Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

# Auditory statistical learning: predictive frequency information affects the deployment of contextually mediated attentional resources on perceptual tasks

Doug Alards-Tomalin, Nicholaus P. Brosowsky & Todd A. Mondor

**To cite this article:** Doug Alards-Tomalin, Nicholaus P. Brosowsky & Todd A. Mondor (2017): Auditory statistical learning: predictive frequency information affects the deployment of contextually mediated attentional resources on perceptual tasks, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2017.1353518

**To link to this article:** http://dx.doi.org/10.1080/20445911.2017.1353518

Published online: 13 Jul 2017.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=pecp21

**Download by:** [Eastern Michigan University] **Date:** 14 July 2017, At: 05:33

JOURNAL OF COGNITIVE PSYCHOLOGY, 2017
https://doi.org/10.1080/20445911.2017.1353518

Routledge
Taylor & Francis Group

Check for updates

# Auditory statistical learning: predictive frequency information affects the deployment of contextually mediated attentional resources on perceptual tasks

Doug Alards-Tomalin[a], Nicholaus P. Brosowsky[b] and Todd A. Mondor[a]

[a]Department of Psychology, University of Manitoba, Winnipeg, MB, Canada; [b]Department of Psychology, The Graduate Center of the City University of New York, New York, NY, USA

**ABSTRACT**
Statistical learning is a general phenomenon in which environmental regularities are implicitly acquired through repeated exposure to those environments. Sometimes, that information can be utilised to affect various aspects of cognitive performance (e.g. reaction time) on tasks that utilise selective attention (e.g. visual search). In the current study, we examined the effect of passively listening to predictive auditory contexts in facilitating attention to a certain frequency or frequency range. In doing so, we found that there is a general tendency for attentional resources to be negatively affected when the context sequences are made novel after context–target associations have been formed (Experiment 1), and when the context no longer reliably cues the previous target (Experiment 2). The experiments are framed to contrast *Associative Learning* and *Memory Hypothesis* perspectives.

**ARTICLE HISTORY**
Received 26 September 2016
Accepted 4 July 2017

**KEYWORDS**
Contextual cueing; auditory attention; pitch perception; implicit memory; statistical learning

Given the limited capacity of human attention and visual working memory mechanisms (Fougnie & Marois, 2006; Huang & Pashler, 2005; Pashler, 1988), the ability to encode and comprehend large quantities of information from the environment is impressively adept. For example, when the circumstances allow for the parallel processing of visual features, the ability to detect a target amongst distractors is nearly instantaneous (Egeth, Jonides, & Wall, 1972; Nakayama & Silverman, 1986; Treisman & Gelade, 1980). Alternatively, conjunctive (serial) searches tend to be slow and consume a great deal of attentional resources (Scholl, 2001; Treisman & Sato, 1990). One means by which attentional resources are preserved during cognitively demanding tasks is via implicit (statistical) learning, in which information about contextual regularities and redundancies in the environment are gradually acquired and used to impact the allocation of limited attentional resources.

It has been recently proposed (see the *Memory Hypothesis of Contextual Control*) that the retrieval of past attentional control settings and filters is largely context dependent, contingent on re-experiencing the context in which that control setting was originally used (Brosowsky & Crump, 2016; Crump, 2016). One outcome of this is that selective attention tasks occurring within a previously experienced context will receive a processing advantage from having a control setting automatically retrieved versus having to generate a new setting. This serves as one account for why acquiring information about patterns and regularities from large quantities of sensory data (i.e. statistical learning) can benefit the deployment of resources on selective attention tasks.

Contextual control can exert a variety of effects on selective attention tasks. For example, the attentional costs related to task switching are offset when some feature of the target display is made predictive of an upcoming task switch (Leboe, Wong, Crump, & Stobbe, 2008). *Stroop* effect sizes can be influenced by varying the total proportion of congruent trials (RED printed in red ink) versus incongruent trials (RED in blue ink) across the experiment (Jacoby, Lindsay, & Hessels, 2003) or on the basis of nominally irrelevant shape/location contextual information (Bugg & Crump, 2012; Corballis & Gratton, 2003; Crump, Brosowsky, & Milliken, 2016; Crump, Gong, & Milliken, 2006; Crump & Milliken, 2009), while

**CONTACT** Doug Alards-Tomalin douglas.alards-tomalin@umanitoba.ca
© 2017 Informa UK Limited, trading as Taylor & Francis Group

number priming effects have been found to be greater in conditions where most of the trials feature congruent, as opposed to incongruent primes (Heinemann, Kunde, & Kiesel, 2009). Finally, while the impact of contextual control is generally thought to be implicit, some types of contextual information appear to require explicit (effortful) processing before having an influence on responding (Alards-Tomalin, Walker, Nepon, & Leboe-McGowan, 2017; Brosowsky & Crump, 2016). Therefore, context can guide the allocation of attentional resources on a variety of selective attention tasks, depending on factors regarding whether the scene is processed globally or locally, whether the contextual information is presented simultaneously or sequentially, and which sensory modalities are involved.

