

Rapid Formation of Robust Auditory Memories: Insights from Noise

Trevor R. Agus,^{1,2,*} Simon J. Thorpe,^{3,4} and Daniel Pressnitzer^{1,2}

¹Laboratoire psychologie de la perception, CNRS & Université Paris Descartes, 45, rue des Saints-Pères, 75006 Paris, France

²Département d'études cognitives, École normale supérieure, 29, rue d'Ulm, 75005 Paris, France

³Université de Toulouse, UPS, Centre de Recherche Cerveau et Cognition, Toulouse, France

⁴CNRS, CerCo, 133, route de Narbonne, 31062 Toulouse Cedex 9, France

*Correspondence: trevor.agus@ens.fr

DOI 10.1016/j.neuron.2010.04.014

SUMMARY

Before a natural sound can be recognized, an auditory signature of its source must be learned through experience. Here we used random waveforms to probe the formation of new memories for arbitrary complex sounds. A behavioral measure was designed, based on the detection of repetitions embedded in noises up to 4 s long. Unbeknownst to listeners, some noise samples reoccurred randomly throughout an experimental block. Results showed that repeated exposure induced learning for otherwise totally unpredictable and meaningless sounds. The learning was unsupervised and resilient to interference from other task-relevant noises. When memories were formed, they emerged rapidly, performance became abruptly near-perfect, and multiple noises were remembered for several weeks. The acoustic transformations to which recall was tolerant suggest that the learned features were local in time. We propose that rapid sensory plasticity could explain how the auditory brain creates useful memories from the ever-changing, but sometimes repeating, acoustical world.

INTRODUCTION

One basic goal of auditory perception is to recognize the plausible physical causes of incoming sensory information. In order to do so, listeners must learn recurring features or templates of complex sounds and associate them with specific sound sources, such as a familiar voice, a piano, or a bird singing. Most of our knowledge on auditory memory is based on simple sounds (for a review, see Cowan, 1984; Demany and Semal, 2008; Jääskeläinen et al., 2007) or speech sounds (for a review, see Baddeley, 1997). How templates emerge from everyday auditory experience with arbitrary complex sounds is currently largely unknown. It has even been suggested that auditory memory for natural sounds may be remarkably poor (Cohen et al., 2009), in contrast to visual memory, which is able to store details of thousands of images after a single exposure (Brady et al., 2008).

Here we used random acoustic waveforms as a tool to observe the creation of new auditory memories. Random noises may be particularly suitable for probing the formation of memories: before learning, random noises all sound like a hiss; yet, they are acoustically diverse, and the listener would not have been exposed to a given sample before the experiment. Noises are meaningless, and no label can readily be attached to different samples, so semantic processing will not interfere with the acoustic memorization process (Cohen et al., 2009). Acoustic details in noise are totally unpredictable, so memorizing them should be particularly challenging and would test any learning mechanism to the extreme, compared to more predictable sounds (Overath et al., 2007). Finally, by collecting behavioral responses after each exposition to the noise, the temporal dynamics of memory formation can be characterized.

The use of noise to probe auditory memory can be traced back to the seminal work of Guttman and Julesz (1963). They introduced a paradigm where a given segment of noise was “frozen” and then repeated identically several times. Discriminating repeated noise from plain random noise requires some form of memory of the repeated waveform. Over a wide range of segment durations and repetition rates, listeners were found to be able to detect repetitions (Guttman and Julesz, 1963; Warren et al., 2001). The repeated-noise paradigm has thus been used to investigate several time constants of nonverbal auditory memory, from tens of milliseconds to tens of seconds (Kaernbach, 2004). Long-term memory traces have also been investigated by using the same noise samples throughout an experimental block (Goossens et al., 2008; Hanna, 1984).

However, some essential characteristics of the everyday learning of new sounds have not been addressed with the noise paradigm. A first issue is that learning should be possible even in an unsupervised fashion. In a typical auditory scene, it is not obvious which segments of the ongoing sounds should be memorized and which can be safely ignored. In the repeated-noise paradigms, listeners knew that repetition would occur within a trial (Guttman and Julesz, 1963) or that a noise token could be repeated within a block (Goossens et al., 2008). A second issue is that intervening sounds are likely to occur between learning and recall. Repetition detection was shown to remain possible when purely random noise segments were inserted between the repeated noise samples (Kaernbach, 2004), but it is unclear whether it may survive the insertion of other sounds that are actively processed. Also, memories of

sounds must be long lasting to be really useful. Memory for noise has not been tested over time ranges extending past the duration of a single experimental block. Finally, recall should be tolerant to some acoustic variations, as the exact same waveform will never be heard twice in real life. If a paradigm using noise provides a good model for everyday auditory learning, then it should display all of these features.

The following experiments investigated the ability of human listeners to learn new sounds by using a new, indirect measure of memory for noise. The general experimental method is illustrated in Figure 1. Listeners were either presented with a 1 s sample of running noise (noise condition, N) or two seamlessly repeated 0.5 s samples of noise (repeated noise, RN). They were asked to detect the repetitions, similar to previous repeated-noise paradigms (Guttman and Julesz, 1963; Kaernbach, 2004; Warren et al., 2001). In the present experiments, however, one particular exemplar of the RN condition reoccurred identically in trials interspersed throughout an experimental block (reference repeated noise, RefRN). Any evolution in performance with repeated exposure would indicate the formation of a memory for the RefRN. Note that learning would have to be unsupervised: listeners were not told that memorizing trials across the block might be beneficial, and, in any case, they could not have identified which trials to memorize without prior learning (no feedback was given). In addition, RefRN was never presented on two consecutive trials, so some interference from intervening trials could be expected. In later experiments, the same noise samples were used on experimental blocks separated by several weeks, which tested for long-term memorization. Finally, we investigated whether learning and recall were robust to a range of acoustic manipulations.

RESULTS

Experiment 1: Fast Learning of Noises

Figure 2A shows the mean sensitivity to repetitions for RN and RefRN conditions for 12 normal-hearing naive listeners. Performance was better for RefRN (mean d' = 1.0) than RN (mean d' = 0.4; t_{11} = 4.12, p = 0.002). This pattern persisted when six of the naive listeners repeated the experiment as “trained” listeners. In fact, training accentuated the difference between RN and RefRN (Figures S1A and S1B). Since RN and RefRN were generated by the same process, this difference cannot be attributed to acoustical differences, but rather to the reoccurrence of the RefRN stimulus.

