Research Article

When Audition Dominates Vision

Evidence From Cross-Modal Statistical Learning

Christopher W. Robinson and Vladimir M. Sloutsky

Department of Psychology and Center for Cognitive Science, The Ohio State University, Columbus, OH, USA

Abstract. Presenting information to multiple sensory modalities sometimes facilitates and sometimes interferes with processing of this information. Research examining interference effects shows that auditory input often interferes with processing of visual input in young children (i.e., auditory dominance effect), whereas visual input often interferes with auditory processing in adults (i.e., visual dominance effect). The current study used a cross-modal statistical learning task to examine modality dominance in adults. Participants ably learned auditory and visual statistics when auditory and visual sequences were presented unimodally and when auditory and visual sequences were correlated during training. However, increasing task demands resulted in an important asymmetry: Increased task demands attenuated visual statistical learning, while having no effect on auditory statistical learning. These findings are consistent with auditory dominance effects reported in young children and have important implications for our understanding of how sensory modalities interact while learning the structure of cross-modal information.

Keywords: attention, cross-modal processing, implicit learning

In many situations (e.g., encountering a speaking person or a honking car), people have to process information coming from different sensory modalities. How does information presented to one modality affect processing in a different modality and are interactions between sensory modalities symmetrical or asymmetrical in nature? There is much evidence suggesting that sensory modalities are interdependent, with stimuli in one modality affecting processing in another modality (Ernst & Bülthoff, 2004; Robinson, Best, Weng, & Sloutsky, 2012; Robinson & Sloutsky, 2010a; Shimojo & Shams, 2001). Under some conditions these effects are facilitative. For example, there is both behavioral and neurophysiological evidence demonstrating that adults are faster at detecting and identifying a target when comparable information is presented to multiple sensory modalities (e.g., a picture of a dog is paired with a dog bark) than when information is presented to a single modality (Colonius & Diederich, 2006; Fort, Delpuech, Pernier, & Giard, 2002; Giard & Peronnet, 1999; Miller, 1982; Molhom, Ritter, Javitt, & Foxe, 2004; Sinnett, Soto-Faraco, & Spence, 2008). However, sometimes cross-modal pairings provide conflicting information (e.g., auditory input elicits one response and visual input elicits a different response), which often results in cross-modal competition (see Robinson et al., 2012; Robinson & Sloutsky, 2010a; Sinnett, Spence, & Soto-Faraco, 2007; Spence, 2009, for reviews).

One interesting component of cross-modal competition is that effects are often asymmetrical: Simultaneously presenting information to multiple sensory modalities attenuates processing in one modality, while having little or no effect on processing in the second modality. For example, there is much evidence for visual dominance in adults (see Spence, 2009, for a review). In one experiment, Colavita (1974) instructed participants to press one button whenever they heard a tone and another button whenever they saw a light. The majority of trials were unimodal and participants exhibited high levels of accuracy. However, sometimes the tone and light were presented simultaneously. On these cross-modal trials, participants tended to respond with the "visual" button, often failing to detect the tone altogether. Visual dominance effects have been replicated using a variety of paradigms examining participants' responses to cross-modal input (see Sinnett et al., 2007; Spence, 2009, for reviews), with many manipulations failing to reverse visual dominance.

In an attempt to explain the phenomenon of visual dominance, Posner et al. (1976) suggested that visual dominance is a result of endogenous attention compensating for the poor alerting abilities inherent in the visual system. The overall idea is that auditory and visual modality may compete for access to the central processor that controls responding. Given the low alerting ability of visual stimuli, participants deliberately focus attention on the visual modality, thus strategically biasing their responding in favor of visual input. Therefore, according to this explanation, visual dominance may reflect response bias, which is particularly likely given that most if not all of the evidence supporting visual dominance comes from research where auditory and visual input are associated with different responses. However, it is also possible that, in addition to response bias, visual dominance stems from visual input attenuating auditory processing.