Research using visual target search tasks have been particularly productive in the field of statistical learning (Corbett & Melcher, 2014; Goujon, Didierjean, & Thorpe, 2015; Oliva & Torralba, 2007). For example, the repetition of visual arrays composed of predictive target–distractor configurations has been found to facilitate target detection times, a form of learning that occurs implicitly, and without feedback (Fiser & Aslin, 2001), a phenomenon more specifically termed *contextual cueing* (Chun, 2000; Chua & Chun, 2003; Chun & Jiang, 1998, 1999; Chun & Turk-Browne, 2007; Goujon, Didierjean, & Marmèche, 2007, 2009; Jiang & Chun, 2001, 2003; Olson & Chun, 2002; Turk-Browne, Scholl, Chun, & Johnson, 2009, 2005).

In the formative paper on this topic, Chun and Jiang (1998) had participants search for a rotated letter *T* amongst a field of rotated letter *L***s**, which functioned as distractors. Across trials, a subset of the target/distractor patterns were repeated. The authors found that reaction times (RTs) on the repeated patterns were facilitated after only several repetitions of a context–target configuration (Chun, 2000) and lasted for as long as a week (Chun & Jiang, 2003), despite participants having no subjective awareness of this learning. Since then, contextual cueing has been shown to be a robust phenomenon in the visual search literature and has been replicated using naturalistic visual environments (Brockmole, Castelhano, & Henderson, 2006; Brockmole, Hambrick, Windisch, & Henderson, 2008; Brockmole & Henderson, 2006a, 2006b), during parallel search tasks (Geyer, Zehetleitner, & Müller, 2010) and even when the predictive context was presented on the trial prior to the scene containing the visual target (Ono, Jiang, & Kawahara, 2005).

In one important recent study, Corbett and Melcher (2014) demonstrated that as long as the spatial configuration of the distractors is invariant across trials, visual contextual cueing occurs even when the target item's location within the scenes varies (alternatively, Makovski & Jiang, 2009; Zellin, Conci, von Mühlenen, & Müller, 2011). This phenomenon is similar, if not identical, to the issue of contextual control discussed earlier (Crump, 2016). If the contextual regularities concerning the task irrelevant (i.e. distracting) stimuli are acquired (via implicit or explicit processes) and stored in memory, then re-experiencing a previous context should facilitate the retrieval of the previous attentional control settings and filters used to process those contextual regularities, permitting more resources to be devoted to the search task. This *Memory Hypothesis* differs substantially from the typical explanation for Contextual Cueing referred to as the *Associative Learning Account* (Chun & Jiang, 1998; Makovski & Jiang, 2009; Rosas, Todd, & Bouton, 2013), which states that spatial associations are formed between distractor positions and information regarding the target's location in an array, such that the repetition of the context serves as predictive cue for the target's location. These theories make fundamentally different predictions about the underlying cause of contextual cueing. From the *Associative Learning Account's* perspective, previously encoded contextual regularities actively direct attention (via automated processes) to the target, whereas the *Memory Hypothesis* states that reinstated attentional control mechanisms facilitate the analysis of the contextual (distracting) elements by releasing resources for the target search task. In the current study, we examine the nature of statistical learning as it relates to basic auditory perceptual judgments, while contrasting these two models.

## Current study

Because the nature of auditory information is such that it unfolds regularly over time, the classic form of visual contextual cueing, in which a target is presented amongst a set of spatially consistent distractors, would not apply to auditory events. As such, any attempt at replicating a visual contextual cueing paradigm in audition would necessitate the extraction of temporal regularities in a sequence of auditory events that are in some way predictive of an upcoming auditory target embedded within that sequence.

While demonstrations of sequential (temporal) statistical learning are less common, there have been examples of it demonstrated in vision (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002) and several in audition (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999) using artificial grammar learning paradigms, and at least one demonstration of visual temporal contextual cueing. In this study, Olson and Chun (2001) demonstrated that people are sensitive to temporal regularities when predicting the onset of an embedded target. For example, in their experiment, participants searched for a target (the letter "k" or "x") amongst a rapidly changing stream of 15 distractor letters (e.g. B-M-Y-G-**X**- …). These letter sequences were repeated over a series of "learning" blocks. After the learning block, participants were presented with a test block which contained only novel sequences. They found a progressive decrease in RT across the learning blocks, followed by a significant increase in RT at test.

In the current study, we presented participants with a series of learning trials in which context sequences were composed of three discrete pure tones, each having a different frequency. While using only three context-sequence tones might appear to be inconsistent with the general framework of previous contextual cueing studies, which often implement a larger number of distractor items, it should be noted that contextual cueing has been demonstrated when as few as two visual distractors in a display are repeated (Brady & Chun, 2007). Furthermore, we employed an auditory perception task rather than the standard search task used in most contextual cueing studies. Therefore, the position of the target in respect to the context items was irrelevant to the target task. Our study parallels prior work examining the influence of context on selective attention (for a review, Bugg & Crump, 2012) in that we investigated whether auditory context can cue the deployment of auditory attention to an associated target sound's frequency.

