Vision Research 99 (2014) 19–36

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Memory and incidental learning for visual frozen noise sequences

Jason M. Gold [a,*], Avi Aizenman [b], Stephanie M. Bond [b], Robert Sekuler [b]

[a] Department of Psychological and Brain Sciences, Indiana University, United States
[b] Volen Center for Complex Systems, Brandeis University, United States

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 7 September 2013
10 September 2013
Available online 25 September 2013

Keywords:
Incidental learning
Frozen noise
Memory
Reverse correlation
Mirror image

ABSTRACT

Five experiments explored short-term memory and incidental learning for random visual spatio-temporal sequences. In each experiment, human observers saw samples of 8 Hz temporally-modulated 1D or 2D contrast noise sequences whose members were either uncorrelated across an entire 1-s long stimulus sequence, or comprised two frozen noise sequences that repeated identically between a stimulus' first and second 500 ms halves ("Repeated" noise). Presented with randomly intermixed stimuli of both types, observers judged whether each sequence repeated or not. Additionally, a particular exemplar of Repeated noise (a frozen or "Fixed Repeated" noise) was interspersed multiple times within a block of trials. As previously shown with auditory frozen noise stimuli (Agus, Thorpe, & Pressnitzer, 2010) recognition performance (d') increased with successive presentations of a Fixed Repeated stimulus, and exceeded performance with regular Repeated noise. However, unlike the case with auditory stimuli, learning of random visual stimuli was slow and gradual, rather than fast and abrupt. Reverse correlation revealed that contrasts occupying particular temporal positions within a sequence had disproportionately heavy weight in observers' judgments. A subsequent experiment suggested that this result arose from observers' uncertainty about the temporal mid-point of the noise sequences. Additionally, discrimination performance fell dramatically when a sequence of contrast values was repeated, but in reverse ("mirror image") order. This poor performance with temporal mirror images is strikingly different from vision's exquisite sensitivity to spatial mirror images.

1. Introduction

Humans are gifted pattern-recognizers, blessed with stunning ability to register, remember and exploit the similarities among sequences of sensory experiences. One especially useful approach to studying pattern recognition is to probe observers' ability to distinguish random stimulus sequences from random sequences onto which some form of structure has been imposed. Because random sequences comprise a homogenous pool of stimuli and can be devoid of semantic content, they put the research focus squarely on pattern-recognition's early stages – sensory processing and memory for features that are challenging to identify and process categorically (Kaernbach, 2004).

Over half a century, multiple researchers have exploited one simple but potentially informative strategy for imposing structure on random stimuli: repetition of a stored ("frozen") noise sample. Among the earliest uses of frozen noise noise, Guttman and Julesz (1963) showed that reiterating the same frozen auditory noise sample multiple times in succession generated characteristic auditory percepts, whose quality varied with the period of reiteration. These observations were instrumental in Neisser's (1967) postulation of an echoic memory, a limited-duration auditory buffer. Later, Kaernbach (2004) showed that even just a single repetition of frozen noise could be discriminated from a non-frozen (that is, nonrepeating) stimulus of equal duration. Recently, Agus, Thorpe, and Pressnitzer (2010) extended this work to explore the formation of auditory memory for sequences of random inputs. Their observers tried to discriminate between (i) 1-s long random sequences of auditory noise ('Random Noise'), and (ii) 1-s long sequences in which a single 500-ms auditory noise sequence was repeated so that it was presented twice in succession with no break between ('Repeated Noise'). Observers' performance discriminating between the two types of stimuli demonstrated their ability to exploit short term auditory memory – memory for the initial 500 ms of the stimulus – that had to be matched against the immediately ensuing 500 ms of the stimulus. Observers had good success in making this discrimination. Importantly, at random times during a block of trials, Agus et al. inserted a trial on which the very same Repeated Noise stimulus was recycled. At issue was whether experience cumulated over multiple trials with the same Repeated Noise stimulus would improve performance. Despite the many other stimuli intervening between successive presentations of a fixed Repeated Noise stimulus, performance

* Corresponding author.
E-mail address: jgold@indiana.edu (J.M. Gold).

0042-6989/$ - see front matter
http://dx.doi.org/10.1016/j.visres.2013.09.005

with that stimulus did improve relative to randomly generated Repeated Noise stimuli. This improvement showed that observers not only formed short term memories, which allowed successive 500 ms sequences on a trial to be compared, but also were simultaneously forming longer-term memories, which cumulated over many trials. Moreover, Agus et al. showed that such learning of a fixed Repeated Noise exemplar was retained over experimental sessions, and was robust in the face of various acoustic transformations.

Agus et al. demonstrated the auditory system's remarkable ability to extract, store, and cumulate structure embedded in an arbitrary random sequence. But do these results reveal something specific to the processing of auditory information, or do they instead reflect some more generalized ability of human sensory systems to process arbitrary patterns, independent of the modality from which the patterns are received? Motivated by Agus et al.'s findings, and by the long-standing controversy about parallels between visual and auditory memory (e.g., Cohen, Horowitz, & Wolfe, 2009; Julesz & Hirsch, 1972; Visscher et al., 2007), we set out to examine vision's ability to support both short term memory and the longer term memory that Agus et al. demonstrated for audition. To do this, we adapted Agus et al.'s paradigm to explore observers' ability to discriminate and learn arbitrary visual noise sequences that are generated by temporally modulating stimulus contrast. We also applied reverse correlation analysis to the complex, temporally modulated stimuli used in our experiments, in order to identify in detail the strategies observers used when making their judgments (Neri & Heeger, 2002; Simoncelli, 2003).

