

Perceptual Learning and Recognition of Random Acoustic Patterns

Hanna Ringer, Erich Schröger & Sabine Grimm

To cite this article: Hanna Ringer, Erich Schröger & Sabine Grimm (2022): Perceptual Learning and Recognition of Random Acoustic Patterns, Auditory Perception & Cognition, DOI: [10.1080/25742442.2022.2082827](https://doi.org/10.1080/25742442.2022.2082827)

To link to this article: <https://doi.org/10.1080/25742442.2022.2082827>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 31 May 2022.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)

Perceptual Learning and Recognition of Random Acoustic Patterns

Hanna Ringer ^{a,b}, Erich Schröger ^b and Sabine Grimm ^{b,c}

^aInternational Max Planck Research School NeuroCom, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; ^bWilhelm Wundt Institute for Psychology, Leipzig University, Leipzig, Germany; ^cPhysics of Cognition Lab, Institute of Physics, Chemnitz University of Technology, Chemnitz, Germany

ABSTRACT

The human auditory system is capable of learning unstructured acoustic patterns that occur repeatedly. While most previous studies on perceptual learning focused on seamless pattern repetitions, our study included several presentation formats, which are more typical for memory tasks (involving temporal delays or irrelevant information between pattern presentations), and probed active recognition of learned patterns more directly. We adapted an established implicit learning paradigm and presented three groups of listeners with the same acoustic patterns in different presentation formats, i.e., either back-to-back, separated by a silent interval or by a masker sound. Participants additionally completed an unexpected memory test after the learning phase. We found substantial learning in all groups, measured indirectly via the increased sensitivity in a perceptual task for patterns that occurred repeatedly (compared to patterns that occurred only once) and more directly via above-chance recognition performance in the memory test. Pattern learning and recognition were robust across presentation formats. Therefore, we propose that similar mechanisms might underlie memory formation for initially unfamiliar sounds in everyday listening situations. Moreover, memories for unstructured acoustic patterns that were acquired implicitly through perceptual learning enable subsequent active recognition.

ARTICLE HISTORY

Received 17 December 2021
Accepted 20 May 2022

KEYWORDS

Perceptual learning; auditory memory; acoustic patterns

Introduction

Listeners make use of their memories of previously heard patterns in the acoustic signal, to meet the challenges of orientation in and interaction with complex auditory environments (Bregman, 1990; Winkler et al., 2009). Rapid recognition of sounds that are familiar helps to channel limited processing resources on components of the auditory input that are unfamiliar and potential carriers of novel, relevant information. Previous exposure also enhances the sensitivity to certain stimulus features that are not retained in detail upon first hearing (McDermott et al., 2013) and improves sound source segregation (Woods & McDermott, 2018). Thus, memories of sounds that have been heard before may facilitate and sharpen auditory perception, especially of largely unstructured

CONTACT Hanna Ringer  hanna.ringer@uni-leipzig.de  International Max Planck Research School NeuroCom, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and spectro-temporally complex acoustic stimuli. Said memory representations are assumed to be established through unsupervised perceptual learning, i.e., an increase in perceptual capacity as a result of growing experience with the stimulus material (Gibson, 1969; Gilbert et al., 2001) – even in the absence of an explicit memory task. That is, repeated exposure to certain sounds improves the efficiency of perception and, for instance, facilitates discrimination of these sounds with regard to different stimulus features such as duration, frequency, or intensity (Wright & Zhang, 2009).

There is compelling evidence that human listeners are capable of learning even random, unstructured acoustic patterns, such as white noise, from repeated exposure (see below). However, most previous studies investigated unsupervised acoustic pattern learning by presenting seamless, often periodic, pattern repetitions within a continuous sound and measured learning indirectly as a performance improvement in a repetition detection task. In the current study, we created an experimental learning context that more closely resembles the demands of everyday listening situations in which patterns repeat in variable formats, including temporal delays or irrelevant information between occurrences. In addition, we included a more direct measure of active pattern recognition after learning.

The Implicit Noise-Learning Paradigm as a Tool to Study Auditory Perceptual Learning

Previous research has often used white noise to study perceptual learning of unfamiliar and meaningless, but spectro-temporally complex acoustic patterns (e.g., Agus & Pressnitzer, 2013, 2021; Agus et al., 2010; Andrillon et al., 2015, 2017; Goossens, 2008; Guttman & Julesz, 1963; Kaernbach, 2004; Luo et al., 2013; Viswanathan et al., 2016; Warren et al., 2001). White noise is perceived as a homogeneous sound devoid of acoustic features that are commonly used to describe and distinguish sounds (Kaernbach, 2004), such as characteristic pitch (contours), transients or amplitude modulations. Its use as stimulus material offers a particular advantage such that it avoids confounding influences of higher-level categorical or semantic processing on sensory-driven learning. Due to the absence of salient physical cues in the signal, different noise samples are (on average) perceptually highly similar and can only be memorized and distinguished based on subtle acoustic idiosyncrasies that they contain by chance. Agus et al. (2010) introduced the so-called noise-learning paradigm to probe the implicit formation of memories for specific white noise tokens that are presented repeatedly to listeners. They asked participants to detect repetitions in noises that either consisted of two seamless presentations of a 500-ms segment (repetition) or of 1000 ms of random noise (no repetition). In addition to the repetitions within trials and unbeknownst to the participants, one specific stimulus (which contained a repetition) reoccurred across trials during an experimental block, while all other stimuli were presented only once. Hit rate in the periodicity detection task gradually increased over the course of the block for the repeatedly presented “reference repeated noise” in comparison to the “repeated noises” that occurred just once, suggesting that listeners formed memories of the reference repeated noise, which in turn facilitated the detection of repetitions (Agus et al., 2010). Learning was characterized as implicit, automatic, fast, resilient to interference, long-lasting and robust against temporal and spectral transformations: substantial learning

was observed after only a few occurrences of the reference noise, without any explicit instruction or even knowledge about the repetition across trials, despite the presentation of irrelevant noises in between, and performance advantages persisted in a re-test after two weeks and for time-compressed or time-reversed versions of the reference noise. Interestingly, learning differed between blocks and reference noises and appeared to occur either fast and (almost) perfectly or not at all (Agus et al., 2010).

The same paradigm was used in several subsequent studies (Agus & Pressnitzer, 2013, 2021; Andrillon et al., 2015, 2017; Luo et al., 2013; Song & Luo, 2017; Viswanathan et al., 2016), which replicated robust learning effects for specific white noise tokens and extended the findings by Agus et al. (2010). In particular, these studies lent further evidence that perceptual learning happens automatically in the absence of attention to the repeating stimuli, as it was also observed while participants performed a distractor task (Andrillon et al., 2015) and even during certain sleep phases (Andrillon et al., 2017). Moreover, electroencephalography (Andrillon et al., 2015, 2017), magnetoencephalography (Luo et al., 2013) and functional magnetic resonance imaging (Kumar et al., 2014) studies identified characteristic neural responses to learned reference noises (or noise-like patterns in the study by Kumar et al. (2014)) that accompanied learning-related behavioral changes.

Toward a Model of Perceptual Learning in More Naturalistic Listening Situations

While research on white noise learning has provided valuable insights into the remarkable learning capacity of the human auditory system by challenging it to the extreme, several further studies attempted to generalize these findings to varying learning contexts and establish their relevance for everyday learning. To this end, these studies manipulated different contextual factors, such as the acoustic properties of the stimulus material, the temporal regularity, and the frequency of occurrence of the to-be-learned reference patterns, to approximate the demands of naturalistic listening situations.

First, there is consistent evidence that perceptual learning of repeatedly presented sounds takes place across a variety of different acoustic patterns besides white noise, including temporal patterns of clicks (Kang et al., 2017, 2018, 2021), sequences of tone pips (Bianco et al., 2020; Herrmann et al., 2021) and “tone clouds” consisting of overlapping, temporally jittered tones of different frequencies (Agus & Pressnitzer, 2021; Kumar et al., 2014). In particular, Agus and Pressnitzer (2021) recently demonstrated that the learning effect is robust and similar in size across tone clouds of varying spectro-temporal complexity, while increased stimulus complexity was only associated with an increase in general task difficulty (i.e., repetition detection became more difficult).

Second, recent studies have shown that although a temporally regular occurrence of repetitions may facilitate learning, the absence of such a regularity does not (fully) prevent the formation of memories for repeatedly presented acoustic patterns (Dauer et al., 2021; Hodapp & Grimm, 2021).

