assumed that participants could compare exactly the first and second halves of each stimulus, even though stimuli were presented continuously. As all trials had equal duration, it is not unlikely that participants learnt the approximate halfway position of the stimuli, and cross-correlation should be reasonably robust to imperfect estimates of the halfway point. Finally, we made no attempt at modeling the added benefit of repeated exposure for RefRC: this advantage was about the same for all conditions, so this did not provide any useful constraint to add a parameter to the model.

The resulting fit was qualitatively correct, but still imperfect. The underestimation of performance at the lowest click rates could be due to issues with computing a meaningful correlation with a small number of intervals. The underestimation of performance at high click rates suggests that participants may have used other strategies (Ross and Houtsma, 1994) than the "holistic" comparison of all intervals implied by the correlation operation. For these high rates, it is plausible that some intervals, for instance the longest ones, were weighted more than others.

Finally, even if the model succeeded in predicting trends in the data, we do not wish to imply that the neural bases of click-train repetition detection literally involve computing cross-correlations over a broad range of time intervals. Auditory models based on auto-correlation could be used to detect repetitions using a similar strategy, but the long duration of our stimuli would place heavy demands on the required delays, and the applicability of such models to random click trains with purely temporal regularities has been challenged (Kaernbach and Demany, 1998). Rather, the present functional model served to identify a general limiting factor for the repetition-detection task, the balance between stimulus variability and internal noise. The underlying neural mechanisms remain as yet unspecified and some possibilities will be discussed in Sec. VI.

## VI. DISCUSSION

The effect of repeated exposure to random auditory temporal patterns was investigated in adult listeners, using an experimental paradigm that had previously been used to demonstrate learning for random spectro-temporal patterns (Agus *et al.,* 2010). In three behavioral experiments, we showed that random auditory time patterns could be learnt rapidly, after only a few exposures. Moreover, the memory for temporal patterns was observed over a surprisingly broad range of inter-click intervals: from sparse click trains, containing about 5 clicks per seconds, to dense click trains, containing more than 50 clicks per second. Such a range encompasses the subjective transition between clicks being heard out individually, like in a rhythm, to clicks being fused together, like in a buzzing sound (Bendor and Wang, 2007; Pressnitzer *et al.,* 2001). Immediate repetition detection was easier for the rhythm-like patterns, but, surprisingly, this did not appear to affect the longer-term memory processes operating across trials, as the gain from increased exposure was comparable for all rates.

There were strong similarities between the characteristics of memory for noise (Agus *et al.,* 2010; Agus and Pressnitzer, 2013) and clicks trains (present experiments). In both cases, learning could be achieved in an unsupervised and implicit manner, with the patterns to be learnt interspersed amongst other statistically matched sounds. The ancillary task used in our paradigm (within-trial repetition detection) was initially challenging. However, as the patterns were heard more than once over the course of a block, performance sometimes improved to reach almost-perfect performance in about ten trials. This change in performance was interpreted as learning, as learnt patterns became perceptually easily distinguishable from novel ones (Agus and Pressnitzer, 2013).

There was, however, an important difference between the memory for noise and click trains. For noise, when a white noise exemplar was time-reversed after having been learnt, an almost-perfect transfer of memory was observed (Agus *et al.,* 2010). Together with other converging evidence reviewed in the Introduction, this suggested that noise was learnt thanks to brief spectral cues. In contrast, here we observed that, when a time pattern was time-reversed after learning, there was no evidence of memory transfer. Rather, the time-reversed pattern had to be learnt afresh. Thus, for click trains, it was the extended series of time intervals and their precise ordering that was learnt.

This leads to an apparent puzzle: auditory memory for noise and click trains relied on qualitatively different cues, but still displayed similar characteristics. One parsimonious interpretation to resolve that puzzle is that both types of cues were recoded into a common neural code before being learnt. Let us first consider the case of noise. Noise contains spectral cues, which are encoded in the auditory system as spatial patterns of activity across frequency-tuned neurons: across auditory nerve fibers in the periphery, or across other neural populations in subsequent tonotopic maps at least up to primary auditory cortex (Bidelman *et al.,* 2014). To form memory traces for such spatial patterns, standard Hebbian models of neural plasticity apply. Indeed, computational models have shown that neurons receiving inputs across a range of afferents, and equipped with well-established plasticity rules such as spike-time dependent plasticity, can develop selectivity to random patterns of near-simultaneously active afferents (Klampfl and Maass, 2013; Masquelier *et al.,* 2008). If the afferents were frequency-selective neurons, then the scheme could apply to the learning of auditory noise (Viswanathan, 2016).