2. Experiment one

Experiment 1 was modeled after the first of Agus, Thorpe, and Pressnitzer (2010)'s experiments, but used visual rather than auditory stimuli. In our experiment, observers tried to detect the presence (or absence) of a repeated sequence of visual contrast noise. Agus et al.'s noise stimuli were sampled and presented at 44 kHz, a value about twice the upper limit of hearing of otologically-normal young adults, but several log units above the temporal resolution of human vision. The many differences between the properties of vision and audition, including differences in temporal resolution, challenge attempts to make fair comparisons between the two (Visscher et al., 2007). In our experiments, we modulated the contrast of our visual stimuli across time as a step function at 8 Hz, a value near the peak of the human temporal contrast sensitivity function (Wilson, 1980). On each trial, the temporal modulation produced a sequence of eight items, each $\sim$133 ms in duration.

On each trial, observers' task was to compare the sequence of the last four contrasts that they saw to the their memory of the sequence of the first four contrasts that they saw. We chose to use stimulus sequences whose units were four items in length because of evidence that visual short-term memory capacity has an upper limit of about four items (Phillips, 1974; Vogel, Woodman, & Luck, 2001). Observers were tested with two different kinds of visual noise: 1D noise, whose contrast was spatially uniform at any moment, but varied over time, at 8 Hz; and 2D noise, whose contrast varied in both time and space. As explained below, the 2D contrast variation in space produced a series of vertical stripes whose contrasts varied independently of one another over time.

2.1. Methods

2.1.1. Observers

Fourteen observers between the ages of 18 and 27 years participated in the experiment for a stipend of $10 per experimental session. All observers had normal or corrected to normal visual acuity, and were naive to the purposes of the experiment.

2.1.2. Apparatus

Unless otherwise specified, the following conditions were maintained across all experiments. Stimuli were presented against a uniform background of average luminance (19.03 cd/m^2) on a CRT monitor (Sony Trinitron UltraScan P780) at a resolution of 1024×768 pixels (33×24.5 cm) and refresh rate 75 Hz. Display contrasts were linearized by means of a calibration-based lookup table. Stimuli were generated and presented by an Apple iMac computer running Matlab (version 7.7) and extensions from the Psychophysics Toolbox (Brainard, 1997). Viewing was binocular through natural pupils. A viewing distance of 57 cm was enforced by means of a chin support. The computer display provided the only source of illumination in the room.

2.1.3. Stimuli

Gaussian white contrast noise was used to generate the contrast levels of all the stimuli in the experiment. Contrast was defined as $(L_{\text{pix}} - L_{\text{bg}})/L_{\text{bg}}$, where L_{pix} is the luminance of a given pixel, and L_{bg} is the background luminance (19.03 cd/m^2). Note that, according to this definition, contrast values could be either positive or negative. For our study, noise contrast levels were sampled from a normal distribution with mean equal to zero contrast and a variance equal to 0.2. Candidate samples more than ±2 standard deviations from zero contrast were replaced by fresh samples, which restricted the range of contrast increments and decrements comprising any sequence. This algorithm for generating stimuli was intended to clamp the distinctiveness of individual sequences so that it would be difficult for observers to identify and explicitly recognize particular sequences.

Each stimulus sequence consisted of eight contrast levels presented in rapid succession to the same $4.1^\circ \times 4.1^\circ$ (128×128 pixels) region of the display. Each contrast level in an entire eight-item sequence was presented for 10 screen refreshes of the CRT display ($\sim$133 ms), which meant that a complete eight-item sequence played out in 1067 ms.

As mentioned earlier, the contrast noise was distributed spatially in two different ways within a 128×128 pixel stimulus square. To generate what we will call 1D noise, for each of the eight stimuli in a sequence, every pixel in a 128×128 pixel stimulus square was assigned the same contrast value. Thus, a 1D noise sequence consisted of a series of eight contrast values (3 bits of information). The other class of stimuli, which we will call 2D noise, was generated by assigning a different noise sample to each column of pixels in any square stimulus in an eight-item sequence. This produced, for each item in a sequence, 128 vertical stripes, each $\sim$2 arcmin wide. The contrast levels of stripes within any item were independent of one another; moreover, the contrast levels varied independently over time, that is, across the eight items in a stimulus sequence. Thus, a 2D noise stimulus comprised a sequence of 128×8 contrast samples, or 10 bits of information, considerably more than in a sequence with our 1D stimuli.

For each kind of noise, 1D and 2D, we applied three different manipulations to the images' statistical structure over the eight items in a sequence (see Fig. 1). The manipulations produced three categories of stimuli, which we term Noise (N), Repeated Noise (RN), and Fixed Repeated Noise (FixRN). For stimuli of category N, the contrasts of the eight items comprising a sequence were independent of one another. In the case of repeated noise (RN), the first four frames ($\sim$533 ms) of the sequence repeated identically during the second half of the stimulus sequence, continuously and with no break in between halves. Finally, in the case of frozen or 'fixed' repeated noise (FixRN), a single randomly chosen RN stimulus was generated anew for each block of trials and was used