Finally, Bianco et al. (2020) demonstrated that learning of specific sequential patterns in rapid tone pip sequences occurred even when the pattern was only repeated much less frequently than in previous studies, i.e., on average every three minutes. Strikingly,

performance benefits for learned patterns were still observable seven weeks after learning, which indicated that perceptual learning in fact supports the formation of long-term memory representations of random auditory patterns (Bianco et al., 2020).

Taken together, these results, along with the above-mentioned automaticity of learning through repeated exposure, strongly suggest that such mechanisms are likely candidates to underlie learning and memory formation in naturalistic everyday listening situations.

The Present Study

One aspect of the learning context that has received little consideration so far is how the acoustic patterns, which are to be learned through repeated exposure, are presented. In most previous studies that applied the noise-learning paradigm (or variants thereof), participants had to detect pattern repetitions that were embedded in a continuous stimulus (and occurred either immediately back-to-back or interspersed within a homogeneous noise or tonal stimulus). The detection of such within-trial pattern repetitions relies on a successful memory comparison itself, and the respective internal representations prepare the ground for higher-order learning of reference patterns that reoccur across multiple trials. The percept associated with pattern repetitions embedded in a continuous sound is known to arise rapidly and automatically (e.g., Andriillon et al., 2015, 2017; Barascud et al., 2016; Chait, 2020). This automatic emergence of a salient (periodic) perceptual event from the unstructured, varying background might be a critical prerequisite for long-term learning. Yet, in naturalistic listening situations (as well as in typical memory tasks in the field of cognitive psychology) it is likely that a recurring pattern would not necessarily repeat seamlessly within one sound, but instead in variable formats that often include a gap between recurring sound events, i.e., some temporal delay or even irrelevant auditory input between presentations. The discrimination of two patterns presented as separate sounds involves more cognitive (instead of rather perceptual) resources as it requires listeners to rely on distinct memory representations of the patterns that do not automatically emerge from the stimulation.

The present study set out to investigate perceptual learning across different presentation formats of the reference pattern. To this end, three groups of listeners were presented with to-be-learned reference patterns either back-to-back, with a silent interval or with a masker sound between them. As stimuli we used random acoustic patterns that were generated from white noise, but transformed so that they matched statistical properties of natural sounds (see, McDermott et al., 2011). That way, we could make sure that the sounds were novel and unfamiliar to the participants, yet comparable to naturalistic auditory input in terms of their spectro-temporal complexity and acoustic structure. It is important to note that, despite matching the statistical properties of natural sounds, our artificially generated sounds still differed from naturalistic everyday sounds in terms of phenomenological properties, i.e., they sounded artificial. However, increasing ecological validity of the stimulus material by using recordings of natural sounds would come at the expense of precise acoustic control between stimuli, which might introduce variance with respect to the memorability of specific sounds. Moreover, listeners' familiarity with specific items or their semantic categories might vary for everyday sounds, while all artificially generated items are equally unfamiliar to them.

We therefore decided to use artificial stimulus material that is carefully controlled and, albeit not fully ecologically valid, closer to naturalistic sounds than the stimuli used in previous research.

Our main goal was to show perceptual learning of random acoustic patterns across different presentation formats, particularly in the absence of immediate (within-sound) repetition, i.e., when the reference patterns are presented with a silent interval or a masker sound between them. Substantial memory formation across presentation formats would point toward a relevance of such learning mechanisms for everyday listening situations.

Additionally, we wanted to explore whether and how perceptual learning of acoustic patterns is modulated by the presentation format. The size of the learning effect as well as certain aspects of learning (such as the number of presentations of the reference pattern required for performance changes to occur) were compared between the three groups of participants who learned the same reference patterns in different presentation formats.

Finally, we aimed to measure learning of specific reference patterns more directly than in previous research. Specifically, we wanted to test whether repeated exposure to a certain reference pattern allows active recognition of that pattern in a subsequent unexpected memory test, beyond the implicit performance benefits in the perceptual task that is used during the learning phase. Participants were asked to complete a short two-alternative forced-choice recognition task at the end of the experiment in which they decided in every trial which out of two sounds they had already heard before. Above-chance recognition of previously learned acoustic patterns would point toward the formation of robust memory representations, which do not only enhance perceptual sensitivity, but can become at least to some degree actively accessible. A positive association between indirect and direct markers of learning would suggest that both learning-related performance benefits in a perceptual task and active recognition of learned patterns rely on the same memory representations.

Materials and Methods

Participants

A total of 74 healthy participants took part in the study, two of whom had to be excluded due to technical issues (no responses were stored for a whole experimental block).¹ The remaining 72 participants (57 of them female, 15 male) who formed the final sample were between 18 and 54 years old ($M = 22.86$ years, $SD = 5.91$ years). Four of them were left-handed, the remaining 68 were right-handed (as assessed with the short form of the Edinburgh Handedness Inventory; Oldfield, 1971). All participants reported normal hearing, normal or corrected-to-normal vision and no history of any neurological or psychiatric disorder. They were naïve regarding the purpose of the study, gave informed consent (by ticking a box in the online form in order to be allowed to proceed to the actual experiment) and received course credits for their participation. The study was conducted at a German university and most participants were psychology undergraduate students. All experimental procedures were in accordance with the Declaration of Helsinki and the study was approved by a local ethics committee.

Stimuli

So-called correlated noise, i.e., randomly generated acoustic patterns that shared statistical properties with natural sounds, served as auditory stimulus material. This type of sounds has been used previously to study auditory perception and is described in detail elsewhere (McDermott et al., 2011). Stimuli with a duration of 500 ms, including 10-ms onset and offset ramps (half-Hanning windows), were created using the Gaussian Sound Synthesis Toolbox (http://mcdermottlab.mit.edu/Gaussian_Sound_Code_for_Distribution_v1.1) in Matlab (version R2021a; The MathWorks Inc., USA). A generative model was used to transform randomly generated white noise tokens into correlated noise stimuli that had a certain correlative structure, such that temporally and spectrally adjacent sampling points in the spectrogram shared similar (correlated) spectral energy values. The strength of the correlative relationship decreased with increasing temporal and spectral distance between sampling points. Constants for this decrease per time or frequency window were chosen such that they matched the correlative structure of natural sounds (e.g., speech stimuli, environmental sounds; as reported in detail by McDermott et al., 2011). Specifically, the strength of the correlation decreased with -0.065 per time window (20 ms) along the temporal dimension and with -0.075 per frequency window (0.196 octaves) along the spectral dimension. In order to avoid discriminability of the patterns solely based on their loudness, time-varying acoustic loudness was determined using the `acousticLoudness` function in Matlab (version R2021a), and amplitudes were adjusted such that the N5 percentile (i.e., the loudness below which were 95% of the sampling points) was within a range of 35 ± 0.1 sones for all sounds. Audio files with example trials can be found in the online supplemental material (https://osf.io/b93h4/?view_only=dca54eec50304f6fa377c50c01821e45).

Procedure

The study was conducted as an online experiment. Thus, participants used their own computer (which was required to have a physical keyboard) and headphones (connected via a jack plug and (where applicable) with disabled noise canceling function). A custom-built program was compiled using the free JavaScript library `jsPsych` (de Leeuw, 2015; <https://www.jspsych.org>) to control stimulus presentation and response registration via the browser (Google Chrome, Mozilla Firefox or Microsoft Edge). The experiment was hosted on a lab-internal server using JATOS (Lange et al., 2015; <https://www.jatos.org>).

Before the actual experiment, participants received instructions with regard to volume settings and their task. After adjusting the volume to an individually comfortable level, they had the chance to familiarize themselves with the task. During a short training block, which consisted of 10 trials, feedback was provided after every trial (correct/incorrect), whereas in the actual experiment they received feedback (percentage of correct responses) only at the end of each block.

The actual experiment consisted of a “learning phase” and a “test phase.” Importantly, participants were not informed about the subsequent test phase in the beginning of the experiment. Instructions were only provided for the task that was to be performed during

the learning phase, along with the information that other two short auditory tasks (the test phase and a headphone screening test; see below) would follow the three experimental blocks. The experimental design is illustrated in Figure 1.