The notion that response bias may account for some of the visual dominance effects has also received indirect support from the developmental literature. In particular, many developmental studies examine infants' looking

DOI: 10.1027/1618-3169/a000177

behaviors, as opposed to examining explicit responses to a stimulus. There is a growing body of evidence indicating that early in development, auditory rather than visual dominance is likely to be found (Robinson & Sloutsky, 2004, 2010b; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008; see also Robinson & Sloutsky, 2010a, for a review). Specifically, when infants and young children were presented with auditory-visual (AV) compounds, they often failed to encode the visual stimulus. This is noteworthy because they ably encode the visual stimulus when it is presented unimodally and there is little evidence suggesting that the visual stimulus attenuates auditory processing. If we assume that young children do not have sufficient control of attention, and auditory input is more likely to automatically engage attention (see Robinson & Sloutsky, 2010a), then it is not surprising that auditory dominance is often observed early in development. At the same time it is possible that auditory dominance effects do not completely disappear across development and often go undetected because adults can strategically bias responses in favor of visual input. Thus, the current study addresses an important question: Will modality dominance effects be found when using a task where participants cannot establish a modality-specific response bias, and will this dominance favor the auditory or visual modality?

One way of addressing this issue is to create a task where: (a) participants are not required to make an explicit response (e.g., passive oddball tasks), (b) explicit responding has no bearing on the critical dependent variable, or (c) the same responses are associated with auditory and visual input, thus, making it impossible to develop a modality-specific response bias. If visual dominance is driven by endogenous attention directed to visual stimuli and a subsequent response bias, then visual dominance should only be observed in tasks where different modalities yield different responses.

One task that does not require modality-specific responding is statistical learning. Statistical learning refers to a process by which an organism detects statistical dependencies in the input. Evidence for statistical learning has been found in auditory, visual, and tactile modalities and mechanisms underlying statistical learning appear to be in place early in development (Conway & Christiansen, 2005, 2006; Fiser & Aslin, 2002a, 2002b; Gomez & Gerken, 2000; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Jungé, & Scholl, 2005). Given that young infants can quickly detect statistical regularities in the environment, statistical learning provides a unique opportunity to capture developmental changes in cross-modal processing while using a comparable task across the life span. The current study begins to address this issue by examining cross-modal statistical learning in adults.

Existing literature on statistical learning yields somewhat contradictory evidence regarding the effects of cross-modal input on statistical learning. For example, Saffran, Newport, Aslin, Tunick, and Barrueco (1997) presented children and adults with two concurrent tasks. The primary task was to perform coloring activities, whereas the secondary task (which was not announced to the participants) was to learn statistical regularities in auditory input that played in the background. Because attention was directed to the visual information, learning of the background auditory sequences would occur either without attention or with divided attention. The researchers found that both children and adults ably learned auditory statistics. These findings present evidence that children can learn auditory sequences, and that this learning can occur even when participants were focusing on a concurrent visual task.

These conclusions are somewhat qualified by Toro, Sinnett, and Soto-Faraco (2005). These researchers presented adults with a statistically structured stream of sounds, which was accompanied by a stream of pictures. Similar to Saffran et al. (1997), there were two tasks, a primary task and a secondary task. In Experiment 2, the primary task was to detect repeated stimuli in the visual stream, and the secondary task (unbeknownst to participants) was to learn auditory statistics. When visual stimuli were presented every 750 ms, participants ably learned auditory statistics. However, learning attenuated substantially when visual stimuli were presented every 500 ms. It was concluded that auditory statistical learning is affected by the difficulty of the primary visual task, thus, providing evidence to the contrary of what was found by Saffran et al. (1997).

However, neither of the above-reviewed studies can evaluate the nature of cross-modal interference (or lack thereof) because they assessed only effects of visual input on auditory statistical learning, whereas, effects of auditory input on visual statistical learning were not assessed. Therefore, it is unclear whether interference effects would be asymmetrical (asymmetry is a signature pattern of modality dominance), and whether this asymmetry would result in auditory or visual dominance.

Given that previous statistical learning results are inconclusive with respect to modality dominance effects, the goal of current research was to further examine these effects using a cross-modal statistical learning task. It has been established that participants can learn statistics in more than one modality (e.g., Seitz, Kim, van Wassenhove, & Shams, 2007). Therefore, presenting participants with a cross-modal statistical learning task and then increasing task demands in one of the modalities may provide important information about modality dominance effects. Suppose that modality A and modality B yield comparable learning performance when participants learn statistics (a) unimodally and (b) when modalities A and B are correlated. Further, suppose that a primary task is introduced in both modalities. If increasing demands in the primary task in modality A attenuates statistical learning in modality B, whereas, increasing demands in modality B does not affect statistical learning in modality A, then asymmetric interference could be inferred. Recall that the signature pattern of modality dominance is that presenting information to multiple sensory modalities attenuates processing in one modality, with little or no attenuation in another modality. The reported experiments were designed to address these issues.