Throughout both experiments, the frequencies of the three context tones were predictive of a perceptual characteristic of the fourth (target) sound. In Experiment 1, the target tone was categorised as either an up-glide (a tone that continuously increased in frequency) or a down-glide (a tone that continuously decreased in frequency). In Experiment 2, participants determined whether there was a 10-ms gap at the midpoint of a target tone frequency. Experiment 1 provides a preliminary demonstration of how auditory statistical learning might impact basic perceptual judgments about a sound in a manner consistent with contextual cueing, whereas Experiment 2 serves to test *Memory Hypothesis* and *Associative Learning* accounts of the contextual cueing phenomenon.

## Experiment 1

### Method

#### Participants

Thirty-two participants enrolled in an Introduction to Psychology course at the University of Manitoba were recruited (20 females, 12 males, mean age = 19.72). All the participants self-reported normal hearing prior to participating in the study and received partial course credit for their time. All the reported experiments received prior approval by the University of Manitoba, Fort Garry Campus, Research Ethics Board (REB).

#### Materials

The data for both Experiments were collected on Dell precision T5400 desktop computers using E-Prime 1.2 system software (Psychology Software Tools, 2012). The same auditory stimuli were used in both Experiments, which were generated in Adobe Audition 1.5 (Adobe Systems Incorporated, 2007) using a sampling rate of 32,000 Hz and were presented binaurally via Sony MDR-600 headphones. The stimuli were 100 ms duration, pure tones, containing 3 ms onset/offset ramps to eliminate onset/offset clicks. The context sequence tones had frequencies of 511, 561, 614, 673, 738, 809, 887, 973, 1067, 1170, 1283, 1407, 1543, 1692, 1855, 2034, 2230, 2445 Hz (all equally spaced in log-frequency). All stimuli were equated for subjective loudness prior to presentation. These 18 sounds were randomly arranged into 6 context sequence patterns, each composed of three context tones followed by a target tone (see Table 1).

Three of the patterns were predictive of the presentation of an up-glide target (which smoothly transitioned in frequency from 511 to 2445 Hz) and three a down-glide target (which smoothly transitioned in frequency from 2445 to 511 Hz). The glides were 140 ms and adjusted to a mean energy intensity level of 90 dB. The inter-stimulus interval (ISI) separating tone 1 from tone 2 (ISI-1) and the interval separating tone 2 from tone 3 (ISI-2) were randomly selected from the intervals

**Table 1.** Experiment 1 context tone frequencies and ISI durations.

| Context pattern | Tone (Hz) 1 | Tone (Hz) 2 | Tone (Hz) 3 | ISI (ms) 1 | ISI (ms) 2 | ISI (ms) 3 | Target (glide type) |
|---|---|---|---|---|---|---|---|
| 1 | 1,855 | 973 | 887 | 85 | 160 | 1,000 | Up |
| 2 | 1,692 | 2,034 | 2,445 | 210 | 260 | 1,000 | Down |
| 3 | 1,407 | 809 | 738 | 135 | 60 | 1,000 | Up |
| 4 | 1,067 | 1,283 | 2,230 | 10 | 335 | 1,000 | Down |
| 5 | 511 | 1,170 | 1,543 | 360 | 35 | 1,000 | Up |
| 6 | 614 | 561 | 673 | 235 | 285 | 1,000 | Down |

10, 35, 60, 85, 135, 160, 210, 235, 260, 285, 335 and 360 ms, while the interval separating tone 3 from the target tone (ISI-3) was 1000 ms.

### *Procedure*

The participants were instructed to listen to the three —task-irrelevant—context tones, and then categorise the fourth (target) tone as an up- or down-glide as rapidly as possible with the keys "1" or "0". The keyboard button/response label mappings were counterbalanced across participants. The participants completed a block of 20 practice trials in which feedback was presented following each response. The participants were required to achieve a baseline threshold of 70% accuracy to move on to the training trials. After successful completion of the practice block, the participants completed three blocks of training trials, each containing 60 randomised repetitions of the context pattern–target configurations provided in Table 1 (10 randomised presentations of each of the three up-glide context sequences and 10 randomised presentations of each of the three down-glide context sequences), for a total of 180 training trials. Following the training trials, the participants were presented with one test block composed of 60 randomly presented trials. During the test block, novel three-tone context patterns were randomly constructed from the tones used during the leaning block and matched with each target. The same target sounds used during the learning blocks were also presented during the test block (30 up-glide and 30 down-glide trials), with the target task (identifying the direction of the glide) being identical. There was no break between training and test blocks, and the participants were not informed of the transition.

### *Results and discussion*

Accuracy on the auditory task approached ceiling across learning and test phases (see Table 2, Experiment 1). Because the participants exhibited such a high level of task accuracy, percentages showed no significant degree of variability across learning and test blocks (all $Fs < 1$). Therefore, only the median RTs for correct responses were reported in the results section of Experiment 1.