Figure 2B shows the mean hit rates plotted in the order of the individual trials within each block, averaged across the four blocks for each of the 12 naive listeners. Initially, the hit rates were the same for RN and RefRN. This was expected because, again, RN and RefRN were generated from the same random process and only distinguishable after learning. After that, an increase in the hit rate for the RefRN was observed. Three-parameter exponential curves were fitted to the RefRN data by the least-squares method. The exponential model provided a significantly better description of the data than a model that assumed a constant hit rate, even taking into account the additional two parameters ($F_{2,47}$ = 8.12, p = 0.001; see Motulsky and Christopoulos, 2004). This demonstrates that the hit rate for

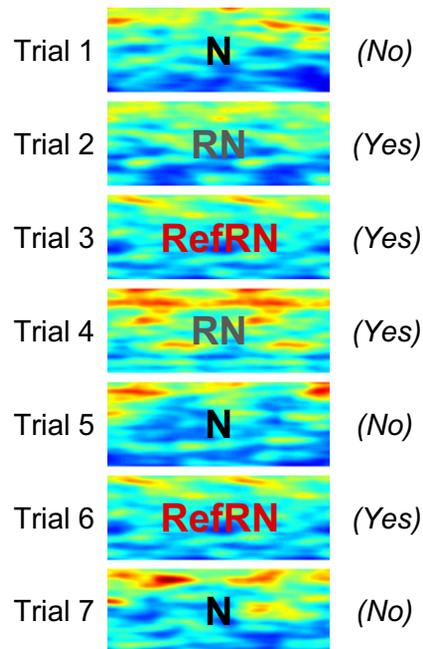


Figure 1. Experimental Method

Samples of Gaussian noise were used as stimuli, illustrated here as schematic spectrograms containing random amplitude fluctuations across time and frequency. Listeners were asked to detect those trials that contained a repetition. The noise (N) trials were formed from segments of noise, so the correct response would be “No” repetition. The repeated-noise (RN) trials were formed from the seamless repetition of a half-duration segment of noise, for which the correct response would be “Yes.” The N and RN trials were generated afresh for each trial. The reference repeated-noise (RefRN) trials also contained a repetition but, importantly, the exact same reference noise sample was used over several trials. Those trials were randomly interspersed within the experimental block and never occurred on two successive trials. Unless a listener formed some memory for the RefRN sample, RN and RefRN conditions would be indistinguishable and equally difficult. If a memory was formed, however, one would expect the RefRN trials to become easier. As an illustration of the memory load imposed by the task, digitally storing a given sample of 0.5 s long RefRN for playback during the experiments required over 10^5 independent bits.

RefRN improved significantly during the block. Based on the time constant of the exponential (mean half-life = 3.9 trials), most of the learning occurred within the first ten presentations of RefRN.

A similar analysis was also performed on the RN data, and it showed that the hit rate for RN *decreased* during the block, with a similar time course (mean half-life = 6.6 trials; $F_{2,47}$ = 8.29, p = 0.001). This was explored further in a supplemental experiment, which compared blocks where only RN had to be detected and blocks containing both RN and RefRN. The inclusion of RefRN did not affect d' sensitivity to RN nor the overall criterion for the block (Figures S1C and S1D). In the framework of signal-detection theory, these findings account for the mirror changes in hit rate observed for RefRN and RN in experiment 1: as sensitivity to RefRN increased during the block through learning, repetitions in RefRN trials became easier to detect. However, listeners maintained a constant criterion, that is, they

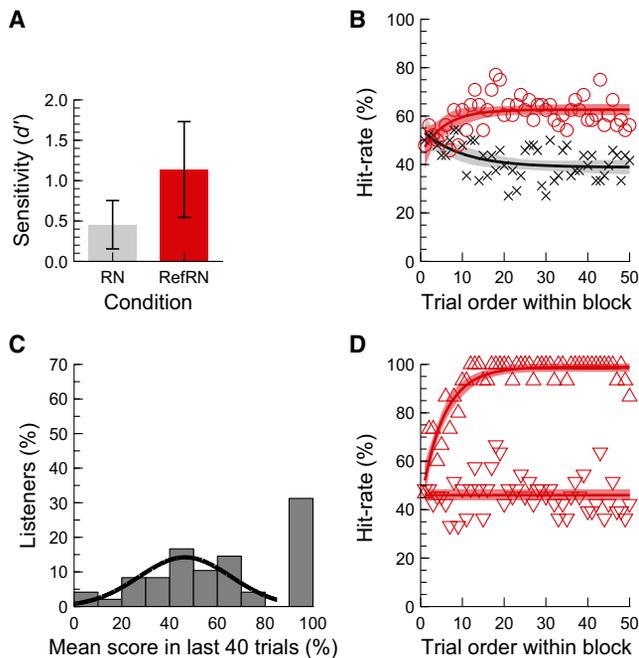


Figure 2. Fast Learning of Noise

(A) Mean repetition-detection sensitivity for the RN and RefRN for the first visit of 12 naive listeners in experiment 1. Error bars are 95% confidence intervals. Sensitivity was significantly better for RefRN than for RN. Trained listeners showed an even larger difference (Figures S1A and S1B).

(B) The variation of RefRN (red) and RN (gray) hit rates throughout the blocks in experiment 1, averaged over all blocks for naive and trained listeners. The lines show the best-fit exponential lines for each condition, and the surrounding bands show the 95% confidence intervals of these fitted exponentials. Hit rates for RefRN increased, while hit rates for RN decreased. This is accounted for by a progressively higher sensitivity to RefRN and a constant overall criterion (a balance of “yes” and “no” responses), leading to a decrease in RN hit rates (Figures S1C and S1D).

(C) The distribution of the hit rates on the last 40 RefRN trials of each block. The line shows a normal curve fitted to the left cluster of mean scores.

(D) The same data shown in red in (B), but separated into the two clusters of blocks highlighted in (C): those in which 90% or greater was scored in the last 40 trials (upward-pointing triangles) and those under 90% (downward-pointing triangles). After a fast initial improvement (half-life = 2.0 trials), performance is almost perfect in the cluster of good blocks.

chose to balance their “yes” and “no” responses so that they reported repetitions on about half of the trials. To do so, they had to increase their response threshold, which resulted in a drop in RN hit rate (and in false-alarm rate, data not shown).

