The detection of repetitions in noise before and after perceptual learning

Trevor R. Agus[a] and Daniel Pressnitzer
Laboratoire Psychologie de la Perception, École normale supérieure, Centre National de la Recherche Scientifique, 29 rue d'Ulm, 75230 Paris Cédex 5, France

(Received 1 December 2012; revised 22 March 2013; accepted 1 May 2013)

In noise repetition-detection tasks, listeners have to distinguish trials of continuously running noise from trials in which noise tokens are repeated in a cyclic manner. Recently, it has been shown that using the exact same noise token across several trials ("reference noise") facilitates the detection of repetitions for this token [Agus *et al.* (2010). Neuron **66**, 610–618]. This was attributed to perceptual learning. Here, the nature of the learning was investigated. In experiment 1, reference noise tokens were embedded in trials with or without cyclic presentation. Naïve listeners reported repetitions in both cases, thus responding to the reference noise even in the absence of an actual repetition. Experiment 2, with the same listeners, showed a similar pattern of results even after the design of the experiment was made explicit, ruling out a misunderstanding of the task. Finally, in experiment 3, listeners reported repetitions in trials containing the reference noise, even before ever hearing it presented cyclically. The results show that listeners were able to learn and recognize noise tokens in the absence of an immediate repetition. Moreover, the learning mandatorily interfered with listeners' ability to detect repetitions. It is concluded that salient perceptual changes accompany the learning of noise. © *2013 Acoustical Society of America*.

[http://dx.doi.org/10.1121/1.4807641]

PACS number(s): 43.66.Lj [ELP] Pages: 464–473

I. INTRODUCTION

Noise is frequently used to mask target stimuli in psychoacoustic experiments but is more rarely the target itself. One notable exception is when noise is repeated cyclically within a trial (Guttman and Julesz, 1963). In this case, over a broad range of repetition rates, listeners are able to distinguish repeated noise from running noise. Furthermore, we showed recently that repetition detection could be enhanced by presenting the exact same repeated noise in several trials, over the course of an experiment (Agus *et al.*, 2010). The enhanced performance for the noise token that was heard in several trials was taken as a sign of perceptual learning.

There are still unresolved questions, however, on the exact mechanisms supporting repetition detection and perceptual learning in those experiments—the two possibly being based on unrelated cues. Here, we first review the available evidence comparing and contrasting repetition detection and learning for noises. Then, we provide further experimental data to disentangle the two elements.

A. Repetition detection for noise

Guttman and Julesz (1963) first noted that when frozen-noise tokens were played cyclically, the perception was different from that of a continuously running noise, even at rates below the pitch range. Along with Warren and Bashford (1981), they described these infrapitch percepts as "motorboating" when the repetition rate of the frozen-noise was between ∼4 and 19 Hz, or "whooshing" for rates between ∼1 and 4 Hz. Slower repetition rates were also found to be detectable with practice, with lower limits variably reported to correspond to rates of 0.1–0.5 Hz (Guttman and Julesz, 1963; Warren and Bashford, 1981; Warren *et al.*, 2001). This corresponds to cyclic presentation of noise tokens as long as 10 s.

Kaernbach (1992, 1993) noted that with continued listening, features are perceived in repeated noises, such as "clanks" and "rasping." Listeners are able to tap in time to such perceptual features, and the same listeners tap in consistent locations for the same noises. Kaernbach (1993) explored the spectral extent of these features by pitting higher and lower spectral regions of the same noises against each other. The spectral locations of the features seemed to depend on the noise and listener, although some listeners appeared to be tapping to the same features in some of the noises. Using the same tapping technique, Warren *et al.* (2001) showed that when tapping to very long frozen noises (10–20 s), listeners were in fact tapping to relatively short sections of the noise (<3 s): When these 3 s frozen noises were presented aperiodically, separately by fresh noises of irregular duration, listeners continued to tap time-locked to the frozen noise. This suggests a percept based on individualized cues from the noises, potentially different from that of the "whooshing" or "motorboating" percept. The nature of such cues remains obscure, however, and there is the possibility that such cues are both listener- and noise-dependent (Kaernbach, 1993; Agus *et al.*, 2010).

B. Discrimination of noise tokens

The discriminability of two noise tokens drawn from the same random distribution has been investigated using

---

[a] Author to whom correspondence should be addressed. Electronic mail: trevor.agus@ens.fr

same-different tasks. This is roughly equivalent to repetition detection for noises presented just twice, except with an intervening silence between the two presentations. This allows for a dissociation of the inter-stimulus interval and the durations of the noise tokens. Surprisingly, presenting longer durations of noise (and thus a greater amount of information) does not improve performance; rather, performance decreases with duration for stimuli beyond approximately 25 ms (Hanna, 1984; Goossens *et al.*, 2008). This suggests that the limiting factor for noise discrimination is not a lack of information available at the auditory periphery, but too much information, somewhat reminiscent of informational masking (Durlach *et al.*, 2003; Goossens *et al.*, 2008). Note that when the tokens are embedded in noise, rather than separated with silence, best performance is observed for longer tokens (200 ms; Kaernbach, 2004). Here, the brief frozen-noise stimuli may be difficult to distinguish from their fresh-noise surrounds, again reminiscent of informational masking. Noise-token discrimination thus seems to depend on successfully isolating a small subset of cues from the large amount of information available.

C. Learning of noise tokens

A final set of experiments using noise as target observed long-term memory traces. In those experiments, the continued use of the same noise token[1] led to increased performance on a given task. For discrimination tasks, an improvement in performance with exposure to the noise token has been observed, at least for experienced listeners (Hanna, 1984, experiment 2b; Goossens *et al.*, 2008, experiment 3). Buus (1990) measured sound-level discrimination for noise bursts and found that difference limens were smaller when the same noise token was used on all trials. Finally, and although noise was not a target, it was found that presenting the same noise token as a masker for all trials in a pure-tone detection task led to lower signal thresholds, compared to when noise was generated afresh for each trial (Pfafflin, 1968). This was interpreted as a form of memory for the noise token used as a masker.