Learning Phase

During the learning phase, we used modified versions of the noise-learning paradigm that was originally introduced by Agus et al. (2010). In every trial, participants were asked to compare two 500 ms noise patterns and indicate via button press whether the same specific noise pattern was repeated (and presented twice) or two different noise patterns were presented. Both cases occurred with the same probability of 50%. Unbeknownst to the listeners, within each experimental block, one noise pattern was presented repeatedly while all other noise patterns were presented only once throughout the experiment. This so-called “reference noise” occurred in half of the trials that contained a repetition of a specific noise pattern. Thus, in 50% of all trials two different noise patterns were presented (“noise”/N), in 25% the same noise pattern was repeated within the trial (“repeated noise”/RN), and in 25% the same noise pattern was not only repeated within the trial, but also across trials within a block (“reference repeated noise”/RefRN). A gradual increase of task performance over the course of the block in the RefRN, but not in the RN

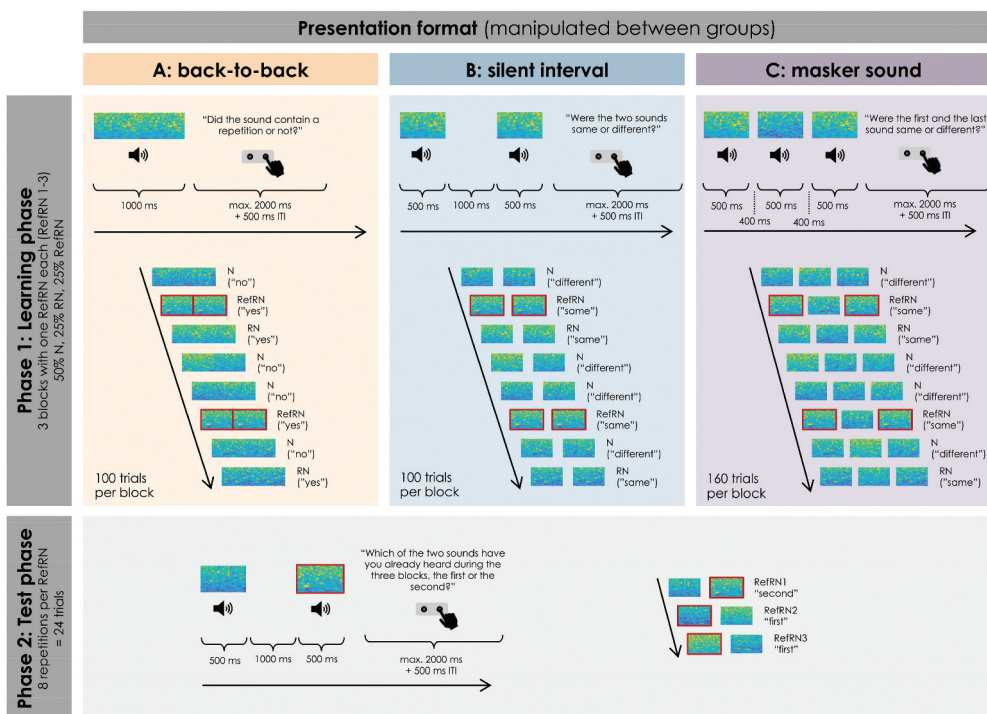


Figure 1. Experimental design. All participants completed a learning phase with three blocks, followed by an unexpected test phase. Presentation format during the learning phase was manipulated across three separate groups of participants. In each block during the learning phase, one specific acoustic pattern (RefRN) reoccurred unbeknownst to the participants and was expected to be learned through repeated exposure.

condition (as reported by previous studies, e.g., Agus et al., 2010; Viswanathan et al., 2016) would be interpreted as an (indirect) indicator of learning of the reference noise. All participants completed three blocks, i.e., they were presented with three different reference noises throughout the experiment.

In addition to the manipulation of the trial type (N , RN , $RefRN$), three groups of participants ($n = 24$ per group) differed with regard to how the two noise patterns within a trial, which were supposed to be compared, were presented. This required slight adaptations of the task and the trial structure between groups.

In group A, patterns were presented *back-to-back*, i.e., the two 500-ms patterns were seamlessly concatenated into one 1000-ms sound without a gap between them, thus following the original presentation format used by Agus et al. (2010). During sound presentation, a fixation cross was displayed on the screen. The listeners' task was to decide whether the sound contained a repetition or not, i.e., whether the first half of the sound was identical to the second half. They gave their response by pressing either key "S" ("repetition") or "L" ("no repetition") on the keyboard of their computer. Response options along with the corresponding key were presented on the screen after the end of the sound until the participant pressed a button or the maximal response interval of 2000 ms expired. The next sound was presented after a silent inter-trial interval of 500 ms.

In group B, patterns were presented with a *silent interval* of 1000 ms between them. Participants were asked to indicate in every trial whether the two sounds were the same or different by pressing either key "S" ("same") or "L" ("different"). The rest of the trial structure including maximal response and inter-trial interval was analogous to group A.

In group C, patterns were presented with a 500-ms *masker sound* between them, with the three sounds separated by inter-sound intervals of 400 ms. As masker sounds, we used correlated noise stimuli that were generated (individually for each trial) according to the same procedures as the two sounds that were to be compared. Listeners had to decide whether the first and the last sound were the same or different. They again gave their response by pressing either key "S" ("same") or "L" ("different"). The rest of the trial structure including maximal response and inter-trial interval was analogous to group A and B.

Within a group, each participant was presented with different reference noises, but the same pool of reference noises was used across groups to avoid random variability in performance due to salient acoustic features that randomly occurred in specific patterns, but not in others. N and RN trials were formed from 500-ms segments that were drawn from the same pool of noise segments in individually randomized combinations and order for each participant, such that each individual heard each segment in exactly one trial throughout the experiment (once if it was part of an N trial, twice if it was part of an RN trial). The pool of 500-ms noise segments used for group A and B was extended by additional segments for group C due to the higher number of trials in this group and the use of an additional masker sound in each trial. Each of the three blocks consisted of 100 trials in group A and B and of 160 trials in group C. As we wanted to keep the online experiment duration as short as possible and previous studies (e.g., Agus et al., 2010) had shown fast learning of the reference noise within the first 10 to 15 presentations, we restricted the number of $RefRN$ trials per block to 25 in group A and B. For the presentation format with a *masker sound* in group C, we

decided to present 40 *RefRN* trials per block, as there were no estimates from the literature with regard to the number of presentations necessary for learning under such a condition, yet we expected that learning might occur with a somewhat slower time course.

Test Phase

During the test phase, the procedure was the same for all three groups. In each trial, participants heard two stimuli that were separated by 1500 ms of silence, one of which was a reference noise that had been presented before during one of the experimental blocks. They were asked to indicate in a two-alternative forced-choice recognition task which of the two sounds they have already heard before during the three blocks. An above-chance performance in this recognition task would be interpreted as a (more direct) indicator that robust memories of the reference noises were formed during the learning phase. Listeners were instructed that this memory test was very difficult, but that previous studies had shown that participants performed surprisingly well if they just relied on their “gut feeling.” They again gave their response by pressing either key “S” (“first sound”) or “L” (“second sound”). The rest of the trial structure including maximal response and inter-trial interval was analogous to the learning phase. Each reference noise was presented eight times during the test phase in a fixed order, such that the reference noises occurred in the same order as in the three blocks during the learning phase and this order (RefRN1, RefRN2, RefRN3) was repeated eight times. This restriction served to avoid that the same reference noise (i.e., target) occurred in two immediately consecutive trials. As non-targets we used three filler noises (Fill1, Fill2, Fill3) that were assigned to trials pseudo-randomly with the restriction that the same stimulus must not occur in two immediately succeeding trials. That way, we made sure that participants could not base their responses solely on item repetition within the test phase, because both targets (i.e., reference noises) and non-targets (i.e., filler noises) repeated and each stimulus occurred equally often (i.e., eight times). The position of the reference noise within the trial was counterbalanced such that it was never presented in the same position in more than three trials in a row. No feedback was provided at the end of the test phase.

The test phase was succeeded by a headphone screening test (Milne et al., 2021) based on the dichotic Huggins Pitch percept (Cramer & Huggins, 1958; see also, Akeroyd et al., 2001; Chait et al., 2006). This illusionary pitch phenomenon should only occur during dichotic presentation of the auditory signal via headphones, but not via loudspeakers (Milne et al., 2021). Stimulus generation as well as the structure of the test block followed the procedures described by Milne et al. (2021).