It was noted that performance for RefRN was highly variable. Figure 2C shows the distribution of the mean hit rates for the last 40 RefRN trials in each block for each listener (i.e., after most of the learning had occurred in the first ten RefRN trials). The distribution is bimodal, with approximately two-thirds of the scores normally distributed around a mean of 47% and with the remaining third of scores exceeding 90%. These near-perfect scores were not limited to a small number of individuals: 7 of the 12 listeners detected 90% or greater in at least one of their blocks, and one listener did so in all four of his blocks. Nor were the high scores an effect of training: three of them

occurred on listeners’ first blocks, and the rest were spread relatively evenly across the remainder.

Figure 2D shows the average hit rate for RefRN over the course of each block for the 15 blocks in which over 90% was scored (upward-pointing triangles) and the equivalent data for the other 33 blocks (downward-pointing triangles). Nearly all of the errors in the high-performance blocks occurred at the start of the blocks and after rapid learning (half-life = 2.0 trials) the participants were able to detect the repetitions in the RefRN nearly perfectly. Whereas learning was clearly evident in the high-performance blocks ($F_{2,47} = 141.49$, $p < 0.001$), no significant learning was observed in the remaining blocks ($F_{2,47} = 1.71$, $p = 0.19$). Thus, the learning seems to be primarily driven by a third of the blocks in which the stimuli were perfectly learned.

Experiment 2: Good Noise, Bad Noise, and Long-Term Memory Traces

Different RefRNs were used for different blocks in experiment 1, so some of the performance variability could have stemmed from acoustic differences between the noises. We used a computational model to investigate whether outstanding acoustic features could predict the high-performance blocks. Noises were passed through a simulation of the auditory periphery, and outstanding features were identified by comparing results for a given noise sample to the distribution observed for Gaussian noises. No correlation was found between behavioral performance and the acoustic features tested (peaks in the amplitude envelope and peaks in the spectro-temporal envelope; see Figures S2A and S2B).

It is possible that other feature extraction techniques would reveal acoustic correlates of performance. However, if some noises are intrinsically easier to learn than others, this should hold true for all participants. We tested this with a new experiment, similar to experiment 1, except that the same reference noises were used across all participants. Furthermore, each participant completed two blocks of each noise, with the original objective of allowing a test-retest comparison. To maximize the potential differences, the RefRNs were selected from those of experiment 1, choosing five noises that were learned well and five that were poorly learned, here termed “good” and “bad” noises.

Figure 3A shows the average d' s observed for the good and bad noises. An independent-samples t test on the d' s measured for each block showed that listeners were more sensitive to the good noises than the bad noises (mean $d' = 2.7$ versus 2.0; $t_{98} = 3.83$, $p < 0.001$). However, the difference in sensitivity to the two groups of noises was relatively small.

The average sensitivity to RefRNs overall was higher than in experiment 1 ($d' = 2.3$ versus 1.0–1.4). The improvement in score seems to have come, not from learning the RefRN faster—it would be difficult to learn the RefRN observably faster than the listeners in experiment 1—but from learning more of the RefRN stimuli. The criterion used in experiment 1 to define a high-performance block ($\geq 90\%$ correct over the last 40 trials) was achieved on 67% of blocks in experiment 2, compared to only 31% of blocks in experiment 1. More experienced listeners thus seem to be able to learn more of the noises.

Figures 3B and 3C show the hit rates for the RefRN and RN noises split between the first blocks and the second blocks

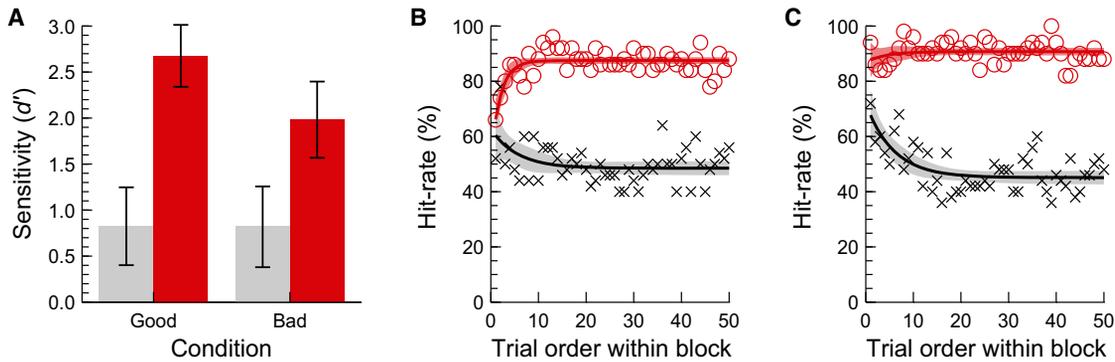


Figure 3. Good Noise, Bad Noise, and Long-Term Memory Traces

(A) Mean sensitivity to good and bad noises in experiment 2 (red) and the mean sensitivity to RN in the same blocks (gray). Error bars are 95% confidence intervals. There is a small but significant advantage for good noises, which was not correlated to outstanding acoustic features in the temporal or spectro-temporal envelope (Figures S2A and S2B). Individual data also failed to reveal a strong consensus across listeners for good and bad noises (Figure S2C).

(B) The variation of RN (red) and RefRN (gray) hit rates throughout the blocks in experiment 2, averaged over the first presentation of the good and bad noises in the first ten blocks. The lines show the best-fit exponential lines for each condition, and the surrounding bands show the 95% confidence intervals of these fitted exponentials.

(C) The equivalent data for the second presentation of the ten noises in the second ten blocks for each listener. There is an advantage for RefRN right from the start of the block, showing that learning carried over between the two presentations (mean interval 16.5 days).

that used the same RefRN samples. In the first presentation, listeners were not significantly more likely to detect repetitions in the first trials of RefRN and RN ($F_{1,94} = 0.75$, $p = 0.39$); which was statistically tested using the “extra sum-of-squares test” (Motulsky and Christopoulos, 2004), comparing the independent fitting of three-parameter exponentials for RefRN and RN with an equivalent model whose initial value was fitted globally for both RefRN and RN. This suggests that the good and bad noises selected were, on average, approximately equivalent to the RN noises in terms of repetition detection before learning. In contrast, in the second presentation, the RefRN hit rate was significantly greater than the RN hit rate from the first trial ($F_{1,94} = 13.60$, $p = 0.004$). In other words, in the second presentation, the listeners had greater sensitivity for the RefRN right from the start of the block. In fact, no significant learning was observed for RefRNs in the second presentation ($F_{2,47} = 0.64$, $p = 0.53$). Both findings show that some learning from the first presentation must have carried over to the second presentation.