To examine the impact of context sequence repetition during the learning phase, median correct RTs for each learning block (*Block 1*, *Block 2*, *Block 3*) were submitted to a one-way repeated measure analysis of variance (ANOVA) treating Block as a three-level within-participant factor and correct median RTs as the dependent measure. Overall, the change in RT across the three learning blocks was non-significant: $F(2, 62) = 1.874$, $p = .162$, $\eta p^2 = .057$ (see Figure 1).

To determine if modifying the context sequence–target associations had any impact on RT, the median RTs in the third learning block ($M = 415$ [13.42]) were compared against the median RTs on the test block ($M = 450$ [16.16]). This analysis revealed that RTs were significantly slower when novel context patterns were presented at test $F(1, 31) = 9.19$, $p = .005$, $\eta p^2 = .23$ (see Figure 1).

Hence, in Experiment 1, while there was no consistent decrease in RT across the learning blocks, the introduction of novel context patterns during the test block did lead to an increase in RTs in comparison to the final training block, evidence that statistical regularities in the context sequence had been acquired across the training blocks. This result demonstrates that even in instances when the acquisition of temporal regularities does not lead to a general decrease in RT across trials, a change in contingencies can still interfere with processing.

One issue with Experiment 1's design was that the context pattern/target pairings were associated, such that within a single block of learning trials, the

**Table 2.** Experiment 1 and 2 accuracy rates.

| | | Block 1 | Block 2 | Block 3 | Block 4 |
|---|---|---|---|---|---|
| Experiment 1 | *M* | .97 | .98 | .98 | .97 |
| | *SE* | (.01) | (.01) | (.01) | (.01) |
| Experiment 2 | *M* | .99 | .99 | .99 | .98 |
| | *SE* | (.004) | (.003) | (.004) | (.01) |

Note: *M* = mean, *SE* = standard error of the mean.

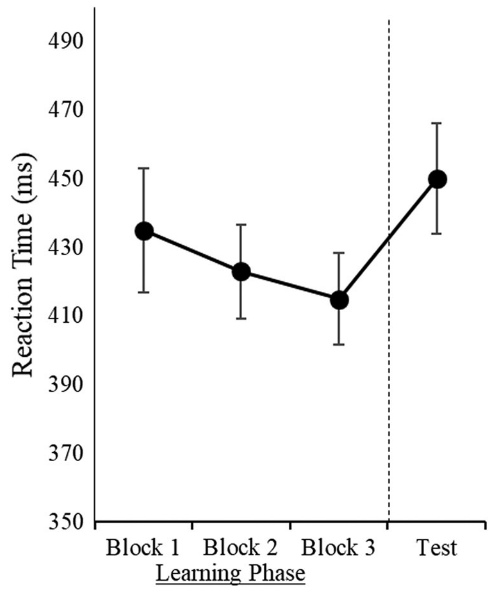


**Figure 1.** Experiment 1 median correct RTs as a function of block. Error bars represent the standard error of the mean.

same three context patterns were 100% predictive of a certain target response (up-glide/down-glide). Therefore, the increase in RT from learning block 3 to the test block may have been due to participants anticipating which target response they were required to generate from the initial context pattern (i.e. response priming), and then having that prediction disconfirmed. To expand, the repetition of context patterns in association with certain target responses could have led to the formation of a stimulus-response (S-R) contingency (e.g. Schmidt & Besner, 2008), or "event file" (Hommel, 1998, 2004), in which a response, continually elicited by a specific stimulus, becomes bound to the stimulus in episodic memory.

In one recent example, Giesen and Rothermund (2015) investigated how the repetition of contextual information about word meaning might affect RTs. In their study, participants viewed words which first appeared in white font, and then changed to either red or green. The participants were tasked with categorising the colour of the word as *red* or *green* as quickly as possible, with word meaning being task irrelevant. The authors established S-R bindings by having conditions in which certain words were always predictive of a "red" or "green" response versus another condition in which the S-R bindings were unreliable. They found that while participants were generally unaware of the established S-R contingencies, the participants were significantly faster when the word was predictive of the appropriate colour response and slower when the contingency was unreliable. The results of Experiment 1 may be due to a similar effect, in which an established S-R contingency between the context tones and target is disrupted at test, increasing RT. Another possibility—as indicated by the *Memory Hypothesis*—is that the presentation of novel context pattern configurations at test may have required the generation of new attentional control sets, causing a temporary depletion of resources prior to the target task.

To address these issues, the task from Experiment 1 was modified in Experiment 2 so that: (1) while the context patterns were predictive of the target tone's frequency, they could not be associated with a specific response. To accomplish this, we changed the target task from a frequency detection to a gap detection task. This was done so that it the correct response on any trial could not be anticipated from the frequencies of the preceding context sounds. To accomplish this, while the context sequences were still predictive of a certain frequency target sound, they were not predictive of whether the target contained or did not contain a short gap at its midpoint (gap present, gap absent), for which both outcomes were paired equally often with each context/target frequency pairing. This ensured that context/target associations could not be used as a basis for forming an S-R contingency. (2) The same context sequences used during the learning phase were reused at test; however, contingencies were altered by pairing each target sound with a different context, thus ensuring no increase in context pattern novelty at test.