In the current experiments, we presented adults with variations of a cross-modal statistical learning task. For some participants auditory and visual sequences were correlated during training, and we tested participants on either unimodal auditory sequences or on unimodal visual sequences. For other participants we increased task demands by randomizing one of the sequences during training, and then we tested participants on the structured sequences (i.e., same sequences that were presented in the correlated conditions). It was hypothesized that learning of the sequences would differ across the two conditions. The differences would stem either from correlated cross-modal cues facilitating learning or from the randomized stream attenuating learning. To distinguish between these two possibilities, we also compared learning in these conditions to the respective unimodal baselines (i.e., only visual or only auditory). Cross-modal facilitation would be inferred if a cross-modal condition exceeds the unimodal baseline, whereas cross-modal interference would be inferred if the unimodal baseline exceeds a cross-modal condition.

If visual dominance effects stem from a response bias, then the current cross-modal statistical learning task should not elicit visual dominance because participants made old/new responses at test, as opposed to making modality-specific responses (i.e., auditory input elicits one response and visual input elicits a different response). However if visual input attenuates processing of auditory input, then it is possible that increased task demands will have a greater cost on auditory processing. At the same time, it is also possible that auditory dominance effects do not completely disappear across development, with auditory input attenuating processing of visual input. If this is the case, then it is possible that increasing task demands will attenuate visual but not auditory statistical learning. These possibilities were tested in the reported experiments. To foreshadow, the results present evidence for auditory dominance: while participants learned auditory and visual sequences when presented unimodally or when cross-modal sequences were correlated during training, increasing the task demands attenuated visual but not auditory statistical learning.

Experiment 1

The goal of this experiment was to examine how increased task demands in one modality affect learning in a different modality. There was a primary task (i.e., detect repeating items during training) and a secondary task (i.e., learning of the statistical structure of the sequences). Because participants were explicitly instructed to detect repeated stimuli, we considered this task as the primary task. In contrast, participants were not instructed that they were going to be tested on the sequences, therefore, we considered statistical learning as the secondary task.

Method

Participants

Eighty-three undergraduate students participated in this experiment. In this and all other experiments reported here, participants were Ohio State University undergraduate students participating for course credit. Forty of the participants were trained on correlated AV sequences, with both sequences conforming to the same statistical pattern. For example, whenever presented with triplet ABC in the visual modality, they always heard triplet XYZ in the auditory modality. At test, participants were tested either on the unimodal auditory sequence (n = 20) or on the unimodal visual sequence (n = 20). For the remaining participants, we randomized one of the sequences during training and then tested them on the structured sequence. Breaking the correlation between auditory and visual sequence requires participants to simultaneously keep track of two different sequences to detect repetitions (i.e., one auditory and one visual), thus, making the primary (and possibly the secondary) task more demanding. Twenty-one participants were trained on an $\text{Auditory}_{\text{structured}}/\text{Visual}_{\text{random}}$ sequence and then we tested them on the auditory sequence, and 22 participants were trained on a $\text{Visual}_{\text{structured}}/\text{Auditory}_{\text{random}}$ sequence and then we tested them on the visual sequence. The structured sequences in the two random conditions were identical to the sequences in the correlated conditions; therefore, any difference in learning across these two conditions can only be accounted for by the sequence presented in the other modality (i.e., structured or random).

Materials

Auditory stimuli consisted of 12 different syllables presented at 65–70 dB using *Presentation* software. Each syllable was produced in isolation by a female experimenter and recorded as a 44.1 kHz wav file. The visual stimuli consisted of 12 different shapes (8 cm × 8 cm), presented centrally on a computer screen (see Figure 1 for timing of AV sequence).

Auditory and visual elements made two sets, each consisting of four triplets (see Figure 2). Each triplet was presented 16 times throughout training, with some participants being trained on Set 1 and some on Set 2. In the correlated conditions, participants were simultaneously presented with two structured sequences, which were perfectly correlated. The sequences were structured in that they were presented as triplets; however, the order of the triplets was pseudorandomized by using the following criteria: triplets could not: (a) occur back to back in succession (e.g., T_1, T_1...) or (b) alternate (e.g., T_1, T_2, T_1, T_2). In the randomized conditions, elements were not presented as triplets, and the order of individual elements was randomized. Participants were presented with a structured sequence in one modality and a randomized sequence in the other modality.

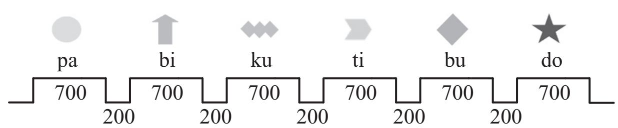

Figure 1. Timing of cross-modal sequence. Values denote time in milliseconds.