The mean time between blocks of identical RefRN stimuli for the same listener was 16.5 days. However, the times between the first and second presentations for a particular RefRN varied greatly with noise and listener, because the noises were presented in a random order and the listeners completed this long experiment in their own time. We split the data in long and short intervals between blocks to evaluate the effect of duration on memory persistence. For the longest interblock intervals for each listener (median 20 days), there was a nonsignificant trend of relearning in the second block ($F_{2,47} = 2.78$, $p = 0.07$), but the small amount of relearning (from a starting point of 86% to an asymptote of 93%, compared to the initial hit rate of 64% in the first blocks) suggests that at most a few noises were forgotten and then relearned. No such trend was present for the shortest interblock intervals (median 10 days). Thus, the listeners sustained memories of several RefRNs for considerably long periods of time.

Experiment 3: Identification of Unrepeated Noises

So far, all stimuli included a repetition. Although the learning of reference *repeated* noises has been observed, this does not necessarily entail that the half-second noise samples that constitute the RefRNs were learned. It is hypothetically feasible that learning was limited to the noise in its repeated form: the listener could be learning the RefRN as a whole or could be sensitive to patterns of slow modulations introduced by the repetition. Experiment 3 tested whether the listeners had learned the individual noise samples in experiment 2 by asking them to report which of a short block of half-second samples corresponded to the RefRN they heard in the immediately preceding block. Thus, each participant completed the blocks of this experiment alternated with blocks from the previous experiment. Given how quickly listeners were able to learn the RefRNs in the previous experiments, there was a risk that listeners could also learn the unrepeated reference noises, even if they were not able to associate them with the RefRNs learned in the preceding block. Thus, a “decoy” noise was also included in each block, which also remained the same from trial to trial, but which had not previously been presented to the listener.

Results show that the average d' for the reference noise ($d' = 1.3$) was significantly greater than that of the decoy noise ($d' = 0.2$; $t_4 = 5.12$, $p = 0.007$). Listeners were able to recognize clearly the 0.5 s reference noise sample when presented on its own. Listeners also had a small but significant tendency to report recognizing the decoy noise itself ($t_4 = 9.95$, $p = 0.001$), despite not being instructed to do so. This suggests that listeners learned the decoy noise to some extent, even though it was never presented in a repeated form.

Experiments 4–6: Robustness to Duration and Acoustic Transformations

In this last series of experiments, we probed the effect of acoustic parameters on the learning mechanism by testing

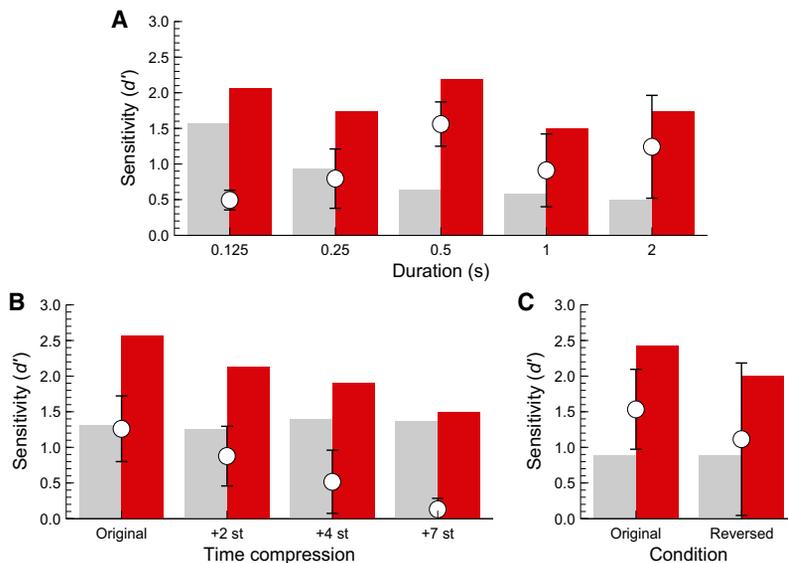


Figure 4. Robustness to Duration and Acoustic Transformations

(A) Results for experiment 4. The mean recognition of RN (gray) and RefRN (red) for noises of different durations and the mean difference (open circles) with error bars showing its 95% confidence interval. An advantage for RefRN is observed for all durations.

(B) Results for experiment 5. Performance after time compressions and frequency shifts (expressed in semitones) of a previously learned RefRN. Format as in (A). The analysis is based on the first ten trials after acoustic transformation to reduce the influence of relearning (the last ten trials are shown in Figure S3). The RefRN advantage diminishes progressively with greater transformation. (C) Results for experiment 6. The acoustic transformation is now a time reversal. Format as in (B), except the mean RN sensitivity is plotted twice for display purposes. Time reversal had no significant effect on the RefRN advantage.

a range of sound durations and by introducing changes after learning.

Experiment 4 was a replication of experiment 1, but with noise samples longer (1 s and 2 s) or shorter (0.125 s and 0.25 s) than the 0.5 s samples used so far. The results are shown in Figure 4A. Sensitivity to RefRN was greater than RN on average ($F_{1,4} = 167.90$, $p < 0.001$), and in fact for all listeners at all durations. There was also a condition-duration interaction ($F_{1,84,7.37} = 6.21$, $p = 0.03$), with more learning observed for the 0.5 s noises than for the shorter noises. A trend for an increase in RN sensitivity with shorter sample durations was also apparent on the average data, indicating that the repetition-detection task may have been easier for short samples and/or faster repetitions. However, there were large between-subject differences at the shortest durations. Here, the repetitions were likely perceived as infrapitch (Warren and Bashford, 1981), which some listeners may not have interpreted as indicative of a repetition. At the other end of the duration range, it is remarkable that learning was as good for the 2 s samples as for the 0.5 s samples.