The two theories mentioned earlier make different predictions, given this second modification. The *Memory Hypothesis* predicts that learning statistical regularities across the context sequences will lead to the creation of attentional control sets that can be effortlessly retrieved upon re-encountering the context. This will provide an overall savings on attentional resources that can then be spent on the target task. As the context sequences used at test were identical to those heard during the learning phase, the *Memory Hypothesis* predicts that while RTs should decrease across the learning trials, due to the retrieval of previously used attentional control sets, there should be no increase in RT at test. Alternatively, the *Associative Learning* account suggests that if the context sequences are being used to predict the target sound frequency, the disconfirmation of that prediction a test will be

**Table 3.** Experiment 2 context tone frequencies, ISI durations and target types for the learning and test phases.

| Context pattern | Tone (Hz) 1 | 2 | 3 | ISI (ms) 1 | 2 | 3 | Target (with/without 10 ms gap) |
|---|---|---|---|---|---|---|---|
| Learning phase | | | | | | | |
| 1 | 2,034 | 511 | 1,170 | 235 | 160 | 1,000 | 887 Hz |
| 2 | 561 | 673 | 1,543 | 210 | 385 | 1,000 | 614 Hz |
| 3 | 973 | 1,407 | 2,230 | 335 | 185 | 1,000 | 1,885 Hz |
| 4 | 1,692 | 809 | 1,067 | 260 | 360 | 1,000 | 1,283 Hz |
| Test phase | | | | | | | |
| 1 | 2,034 | 511 | 1,170 | 235 | 160 | 1,000 | 1,855 Hz |
| 2 | 561 | 673 | 1,543 | 210 | 385 | 1,000 | 1,283 Hz |
| 3 | 973 | 1,407 | 2,230 | 335 | 185 | 1,000 | 887 Hz |
| 4 | 1,692 | 809 | 1,067 | 260 | 360 | 1,000 | 614 Hz |

disruptive, and cause RTs to increase relative to the learning phase.

## Experiment 2

Experiment 2 followed the same paradigm as Experiment 1, three-tone context patterns were paired with a specific frequency target tone that may, or may not, have a 10-ms gap at its midpoint. The presentation of a gap or non-gap target was randomised across trials and could not be inferred from the preceding context pattern, eliminating the possibility of participants forming an S-R binding during the learning phase.

### *Method*

#### *Participants*

Twenty-four participants enrolled in an Introduction to Psychology course at the University of Manitoba were recruited (14 females, 10 males, mean age = 21.21).

#### *Procedure*

The general procedure differed from Experiment 1 in several respects. First, the number of learning phase trials was reduced from 60 to 40 trials per block by eliminating two context–target configurations (a total of four context patterns were used instead of six). To accomplish this, four target tone frequencies were selected (614, 887, 1283 and 1885 Hz) and paired randomly with three context tones. The reason for this had to do with improving the acquisition of the pattern information by increasing overall contextual redundancy within the learning block (Corbett & Melcher, 2014). On each of three learning blocks, the four context–target configurations were randomly presented 10 times, 5 times with a 10-ms gap at the midpoint and 5 times with no gap (see Table 3).

Second, the target task differed from Experiment 1 in that participants judged whether the target sound contained a 10-ms gap. However, like Experiment 1, immediately following the learning phase, a test phase occurred in which the context patterns from the learning phase were associated with different frequency targets. The test block was composed of 40 trials.

### *Results and discussion*

As in Experiment 1, task accuracy approached ceiling across both learning and test phases (see Table 2, Experiment 2), showing no significant degree of variability across learning and test blocks (all $Fs < 1$). Therefore, only the median RTs for correct responses were reported in the results section of Experiment 2.

To ascertain the effect of context repetition across the learning blocks, the median correct RTs for each learning block (Block 1, Block 2, Block 3) were submitted to a one-way repeated measures ANOVA, treating Block as a three-level, within-participant's independent variable, and the median correct RTs as the dependent measure. This analysis revealed a significant main effect of Block $F(2, 46) = 9.720$, $p < .001$, $\eta p^2 = .297$, further evidenced by a significant linear trend $F(1, 23) = 13.104$, $p < .001$, $\eta p^2 = .363$. Therefore, in Experiment 2 there was a continuous reduction in RT on the gap detection task across the learning trials (see Figure 2). Post-hoc, pairwise contrasts further revealed that the difference between Block 1 and Block 2 was statistically significant $F(1, 23) = 4.90$, $p = .037$, $\eta p^2 = .18$, as was the difference between Block 2 and Block 3, $F(1, 23) = 7.69$, $p = .011$, $\eta p^2 = .25$. Overall, from Block 1 to Block 3 of the learning trials, there was a 54-ms decrease in RT on the gap detection task $F(1, 23) = 13.10$, $p = .001$, $\eta p^2 = .36$.