For auditory time patterns, the learning was dependent on details of successive time intervals and their order, and thus required the integration of non-simultaneous information. In the auditory periphery, at the level of the auditory nerve, time intervals are not encoded as a spatial map, but rather transmitted explicitly as inter-spike intervals (Cariani and Delgutte, 1996). Neural spikes synchronized to clicks have been observed up to at least primary auditory cortex for time-jittered click trains resembling our stimuli (Bendor and Wang, 2007; Lu *et al.,* 2001). However, there is also evidence that inter-spike intervals are recoded into a rate-place code: a “synchronized” neural population co-exists with “non-synchronized” populations, which encode click-rate through neural discharge rate in primary (Gao *et al.,* 2016; Lu *et al.,* 2001) and non-primary auditory cortex (Bendor and Wang, 2007) depending on the click-rates (Petkov and Bendor, 2016).

A rate-place code for click rate should be able to produce neural spatial maps distinguishing click trains with different rates. However, it has not been tested whether such a representation would support the learning of a specific pattern amongst other patterns with identical average rate, as was the case in our experiments. To do so, a generic class of computational models have been put forward: state-dependent networks, which, thanks to the complex temporal dynamics imposed by individual neurons and by the network topography, can produce a large number of unique network states depending on the exact temporal succession of input spikes (Karmarkar and Buonomano, 2007; Lim *et al.,* 2016). Intuitively, such a network state is simply a place pattern of neural activity that depends on the past and present input, with the potential to distinguish between temporal patterns even when they share the same average rate. Such a model has been used to account for behavioral context effects in auditory temporal pattern perception in human listeners (Karmarkar and Buonomano, 2007). Moreover, recent physiological and behavioral evidence for a conversion from time patterns to place patterns have been found in the avian auditory system. Using random click trains with time intervals between 11 and 40 ms—intended to mimic the timing characteristics of bird sounds, but not unlike our own stimuli—Lim *et al.* (2016) recorded neural activity in primary and secondary auditory regions of the anaesthetized zebra finch forebrain. They observed distinctive neural place patterns of activity in secondary regions only, which could support the behavioral discrimination between acoustic temporal patterns. They also tested for memory transfer after time-reversal of the click trains: there was no transfer in a behavioral task, consistent with observed changes in neural place maps after time-reversal.

In summary, such a general scheme would in principle be consistent with all aspects of the present results. Somewhere along the auditory pathways, time patterns could be converted to neural place patterns, unique to each input sequence and sensitive to time-reversal. Once recoded as place patterns, unsupervised learning of acoustic temporal patterns could be achieved through standard neural plasticity processes (Klampfl and Maass, 2013; Masquelier *et al.*, 2008). As the spectral cues present in noise or tone clouds (Agus *et al.*, 2010; Kumar *et al.*, 2014) could also be coded as neural place patterns, the same plasticity processes apply, hence explaining the similarities observed for the learning of all types of auditory stimuli. There are of course many aspects of our findings that would need to be tested more thoroughly against this speculative account. In particular, we observed learning over a broad range of time scales, straddling the divide between "synchronized" and "non-synchronized" neural populations (Lu *et al.*, 2001; Petkov and Bendor, 2016). The time scales involved in the state-dependent models (Karmarkar and Buonomano, 2007; Lim *et al.*, 2016) are broadly consistent with the stimulus range we tested, but differences between stimuli and species prevent a more detailed comparison. The locus, or loci, of this putative recoding and subsequent learning remain to be ascertained in humans. Interestingly, the available fMRI (Kumar *et al.*, 2014) and EEG (Andrillon *et al.*, 2015) data for the learning of noise suggest neural correlates in secondary auditory areas, which would be consistent with the animal data (Gao *et al.*, 2016; Lim *et al.*, 2016; Lu *et al.*, 2001).

## VII. CONCLUSIONS

Using random click trains in an auditory memory paradigm (Agus *et al.*, 2010), we showed that auditory temporal patterns could be learnt by adult human listeners. We tested average click rates from about 5 clicks per seconds up to about 50 clicks per second. Even though this range covered patterns sounding like rhythms for low rates or irregular buzzing sounds for high rates, learning was observed in all cases. We also varied the higher-order statistics of the random distribution of time intervals. A higher variance led to better within-trial repetition detection, but did not affect much the across-trial longer term memory effect. In all cases, learning was fast and occurred in an unsupervised fashion, with the patterns to be learnt interspersed among statistically matched patterns. The learning depended on the precise time-series of interval and their presentation order, as time-reversal disrupted performance. We suggest that auditory temporal patterns may be recoded as neural spatial patterns along the auditory hierarchy, which are then amenable to standard neural plasticity mechanisms. Overall, the experimental evidence now shows that audition benefits from remarkably fast and efficient unsupervised learning processes for noises (Agus *et al.*, 2010), other random spectro-temporal patterns like tone clouds (Kumar *et al.*, 2014), and also purely temporal patterns such as click trains (present data). This is likely to cover any sort of acoustic cues present in natural sounds, e.g., the timbre of a novel sound source (Pressnitzer *et al.*, 2015). More generally, such an almost-inevitable emergence of memory traces after repeated exposure, for different kinds of auditory cues and several time scales, would be consistent with views suggesting that memory may be viewed as an integral part of sensory processing (Hasson *et al.*, 2015).