Recently, Agus *et al.* (2010) investigated again the memory for noise, this time by combining a repetition-detection task with multiple presentations of a same noise token over different trials. Listeners had to distinguish trials made of the seamless repetition of two identical segments of noise from trials where a single noise segment (of double duration) was presented. The task was thus to detect noise repetitions *within a trial*, similar to what was described in Sec. I A. However, an important difference with previous investigations was how trials containing a repetition were constructed. Half of those used noise tokens generated afresh for each trial (repeated noise, "RN"), so the repetition was purely within trial. The other half re-used the exact same noise token over the course of an experimental block (reference repeated noise, "RefRN"), so the repetition was *both within-trial and also across several trials*. Finally, the noise trials that did not contain any repetition ("N") used noise samples generated afresh for each trial. Listeners were not told of the presence of RefRNs. Moreover, between presentations of RefRNs, there were always intervening trials that contained Ns and/or RNs, which had to be processed actively and responded to. Thus, listeners did not know there was anything to learn in the experiment, nor could they have guessed which trials would have had to be learned. Still, listeners reported repetitions in the RefRN considerably more often than for the RN. Furthermore, when this occurred, the performance improvement was rapid (within a few presentations) and repetition detection became near perfect for the RefRN. This was interpreted as fast perceptual learning for noise.

D. What is learned in repeated noises?

In Agus *et al.* (2010), the effect of learning was clear: Listeners could report the RefRNs, which they had heard before, more often than the RNs, which were novel on each trial. However, it is not clear how listeners obtained this improved performance for RefRN. Indeed, the repetition-detection task was used as an ancillary measure of auditory memory. By asking listeners to detect repetitions within a trial, it was possible to obtain a performance measure without explicitly asking for the recognition of a noise token. This had two benefits. First, listeners were not told that there was anything to learn in the experiment, so unsupervised learning (as useful in realistic situations) could be investigated. In addition, piloting suggested that being asked to recognize a noise among other noises was confusing for naïve listeners: Most often, listeners would report that all noises sounded the same, which they probably initially did. However, the design of Agus *et al.* (2010) had the additional consequence that all RefRNs contained a within-trial repetition of the reference noise token. As a result, there are at least two possible accounts of the improved performance, which are not mutually exclusive.

A first possibility is that upon hearing a RefRN on multiple trials, listeners learned the corresponding reference noise token itself. This could have improved performance on the task in several ways. At one extreme, listeners may have learned to recognize the noise token and used this recognition as a surrogate for the repetition detection task, eventually reporting the familiar token irrespective of whether or not they perceived its repetition in a given trial. A weaker version of the token-learning account is that listeners were sensitized to some features in the reference noise token, which they could then spot more easily as repeating within a trial. In both versions of this interpretation, learning affects the perception of the noise token itself.

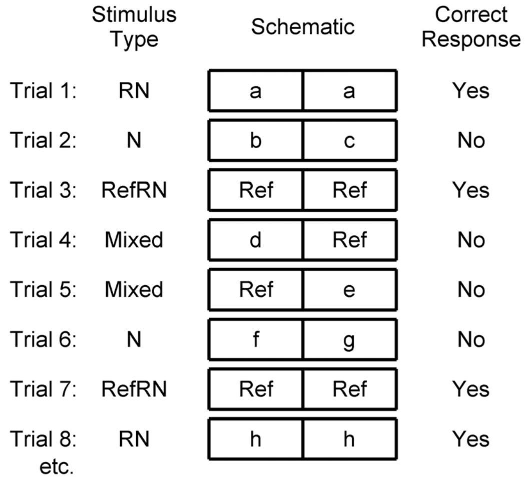
A second possibility is that listeners learned the pattern of modulation associated with the within-trial repetition associated to a RefRN. Consider what happens when a 0.5-s noise token is repeated to produce a one-second stimulus. Such a stimulus contains amplitude modulations at 2 Hz and its harmonics. Thus, repeated noises could be distinguished from running noise on the basis of the modulation spectrum alone. As RefRN trials re-occurred over the course of an experimental block, listeners may have been able to learn the specifics of the modulation spectrum for a given RefRN and hence improve their performance on the repetition detection task. Note that this is qualitatively different from learning the noise tokens themselves: In principle, learning could occur on a much-reduced representation of the noise, such as for instance the output of a 2-Hz filter in a modulation filterbank (Dau *et al.*, 1997). Also, a direct prediction of this hypothesis is that no learning nor retrieval should be possible when there is no immediate within-trial repetition of the reference noise tokens.

Deciding between the two possibilities has important consequences for the interpretation of both the classic noise repetition-detection tasks and the more recent findings about the memory for noise. In the following experiments, we use variants of the task of Agus *et al*. (2010) to clarify the nature of the cues that are used in a noise repetition-detection task, and how such cues may be shaped by perceptual learning.

II. EXPERIMENT 1: MIXED STIMULI, NAÏVE LISTENERS

Naïve listeners were asked to report repetitions in noise. As for Agus *et al*. (2010), half of the trials to be reported used a noise token generated afresh for each trial (RN), whereas the other half used the same reference noise token throughout an experimental block (RefRN). Here, in addition, half of the trials without any repetition incorporated the reference noise token that formed each half of the RefRN. These "Mixed" stimuli were thus formed by concatenating the 0.5-s reference noise token to 0.5 s of fresh noise that differed on each trial. Figure 1 shows a representative sequence of eight trials, with schematics of the four types of stimuli involved.

Thus, listeners heard the reference noise token in the absence of an immediate repetition. If listeners used recognition of the reference noise token to better report repetitions, then they might mistake the Mixed stimulus for the RefRN. This could be observed as a greater false-alarm rate for Mixed stimuli relative to N stimuli. On the other hand, if listeners were merely more sensitized to the repetitions in the RefRN, then the responses to the Mixed stimuli should be the same, on average, as the responses to the N stimuli. A third possibility is that listeners compared the first and second half of the noises to detect repeats. In this case, learning of the reference noise token could be used to more clearly identify the Mixed stimuli as non-repeating, observable as a reduction in false alarms to the Mixed stimuli.



	Stimulus Type	Schematic		Correct Response
Trial 1:	RN	a	a	Yes
Trial 2:	N	b	c	No
Trial 3:	RefRN	Ref	Ref	Yes
Trial 4:	Mixed	d	Ref	No
Trial 5:	Mixed	Ref	e	No
Trial 6:	N	f	g	No
Trial 7:	RefRN	Ref	Ref	Yes
Trial 8:	RN	h	h	Yes
etc.				