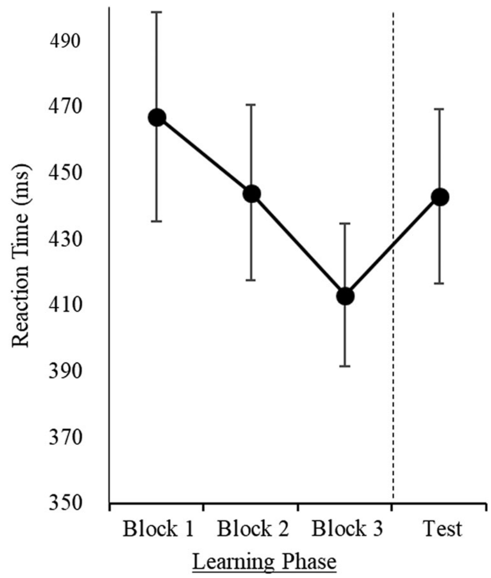


**Figure 2.** Experiment 2 median correct RTs as a function of block. Error bars represent the standard error of the mean.

Finally, we compared median RTs on the third block of learning trials against the test block. This analysis revealed a statistically significant performance deficit of 30 ms in RT from Block 3 to the test block $F(1,\ 23) = 6.23$, $p = .02$, $\eta p^2 = .21$. Therefore, the results appear to be more consistent with the *Associative Learning* account than the *Memory Hypothesis*.

To summarise, RTs on the target task decreased as participants acquired experience with the context pattern–target configurations, which increased on the Test Block. This suggests that auditory statistical learning can impact basic auditory perceptual tasks, and this effect cannot be attributed to the release of attentional resources due to the retrieval of a past attentional control set. We conclude that because the target task in Experiment 2 (gap present/gap absent) could not be directly inferred from the context sequence tone frequencies, the decrease in RT across learning blocks resulted due to an overall attentional savings incurred from being able to predict the target tone's frequency rather than having to explicitly process it. In a sense, this confirms a form of *Memory Hypothesis*, however, the reinstated attentional control set is related to processing the target, not the context, and is cued by the association of the context frequencies with a specific target frequency. When that contingency was suddenly modified (at test), discrepancies between the predication and the outcome led to a cost because the context sequences cued the retrieval of an attentional set originally used for processing a different frequency tone, which impeded RTs on the secondary target task (identifying the presence of a gap).

## General discussion

To briefly summarise, the form of auditory statistical learning we have demonstrated, like visual contextual cueing, is one in which participants acquired contextual regularities over time, and then formed predictive inferences about those contextual regularities relative to the target. While the acquisition of those regularities can lead to general attentional savings over learning trials (see Experiment 2), this is not necessarily always the case (see Experiment 1). However, disrupting those predictions at test tends to result in processing deficits (see Experiments 1 and 2).

To date, auditory contextual cueing has generally received less investigation than its visual counterpart. Some preliminary work in our lab (Doan, 2014) has suggested the existence of auditory temporal contextual cueing similar to that demonstrated by Olson and Chun (2001) in vision, namely, the temporal regularities in a sequence of, auditorily presented, spoken letters can be used to predict the onset of an auditory target embedded within that sequence. In the current study, we have found evidence to suggest that regularities in an auditory sequence leading up to the presentation of a temporally invariant target are acquired over trial repetitions, and then used to make predictions about the target's frequency, a sub-type of contextual cueing.

Furthermore, the findings are supportive of the *Associative Learning* model, over S-R binding (event file) and *Memory Hypothesis* models. Despite this, we do not dispute the validity of the *Memory Hypothesis of Contextual Control* and suggest that a form of memory-based contextual control is likely operating; however, the attention sets being retrieved are only applicable to processing the target tone frequency and not those of the context tones preceding the target. Hence, the presentation of a predictive context cues the retrieval of an attentional set used to process a certain frequency target tone, and when that prediction is disrupted, forming a new attentional set detracts resources from the gap detection task (Experiment 2).

### *The temporal nature of auditory statistic learning*

The question has arisen regarding whether statistical learning comprises a domain-general phenomenon or is domain specific. For example, statistical learning might be a general property of the nervous system that operates equivalently across the senses (Kirkham et al., 2002) or each modality might use similar, but independent, computational principles. Research has shown evidence for both accounts. In the case of the domain-general hypothesis, people can acquire information concerning contextual regularities in one modality and use those regularities to facilitate search in a different modality (Kawahara, 2007; Nabeta, Ono, & Kawahara, 2003).

Alternatively, constraints can impact the type of contextual information best suited for statistical learning in a specific modality. For example, some modalities (e.g. vision) are more sensitive to the acquisition of spatial regularities and other modalities (e.g. audition), the acquisition of temporal regularities (Conway & Christiansen, 2005, 2006; Kubovy, 1988; Lechelt, 1975; Saffran, 2002). In one interesting example, Conway and Christiansen (2005) examined the existence of modality constraints in statistical learning, including touch, vision and audition. On the task, the participants were presented with three to six sequential events on each trial, which were generated using an artificial grammar (and thus were *rule conforming*). These events involved: (a) vibrations on their fingers, (b) sequences of black squares presented at different horizontal locations or (c) sequences of pure tones differing in pitch. At test the participants were asked to distinguish between novel *rule conforming* sequences and *non-conforming* sequences. They found that the participants performed above change on all modalities and performed substantially better in the auditory modality (75% accuracy) than in either touch or vision (62% and 63% respectively). Thus, auditory statistical learning is unique with respect to visual statistical learning and best serves the acquisition of contextual information that unfolds temporally rather than spatially. As this is the case, more research is needed on the limits of statistical learning and contextual cueing phenomenon in each sensory modality.