Experiments 5 and 6 introduced acoustic transformations to a noise sample that had just been learned, to test whether recall would tolerate these changes. After 80 trials containing 20 RefRN presentations in which listeners had a chance to learn a given RefRN (same procedure as experiment 1), trials with acoustic transformations were randomly inserted in the experimental block, without telling the listeners and without feedback. In experiment 5, the transformation was a time compression: all features putatively learned in the RefRN were presented in faster succession with an accompanying frequency shift (0, +2, +4, or +7 semitones, depending on the condition). To make sure that listeners could not identify the transformed trials on the basis of duration, all stimuli were padded to 0.5 s. Results showed that time-compressed RefRNs were significantly better detected than RNs for all conditions ($t_4 = 4.07$ – 10.69 , $p = 0.001$ – 0.02), but sensitivity to RefRN depended on the extent of the shift ($F_{3,12} = 3.72$, $p = 0.04$). This stimulus-shift interaction was driven by the first ten trials after transformation ($F_{3,12} = 8.14$, $p = 0.003$;

Figure 4B), with no such effect over the last ten trials ($F_{3,12} = 0.24$, $p = 0.87$; Figure S3A). Thus, only the large time compressions impaired recall for previously learned RefRN, which were then relearned over the course of an experimental block. In experiment 6, the acoustic transformation was a time reversal of the previously learned RefRN. Perhaps surprisingly, this major acoustic transformation only resulted in a nonsignificant trend in reduction in sensitivity compared to the original RefRN ($d' = 2.1$ versus 2.3, $t_4 = 0.66$, $p = 0.55$; Figures 4C and S3B). Both experiments therefore show that recall is tolerant to a range of spectral and temporal transformations.

DISCUSSION

Characteristics of Learning: Unsupervised, Fast-Acting, and Long-Lasting

In the experiments described above, listeners were better able to detect repetitions in noise samples that were presented several times throughout an experimental block, compared to noise samples that were generated afresh for each trial. These two types of stimuli (RefRN and RN) were generated in the same manner, the only difference being that identical copies of the RefRN stimulus reoccurred within a block. Thus, repeated exposure to a random waveform, up to 2 s long, resulted in the learning of acoustic details of the waveform.

In the paradigm introduced here, listeners could not have known, a priori, which of the trials to learn. These could only be distinguished after a memory trace of the particular RefRN presented in the experimental block occurred. The listeners thus learned the RefRN in the absence of explicit exemplars and without feedback, so the learning was unsupervised. In another sense, the learning was “implicit,” in that the listeners were not made aware that the same frozen noises would reappear throughout the block, so they had no incentive to learn individual noises. The listeners were, however, encouraged to process the sound. Depth of processing is known to aid the formation of memories (Craik and Lockhard, 1972) even when

the stimulus to be learned simply co-occurs with increased attention to a task (Seitz and Dinse, 2007) or even when it is paired with a reward in the absence of a task (Seitz et al., 2009). Still, the lack of any instruction to the participants to memorize the stimuli reduces the likelihood that listeners could have benefited from some learning strategies, such as mental rehearsal. Furthermore, noise is devoid of obvious features that could provide a basis for mental rehearsal. Rather, the memories must have been stored relatively automatically.

When learning occurred, it occurred surprisingly fast. In many cases, listeners tended toward perfect performance with a half-life of about two trials. After these two trials (four presentations of the noise), enough acoustic details were memorized for near-perfect performance. Fast perceptual learning has been reported before for frequency discrimination (Hawkey et al., 2004), interaural time differences (Ortiz and Wright, 2009), visual texture segregation (Karni and Sagi, 1993), and visual spatial discrimination (Poggio et al., 1992). However, the noise learning exhibited here occurred on a much shorter timescale than was observed in these previous studies.

In addition to being fast, the learning was robust and long-lasting. Once listeners learned a noise in experiment 1, they maintained almost perfect performance, despite the interference of intervening RN and N trials, which also had to be actively processed and which contained sounds statistically similar to the sample to be learned. Long-term auditory memories were seen in experiment 2, in which listeners could remember multiple noises they had heard in previous blocks on different days. Although the time between blocks varied, the average time interval between repeats was more than 2 weeks. Listeners thus retained memories for multiple individual noises over the course of days or even weeks, despite the potential interference of everyday listening.

The characteristics of the learning we observed for noise samples (unsupervised, fast-acting, resistant to interference, and long-lasting) would be highly desirable for learning about the structure of the acoustic environment in realistic situations. To our knowledge, this is the first time that auditory learning with such ecologically relevant characteristics has been revealed by means of a psychophysical paradigm. The paradigm uses meaningless sounds, so it may also be useful for investigations with animal models.

Our results are also reminiscent of the phenomenon of “insight” (Köhler, 1925), a sudden improvement in task performance that is often long lasting. Although insight is normally associated with high-level cognitive tasks, perceptual insight has been demonstrated for stimulus-specific features in vision (Rubin et al., 1997, 2002). In our experiments, after listeners learned a given noise sample, they were able to detect it on almost every trial. This shows that the differences between the noises were highly salient after learning, even though the exact same features were initially largely ignored by the listeners. Our observations may thus be the first demonstration of insight for “low-level” acoustic details of auditory signals.

Features in Noise

Noise has much less structure and predictability than natural sounds. It thus seems unlikely that listeners memorized the

whole waveform that they were able to recognize. Rather, a subset of discriminative features may have been used. What could these features be?

Evaluating the performance of different listeners on the same set of noise samples is a first way to address this question. In experiment 2, we found that some noises were indeed easier to learn than others, but this difference was small. Most of the more experienced listeners are able to learn most of the noises. In addition, acoustical analyses failed to reveal any obvious differences between good and bad noises. These observations are consistent with the data of Kaernbach (1993), who measured the ability of listeners to tap in time to repetitions in frozen noise. He observed both self-consistency and interindividual differences in the tapping. Given the present data and Kaernbach’s, at least some of the features learned in noise are likely to be idiosyncratic; that is, they will not always be the same for different listeners and the same noise.

The effect of acoustic transformations on recall further constrains the nature of features. Time reversal of a previously learned sample had a surprisingly small influence on recall, whereas increasingly severe spectro-temporal transformations eventually led to poorer recall (even though some tolerance to this transformation was also observed). This pattern of results would be predicted if the features learned in noise were local in time: for instance, listeners may have isolated a few random short-term spectral shapes from each sample (Guttman and Julesz, 1963). It is unclear yet whether this finding is a general characteristic of auditory learning or whether it derives from the statistics of the noise which, by definition, does not contain long-term spectro-temporal regularities.