FIG. 1. A schematic showing the possible stimuli over the first eight trials of a block. Each stimulus is formed from two 0.5-s segments. The segments labeled "a" to "h" represent "fresh" noises that are presented once then never again, as opposed to the reference noise, labeled "Ref," which was the exact same noise token throughout a block. Note that some of the stimuli (RN and RefRN) contain a within-trial repetition, as their first and second segments are identical. The others (N and Mixed) do not contain any within-trial repetition, as their first and second segments are different. The reference noise token occurred in stimuli with (RefRN) and without (Mixed) a within-trial repetition.

A. Listeners

There were six listeners (3 male, 3 female) aged 21–28 yr old ($M=24$ yr), each with self-reported normal hearing. All participants were naïve as to the learning element of the experiment and had not previously participated in psychoacoustic experiments.

B. Stimuli

The stimuli were generated from Gaussian noise, generated as sequences of normally distributed random numbers at a sample-rate of 44.1 kHz and a 16-bit amplitude resolution. There were four types of stimuli, referred to as N, RN, RefRN, and Mixed. The N stimuli consisted of 1 s of noise. The RN stimuli were repeated noises, formed from a 0.5-s noise concatenated to an identical copy of itself without any intervening silence. Both the N and RN stimuli were generated afresh for each trial. In contrast, the RefRN stimuli were generated in exactly the same manner as the RN stimuli but were then presented identically on each RefRN trial throughout a block (then replaced by another RefRN at the start of the next block). The Mixed stimuli were formed from the same 0.5-s noise token used to generate the RefRN, but concatenated (either before or after) to 0.5 s of noise which was generated afresh for each trial. Thus neither the N nor the Mixed stimuli contained any within-trial repetition, whereas both the RN and RefRN contained a within-trial repetition. However, the Mixed and RefRN both contained the same reference noise token, which the listeners had the opportunity to hear across several trials over the course of a block.

C. Procedure

The experiment was run in a single session, consisting of training on the repetition-detection task followed by 12 experimental blocks, with a different RefRN stimulus for each block and each listener.

The training lasted approximately 30 min, during which listeners were trained to detect repetitions in noise, starting with ten repetitions of 0.5-s samples, then reducing the number of repetitions in stages to just three. During training, each stimulus was generated afresh for each trial; nothing analogous to the RefRN or Mixed stimuli was presented before the start of the main experiment.

During the main experiment, each block consisted of 20 N trials, 20 RN trials, 20 RefRN trials, and 20 Mixed trials (for those, in ten trials the reference noise token was presented before the fresh noise; and in ten trials, after). The proportions of stimulus types used in different experiments are illustrated in Fig. 2, including the original Agus *et al.* (2010) design (top panel) and the design used in the current experiment (middle panel). The trials were ordered pseudo-randomly, with the only restriction that the RefRN stimuli never occurred on subsequent trials. Listeners were asked to detect within-trial repetitions. They were not informed of the introduction of RefRN and Mixed stimuli until after the experiment.

The noise was presented at an overall level of 70 dB(A) in a double-walled sound-treated booth, through an RME Fireface UC soundcard and Sennheiser HD580 Precision headphones. After each trial presentation, listeners responded by a keypress.

D. Results

Figure 3 shows the mean hit rates and false-alarm rates for the RefRN, RN, N and Mixed stimuli (referred to as

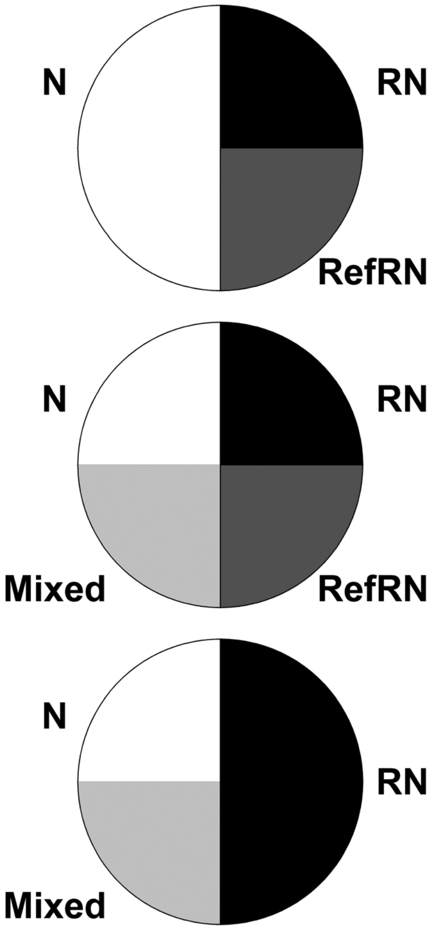


FIG. 2. A summary of the designs for the three experiments, illustrating the proportions of each stimulus. Note that half of the trials in each block do not contain any within-trial repetition (N and Mixed; left) and half contain a within-trial repetition (RN and RefRN; right). (top) The original design of Agus *et al.* (2010), also used in experiment 3. (middle) The design of experiments 1 and 2, incorporating the Mixed stimuli. (bottom) The design of experiment 3's Experimental condition, still including Mixed stimuli, but with no RefRN. The Baseline condition was the same as the Agus *et al.* (2010) design, as shown here in the top panel.

H_{RefRN}, H_{RN}, F_N, and F_{Mixed} throughout). Listeners were able to perform the primary task of within-trial repetition detection, in that H_{RN} was greater than F_N (35% vs 23%; $t_5 = 4.71$, $p = 0.005$), although the low d' ($M = 0.4$) highlights the difficulty of the task. Furthermore, listeners showed some learning of the RefRN stimuli, in that H_{RefRN} was greater than H_{RN} (70% vs 35%; $t_5 = 6.86$, $p = 0.001$). In fact, a preference for RefRN over RN was observed in 45 of the 72 individual blocks, defined as H_{RefRN} being significantly greater than H_{RN} as measured by Fisher's exact test (without corrections for multiple comparisons); the opposite result was observed significant in only 2 of the 72 blocks.

For the Mixed condition, F_{Mixed} was greater than F_N (44% vs 23%; $t_5 = 4.79$, $p = 0.005$). In other words, the inclusion of the reference noises *increased* the rate at which listeners incorrectly reported that they had detected a repetition. There were 26 blocks out of 72 blocks in which F_{Mixed} significantly exceeded F_N (again by Fisher's exact test) and none in the other direction.