Finally, participants were asked to report their experiences during the experiment in a short questionnaire. This gave us the chance to assess potential problems that may have occurred and would usually become apparent when talking to the participants in person in the lab. In this questionnaire, participants were also asked to rate the perceived difficulty of the task during the learning phase (on a scale ranging from 1 = “not difficult at all,” to 5 = “very difficult”) and the subjective confidence with regard to their responses (on a scale ranging from 1 = “not confident at all” to 5 = “very confident”). The total duration of the online experiment was around 25 minutes for group A, 30 minutes for group B and 40 minutes for group C.

Data Analysis and Statistical Inference

Data analysis was mainly done in RStudio (version 4.0.2; RStudio Inc., USA), except for the cluster-based permutation tests, which were computed in Matlab (version R2021a; The MathWorks Inc., USA).

Learning Phase

Groupwise Analysis. Data from the learning phase were first analyzed separately within each group. Performance in the respective perceptual task was evaluated within the framework of signal detection theory, which is widely used to quantify and disentangle response accuracy and response bias in perceptual categorization tasks (Macmillan, 2001). For each participant, the sensitivity index d' was computed from hit rates and false alarm rates for *RefRN* and *RN* condition, respectively, applying the so-called log-linear transformation (Hautus & Lee, 2006) to avoid infinite values of d' in the case of extreme hit or false alarm rates. One-sided paired t-tests were used to test whether the performance was higher in the *RefRN* compared the *RN* condition, which would indicate that learning of repeatedly presented reference noises took place. Statistical significance was defined by the standard .05 alpha criterion and Cohen's d was reported as a measure of effect size. In addition to the frequentist tests, we conducted complementary Bayesian t-tests and computed Bayes Factors (BF_{10}) using the package "BayesFactor" in RStudio (Morey & Rouder, 2011; Morey et al., 2018; Rouder et al., 2009). In accordance with widely used conventions (Lee & Wagenmakers, 2014), $BF_{10} > 3$ was considered moderate evidence for the alternative hypothesis and $BF_{10} < 0.33$ was considered moderate evidence for the null hypothesis, while values in between were deemed inconclusive. $BF_{10} > 10$ or $BF_{10} < 0.1$ were interpreted as strong evidence for the alternative or null hypothesis.

To illustrate changes in performance over the course of the block, hit rates and false alarm rates were averaged across listeners at each position within the block (for hit rates: 1 to 25 in group A and B, and 1 to 40 in group C; for false alarm rates: 1 to 50 in group A and B, and 1 to 80 in group C). Trajectories of hit rates in the *RefRN* and *RN* condition and false alarm rates were estimated and plotted by fitting a local regression (loess curve) with the formula $y \sim x$.

Comparison Between Groups. In a second step, data from the learning phase were compared between groups to explore whether perceptual learning was modulated by the presentation format, which differed between groups. In order to use an equal number of trials across groups, only the first 25 *RefRN* and *RN* trials (and the first 50 *N* trials) were included for group C. Difference scores were computed for each participant by subtracting d' in the *RN* condition from the *RefRN* condition as a measure of the size of the learning effect across the whole block. We compared these d' differences between groups by means of a one-way (non-repeated measures) ANOVA with the three-level factor Presentation Format (*back-to-back*, *silent interval*, *masker sound*) using the package "ez" (Lawrence, 2016). Greenhouse-Geisser correction was applied to correct for non-sphericity (as indicated by a significant Mauchly's test with $p < .05$). A significant effect of Presentation Format was followed up with pairwise two-sided independent-sample tests between the three groups. P-values of these post-hoc t-tests were adjusted based on

the false discovery rate (Benjamini & Hochberg, 1995). A complementary Bayesian ANOVA was computed, again using the package “BayesFactors” (Morey et al., 2018; Rouder et al., 2012).

To shed further light on the more subtle influences of the presentation format on certain aspects of learning, we subsequently compared the time course of the learning effect as well as the proportion of (near-) perfectly learned reference noises between groups.

For the analysis of time courses, differences in hit rate between *RefRN* and *RN* condition were computed for each listener at each occurrence (1 to 25) of *RefRN* and *RN* trials within the block. These differences were compared between the three groups using pairwise Mann-Whitney U-tests. We chose the non-parametric equivalent of independent-sample t-tests because within-subject averages were calculated from just three blocks, thus could only adopt a very limited number of values. Tests were computed for each of the 25 trial positions within the block, which resulted in a time series of U-values. We used a cluster-based permutation approach to control for multiple comparisons and identify clusters of significant differences between groups at adjacent positions. To this end, clusters of at least two adjacent positions with $p < .05$ were identified and the sum of U-values was calculated within each cluster. To define the threshold for significance, assignment of trials to the groups was permuted and the maximum of the cluster-level summed U-values were extracted from each of 1000 random permutations. The resulting random distribution of cluster U-values was then used to determine the cluster threshold for statistical significance at $p = .05$.

For the analysis of the proportion of (near-) perfectly learned reference noises, average hit rates were computed for the last ten *RefRN* trials (i.e., after most of the learning has presumably already taken place) within each block and each participant. Based on these hit rate scores, blocks were categorized either as learned (when performance reached a near-perfect level of at least 90% hit rate) or non-learned (when performance remained below 90%). Pairwise chi-squared tests were used to compare the proportion of learned and non-learned reference noises (corresponding to the number of blocks) between the three groups. For illustration purposes, trajectories of hit rates over the course of the block were plotted for each group only for the blocks in which a near-perfect performance was reached after learning.

Test Phase

Because the number of trials differed between groups during the learning phase, data from the test phase was analyzed separately within each group only. Active recognition performance, i.e., the percentage of correct responses during the test phase, was computed for each listener as a more direct measure of learning. To assess whether participants recognized previously learned patterns above-chance, performance was then tested against 50% chance level by means of a one-sided one-sample t-test in each group. Again, corresponding Bayesian t-tests were calculated to complement the frequentist analysis. Finally, we tested a potential association between the indirect and more direct measure of learning. Specifically, Spearman’s rank correlations (Shapiro-Wilk normality tests revealed violations of the assumption of normality: W ’s < 0.79 , uncorrected p ’s $< .001$)

between the mean hit rate in the last ten *RefRN* trials of the learning phase and the percentage of correct responses during the test phase were computed in each group and correlation coefficients were tested against zero.

Results

The headphone screening test revealed that 19 (out of 72) participants, did not pass a moderate criterion of at least five (out of six) correct responses (corresponding to a result obtained with a probability of only 1.78% if listeners were just guessing). Please note that there was an unexpected high number of missing responses in the headphone screening test. This suggests that, instead of responding incorrectly, participants might have failed to respond within a 2000-ms time window, or that there might have been technical problems (in recording the participant's response) in some trials. Thus, the test might overestimate the number of individuals identified as not passing the test. Therefore, we conducted all reported analyses twice, once including data from all participants irrespective of their headphone screening test result, and once including only data from the 53 participants who did pass the test. As both procedures yielded the same pattern of results, we decided to report data from all 72 participants for the sake of statistical power as well as an equal number of listeners in each group.

Learning Phase

Groupwise Analysis: Significant Learning Effect in Each Group

Groupwise performance during the learning phase (in terms of the sensitivity index d') is depicted in Figure 2. An additional figure that shows hit rates and false alarm rates separately is included in the online supplemental material (https://osf.io/b93h4/?view_only=dca54eec50304f6fa377c50c01821e45). On average, participants across all groups performed above chance in both *RN* and *RefRN* trials, which indicated that they were able to successfully solve the respective task (i.e., detect repetitions within sounds or discriminate two sounds). Crucially, we observed substantial learning of reference noises for all three presentation formats, as reflected in a higher sensitivity in the *RefRN* compared to the *RN* condition (*back-to-back*: $t(23) = 5.55, p < .001, d = 1.13, BF_{10} = 3641.32$; *silent interval*: $t(23) = 3.76, p = .001, d = 0.77, BF_{10} = 69.48$; *masker sound*: $t(23) = 2.59, p = .008, d = 0.36, BF_{10} = 6.41$).

As visible in the middle row of Figure 2, hit rate increased over the course of the block in the *RefRN* condition for all groups. In the *RN* condition, the time course of the performance was less consistent across groups: While the hit rate seemed to decrease rather early in the *back-to-back* group, it slightly increased along with performance in the *RefRN* condition in the *silent interval* and *masker sound* groups in the beginning of the block, possibly reflecting an unspecific improvement as a result of growing experience with the task. Nevertheless, hit rate remained higher throughout the whole block and the performance increase was larger in the *RefRN* than in the *RN* condition, which resulted in a divergence of the two curves in all groups (albeit only after a different number of trials in each group). False alarm rates were largely unchanged over the course of the block and only showed a slight decrease in the *back-to-back* group (see lower row of Figure 2).