As noise contains an arbitrarily large number of random features, a selection process is needed to single out features distinctive enough to support almost-perfect performance after learning. Automatic saliency-detection mechanisms may be involved (Itti et al., 1998), but they alone are unlikely to account for our results: some features seem to be idiosyncratic, and current auditory-saliency models (Kayser et al., 2005) would not predict any highly salient features in noise. Feature reduction could also be implemented by a form of competition within feature space (Desimone, 1996). Repetition suppression (Grill-Spector et al., 2006) or stimulus-specific adaptation (Ulanovsky et al., 2003) are widely observed phenomena that may reflect the changes in neural responses for selected and repeated features.

Top-down processing could also accelerate the learning of noise, by actively picking out subsets of features. Reverse-hierarchy theory (Ahissar et al., 2009; Hochstein and Ahissar, 2002) suggests that listeners are initially only aware of the “gist” of a sound (e.g., that it is a noise) but can then access lower-level features of the sound if required by the task. In the course of this “perceptual mining” of the noise samples, perhaps encouraged by our ancillary repetition-detection task, a subset of the low-level features could be selected and committed to long-term memory.

Possible Neural Mechanisms

How might the brain detect the presence of repeating patterns of sensory activity in the auditory system? Tentatively, several characteristics of the data are consistent with a modulation of

synaptic weights following a spike-timing-dependent plasticity (STDP) rule (Markram et al., 1997). Recent modeling studies have shown that neurons equipped with a simple STDP rule reliably become selective to random patterns of afferent activity, provided that these patterns are presented repeatedly (Masquelier et al., 2008, 2009). Importantly for the present experiments, which used stationary broadband noise, the selectivity emerged even when no change in firing rate was associated with the repeated pattern. In addition, STDP detected coincident activity among afferents, which would be largely preserved for time-compressed or time-reversed versions of the afferent pattern. Finally, given the plausible assumption that auditory afferents have relatively broad frequency tuning, the selectivity would be tolerant to moderate frequency shifts. These predictions are consistent with the effects of repetition and acoustic transformations measured here.

The extreme speed of the learning we observed remains an issue for any plasticity rule. However, the efficiency (and speed) of STDP is highest when spike timing is reliable and precise (Markram et al., 1997). Many stages in the auditory pathways display highly accurate and reproducible spike-timing characteristics, including auditory cortex (Elhilali et al., 2004). Combined with a sparse representation of complex signals (Hromádka et al., 2008), this may be the key to fast plasticity for repeated auditory events.

The neural substrate of the plastic changes supporting the memory of noise is another important outstanding issue. The current consensus, in several sensory modalities, is that memory encoding and retrieval recruit a widely distributed network (Alain et al., 1998; Harris et al., 2001). Rapid adaptive plasticity, defined as task-dependent changes in the feature selectivity of sensory neurons, is well documented in primary and secondary auditory cortex (Atiani et al., 2009; Fritz et al., 2003). However, the neural substrate of auditory memory formation may also encompass subcortical areas, as plasticity (Tzounopoulos and Kraus, 2009) and adaptive coding (Dean et al., 2005) have been observed in the auditory brainstem.

EXPERIMENTAL PROCEDURES

Experiment 1

Participants

There were 12 listeners with self-reported normal hearing, aged between 19 and 55. They had not previously participated in experiments involving repeated-noise stimuli. Six of these “naive listeners” returned as “trained listeners” and repeated the experiment with different RefRNs.

Stimuli

The stimuli were formed from Gaussian noises, generated as sequences of normally distributed random numbers at a sample rate of 44.1 kHz and a 16 bit amplitude resolution. Each N stimulus was a 1 s sample of noise; each RN or RefRN stimulus was formed from a 0.5 s sample of noise, concatenated to an identical copy of itself without any intervening silence. Both N and RN were generated afresh for each presentation, but RefRN was identical within an experimental block.

Procedure

Each block consisted of 100 N trials, 50 RN trials, and 50 RefRN trials. Thus, half of the trials featured repeated noise, and half were unrepeated noise. The trials were pseudorandomly ordered, with the restriction that the RefRN was never presented on two consecutive trials. After each stimulus presentation, listeners had to report whether or not they heard a repetition in a “yes-no” task. Each listener completed four of these blocks, with each block based on

a different RefRN. No feedback was given. As part of experiment 1, minimal training was provided: the listeners were initially presented with demonstration sounds using ten repetitions of half-second noise samples, resulting in 5 s stimuli. The number of repetitions was then incrementally decreased. The listeners did not experience a 1 s, repeated stimulus until the first experimental block. The training and subsequent four blocks were presented in a single session.

Analysis

Hit rates and false-alarm rates were analyzed in terms of the sensitivity (d') and criterion (c) measures of signal detection theory (MacMillan and Creelman, 2001). Specifically, the d' s for RN and RefRN were calculated from their respective hit rates (H_{RN} and H_{RefRN}) and the false-alarm rate for unrepeated noise (F):

$$d'_{RN} = z(H_{RN}) - z(F)$$

$$d'_{RefRN} = z(H_{RefRN}) - z(F)$$

Note that RN and RefRN stimuli were presented in the same block, so they share a common false-alarm rate. Because of this, it is not possible to measure two d' s and two criteria independently. However, since participants are known to be unable to maintain separate criteria (Gorea and Sagi, 2000), we consider a single measure of criterion, based on the average of the two hit rates.

$$c = \frac{-1}{2} \left(z \left(\frac{H_{RN} + H_{RefRN}}{2} \right) + z(F) \right)$$

The d'_{RN} , d'_{RefRN} , and c were calculated for each block based on the 50 RN, 50 RefRN, and 100 N trials, then averaged to get a mean value for each listener where required.

Apparatus

Stimuli were played through an RME Fireface digital-to-analog converter with 16 bit resolution at 44.1 kHz sample rate. They were presented to both ears simultaneously through Sennheiser HD 250 Linear II headphones. The presentation level was 70 dB(A). Listeners were tested individually in a double-walled IAC sound booth. The apparatus remained identical for all subsequent experiments.

Experiment 2

Participants

There were five listeners with self-reported normal hearing, aged between 21 and 30, including the first author. Three listeners and the first author had previously participated in experiment S1 (see Supplemental Information); all listeners but the author had participated in experiment 1.