The time-course of learning can be observed by looking at the average responses to the first presentations of each type of stimulus across different blocks and listeners, and comparing these to subsequent presentations. Figure 3 shows the proportion of repetitions reported for the *n*th presentation of each type of stimulus. H_{RefRN} and H_{RN} were initially similar, showing that there was nothing inherently more detectable about the repetitions in the RefRN stimuli. However, by the third presentation of RefRN, its hit rate was clearly greater than for RN, because of an increased hit rate for the RefRN but also a reduced hit rate for RN. Most of the increase had occurred by the 12th presentation. In parallel, the false-alarm rates for Mixed and N were initially similar, before diverging on a time-scale similar to that of the RN and RefRN stimuli.

E. Discussion

Naive listeners were able to make use of the re-occurrence of the RefRN to report repetitions more often than for RN. This replicates the primary result of Agus *et al.* (2010). The rate of the learning was also similar, with clear learning observed within just a few presentations of RefRN and most learning occurring within ten presentations. Learning was observed in a larger proportion of blocks (61%) than for naïve listeners in Agus *et al.* (2010) (33%). This could have been because the reference noise token was presented more frequently, since it also occurred as part of the Mixed stimuli. However, there are other differences between the two experiments that could have affected learning (including the set of listeners, a larger number of blocks per listener, and a smaller number of trials per block).

It is a curious result that listeners are more likely to report (incorrectly) repetitions for Mixed than N. These two conditions only differed in that the Mixed stimulus included a single presentation of the reference noise token that listeners had the opportunity to learn. This shows that the learning of the RefRN is not specific to its repeated form. Rather, the reference noise token was reported without any immediate repetition in the Mixed stimuli. Furthermore, the fact that the

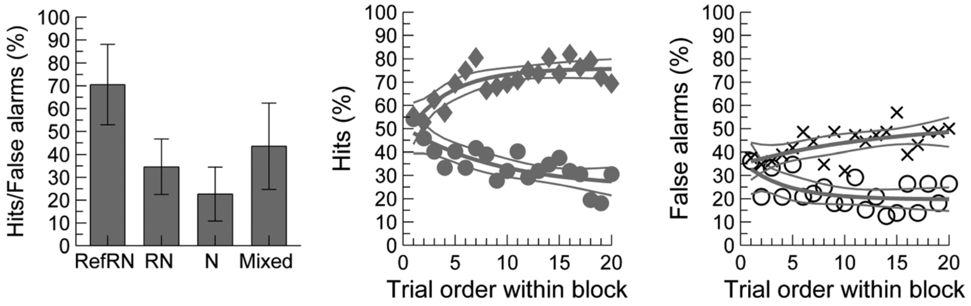


FIG. 3. (left) The mean hit rates and false-alarm rates for the four types of stimulus in experiment 1. Error bars are 95% confidence intervals centered on the mean. (middle) The time course of hit rates for RefRN (closed diamonds) and RN (closed circles) throughout the blocks in experiment 1. The thick gray lines show the best-fit exponential lines for each condition, and the surrounding thin lines show the 95% confidence intervals for these fitted exponentials. (right) Equivalent time courses of false-alarm rates for Mixed (crosses) stimuli and N (open circles).

presence of the reference noise token was erroneously reported as a repetition suggests that listeners were relying on the recognition of the reference noise token to provide their answers, instead of performing the task they were trained on and instructed to do, that is, a pure repetition-detection task. This task substitution may have been encouraged by the difficulty of the baseline task of detecting repetitions in RNs.

From the time-course of the learning, it is apparent that the learning resulted not just in more hits and false alarms for stimuli containing the reference noise, but also in fewer hits and false alarms for the RN and N stimuli. This phenomenon was also observed by Agus *et al.* (2010; see supplemental experiment S1), who interpreted it as a criterion effect. Essentially, listeners responding with a relatively constant proportion of yeses and nos will have to compensate the increased number of yes responses for RefRN by a decreased number of yes for RN. Consistent with this interpretation, here listeners' rate of reporting repetitions remained constant during a block: It was 42% at the start of a block and 43% by the end. The sensitivity to novel repetitions, calculated from RN and N responses alone, only decreased a little from $d' = 0.4$ at the start of a block to $d' = 0.2$ at the end (based on the fitted exponential curves shown in Fig. 3), but most of the decrease in hit rates and false-alarm rates was accounted for by the criterion increasing from $c = 0.2$ to $c = 0.7$.

III. EXPERIMENT 2: EXPERIENCED PARTICIPANTS

In the previous experiment, listeners were not told about the RefRN or Mixed stimuli. As such, recognition of the reference noise token may have caused confusion. Listeners were thus invited to repeat the experiment as "experienced listeners." This time, they were fully briefed on the design of the experiment: That there were RefRN and Mixed stimuli, and in particular that the Mixed stimuli should not be reported as containing a repetition. It was hypothesized that listeners would now be able to use their recognition of the reference noise token to generate fewer false alarms for the Mixed stimuli than the N stimuli, reversing the pattern of results observed in experiment 1.

A. Listeners

There were five listeners (2 male, 3 female) aged 21–28 yr old ($M = 23$ yr), each with self-reported normal hearing. All had previously participated as naïve listeners in experiment 1 but had since been debriefed as to the design and about the study. The sixth listener from experiment 1 was unavailable.

B. Stimuli and procedure

The stimuli were equivalent to those of experiment 1. The training was adapted for the experienced listeners. First, the four types of stimuli (N, RN, RefRN, and Mixed) were described to them, including the presence of the reference noise token and the repetitions. They were briefly tested verbally to confirm that they could label each stimulus correctly given a description of the components used to generate it, and that they also knew the appropriate response to each stimulus ("yes" for RN and RefRN; "no" for N and Mixed). It was emphasized to the listeners that if a noise sample was recognized but not repeating, this would point toward "Mixed" stimulus, for which the appropriate answer was "no." Listeners then completed a short block discriminating longer RN (4 repeats of 0.5-s noises) from N (2-s noises), all generated afresh for each trial, to remind them of the repetition-detection task.

Listeners then each completed 12 experimental blocks equivalent to those in experiment 1, but with new reference noises generated for each block and listener. The equipment and presentation levels were the same as those of experiment 1.