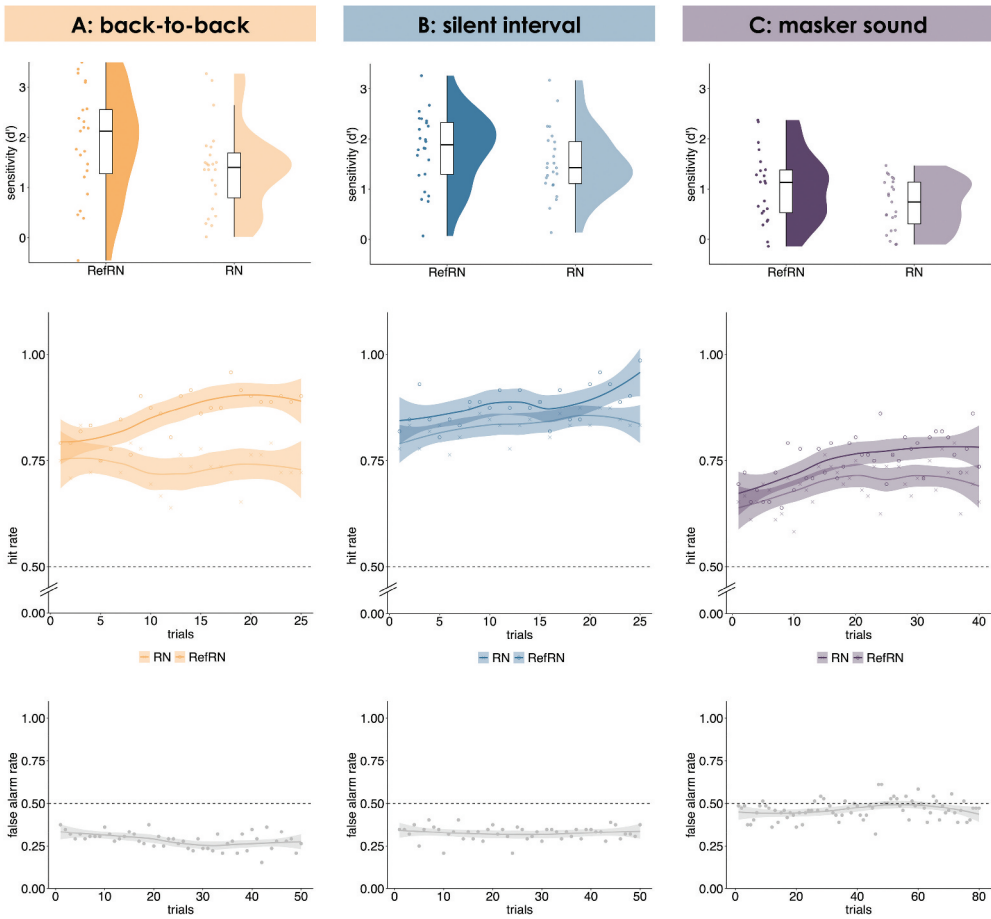


Figure 2. Groupwise performance in the learning phase. Upper row: Mean sensitivity index d' in RefRN and RN condition. Single data points correspond to individual participants. Middle row: Mean hit rate in RefRN and RN condition at each position within the block. Curves correspond to local regressions (loess curves) with the formula $y \sim x$. Shaded areas indicate the 95% confidence interval. Lower row: Mean false alarm rate at each position within the block.

Comparison Between Groups: Differences in Certain Aspects of Learning

Overall, performance appeared to be weaker, i.e., d' was smaller, in group C compared to the other two groups, suggesting that the presence of a masker sound between two correlated noise tokens that were to be compared increased the task difficulty. In fact, listeners reported an increased perceived task difficulty ($F(2, 69) = 6.15, p = .003, \text{partial } \eta^2 = .15, \text{BF}_{10} = 11.48$; A vs. B: $t(46) = -0.13, \text{adjusted } p = .883, d = 0.04, \text{BF}_{10} = 0.29$; A vs. C: $t(46) = -3.26, \text{adjusted } p = .006, d = 0.94, \text{BF}_{10} = 16.87$; B vs. C: $t(46) = -3.28, \text{adjusted } p = .006, d = 0.95, \text{BF}_{10} = 17.66$) and a reduced subjective confidence with regard to their responses ($F(2, 69) = 6.56, p = .002, \text{partial } \eta^2 = .16, \text{BF}_{10} = 15.41$; A vs. B: $t(46) = 0.32, \text{adjusted } p = .730, d = 0.09, \text{BF}_{10} = 0.30$; A vs. C: $t(46) = 3.37, \text{adjusted } p = .005, d = 0.97, \text{BF}_{10} = 21.94$; B vs. C: $t(46) = 3.15, \text{adjusted } p = .007, d = 0.91, \text{BF}_{10} = 13.05$).

The comparison of d' difference scores (*RefRN* minus *RN*) between groups revealed that the size of the learning effect for reference noises differed significantly between presentation formats (see, [Figure 3\(a\)](#)). Concretely, the ANOVA yielded a significant main effect of Presentation Format ($F(2, 69) = 5.36, p = .007$, partial $\eta^2 = .13$, $BF_{10} = 6.51$). Post-hoc contrasts showed that the learning effect was larger in the *back-to-back* group than in both the *silent interval* ($t(46) = 2.58$, adjusted $p = .017$, $d = 0.74$, $BF_{10} = 3.93$) and the *masker sound* ($t(46) = 2.77$, adjusted $p = .010$, $d = 0.80$, $BF_{10} = 5.73$) group, while it did not differ between the latter two ($t(46) = 0.48$, adjusted $p = .668$, $d = 0.14$, $BF_{10} = 0.32$). This difference in the size of the learning effect is likely accounted for by differences in the time course of the learning effect and in the proportion of (near-) perfectly learned reference noises between groups.

As shown in [Figure 3\(b\)](#), learning had a distinct time course within the block for the different presentation formats. While all curves overlapped at the beginning and converged toward the end, they clearly diverged in the middle of the block. Specifically, the curve began to rise earlier, after just about five presentations of the reference noise, in the *back-to-back* group, while the increase occurred only after about 15 to 20 trials in the *silent interval*

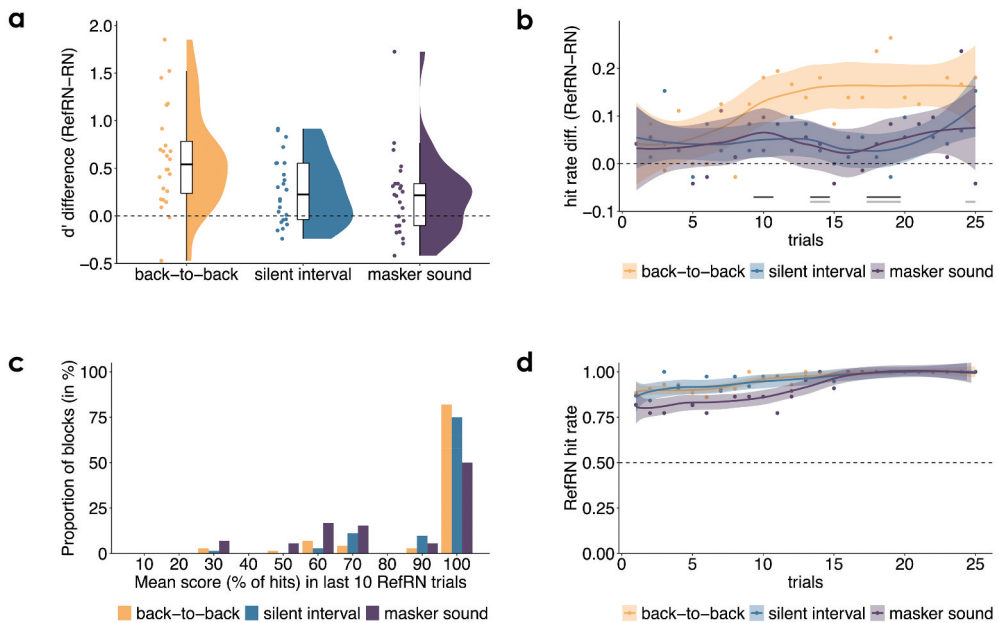


Figure 3. Performance differences between groups in the learning phase. **a:** Mean difference in sensitivity index d' of RefRN minus RN condition. Single data points correspond to individual participants. **b:** Mean difference hit rate of RefRN minus RN condition at each position within the block. Curves correspond to local regressions (loess curves) with the formula $y \sim x$. Shaded areas indicate the 95% confidence interval. Horizontal bars indicate a significant difference between the back-to-back and the silent interval group (dark gray) and between the back-to-back and the masker sound group (light gray) as revealed by pairwise cluster-based permutation tests (no clusters of significant differences were found between the silent interval and the masker sound group). **c:** Distribution of hit rates in the last ten RefRN trials of each block. **d:** Mean hit rate in RefRN condition at each position within the block for blocks in which a hit rate of at least 90% is reached in the last ten RefRN trials.