Stimuli

The N and RN stimuli were generated as in experiment 1, while the RefRN stimuli were identical to ten of the RefRN stimuli from experiment 1. Specifically, five “good noises” and five “bad noises” were chosen. The five good noises were those that elicited the highest d' s in experiment 1, selecting at most one noise from each listener. This restriction was to promote the selection of most learnable noises rather than the noises used by the best learners of noise. The five bad noises were those that elicited the lowest d' s in experiment 1, again selecting at most one noise from each listener, with two further restrictions: (1) the listeners had to have already scored $d' \geq 1$ for a RefRN in a preceding block, to show that they were capable of learning the RefRN stimulus, and (2) the listener had to have scored $d' \geq 0.5$ for the RN in the same block, to show that the poor learning of the RefRN was not due to a lack of concentration. The d' s observed in experiment 1 were 2.9–3.9 for the good noises and 0.5–1.2 for the bad noises. Note that the reuse of the frozen noises from experiment 1 meant that some listeners had already heard some of the RefRNs of experiment 2 ~4 months previously. Such long-term memories were not anticipated when designing the experiment but could have affected the initial RefRN hit rate in 9 of the 50 participant-noise combinations. However, any effect did not lead to a significant difference in the initial values of the RN and RefRN hit rates ($F_{1,94} = 0.75$, $p = 0.39$).

Procedure

The procedure was the same as for experiment 1, except that there were 20 blocks, completed in sessions with durations of the listeners' preference.

The first ten blocks included the five good noises and five bad noises as RefRNs, presented in a random order for each listener. The last ten blocks were based on the same ten RefRNs in a different random order. Each of the 20 blocks was succeeded by a short block testing the recognition of the frozen RefRN sample presented without repetition, here described as experiment 3.

Experiment 3

Participants

The participants were the same as for experiment 2.

Stimuli

Each stimulus was a 500 ms sample of Gaussian noise, just like the ones used to form the repeated noises of the previous experiments. Fresh noise trials were generated anew for each trial, with 500 ms duration. The reference noise was always identical to the un-repeated half of the RefRN from the preceding block (see experiment 2). The decoy noise was generated like a fresh noise in the first instance, but then reoccurred throughout a block unchanged, like the reference noise.

Procedure

Each block was formed from 20 reference noises, 40 fresh noises, and 20 decoy noises. The 80 trials were pseudorandomly ordered such that no two subsequent trials were formed from either two reference noises or two decoy noises. There were 20 such blocks, each based on a reference noise taken from the RefRN used in the preceding repetition-detection blocks (see experiment 2). Thus, these data were collected at the same time as the data for experiment 2. The five participants were asked, on each trial, to report whether or not they thought the presented noise was the noise from the preceding block. They were not told that there would also be a decoy noise, but two of the listeners asked why they were hearing a second noise recurring throughout the trial; the concept of the decoy noise was then explained to them, and they were instructed not to report the decoy noise as a target.

Analysis

Here, because there were never more than 20 target trials per block, hit rates and false-alarm rates were averaged across listeners before d' was calculated.

Experiment 4

Participants

There were five new naive listeners with self-reported normal hearing, aged between 21 and 28. A sixth listener was excluded whose RN sensitivity never exceeded $d' = 0.1$, although some learning for RefRN was still observed for this listener at all durations. None of the listeners had previously participated in experiments involving repeated-noise stimuli.

Stimuli

The RN and RefRN stimuli were created as in experiment 1, but with each segment of RN or RefRN being 0.125, 0.25, 0.5, 1, or 2 s in duration. The N stimuli were double those lengths, so that all stimuli had the same total durations within a block.

Procedure

Each block was formed in the same manner as experiment 1, except that the durations of the stimuli varied from block to block. Each of the five durations was presented in five separate blocks (in a random order), and this process was repeated four times to form a total of 20 blocks. As for experiment 1, listeners were minimally trained to detect repetitions without use of any RefRN, but here the training used the full range of durations in the main experiment.

Analysis

As for experiment 1, d' and c were calculated for each block then averaged across blocks of equal duration for the same listener.

Experiment 5

Participants

There were five listeners with self-reported normal hearing, aged between 24 and 31. All listeners had previously participated in experiment 6 and some of the preceding experiments.

Stimuli

N, RN, and RefRN were created as for experiment 1. Each listener was presented the same six RefRN stimuli. Time compressions were applied to N,

RN, and RefRN stimuli. Stimuli were resampled at ratios of 8:9, 4:5, or 2:3, resulting in frequency shifts of +2, +4, or +7 semitones. Before resampling, the first and second halves of each stimulus were padded with additional samples of Gaussian noise (repeated for RN or RefRN and different for N) so that the processed stimuli had the same duration as the original stimuli. The padding was fresh for each trial, including the RefRN trials.

Procedure

Each block started with a "learning phase", which was formed in the same manner as experiment 1, but with just 20 RefRN, 20 RN, and 40 N trials. A "testing phase" immediately followed, unbeknownst to the listeners, as part of the same block. In the "testing phase", there were 20 RN, 20 RefRN, and 40 N for each of the four time compression ratios. The ordering within the testing phase was pseudorandom, such that no two subsequent trials were identical, but time-compressed counterparts of the same RefRN could be presented on subsequent trials. Note that listeners were asked to report whether they had detected a repetition, not whether they had recognized the sound.

Analysis

As for experiment 3, hit rates and false-alarm rates were averaged within listener before d' and c were calculated. Six blocks were excluded because insufficient RefRN learning was observed (defined as less than 15 out of 20 RefRN trials in the learning phase successfully reported as containing repetitions). These excluded blocks were reasonably evenly distributed across the listeners.

Experiment 6

Participants

The participants were the same as for experiment 5, although they completed experiment 6 first. All listeners had previously participated in some of the earlier experiments.

Stimuli

N, RN, and RefRN were created as for experiment 1. Each listener was presented the same six RefRN stimuli. A reversed RefRN was also generated, by simply reversing the order of the waveform samples.

Procedure

Each block started with a "learning phase," which was formed in the same manner as in experiment 5. The testing phase was formed from 20 RefRN, 20 reversed RefRN, 40 RN, and 80 N trials. The ordering within the testing phase was again pseudorandom, such that no two subsequent trials were identical, but RefRN and its reversed counterpart could be presented on subsequent trials.