C. Results

Figure 4 shows the hit rates and false-alarm rates for the four conditions. Again, listeners reported more repetitions for RN than N (37% vs 14%; $t_4 = 8.00$, $p = 0.001$). Although the hit-rate for RefRN was higher than for RN (61% vs 37%), this difference was not significant across listeners ($t_4 = 2.39$, $p = 0.08$). Neither were listeners' false-alarm rates for the Mixed stimuli significantly different from the N stimuli ($t_4 = 1.70$, $p = 0.17$). If anything, the trend was for H_{Mixed} to be higher. Only one listener had an average F_{Mixed} lower than F_N, and the same listener was also the only listener to have a lower H_{RefRN} than H_{RN}.

Since all five of the experienced listeners had participated in experiment 1 as naïve listeners, we can make within-subject comparisons. It seems that the experienced listeners'

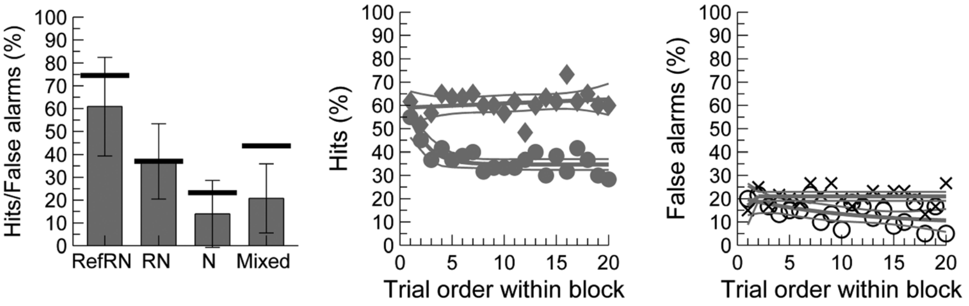


FIG. 4. Equivalent to Fig. 3, but for the results from experiment 2. The additional solid thick lines in the left panel indicate the mean hit rates and false-alarm rates for the same five participants in experiment 1, omitting its sixth participant who did not take part in experiment 2.

response strategy changed from when they performed the same task as naïve listeners, with an overall lower probability of reporting that a noise contained a repetition (33% compared to 45% in experiment 1). A 2×4 repeated-measures analysis of variance (ANOVA) on the hit rates showed a significant reduction in overall hit rates ($F_{1,4} = 19.83$, $p = 0.01$) and an effect of condition ($F_{3,12} = 33.35$, $p < 0.001$) and, importantly, there was no significant interaction ($F_{3,12} = 2.48$, $p = 0.11$).

Despite their experience, the listeners showed learning in the form of higher hit rates for RefRN than RN in only 26 blocks (43%), which was significantly fewer than the 39 blocks (65%) for the same listeners in the first experiment (Fisher's exact test; $p = 0.03$). This may be explained by a reduction in power due to the overall reduced hit rate The opposite pattern was only observed significantly in 5 blocks (8%). F_{Mixed} significantly exceeded F_{N} in 8 blocks, while the opposite was never observed.

The effects of the listeners' new response strategies can be seen more clearly in the time course of their responses. For the repeated stimuli (Fig. 4, middle panel), the learning of the RefRN is again observable as a bifurcation from the hit rates of the RN. But in contrast to experiment 1, the bifurcation is primarily due to a decreasing hit rate for the RN; the RefRN hit rate changes little across the block. This suggests that the criterion shifted while learning occurred. Likewise, for the stimuli without any within-trial repetition, the false-alarm rates for N decreased, while they remained the same for the Mixed stimuli. However, this occurred much closer to the floor than for the naïve listeners (see Fig. 3, right panel).

D. Discussion

The experienced listeners showed a largely similar pattern of results to their performance as naïve listeners, except with a generally reduced hit-rate. The more detailed instructions, including a description of the RefRN and Mixed stimuli, seems to have changed their response strategy. In particular, they were aware that, as naïve listeners, they had incorrectly reported many Mixed stimuli as containing a within-trial repetition. Their efforts to avoid this mistake may have contributed to the more conservative nature of their response strategy.

However, and most importantly for the hypothesis tested by experiment 2, listeners were unable to make use of any recognition of the Mixed stimuli to reduce their false-alarm rate relative to the N condition. Certainly, the experienced listeners' false-alarm rates for the Mixed stimuli were reduced in comparison to their performance as naïve listeners, but there was no significant interaction between condition and hit rate, suggesting that all observed effects stemmed from the main effect of reduced hit rates.

IV. EXPERIMENT 3: LEARNING NOISES TOKEN WITHOUT IMMEDIATE REPETITIONS

The first experiment showed that naïve listeners had a tendency to report that Mixed stimuli included a repetition, even though they only contained a single presentation of the reference noise token abutting with a different noise token. One reason proposed for this unexpected behavior was that listeners learned that the reference noise signaled a repetition in the RefRN stimuli, then they confused Mixed stimuli with the RefRN. If this was the case, then we should expect naïve listeners to avoid this mistake if they are not exposed to the reference noise token in trials containing a repetition. The current experiment tested this hypothesis by presenting listeners with N, RN, and the Mixed stimulus, omitting the RefRN stimulus altogether.

It should be noted that there is a second hypothesis implicitly tested here as part of the same design: That listeners can learn a reference noise token, even when it is not repeated within the same trial in the form of a RefRN. This may not be the case. The immediate repetition in the RefRN meant the listeners heard the reference noise token twice within 1 s, which is within the range of what has sometimes been termed the short auditory store (Cowan, 1984). If such a short auditory store plays a critical role in noise learning, then we would see no learning of the reference noise token presented only in the Mixed stimulus.

Since there was a possibility that no learning would be observed in the absence of the RefRN, a control condition was included to check that each listener was capable of learning noise, based on the original Agus *et al.* (2010) design (blocks formed from N, RN, and RefRN trials). Note that different reference noise tokens were used for each block, so that listeners would not hear the same reference noises in both contexts.

A. Listeners

There were six listeners (4 male, 2 female) aged 23–36 yr old ($M = 28$ yr), each with self-reported normal hearing. All participants were naïve as to the learning element of the experiment and had not previously participated in psychoacoustic experiments. Two further listeners were excluded: One had self-reported noise-induced hearing loss;

the other's results suggested she was unable to detect single repetitions in noise at all ($d' < 0.1$).

B. Stimuli and procedure

The stimuli were equivalent to those of experiment 1, with different reference noises.