and the *masker sound* group. Cluster-based permutation tests supported this observation and revealed significant clusters of differences in the time courses between the *back-to-back* group and both the *silent interval* ($z_{\text{cluster}} = 5.62, p_{\text{cluster}} = .002$) and the *masker sound* group ($z_{\text{cluster}} = 2.43, p_{\text{cluster}} = .046$). No significant cluster of differences was found between the *silent interval* and the *masker sound* group ($z_{\text{cluster}} = 2.16, p_{\text{cluster}} = .103$).

Figure 3(c) shows the proportion of blocks in which a certain performance was reached within the last ten *RefRN* trials of the block. Near-perfect performance (i.e., a hit rate of at least 90%) was reached in more than 75% of the blocks in the *back-to-back* and the *silent interval* group, while this was only the case for about 50% of the blocks in the *masker sound* group. Chi-squared tests supported that the proportion of learned reference noises (with near-perfect performance toward the end of the block) was significantly smaller in the *masker sound* group compared to both the *back-to-back* ($X^2(1) = 14.97, p < .001$) and the *silent interval* group ($X^2(1) = 8.56, p = .003$), while there was no significant difference between the latter two ($X^2(1) = 0.66, p = .417$). In Figure 3(d), the time course of learning is depicted selectively for the blocks in which the reference noise was successfully learned. While the performance was initially lower in the *masker sound* group compared to the other two, a steeper rise of the *RefRN* hit rate throughout the first 15 presentations of the reference noise in this group resulted in an approximation of all curves toward the end of the block (at ceiling performance).

Test Phase

In addition to performance changes during the learning phase as an indirect correlate of learning, we measured learning more directly via the active recognition performance after learning (see, Figure 4(a)). During the unexpected test phase, listeners recognized previously learned reference noises in 58.33% of the trials in the *back-to-back* group, in 60.07% in the *silent interval* group and in 63.54% in the *masker sound* group. Albeit far from perfect, recognition performance was significantly above chance level (i.e., 50%) in all groups (*back-to-back*: $t(23) = 2.58, p = .008, d = 0.53, \text{BF}_{10} = 6.22$; *silent interval*: $t(23) = 3.25, p = .002, d = 0.66, \text{BF}_{10} = 23.25$; *masker sound*: $t(23) = 4.58, p < .001, d = 0.93, \text{BF}_{10} = 416.13$).

As shown in Figure 4(b), the hit rate in the last ten *RefRN* trials of the learning phase was positively associated with the percentage of correct responses during the test phase across all groups (*back-to-back*: $r = .463, p = .023$; *silent interval*: $r = .227, p = .286$; *masker sound*: $r = .438, p = .032$). This correlation suggests that listeners who showed a higher task performance for the reference noises at the end of the learning phase tended to recognize the learned reference noises better during the subsequent test phase. It may be plausible to assume that the rather little inter-individual variability in hit rate in the *silent interval* group (relative to the other two groups) led to a decrease of the correlation coefficient in this group, which fell short of statistical significance.

Discussion

The main goal of the present behavioral study was to test whether perceptual learning of random acoustic patterns occurs across different presentation formats, i.e., not only when patterns are repeated immediately within the same sound, but

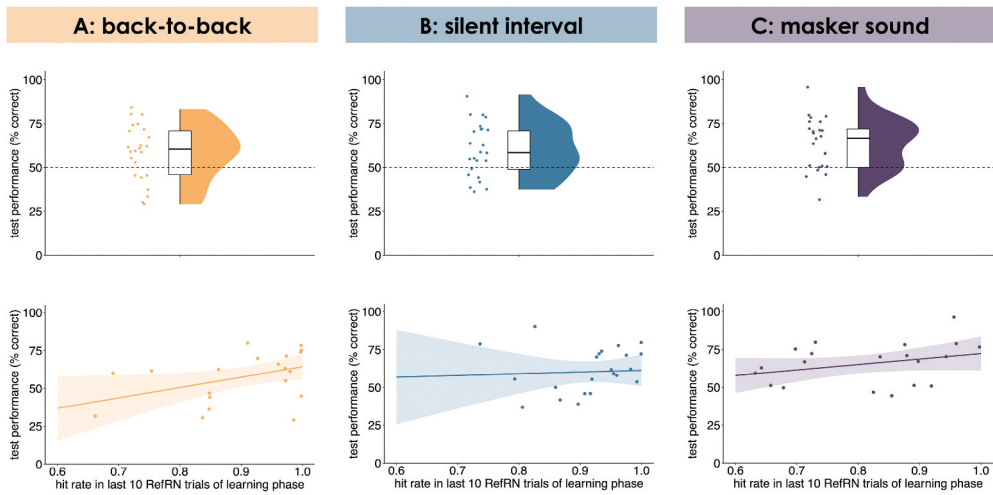


Figure 4. Groupwise performance in the test phase. Upper row: Mean percentage of correct responses in the memory test. Single data points correspond to individual participants. Lower row: Correlations between the hit rate in the last ten RefRN of each block in the learning phase and the percentage of correct responses in the test phase. Single data points correspond to individual participants. Shaded areas indicate the 95% confidence interval.

also when they are presented as two separate sounds with a silent interval or a masker sound in between – as it might happen in naturalistic listening situations. To this end, we asked three groups of listeners to compare random acoustic patterns that were presented either back-to-back, with a silent interval or with a masker sound between them, while certain to-be-learned reference patterns occurred repeatedly without participants’ knowledge about the repetitions. Another goal was to test whether memories that were built up implicitly through perceptual learning enable subsequent recognition of previously learned patterns. Thus, an unexpected two-alternative forced-choice memory test at the end of the experiment probed participants’ active recognition performance of the reference patterns as a more direct measure of memory formation.

Our data showed that performance during the learning phase increased over the course of the block for reference patterns that were presented repeatedly in comparison to patterns that were presented only once. Crucially, substantial learning of reference patterns was found across all three groups, i.e., across different presentation formats. Subsequent exploration of differences between the groups revealed that the presentation format had subtle influences on certain aspects of learning. Concretely, back-to-back repetition seemed to decrease the number of presentations that are necessary for a reference pattern to be learned, while the presence of a masker sound seemed to reduce the proportion of (near-) perfectly learned reference patterns. Finally, listeners were able to actively recognize reference patterns above chance in the test phase, regardless of the presentation format of the stimuli during the learning phase, and performance during learning and test phase were positively associated.