Analysis

As for experiments 3 and 5, hit rates and false-alarm rates were averaged within listener before d' and c were calculated. Six blocks were excluded because insufficient RefRN learning was observed (defined as less than 15 out of 20 RefRN trials in the learning phase reported as containing repetitions). These excluded blocks were reasonably evenly distributed across the listeners.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and Supplemental Experimental Procedures, including Supplemental Experiment S1 and computational models, and can be found with this article online at [doi:10.1016/j.neuron.2010.04.014](https://doi.org/10.1016/j.neuron.2010.04.014).

ACKNOWLEDGMENTS

This work was supported by grant ANR-06-NEUR-O22-01. We thank Laurent Demany, Israel Nelken, and Misha Tsodyks for comments on earlier versions of this manuscript. We also thank the reviewers for suggestions that led to the design of experiments 4–6. Parts of the data have previously been reported at the International Symposium of Hearing (Agus et al., 2009).

Accepted: April 2, 2010

Published: May 26, 2010

REFERENCES

- Agus, T.R., Beauvais, M., Thorpe, S.J., and Pressnitzer, D. (2009). The implicit learning of noise: Behavioral data and computational models. In 15th International Symposium on Hearing (June 2009), E.A. Lopez-Poveda, A.R. Palmer, and R. Meddis, eds. (Salamanca, Spain: Springer-Verlag).
- Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 285–299.
- Alain, C., Woods, D.L., and Knight, R.T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Res.* 812, 23–37.
- Atiani, S., Elhilali, M., David, S.V., Fritz, J.B., and Shamma, S.A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* 61, 467–480.
- Baddeley, A.D. (1997). *Human Memory: Theory and Practice*, Revised edition (Boston, MA: Allyn and Bacon).
- Brady, T.F., Konkle, T., Alvarez, G.A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. USA* 105, 14325–14329.
- Cohen, M.A., Horowitz, T.S., and Wolfe, J.M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proc. Natl. Acad. Sci. USA* 106, 6008–6010.
- Cowan, N. (1984). On short and long auditory stores. *Psychol. Bull.* 96, 341–370.
- Craik, F.I.M., and Lockhard, R.S. (1972). Levels of processing: A framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684.
- Dean, I., Harper, N.S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* 8, 1684–1689.
- Demany, L., and Semal, C. (2008). The role of memory in auditory perception. In *Auditory Perception of Sound Sources*, W.A. Yost, A.N. Popper, and R.R. Fay, eds. (New York: Springer Verlag), pp. 77–113.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. USA* 93, 13494–13499.
- Elhilali, M., Fritz, J.B., Klein, D.J., Simon, J.Z., and Shamma, S.A. (2004). Dynamics of precise spike timing in primary auditory cortex. *J. Neurosci.* 24, 1159–1172.
- Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223.
- Gorea, A., and Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proc. Natl. Acad. Sci. USA* 97, 12380–12384.
- 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.
- Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23.
- Guttman, N., and Julesz, B. (1963). Lower limits of auditory analysis. *J. Acoust. Soc. Am.* 35, 610.
- Hanna, T.E. (1984). Discrimination of reproducible noise as a function of bandwidth and duration. *Percept. Psychophys.* 36, 409–416.
- Harris, J.A., Petersen, R.S., and Diamond, M.E. (2001). The cortical distribution of sensory memories. *Neuron* 30, 315–318.
- Hawkey, D.J., Amitay, S., and Moore, D.R. (2004). Early and rapid perceptual learning. *Nat. Neurosci.* 7, 1055–1056.
- Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.
- Hromádka, T., Deweese, M.R., and Zador, A.M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* 6, e16.
- Itti, L., Koch, C., and Nieber, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259.
- Jääskeläinen, I.P., Ahveninen, J., Belliveau, J.W., Raij, T., and Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends Neurosci.* 30, 653–661.
- Kaernbach, C. (1993). Temporal and spectral basis of the features perceived in repeated noise. *J. Acoust. Soc. Am.* 94, 91–97.
- Kaernbach, C. (2004). The memory of noise. *Exp. Psychol.* 51, 240–248.
- Karni, A., and Sagi, D. (1993). The time course of learning a visual skill. *Nature* 365, 250–252.
- Kayser, C., Petkov, C.I., Lippert, M., and Logothetis, N.K. (2005). Mechanisms for allocating auditory attention: an auditory saliency map. *Curr. Biol.* 15, 1943–1947.
- Köhler, W. (1925). *The Mentality of Apes* (London, New York: K. Paul, Trench, Trubner & co.; Harcourt, Brace & Company, inc.).
- MacMillan, N.A., and Creelman, C.D. (2001). *Detection Theory: A User's Guide* (Mahway, NJ: Lawrence Erlbaum Associates, Inc.).
- Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215.
- 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.
- Masquelier, T., Guyonneau, R., and Thorpe, S.J. (2009). Competitive STDP-based spike pattern learning. *Neural Comput.* 21, 1259–1276.
- Motulsky, H., and Christopoulos, A. (2004). *Fitting Models to Biological Data using Linear and Nonlinear Regression* (New York: Oxford University Press).
- Ortiz, J.A., and Wright, B.A. (2009). Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 188–194.
- Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J.D., Grube, M., Carlyon, R.P., and Griffiths, T.D. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biol.* 5, e288.
- Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science* 256, 1018–1021.
- Rubin, N., Nakayama, K., and Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Curr. Biol.* 7, 461–467.
- Rubin, N., Nakayama, K., and Shapley, R. (2002). The role of insight in perceptual learning: Evidence from Illusory Contour Perception. In *Perceptual Learning*, M. Fahle and T. Poggio, eds. (Cambridge, MA: MIT Press).
- Seitz, A.R., and Dinse, H.R. (2007). A common framework for perceptual learning. *Curr. Opin. Neurobiol.* 17, 148–153.
- Seitz, A.R., Kim, D., and Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron* 61, 700–707.
- Tzounopoulos, T., and Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469.
- Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398.
- Warren, R.M., and Bashford, J.A., Jr. (1981). Perception of acoustic iterance: pitch and infrapitch. *Percept. Psychophys.* 29, 395–402.
- Warren, R.M., Bashford, J.A., Jr., Cooley, J.M., and Brubaker, B.S. (2001). Detection of acoustic repetition for very long stochastic patterns. *Percept. Psychophys.* 63, 175–182.