It should be noted that auditory statistical learning has been previously demonstrated in both infants and adults. For example, 8-month-old infants are capable of extracting rule-based information regarding the relationships of neighboring speech sounds to determine whether a sound comprises a word (Saffran et al., 1996). Furthermore, in a replication of this experiment, the authors created an artificial, *non-linguistic*, “musical language” in which the phonological regularities previously applied to nonsense syllables were now applied to musical notes (Saffran et al., 1999). Similarly, the authors found that participants (infants and adults) could reliably identify rule conforming exemplars from non-rule conforming exemplars implicitly after exposure to enough rule conforming samples.

Therefore, the question is not whether statistical learning occurs in audition, but the extent of that learning. Auditory statistical learning plays an important role in the implicit formation of linguistic (Saffran et al., 1996) and non-linguistic auditory categories (Conway & Christiansen, 2005; Saffran et al., 1999). Notwithstanding this - the current study - we asked whether the acquisition of auditory regularities be used as a contextual control mechanism for selective attention? In other words, can the formation of statistical regularities in an auditory context benefit auditiory selective attention in the same way that the acquisition of visual statistical regularities can benefit visual search?

### *Conclusion*

The results of the current study support the notion of flexibility in statistical learning, such that it occurs not merely in learning the regularities of visual global patterns to predict a spatial location or learning the temporal regularities in visual/auditory sequences to predict target’s onset, but also impact the efficiency of basic auditory perceptual tasks like those involved in pitch perception. In the current study, we found that regularities in the frequencies across a sequence of three discrete context tones are used to facilitate predictions about the perceptual characteristics (e.g. frequency) of an associated target sound, and disruptions to these predictions cause lead to processing deficits — possibly because the retrieved attentional set used to process the target are no longer valid.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The research was supported by a grant from the Natural Sciences and Engineering Council of Canada to TAM. Thanks to Heather Rodd, Adriana Chipelski and Dana Bernier for their assistance in the preparation of this manuscript.

## References

Adobe Systems Incorporated. (2007). *Adobe Audition*. San Jose, CA: Adobe Systems Incorporated.

Alards-Tomalin, D., Walker, A. C., Nepon, H., & Leboe-McGowan, L. C. (2017). Dual-task interference effects on cross-modal numerical order and sound intensity judgments: The more the louder? *The Quarterly Journal of Experimental Psychology*, *70*(9), 1943–1963. doi:10.1080/17470218.2016.1216139

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 798–815. doi:10.1037/0096-1523.33.4.798

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 699–706. doi:10.1037/0278-7393.32.4.699

Brockmole, J. R., Hambrick, D. Z., Windisch, D. J., & Henderson, J. M. (2008). The role of meaning in contextual cueing: Evidence from chess expertise. *Quarterly Journal of Experimental Psychology*, *61*(12), 1886–1896. doi:10.1080/17470210701781155

Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *59*(7), 1177–1187. doi:10.1080/17470210600665996

Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. doi:10.1080/13506280500165188

Brosowsky, N. P., & Crump, M. J. C. (2016). Context-specific attentional sampling: Intentional control as a pre-requisite for contextual control. *Consciousness and Cognition*, *44*, 146–160. doi:10.1016/j.concog.2016.07.001

Bugg, J. M., & Crump, M. J. C. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*(SEP), 1–16. doi:10.3389/fpsyg.2012.00367

Chua, K.-P., & Chun, M. M. (2003). Implicit scene learning is viewpoint dependent. *Perception & Psychophysics*, *65*(1), 72–80. doi:10.3758/BF03194784

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. doi:10.1016/S1364-6613(00)01476-5

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. doi.10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365. doi:10.1111/1467-9280.00168

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 224–234. doi:10.1037/0278-7393.29.2.224

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. doi:10.1016/j.conb.2007.03.005

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. doi:10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities. *Psychological Science*, *17*(10), 905–912. doi:10.1111/j.1467-9280.2006.01801.x

Corballis, P. M., & Gratton, G. (2003). Independent control of processing strategies for different locations in the visual field. *Biological Psychology*, *64*, 191–209. doi:10.1016/S0301-0511(03)00109-1

Corbett, J. E., & Melcher, D. (2014). Stable statistical representations facilitate visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1915–1925. doi:10.1037/a0037375

Crump, M. J. C. (2016). Learning to selectively attend from context-specific attentional histories: A demonstration and some constraints. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*, 59–77. doi:10.1037/cep0000066

Crump, M. J. C., Brosowsky, N. P., & Milliken, B. (2016). Reproducing the location-based context-specific proportion congruent effect for frequency unbiased items: A reply to Hutcheon and Spieler (2016). *The Quarterly Journal of Experimental Psychology*, 1–16. doi:10.1080/17470218.2016.1206130

Crump, M. J. C., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review*, *13*(2), 316–321. doi.10.3758/BF03193850.