There were two conditions. A Baseline condition was designed to replicate the results of Agus *et al.* (2010): Each block contained 80 N trials, 40 RN trials, and 40 RefRN trials (without Mixed stimuli). The Experimental condition presented the reference noise only in the form of Mixed stimuli: There were 40 N trials, 40 Mixed trials, and 80 RN trials (without RefRN stimuli; see Fig. 2). Thus, in each condition, there were as many noises that contained within-trial repetitions as otherwise, but the reference noise tokens were presented in the context of within-trial repetitions for Baseline blocks, and without within-trial repetition for Experimental blocks. Note that there are twice as many trials per block in this experiment compared to experiment 1: The reasoning was that since the reference noise was only presented once for every four trials, any learning might have required a greater number of trials.

Each listener completed eight of these blocks in a single session, consisting of four blocks of the Baseline condition and four of the Experimental condition, interleaved in randomly ordered pairs.

C. Results

Figure 5 shows the hit rates and false-alarm rates for the four conditions presented in the two types of block. On the left, in the Baseline condition, there were more hits for the RefRN than the RN stimuli, both greater than the false-alarm rate ($F_{2,10} = 39.23$, $p < 0.001$). On the right, for the Experimental condition, there were also significant effects of stimulus ($F_{2,10} = 10.44$, $p = 0.004$). Listeners were more likely to report repetitions in RN than N ($t_5 = 4.18$, $p = 0.009$). Again, false-alarm rates were higher for Mixed than N (32% vs 23%, $t_5 = 2.42$, $p = 0.01$).

Although on average F_{Mixed} was greater than F_N for five out of the six listeners, the Fisher test found this trend to be significant in only four of the 24 individual blocks. The opposite effect was found in only one block, as would be expected by chance alone.

Despite the small size of the effect, it can also be seen in the time course of the responses (Fig. 5, right panel). Initially, false-alarm rates for the Mixed stimuli (crosses) and the N (open circles) were similar, but thereafter, the false-alarm rates of the Mixed stimuli tended to be greater.

D. Discussion

The critical result, as far as the hypotheses are concerned, is that listeners reported more within-trial repetitions when they heard a reference noise token that only re-occurred over distinct trials: More false alarms were observed for the Mixed stimuli than the N stimuli. First, this shows that listeners were able to learn a reference noise token, even when it was presented in the absence of an immediate repetition. Second, these naïve listeners showed a tendency to report the Mixed stimuli as containing a within-trial repetition, as for the naïve listeners in experiment 1.

A possible criticism of the design is that listeners were all presented with RefRNs, and this might have led them to associate recognized noises with "yes" responses in general, even though subsequent Mixed stimuli used different reference noises tokens. However, if we consider the four blocks in which F_{Mixed} exceeded F_N (by the Fisher exact test), three of them were the very first blocks for three different listeners. The fourth of the blocks, from a fourth listener, was preceded by two blocks, one of which included RefRN, but no learning was observed in either of these two blocks (by the Fisher exact test). Thus, the tendency to report Mixed cannot be attributed to any association with RefRN. If anything, listeners may have learned not to report Mixed stimuli as repeating: It seems unlikely that they stopped learning the reference noises after their first block of it. This cannot be attributed to a loss in the listeners' concentration: They continued to learn RefRNs, with five out of six listeners learning RefRNs in their last blocks that included it.

V. GENERAL DISCUSSION

A. Response strategies

The main result, observed in all three experiments, is that after hearing the same noise token across different trials

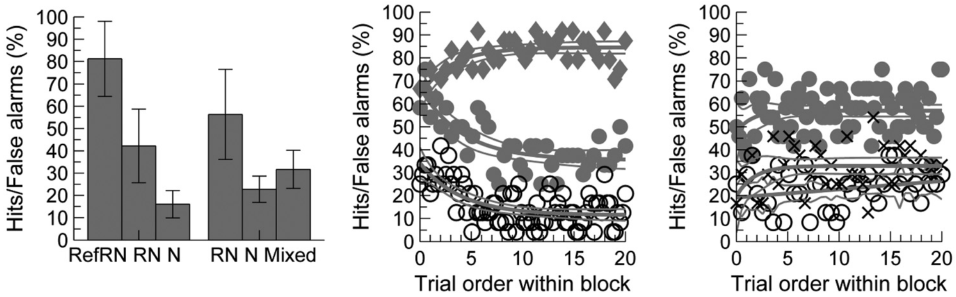


FIG. 5. (left) The mean hit rates and false-alarm rates for the four types of stimuli in experiment 3 split over the Baseline condition (left group of bars) and the Experimental condition (right group of bars). Error bars are 95% confidence intervals centered on the mean. (middle) The time courses of hit rates and false-alarm rates in the Baseline condition, with RefRN (closed diamonds), RN (closed circles), and N (open circles). For each stimulus there is also a fitted exponential curve (thick gray lines) and their 95% confidence intervals centered on the mean (thin gray lines). Note that there are in fact twice the number of N trials, which is not reflected in the values shown on the x axis. (right) Equivalent time courses of hit rates and false-alarm rates for the Experimental condition, including for RN (closed circles), Mixed stimuli (crosses) and N (open circles). Note that there are in fact twice the number of RN trials, which is not reflected in the x axis.

in an experimental block, listeners had a tendency to report stimuli including this token as containing a within-trial repetition, even when this was actually not the case (Mixed stimuli). But a reasonable *a priori* hypothesis could have predicted the opposite: Once listeners had learned a reference noise, they would have been able to decide with more certainty that the Mixed stimuli only contained a single copy of it. This did not happen.

We suggest two reasons why the listeners may have reported the Mixed stimulus as containing a within-trial repetition. The first reason is that the listener may correctly hear the reference noise token presented just once in a Mixed trial, but incorrectly interpret this percept as signaling a repetition. The use of alternative response strategies could have been encouraged by the difficulty of the baseline repetition-detection task that listeners were asked to perform. Such a change in strategy could explain some of the responses of the naïve listeners (experiment 1) but the listeners briefed about Mixed stimuli (experiment 2) should have been able to refrain from using a strategy that ultimately resulted in more errors. The second proposed reason to erroneously report a within-trial repetition is that errors could be made judging whether the Mixed stimuli included a second presentation of the reference noise token. As such, Mixed stimuli would be more likely to be confused with RefRN than a stimulus without any reference noise token, such as N. Both naïve and experienced listeners can be expected to produce this second type of error. Thus, perhaps counter-intuitively, if listeners used the recognition of reference noise token to increase their hit-rate for RefRN, this would also in general lead to an increased false-alarm rate for the Mixed stimuli. Conversely, a deliberate attempt to reduce the false-alarm rate of the Mixed stimuli would also result in a reduced hit-rate for RefRN stimuli. These two reasons for reporting Mixed stimuli as repeating are rephrased, in the Appendix, in a more mathematical form, showing that they do entail an increased false-alarm rate for the Mixed stimulus over a wide range of parameters.