## ACKNOWLEDGMENTS

We would like to thank Christian Kaernbach and an anonymous reviewer for helpful comments on a previous version of the manuscript. This work was funded by the ERC Grant "ADAM" No. 295603, the ANR-10-LABX-0087 IEC, and ANR-10-IDEX-0001-02 PSL*.

## APPENDIX

In all experiments, we distinguished good blocks, where learning of RefRC was observed, from bad blocks, where no learning of RefRC was observed. Here, we further analyze the temporal patterns that comprised the RefRCs, to test whether there were systematic acoustic differences between the two kinds of blocks.

Even though generative parameters were the same for all blocks, it is possible that by chance good blocks were associated with fewer time intervals to learn, or perhaps more distinctive ones. For each stimulus used in the experiments, we computed the mean ICI, $M$, inversely related to the number of intervals in the pattern as duration was kept constant. We also computed the standard deviation of ICIs, $SD$. In the main results, higher $M$s and $SD$s were associated with better performance. Finally, we computed the maximum of the absolute value of $z$-scored ICIs (using an estimation of standard deviation of all RefRC intervals within a condition). This later index should be sensitive to exceptionally long or exceptionally short ICIs within a pattern, which could provide distinctive cues.

All indices were then averaged within rate conditions for experiments 1 and 2. To assess the reliability of the differences between good and bad blocks, we used Welch's two sample $t$-tests, as samples had unequal sizes. We report

TABLE I. Observed characteristics of RefRCs corresponding to good and bad blocks. For each RefRC pattern, three indices were computed: $M$, the average ICI; $SD$, the standard deviation of ICIs; Max($|z|$), the maximum absolute value of $z$-scored ICIs. Indices were then averaged for each rate condition and experiment. Values in bold and italics indicate good/bad contrasts reaching a statistical significance criterion (Welch's $t$-test, uncorrected. Bold: $p < 0.01$. Italics: $p < 0.05$).

| | Rate | 5 | | 10 | | 18 | | 33 | | 57 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Block group | Good | Bad | Good | Bad | Good | Bad | Good | Bad | Good | Bad |
| Experiment 1 | Number of blocks | 36 | 20 | 31 | 25 | 23 | 33 | 22 | 34 | 13 | 43 |
| | $M$ | *0.218* | *0.255* | *0.101* | *0.114* | 0.056 | 0.056 | 0.029 | 0.031 | 0.018 | 0.018 |
| | $SD$ | 0.101 | 0.088 | 0.055 | 0.054 | 0.025 | 0.025 | 0.012 | 0.012 | 0.005 | 0.004 |
| | Max($\|z\|$) | 1.097 | 0.872 | 1.540 | 1.430 | 1.759 | 1.723 | 1.763 | 1.759 | 1.735 | 1.747 |
| Experiment 2 | Number of blocks | 44 | 12 | 51 | 5 | 35 | 21 | 32 | 24 | 21 | 35 |
| | $M$ | **0.182** | **0.285** | 0.110 | 0.136 | 0.058 | 0.057 | 0.030 | 0.031 | 0.018 | 0.018 |
| | $SD$ | 0.194 | 0.147 | 0.089 | 0.124 | 0.046 | 0.041 | 0.021 | 0.019 | 0.008 | 0.008 |
| | Max($\|z\|$) | 1.255 | 1.406 | *1.745* | *2.574* | 2.737 | 2.276 | *3.169* | *2.672* | 3.758 | 3.529 |

both exploratory, non-corrected $p$-values with a significance criterion set at $p < 0.05$, and $p$-values corrected for multiple comparisons using the Bonferroni method, with the same significance criterion.