Robust Learning and Recognition of Acoustic Patterns across Presentation Formats

Our results from the learning phase replicate and extend previous findings on perceptual learning of different types of random acoustic patterns, including white noise (Agus et al., 2010; Andrillon et al., 2015; Luo et al., 2013; Song & Luo, 2017; Viswanathan et al., 2016), temporal patterns of clicks (Kang et al., 2017, 2018, 2021), sequences of tone pips (Bianco et al., 2020; Herrmann et al., 2021) and “tone clouds” (Agus & Pressnitzer, 2021; Kumar et al., 2014). In particular the data from the back-to-back group are consistent with earlier studies that also involved immediate (within-sound) repetitions of the to-be-learned reference patterns and reported sensitivity increases for the recurring patterns (relative to patterns that were presented only once), along with characteristic neural responses (Andrillon et al., 2015; Luo et al., 2013). Importantly, perceptual learning of reference patterns was not restricted to the back-to-back presentation format, but also occurred in the groups in which reference patterns were presented with a silent interval or a masker sound between them. One previous study has already suggested that immediate repetition of the reference pattern is not a prerequisite for learning, as indirectly inferred from the observation that trials that contained the reference noise were disproportionately likely to be misclassified as containing a repetition if they actually did not (Agus & Pressnitzer, 2013). Another study showed an increase in repetition detection performance for reference patterns in random tone pip sequences even if the repetitions within the sequence were non-adjacent, although not to the same extent as for adjacent repetitions (Bianco et al., 2020; Experiment 2). The performance increase for reference patterns that we found across presentation formats supports these findings and lends further evidence for learning without back-to-back repetition. Moreover, the current study reports learning of stimulus material that was random and meaningless to the listeners, but more similar to naturalistic sounds in terms of spectro-temporal complexity and statistical properties than stimuli used in previous studies. Together, this points toward a flexible learning mechanism that may be relevant for perceptual learning in naturalistic listening situations in which specific acoustic patterns repeat in variable formats that do not necessarily involve immediate repetition. In line with previous interpretations (e.g., Agus et al., 2010), we assume that repeated exposure to an initially unfamiliar reference pattern leads to the formation of a memory representation, which in turn improves perceptual sensitivity. Our study suggests that such memory representations can also be built up in contexts in which the repeating pattern occurs in isolation and its percept does not automatically emerge through back-to-back repetition within a continuous sound. Nevertheless, it should be noted that the experimental learning contexts created in the current study remain dissimilar from real-life listening situations and our variation of the presentation format constitutes only one possible step toward ecologically valid auditory learning contexts. To further increase ecological validity, future studies may, for instance, incorporate temporally varying, unpredictable intervals between pattern repetitions, non-identical pattern repetitions and interference from concurrent irrelevant sound streams, which are all natural characteristics of complex auditory environments.

Remarkably, implicit learning of reference pattern did not only increase the sensitivity in a perceptual task, but also resulted in above-chance recognition of the learned patterns in a subsequent memory test. This suggests that memories for the reference noises were

successfully established during the learning phase and could be (at least to some degree) actively accessed in a recognition task several minutes after learning. Unlike the modification of behavior during the learning phase, which is assumed to happen automatically on a rather perceptual level, the memory test requires more cognitive effort, a more deliberate access to memory representations and an active selection of the recognized pattern. The procedure of the unexpected test phase was similar to a recent study, which showed substantial recognition of (less acoustically well-controlled) environmental sounds that had been encoded incidentally while participants performed a visual distractor task (Hutmacher & Kuhbandner, 2020). To the best of our knowledge, the current study is the first auditory perceptual learning study that captured an additional, more direct measure of recognition of learned patterns besides performance changes as an indirect measure. The positive association between the direct and indirect measures of learning and recognition suggests that the same memory representations, which are formed through repeated exposure, may underlie performance in both tasks.

Influences of the Presentation Format on Certain Aspects of Learning

Averaged over the whole duration of all blocks, the learning effect appeared to be larger in the group with back-to-back presentation compared to the groups in which reference patterns were presented with a silent interval or a masker sound between them. This difference between groups is likely driven by two aspects of learning, that is the time course of the learning effect and the proportion of (near-) perfectly learned reference patterns. Specifically, fewer presentations were necessary in the back-to-back group before performance notably diverged between reference patterns and patterns that occurred only once. The earlier divergence of the two performance curves consequently resulted in a larger difference when averaging over the whole block. This observation is in line with the idea that immediate (and/or periodic) repetition of a particular sound segment increases sensitivity to subtle acoustic features and makes them more memorable, which was put forward earlier (Agus et al., 2010; Andrillon et al., 2015; McDermott et al., 2013). Based on this assumption, it is plausible to assume that back-to-back presentation of the patterns does not only facilitate the task overall, but also improves, more precisely speeds up, learning. In addition to these differences with regard to the time course of learning, the proportion of (near-) perfectly learned reference noises was decreased in the group in which a masker sound was inserted between presentations of the reference noise. Besides increasing task difficulty in general, the presence of a masker sound, presumably, also reduced the capacity to memorize specific patterns (such that, for example, only some patterns were learned, which, coincidentally, contained a perceptually salient feature). Yet, for the reference patterns for which a (near-) perfect performance was reached toward the end of the block, the learning curve was similar across all groups, consistent with the earlier finding that patterns are either learned perfectly or not at all (Agus et al., 2010).

The focus of the current study was the specific learning effect for certain reference patterns, defined, in accordance with the literature (e.g., Agus et al., 2010), as the increase in perceptual sensitivity for these repeatedly presented patterns (*RefRN*) compared to other patterns that were presented only once (*RN*). Beyond that, we observed performance differences between groups that affected both the *RefRN* and the *RN* condition

and likely reflect more general effects of the different presentation formats. Performance was overall decreased (i.e., sensitivity and hit rates were decreased, while false alarm rates were increased) in the *masker sound* group, which is in line with participants' reports of enhanced perceived task difficulty and reduced subjective confidence. Perhaps counter-intuitively, overall performance was not decreased in the *silent interval* compared to the *back-to-back* condition although the time delay between the pattern presentations could have enhanced sensory memory demands because auditory information needed to be retained for a longer period of time. Instead, the silent gap between the two patterns could also act as a cue for the onset of the second pattern, which is less salient when patterns are presented *back-to-back*, and ameliorate the potential negative effects of the longer retention interval (for related results, see, Goossens et al., 2009). Moreover, in the *silent interval* and in the *masker sound* group, hit rates showed an increase in the first half of the block not only in the *RefRN*, but also in the *RN* condition, which can possibly be attributed to an unspecific improvement due to growing experience with the task. This trajectory is reminiscent of the course of learning-dependent performance changes shown in a previous study (Woods & McDermott, 2018; Experiment 5). In contrast, in the *back-to-back* condition, hit rates only increased in the *RefRN* condition, but decreased in the *RN* condition. This inverse pattern of performance changes between conditions is consistent with what was found in earlier studies and has been related to balancing the two response options once the decision became easier in trials that included the learned reference pattern (Agus & Pressnitzer, 2021; Agus et al., 2010). It is plausible to assume that an unspecific experience-related improvement would also occur in the *RN* condition in the *back-to-back* group, but is superposed by the counteractive performance decrease as a consequence of specific learning of the reference pattern, which occurs earlier when the presentation format involves immediate repetitions.

In the test phase, we did not compare recognition performance statistically between groups due to an unequal number of trials in the preceding learning phase. This methodological limitation precludes a clear conclusion as to whether or not the presentation format during the learning phase modulates listeners' ability to actively recognize learned patterns in a subsequent memory test. Thus, this question cannot be answered based on the current data and would require additional experiments. However, it is important to note that neither a significant difference between groups nor the absence of a significant difference could be interpreted straightforwardly even if the number of trials during the learning phase was equal. Any effects could not be attributed to our manipulation of the presentation format during the learning phase alone, because groups also differed with respect to changes in the presentation format from learning to test phase, which might have biased test performance as well. Nevertheless, our study provides an important proof of concept with regard to active recognition of implicitly learned acoustic patterns: Recognition performance was above chance in all groups, which suggests that at the end of the learning phase sufficiently robust memories of the reference patterns were formed across presentation formats.

Conclusions

In summary, the key finding of the current study is that substantial learning and subsequent active recognition occurred across presentation formats, despite subtle differences with regard to certain aspects of learning. This points toward robust learning

mechanisms that would be suited to underlie auditory memory formation in everyday listening situations in which the recurrences of a certain acoustic pattern are not necessarily seamless, but often separated by a time delay or irrelevant auditory input.

Note

1. Prior to our main study, we conducted a pilot study to ensure the feasibility of applying the noise learning paradigm in an online setting. This pilot study was a replication of the main experiment of Agus et al. (2010) on learning of white noise with the modification that it was run as an online experiment with a reduced number of trials per block (120 instead of 200) to shorten the experiment duration. In a group of 24 participants, we found a significant learning effect (*RefRN* vs. *RN* difference) with an effect size that was smaller than in the original study ($d = 0.64$ vs. $d = 1.12$). According to sample size calculations, with 24 participants per group, a within-group effect of at least a size of $d = 0.62$ can be detected with a power of .90 (with a one-sided paired *t*-test at the standard .05 alpha error probability). Therefore, we planned our study with 24 participants per group to be able to detect an effect in the size of the learning effect observed in our pilot study.

Acknowledgments

The authors thank Frederike Obereigner for help with data collection.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the German Research Foundation (www.dfg.de) with the project number [GR 3412/2-2].