The question remains as to why a listener trained to detect repetitions in noise would start to class learned noises as "repeated" in the absence of a within-trial repetition. This is especially puzzling for experiment 3, where the false-alarm rates for Mixed stimuli were elevated even before listeners had a chance to associate RefRN stimuli with repetitions. We speculate that this tells us about how the experimental task was actually performed. During the training session, naïve listeners started with noises that were repeated ten times in a cyclic manner, such that regular perceptual features likely emerged ("whooshes," "cranks," and "rasping"; Guttman and Julesz, 1963; Kaernbach, 1992, 1993). Thus, the listeners may have equated unusual features in the noise with repetitions. For the main experiment, the difficult repetition-detection task may thus have been traded for a feature-detection task, especially if reference noise tokens acquired salient features through learning.

B. Noise to probe auditory perceptual learning

Previously, auditory learning of noises had only been shown for noises that were immediately repeated (Agus *et al.*, 2010) or with highly trained listeners (Hanna, 1984; Goossens *et al.*, 2008). If these were the only circumstances under which perceptual learning of noise occurred, then such a psychophysical paradigm would not be relevant to most realistic situations, where the repetitions may occur irregularly and infrequently. Luo *et al.* (2013) observed learning of unrepeated noises in magnetoencephalographic (MEG) phase responses, although not in the behavioral responses. Here, we found rapid learning for noise samples that were presented irregularly and infrequently. Experiment 3 in particular showed that listeners were able to learn the reference noise token when it was only presented without any immediate repetition, in the Mixed stimuli. It is also noticeable that such an outcome was observed even though listeners were discouraged to report the Mixed stimuli through instructions. If anything, our results thus underestimate the amount of learning that is achievable with multiple exposures to a sample of noise.

Along with those of Agus *et al.* (2010), the present results suggest that the noise learning paradigm taps into a powerful auditory learning mechanism. Learning is unsupervised, robust to interference, and it occurs with or without immediate repeats. Learning is rapid, observed within a few presentations of a sound. Intriguingly, most of the learning occurred within 10 trials, a time-scale that is very similar to adaptation to time-compressed speech (Dupoux and Green, 1997; Peelle and Wingfield, 2005). Moreover, our paradigm clearly distinguishes perceptual learning from procedural learning, in that the effect is observed as a within-block, within-listeners contrasts of two categories of stimuli (RN and RefRN) on which the listeners perform the same task. Distinguishing procedural and perceptual learning for auditory tasks has previously required more trials and a design that compares the effects of different types of training (e.g., Demany, 1985; Hawkey *et al.*, 2004; Adank and Janse, 2009; Kumpik *et al.*, 2009; Miyazono *et al.*, 2010). Finally, the fact that listeners chose to perform a feature-detection task rather than a repetition-detection task, as they were instructed and trained on, strongly suggests that learning has a salient perceptual effect on noise. A noise token presumably acquires distinctive features through learning, and this is what is measured by our paradigm.

The learning of noise may thus provides an efficient psychophysical tool for observing in real time the emergence of memory traces for complex sounds, as relevant for acquiring mnesic traces for novel sounds in everyday situations.

ACKNOWLEDGMENTS

This work was supported by the CNRS and the Agence Nationale de la Recherche, grants ANR-08-BLAN-0167-01 and ANR-2010-BLAN-1906.

APPENDIX

The model outlined here assumes that listeners distinguish between RefRN, Mixed and N stimuli by counting the number of times they detected the learned reference noise token (twice, once, or not at all, respectively). The aim of the modeling exercise is to test whether such a proposition can be made to fit with the experimental data, as well as to cast our interpretations in a signal-detection theory framework. To simplify the presentation, we will initially ignore the RN stimulus and other strategies for detecting repetitions. We describe listeners' response strategies with two variables, H, the hit-rate for correctly detecting a reference noise token, and F, the false-alarm rate for detecting the reference noise when a Mixed or N stimulus is presented.

A reference noise token could be presented in the first or second half of each trial. Thus, H and F do not refer directly to any of the hit rates or false-alarm rates measured in the experiments. However, given H and F, we can calculate the probability of hearing the reference noise once or twice upon hearing a RefRN, Mixed, or N stimulus. For example, to "correctly" perceive a Mixed stimulus as a having one repeat of the reference noise token, the listener could *either* correctly detect the reference noise token and reject the other noise, *or* incorrectly reject the reference noise token and incorrectly classify the other noise as a reference noise token:

$$P(\text{one reference noise perceived}|\text{Mixed}) = H(1-F) + (1-H)F.$$

The probabilities for each pair of stimulus and percept are shown in Table I.

Counting the number of repetitions of the reference noise token is only the first step. The model must then decide how to respond given the number of reference noises perceived in a trial. For two reference noises, the required response seems trivially to be that the noise is repeated. Likewise for no detected reference noises, the response should always be that the noise is not repeated within the trial. However, the experimental results suggest that listeners can have different strategies for responding to the case when one reference noise is perceived: This could be reported as a repetition, perhaps because listeners associated the reference noise token to trials including a repetition. Thus, depending on the response strategy, the probability of reporting a repetition for Mixed could range from the probabilities given in the top row of Table I (the reference noise detected twice) to the sum of probabilities in the top two rows of Table I (the reference noise detected once or twice). These calculations are summarized in Table II. Note that although the response strategy deals with how to respond when the stimulus is perceived as a Mixed stimulus, it affects the responses to all stimuli.