Results are shown in Table I. Most comparisons did not reach our lenient uncorrected significance criterion. For experiment 1, only the contrast between mean ICIs at rates 5 Hz [$t(39.64) = 2.2$, $p < 0.05$] and 10 Hz [$t(41.67) = 2.28$, $p < 0.05$] reached significance. For experiment 2, only the contrasts between mean ICIs at 5 Hz [$t(14.47) = 3.93$, $p < 0.01$] and between Max($|z|$) at 10 Hz [$t(6.85) = 3.48$, $p < 0.05$] and 33 Hz [$t(51.03) = -2.04$, $p < 0.05$] reached significance. When applying the correction for multiple comparisons, only the contrast between mean ICIs at 5 Hz in experiment 2 remained significant ($p < 0.05/15$).

The only significant difference that survived the multiple comparisons correction was found for low click rates. This may have been expected, as low rates correspond to small numbers of intervals, and hence a greater likelihood to obtain discrepancies between generative and observed characteristics for a given random draw. In this case, good blocks were actually associated with higher observed rates. This is in apparent contradiction with the general pattern of results, where lower rates led to better performance. It is possible that for the lowest rate here, the limiting factor was that patterns became too sparse to be distinguishable.

In any case, we interpret Table I as suggesting that, over all experimental conditions, good and bad blocks were not systematically different in terms of mean ICIs, standard deviation of ICIs, and unusually short or long ICIs.

Agus, T. R., Carrión-Castillo, A., Pressnitzer, D., and Ramus, F. (**2014**). "Perceptual learning of acoustic noise by individuals with dyslexia," J. Speech Lang. Hear. Res **57**, 1069–1077.

Agus, T. R., and Pressnitzer, D. (**2013**). "The detection of repetitions in noise before and after perceptual learning," J. Acoust. Soc. Am **134**, 464–473.

Agus, T. R., Suied, C., Thorpe, S. J., and Pressnitzer, D. (**2012**). "Fast recognition of musical sounds based on timbre," J. Acoust. Soc. Am **131**, 4124–4133.

Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (**2010**). "Rapid formation of robust auditory memories: Insights from noise," Neuron **66**, 610–618.

Andrillon, T., Kouider, S., Agus, T., and Pressnitzer, D. (**2015**). "Perceptual learning of acoustic noise generates memory-evoked potentials," Curr. Biol. **25**, 2823–2829.

Andrillon, T., Pressnitzer, D., Léger, D., and Kouider, S. (**2017**). "Formation and suppression of acoustic memories during human sleep," Nat. Commun. **8**, 179.

Bendor, D., and Wang, X. (**2007**). "Differential neural coding of acoustic flutter within primate auditory cortex," Nat. Neurosci. **10**, 763–771.

Bidelman, G. M., Schug, J. M., Jennings, S. G., and Bhagat, S. P. (**2014**). "Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians," J. Acoust. Soc. Am **136**, EL33–EL39.

Cariani, P. A., and Delgutte, B. (**1996**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698–1716.

Demany, L., and Semal, C. (**2007**). "The role of memory in auditory perception," in *Auditory Perception of Sound Sources, Springer Handbook of Auditory Research* (Springer U.S., Boston, MA), Vol. 29, pp. 77–113.

Gao, L., Kostlan, K., Wang, Y., and Wang, X. (**2016**). "Distinct subthreshold mechanisms underlying rate-coding principles in primate auditory cortex," Neuron **91**, 905–919.

Goossens, T., van de Par, S., and Kohlrausch, A. (**2008**). "On the ability to discriminate Gaussian-noise tokens or random tone-burst complexes," J. Acoust. Soc. Am **124**, 2251–2262.

Grondin, S. (**2010**). "Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions," Atten. Percept. Psychophys. **72**, 561–582.

Guttman, N., and Julesz, B. (**1963**). "Lower limits of auditory periodicity analysis," J. Acoust. Soc. Am **35**, 610.

Hasson, U., Chen, J., and Honey, C. J. (**2015**). "Hierarchical process memory: Memory as an integral component of information processing," Trends Cogn. Sci. **19**, 304–313.

Hirsh, I. J., Monahan, C. B., Grant, K. W., and Singh, P. G. (**1990**). "Studies in auditory timing: 1. Simple patterns," Percept. Psychophys. **47**, 215–226.

Kaernbach, C. (**1993**). "Temporal and spectral basis of the features perceived in repeated noise," J. Acoust. Soc. Am. **94**, 91–97.

Kaernbach, C. (**1999**). "A behavioral reverse correlation technique to decipher early auditory feature coding," J. Acoust. Soc. Am **105**, 966.

Kaernbach, C. (**2000**). "Early auditory feature coding," in *Contributions to Psychological Acoustics: Results of the Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, M. Meister, and C. Reckhardt, Oldenburg, pp. 295–307.