Crump, M. J. C., & Milliken, B. (2009). The flexibility of context-specific control: Evidence for context-driven generalization of item-specific control settings. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1523–1532. doi:10.1080/17470210902752096

Doan, L. A. (2014). Winnipeg, MB, Canada: University of Manitoba.

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology*, *3*(4), 674–698. doi:10.1016/0010-0285(72)90026-6

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. doi:10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. doi:10.1037/0278-7393.28.3.458

Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, *17*(6), 526–534. doi:10.1177/0963721411415340

Geyer, T., Zehetleitner, M., & Müller, H. J. (2010). Contextual cueing of pop-out visual search: When context guides the deployment of attention. *Journal of Vision*, *10*(5), 20. doi:10.1167/10.5.20

Giesen, C., & Rothermund, K. (2015). Adapting to stimulus–response contingencies without noticing them. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1475–1481. doi:10.1037/xhp0000122

Goujon, A., Didierjean, A., & Marmèche, E. (2007). Contextual cueing based on specific and categorical properties of the environment. *Visual Cognition*, *15*(3), 257–275. doi:10.1080/13506280600677744

Goujon, A., Didierjean, A., & Marmèche, E. (2009). Semantic contextual cuing and visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 50–71. doi:10.1037/0096-1523.35.1.50

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences*, *19*(9), 524–533. doi:10.1016/j.tics.2015.07.009

Heinemann, A., Kunde, W., & Kiesel, A. (2009). Context-specific prime-congruency effects: On the role of conscious stimulus representations for cognitive control. *Consciousness and Cognition*, *18*(4), 966–976. doi:10.1016/j.concog.2009.08.009

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*(1–2), 183–216. doi:10.1080/713756773

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500. doi:10.1016/j.tics.2004.08.007

Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition*, *94*, B101–B111. doi.10.1016/j.cognition.2004.06.006

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*(3), 638–644.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1105–1124. doi:10.1080/0272498004200051

Jiang, Y., & Chun, M. M. (2003). Contextual cueing: Reciprocal influences between attention and implicit learning. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 277–296). Philadelphia, PAUSA: John Benjamins Publishing.

Kawahara, J.-I. (2007). Auditory – visual contextual cuing effect. *Perception & Psychophysics*, *69*(8), 1399–1408. doi.10.3758/BF03192955

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. doi:10.1016/S0010-0277(02)00004-5

Kubovy, M. (1988). Should we resist the seductiveness of the space is to time as vision is to audition analogy? *Journal of Experimental Psychology*, *14*(2), 318–320. doi:10.1037//0096-1523.14.2.318

Leboe, J. P., Wong, J., Crump, M. J., & Stobbe, K. (2008). Probe-specific proportion task repetition effects on switching costs. *Perception & Psychophysics*, *70*(6), 935–945. doi:10.3758/PP

Lechelt, E. C. (1975). Temporal numerosity discrimination: Intermodal comparisons revisted. *British Journal of Psychology*, *66*(1), 101–108. doi.10.1111/j.2044-8295.1975.tb01444.x

Makovski, T., & Jiang, Y. V. (2009). Contextual cost: When a visual-search target is not where it should be. *Quarterly Journal of Experimental Psychology*, *63*(2), 216–225. doi.10.1080/17470210903281590

Nabeta, T., Ono, F., & Kawahara, J. I. (2003). Transfer of spatial context from visual to haptic search. *Perception*, *32*(11), 1351–1358. doi:10.1068/p5135

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, doi:10.1038/320264 a0

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. doi:10.1016/j.tics.2007.09.009

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302. doi:10.1080/13506280042000162

Olson, I. R, & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1299–1313. doi.10.1037//0278-7393.27.5.1299

Ono, F., Jiang, Y., & Kawahara, J. (2005). Intertrial temporal contextual cuing: Association across successive visual search trials guides spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 703–712. doi:10.1037/0096-1523.31.4.703

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. doi:10.3758/BF03210419

Psychology Software Tools. (2012). *E-Prime*. Sharpsburg, PA: Psychology Software Tools, Inc.

Rosas, J. M., Todd, T. P., & Bouton, M. E. (2013). Context change and associative learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 237–244. doi:10.1002/wcs.1225

Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196. doi:10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, NY)*. doi:10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. doi:10.1016/S0010-0277(98)00075-4

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 514–523. doi:10.1037/0278-7393.34.3.514

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1–2), 1–46. doi:10.1016/S0010-0277(00)00152-9

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi:10.1016/0010-0285(80)90005-5

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 459–478. doi:10.1037/0096-1523.16.3.459

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. doi:10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. doi:10.1162/jocn.2009.21131.Neural

Zellin, M., Conci, M., von Mühlenen, A., & Müller, H. J. (2011). Two (or three) is one too many: Testing the flexibility of contextual cueing with multiple target locations. *Attention, Perception & Psychophysics*, *73*(7), 2065–2076. doi:10.3758/s13414-011-0175-x