In the experimental data, the false-alarm rates for the Mixed stimuli were generally higher than the false-alarm rates for the N. This can be rewritten as $F_{\text{Mixed}} - F_N \geq 0$. In the model, the value of $F_{\text{Mixed}} - F_N$ depends on the listeners strategy. From Table II, if the model listener reports the percept of a single reference noise as a repetition, then

$$\begin{aligned} F_{\text{Mixed}} - F_N &= (H + F - HF) - [F(2-F)] \\ &= H - F(1 - H - 2F) \geq H - F \geq 0, \end{aligned}$$

because by definition, H and F are positive and $H \geq F$ for any listener performing above chance. So, unsurprisingly, the model supports the idea that $F_{\text{Mixed}} \geq F_N$ whenever the listener chooses to report Mixed stimuli as repeating. Less intuitively, however, the same is found to hold true even when the listener interprets a single perceived reference noise as non-repeating. Here,

$$F_{\text{Mixed}} - F_N = HF - F^2 = F(H-F) \geq 0,$$

because again $F \geq 0$ and $H \geq F$. As such, in this simple model, $F_{\text{Mixed}} \geq F_N$ irrespective of the listeners strategy for reporting the Mixed stimuli. This could explain the major feature of the experimental data.

Now we complete the model by introducing the RN stimulus. The listener then has two separable sub-tasks: (1) To count the number of times the reference noise token was presented; (2) to judge whether a noise was immediately repeated or not, without recognition of the reference noise token. Here the number of possible response strategies increases greatly, and for some parameters, it is possible that F_{Mixed} would be less than F_{N}, even though such a pattern was not observed experimentally.

A full model was implemented and fitted to the whole dataset. We assumed that listeners first judged the number of times the reference noise was presented (once, twice, or not at all), and if the reference noise was not heard, only then did the listener judge whether or not the noise was repeated. This gives us already four free parameters, namely, the hit rates and false-alarm rates for the reference noise token as before (H and F), but also for "fresh" noise repetitions: H_{rep} and F_{rep}. A further free parameter (M) was added to reflect the tendency to report detection of a single reference token as repeated, as summarized in Table II where M is 0 or 1. In the fitting, the parameter was treated as continuous between these two extreme strategies. Listeners were assumed to change their strategy from experiment 1 to experiment 2, after being briefed on the experimental design, so M was broken into two free parameters, $M_{\text{naïve}}$ and $M_{\text{experienced}}$. It was assumed that the other free parameters, H, F, H_{rep}, F_{rep} remained constant throughout the experiments. The six free parameters were then fitted to the overall average data from experiments 1 and 2, using the least squares method. Unsurprisingly, given eight data points and six free

TABLE I. Probabilities of the number of times the reference noise token will be perceived given the stimulus presented.

Perception	RefRN	Mixed	N
2 reference noises	H^2	HF	F^2
1 reference noise	$2H(1-H)$	$H(1-F)+(1-H)F$	$2F(1-F)$
0 reference noises	$(1-H)^2$	$(1-H)(1-F)$	$(1-F)^2$

TABLE II. Probabilities of reporting a repetition for each stimulus given two extreme strategies for responding to a stimulus in which the reference noise was only heard once.

Perception	RefRN	Mixed	N
P(Repeated|1 ref. noise ⇒ No)	H^2	HF	F^2
P(Repeated|1 ref. noise ⇒ Yes)	$H(2-H)$	$H+F-HF$	$F(2-F)$

TABLE III. Model parameters fitted to the data of experiments 1 and 2 and their signal-detection theory measures.

Parameter	Value
H	72%
F	8%
H_{rep}	37%
F_{rep}	16%
M_{naive}	50%
$M_{\mathrm{experienced}}$	16%
$d'_{\mathrm{reference}}$	2.0
$c_{\mathrm{reference}}$	0.4
d'_{rep}	0.7
c_{rep}	0.7

parameters, a reasonable fit was found (RMS error = 5%). The fitted parameters are shown in Table III, along with some signal-detection theory measures to aid interpretation.

The d' sensitivity measure for repetition detection in general was rather low $(d' = 0.7)$ but similar to previous more direct measures of it (e.g., 0.5; Agus *et al.*, 2010). Listeners were much more sensitive to the reference noise $(d' = 2.0)$. Criteria were positive for both the detection of the reference noise and the detection of repetitions. With experience, the listeners' strategy parameter M for reporting perceptually Mixed stimuli dropped from 50% to 16%, in line with their intervening explicit instructions that the Mixed stimuli should not be reported as repeated. Thus, although the model is oversimplified, it captures and quantifies many aspects of the data and our interpretation of it. Importantly, it also demonstrates that the pattern of results fits with the idea that listeners performed a recognition task on the reference noise tokens, the most parsimonious hypothesis to explain the outcome of experiment 3.

---

[1] The noise reoccurring on different trials is sometimes termed "frozen noise" (Goossens *et al.*, 2008). We will refrain from using this denomination here because of the possible confusion with within-trial repeated noise, also termed frozen noise by other authors (Warren *et al.*, 2001).

Adank, P., and Janse, E. (**2009**). "Perceptual learning of time-compressed and natural fast speech," J. Acoust. Soc. Am. **126**, 2649–2659.

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

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

Cowan, N. (**1984**). "On short and long auditory stores," Psychol. Bull. **96**, 341–370.

Dau, D., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Demany, L. (**1985**). "Perceptual learning in frequency discrimination," J. Acoust. Soc. Am. **78**, 1118–1120.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. **23**, 914–927.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

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.

Guttman, N., and Julesz, B. (**1963**). "Lower limits of auditory periodicity 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.

Hawkey, D. J., Amitay, S., and Moore, D. R. (**2004**). "Early and rapid perceptual learning," Nat. Neurosci. **7**, 1055–1056.

Kaernbach, C. (**1992**). "On the consistency of tapping to repeated noise," J. Acoust. Soc. Am. **92**, 788–793.

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.

Kumpik, D., Ting, J., Campbell, R. A., Schnupp, J. W., and King, A. J. (**2009**). "Specificity of binaural perceptual learning for amplitude modulated tones: A comparison of two training methods," J. Acoust. Soc. Am. **125**, 2221–2232.

Luo, H., Tian, X., Song, K., Zhou, K., and Poeppel, D. (**2013**). "Neural response phase tracks how listeners learn new acoustic representations," Current Biology (in press).

Miyazono, H., Glasberg, B. R., and Moore, B. C. J. (**2010**). "Perceptual learning of fundamental frequency discrimination: Effects of fundamental frequency, harmonic number, and component phase," J. Acoust. Soc. Am. **128**, 3649–3657.

Peelle, J. E., and Wingfield, A. (**2005**). "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech," J. Exp. Psychol. **31**, 1315–1330.

Pfafflin, S. M. (**1968**). "Detection of auditory signal in restricted sets of reproducible noise," J. Acoust. Soc. Am. **43**, 487–490.

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.