ORCID

Hanna Ringer  <http://orcid.org/0000-0001-6115-7878>

Erich Schröger  <http://orcid.org/0000-0002-8321-6629>

Sabine Grimm  <http://orcid.org/0000-0002-9071-5944>

Data Availability

Single-subject data and supplemental material can be found here: https://osf.io/b93h4/?view_only=dca54eec50304f6fa377c50c01821e45

References

- Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66(4), 610–618. <https://doi.org/10.1016/j.neuron.2010.04.014>
- Agus, T. R., & Pressnitzer, D. (2013). The detection of repetitions in noise before and after perceptual learning. *The Journal of the Acoustical Society of America*, 134(1), 464–473. <https://doi.org/10.1121/1.4807641>

- Agus, T. R., & Pressnitzer, D. (2021). Repetition detection and rapid auditory learning for stochastic tone clouds. *The Journal of the Acoustical Society of America*, 150(3), 1735–1749. <https://doi.org/10.1121/10.0005935>
- Akeroyd, M. A., Moore, B. C. J., & Moore, G. A. (2001). Melody recognition using three types of dichotic-pitch stimulus. *The Journal of the Acoustical Society of America*, 110(3), 1498–1504. <https://doi.org/10.1121/1.1390336>
- Andrillon, T., Kouider, S., Agus, T., & Pressnitzer, D. (2015). Perceptual learning of acoustic noise generates memory-evoked potentials. *Current Biology*, 25(21), 2823–2829. <https://doi.org/10.1016/j.cub.2015.09.027>
- Andrillon, T., Pressnitzer, D., Léger, D., & Kouider, S. (2017). Formation and suppression of acoustic memories during human sleep. *Nature Communications*, 8(1), 179. <https://doi.org/10.1038/s41467-017-00071-z>
- Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 113(5), E616–E625. <https://doi.org/10.1073/pnas.1508523113>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1017/CBO9781107415324.004>
- Bianco, R., Harrison, P. M., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *ELife*, 9, e56073. <https://doi.org/10.7554/eLife.56073>
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.
- Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex*, 16(6), 835–848. <https://doi.org/10.1093/cercor/bhj027>
- Chait, M. (2020). How the brain discovers structure in sound sequences. *Acoustical Science and Technology*, 41(1), 48–53. <https://doi.org/10.1250/ast.41.48>
- Cramer, E. M., & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *The Journal of the Acoustical Society of America*, 30(5), 413–417. <https://doi.org/10.1121/1.1909628>
- Dauer, T. D., Henry, M. J., & Herrmann, B. (2021). Auditory perceptual learning depends on temporal regularity and certainty [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/qm2se>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.
- Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31(5), 681–697. [https://doi.org/10.1016/S0896-6273\(01\)00424-X](https://doi.org/10.1016/S0896-6273(01)00424-X)
- Goossens, T. (2008). On the ability to discriminate Gaussian-noise tokens or random tone-burst complexes. *The Journal of the Acoustical Society of America*, 124(4), 2251–2262. <https://doi.org/10.1121/1.2973184>
- Goossens, T., van de Par, S., & Kohlrausch, A. (2009). Gaussian-noise discrimination and its relation to auditory object formation. *The Journal of the Acoustical Society of America*, 125(6), 3882–3893. <https://doi.org/10.1121/1.3119626>
- Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *The Journal of the Acoustical Society of America*, 35(4), 610. <https://doi.org/10.1121/1.1918551>
- Hautus, M. J., & Lee, A. (2006). Estimating sensitivity and bias in a yes/no task. *British Journal of Mathematical and Statistical Psychology*, 59(2), 257–273. <https://doi.org/10.1348/000711005X65753>
- Herrmann, B., Araz, K., & Johnsrude, I. S. (2021). Sustained neural activity correlates with rapid perceptual learning of auditory patterns. *NeuroImage*, 238, 118238. <https://doi.org/10.1016/j.neuroimage.2021.118238>
- Hodapp, A., & Grimm, S. (2021). Neural signatures of temporal regularity and recurring patterns in random tonal sound sequences. *European Journal of Neuroscience*, 53(8), 2740–2754. <https://doi.org/10.1111/ejn.15123>

- Hutmacher, F., & Kuhbandner, C. (2020). Detailed long-term memory for unattended, irrelevant, and incidentally encoded auditory information. *Journal of Experimental Psychology: General*, 149(2), 222–229. <https://doi.org/10.1037/xge0000650>
- Kaernbach, C. (2004). The memory of noise. *Experimental Psychology*, 51(4), 240–248. <https://doi.org/10.1027/1618-3169.51.4.240>
- Kang, H., Agus, T. R., & Pressnitzer, D. (2017). Auditory memory for random time patterns. *The Journal of the Acoustical Society of America*, 142(4), 2219–2232. <https://doi.org/10.1121/1.5007730>
- Kang, H., Lancelin, D., & Pressnitzer, D. (2018). Memory for random time patterns in audition, touch, and vision. *Neuroscience*, 389, 118–132. <https://doi.org/10.1016/j.neuroscience.2018.03.017>
- Kang, H., Macherey, O., Roman, S., & Pressnitzer, D. (2021). Auditory memory for random time patterns in cochlear implant listeners. *The Journal of the Acoustical Society of America*, 150(3), 1934–1944. <https://doi.org/10.1121/10.0005728>
- Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., & Griffiths, T. D. (2014). Representations of specific acoustic patterns in the auditory cortex and hippocampus. *Proceedings of the Royal Society B: Biological Sciences*, 281(1791), 20141000. <https://doi.org/10.1098/rspb.2014.1000>
- Lange, K., Kühn, S., & Filevich, E. (2015). “Just another tool for online studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, 10(6), e0130834. <https://doi.org/10.1371/journal.pone.0130834>
- Lawrence, M. A. (2016). *ez: Easy analysis and visualization of factorial experiments* (R package version 4.4-0). <https://CRAN.Rproject.org/package=ez>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Luo, H., Tian, X., Song, K., Zhou, K., & Poeppel, D. (2013). Neural response phase tracks how listeners learn new acoustic representations. *Current Biology*, 23(11), 968–974. <https://doi.org/10.1016/j.cub.2013.04.031>
- Macmillan, N. A. (2001). Signal Detection Theory. In Smelser, N. J., & Baltes, P. B. (Eds.). *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon, 14075–14078. <https://doi.org/10.1016/b0-08-043076-7/00677-x>
- McDermott, J. H., Wroblewski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences*, 108(3), 1188–1193. <https://doi.org/10.1073/pnas.1004765108>
- McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, 16(4), 493–498. <https://doi.org/10.1038/nn.3347>
- Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, 53(4), 1551–1562. <https://doi.org/10.3758/s13428-020-01514-0>
- Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, 16(4), 406–419. <https://doi.org/10.1037/a0024377>
- Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2018). *BayesFactor: Computation of Bayes factors for common designs* (R Package Version, 0.9.12–4.2). <https://cran.r-project.org/package=BayesFactor>
- Oldfield, R. C. (1971). The Edinburgh handedness inventory. *Neuropsychologia*, 9(1), 97–111. <https://doi.org/10.1002/mus.10529>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Song, K., & Luo, H. (2017). Temporal organization of sound information in auditory memory. *Frontiers in Psychology*, 8, 999. <https://doi.org/10.3389/fpsyg.2017.00999>

- Viswanathan, J., Rémy, F., Bacon-Macé, N., & Thorpe, S. J. (2016). Long term memory for noise: Evidence of robust encoding of very short temporal acoustic patterns. *Frontiers in Neuroscience*, *10*, 490. <https://doi.org/10.3389/fnins.2016.00490>
- Warren, R. M., Bashford, J. A., Cooley, J. M., & Brubaker, B. S. (2001). Detection of acoustic repetition for very long stochastic patterns. *Perception & Psychophysics*, *63*(1), 175–182. <https://doi.org/10.3758/BF03200511>
- Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. <https://doi.org/10.1016/j.tics.2009.09.003>
- Woods, K. J. P., & McDermott, J. H. (2018). Schema learning for the cocktail party problem. *Proceedings of the National Academy of Sciences*, *115*(14), E3313–E3322. <https://doi.org/10.1073/pnas.1801614115>
- Wright, B. A., & Zhang, Y. (2009). Insights into human auditory processing gained from perceptual learning. In Gazzaniga, M. S.(Ed.). *The cognitive neurosciences* (4th ed., pp. 353–365). Massachusetts Institute of Technology.