Kaernbach, C. (**2004**). "The memory of noise," Exp. Psychol. **51**, 240–248.

Kaernbach, C. (**2017**). (private communication).

Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," J. Acoust. Soc. Am **104**, 2298–2306.

Karmarkar, U. R., and Buonomano, D. V. (**2007**). "Timing in the absence of clocks: Encoding time in neural network states," Neuron **53**, 427–438.

Klampfl, S., and Maass, W. (**2013**). "Emergence of dynamic memory traces in cortical microcircuit models through STDP," J. Neurosci. **33**, 11515–11529.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., and Griffiths, T. D. (**2014**). "Representations of specific acoustic patterns in the auditory cortex and hippocampus," Proc. Biol. Sci. **281**, 20141000.

Lakens, D. (**2013**). "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," Front. Psychol. **4**, 863.

Lim, Y., Lagoy, R., Shinn-Cunningham, B. G., and Gardner, T. J. (**2016**). "Transformation of temporal sequences in the zebra finch auditory system," eLife **5**, e18205.

Lu, T., Liang, L., and Wang, X. (**2001**). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates," Nat. Neurosci. **4**, 1131–1138.

Luo, H., Tian, X., Song, K., Zhou, K., and Poeppel, D. (**2013**). "Neural response phase tracks: How listeners learn new acoustic representations," Curr. Biol. **23**, 968–974.

Masquelier, T., Guyonneau, R., and Thorpe, S. J. (**2008**). "Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains," PLoS One **3**, e1377.

Motulsky, H., and Christopoulos, A. (**2004**). *Fitting Models to Biological Data using Linear and Nonlinear Regression* (Oxford University Press, New York).

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (**2015**). "Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition," Neuron **88**, 1281–1296.

Petkov, C. I., and Bendor, D. (**2016**). "Neuronal mechanisms and transformations encoding time-varying signals," Neuron **91**, 718–721.

Pollack, I. (**1968a**). "Periodicity discrimination for auditory pulse trains," J. Acoust. Soc. Am. **43**, 1113–1119.

Pollack, I. (**1968b**). "Detection and relative discrimination of auditory 'jitter,' " J. Acoust. Soc. Am. **43**, 308–315.

Povel, D.-J., and Essens, P. (**1985**). "Perception of temporal patterns," Music Percept.: Interdisc. J. **2**, 411–440.

Pressnitzer, D., Agus, T., and Suied, C. (**2015**). "Acoustic timbre recognition," in *Encyclopedia of Computational Neuroscience* (Springer, New York), pp. 128–133.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

Rajendran, V. G., Harper, N. S., Abdel-Latif, K. H. A., and Schnupp, J. W. H. (**2016**). "Rhythm facilitates the detection of repeating sound patterns," Front. Neurosci. **10**, 464–467.

Ross, J., and Houtsma, A. J. (**1994**). "Discrimination of auditory temporal patterns," Percept. Psychophys. **56**, 19–26.

Rothschild, G., Nelken, I., and Mizrahi, A. (**2010**). "Functional organization and population dynamics in the mouse primary auditory cortex," Nature **13**, 353–360.

Roye, A., Schröger, E., Jacobsen, T., and Gruber, T. (**2010**). "Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans," J. Neurosci. **30**, 7310–7313.

Royer, F. L., and Garner, W. R. (**1966**). "Response uncertainty and perceptual difficulty of auditory temporal patterns," Percept. Psychophys. **1**, 41–47.

Shamma, S., and Lorenzi, C. (**2013**). "On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system," J. Acoust. Soc. Am **133**, 2818–2833.

Sorkin, R. D. (**1990**). "Perception of temporal patterns defined by tonal sequences," J. Acoust. Soc. Am. **87**, 1695–1701.

Teki, S., and Griffiths, T. D. (**2014**). "Working memory for time intervals in auditory rhythmic sequences," Front. Psychol. **5**, 1329.

Viswanathan, J. (**2016**). "When noise enlightens: An investigation into the mechanisms of perception and long- term memory for meaningless auditory stimuli," edited by S. J. Thorpe, Doctoral dissertation, Université Paul Sabatier, Toulouse.

Viswanathan, J., Rémy, F., Bacon-Macé, N., and Thorpe, S. J. (**2016**). "Long term memory for noise: Evidence of robust encoding of very short temporal acoustic patterns," Front. Neurosci. **10**, 610–611.

Warren, R. M., Bashford, J. A., Cooley, J. M., and Brubaker, B. S. (**2001**). "Detection of acoustic repetition for very long stochastic patterns," Percept. Psychophys. **63**, 175–182.