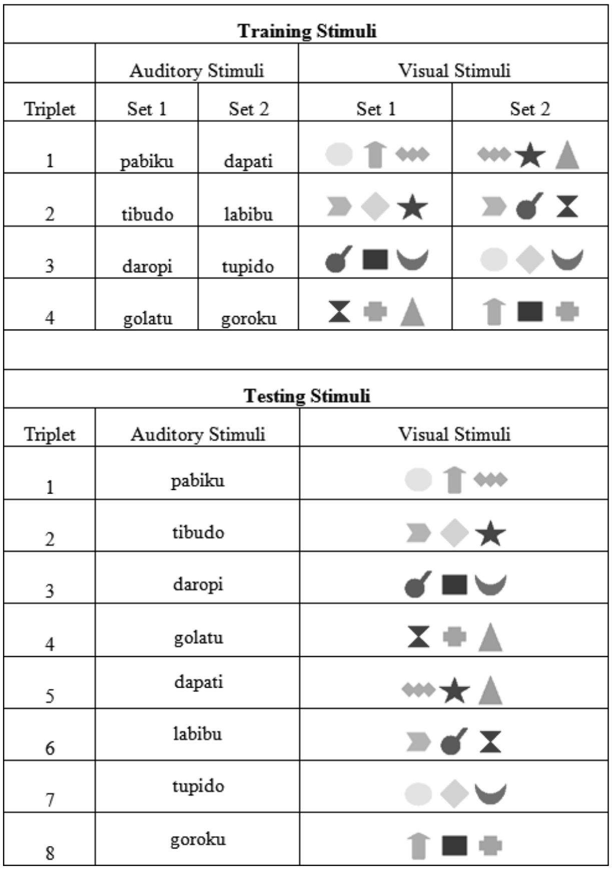

Figure 2. Stimuli presented in Experiments 1 and 2.

The randomized sequence was constrained by the following criteria: (a) we equated the frequency of each element (each element was presented 16 times) and (b) as in the structured sequences, there were eight repeating elements throughout the 3-minute sequence (see Procedure section for details).

At test, participants were tested on one of the structured sequences. The four triplets from each training set were presented, with each triplet repeating twice, thus, resulting in 16 test trials. For some participants, the four triplets from Set 1 were familiar and the four triplets from Set 2 were novel, whereas for others, the reverse was the case.

Procedure

During training, participants were presented with a continuous 3-minute-long AV sequence. Participants were instructed to press the spacebar when they heard two identical sounds in a row or when they saw two identical shapes in a row (cf. Toro et al., 2005; Turk-Browne et al., 2005). For structured sequences, the repeating elements always occurred between two triplets (e.g., ABC-C-DEF), rather than being embedded within a triplet. Therefore, the button press was in response to items that were not part of the to-be-learned statistics. In the random conditions, repeating elements could appear at any point in the stream. As mentioned above, we equated the number of repeating elements in the structured and random sequences. During testing, participants had to determine if a three-shaped or three-sound sequence was presented during training. No feedback was provided during testing.

Results and Discussion

Analyses focused on both the primary task – the detection of repeating items, and on the secondary task – statistical learning. Performance on the primary task was assessed on the basis of responses provided during training, whereas, performance on the secondary task was assessed on the basis of responses provided at test. In the correlated conditions, there were 192 non-repeating items and eight repeating items. Recall that the auditory and visual streams were correlated; thus, when participants encountered a repeating element in one modality, they also encountered a repeating element in the other modality. In the random conditions, there were 16 repeating elements (eight in the structured sequence and eight in the random sequence). A correct detection of a repeating item occurred (i.e., hit) if the participant made a response during the item presentation or during the subsequent interstimulus interval. All other button presses during training were considered false alarms. Repetition detection accuracies (proportion of hits-proportion of false alarms) across all conditions and experiments are presented in Table 1.[1] In this and all other experiments reported here, participants detected the repeating items with above-chance accuracy (all one-sample *t*s > 7.52, *p*s < .001).

The main results of Experiment 1 are presented in Table 1 and Figure 3. To examine how correlated cross-modal cues and noise (i.e., a randomized sequence) affect the detection of repeating items in the structured sequences, we submitted repetition detection accuracy of the structured sequences to a Modality (Auditory vs. Visual) × Condition (Correlated vs. Random) between-subjects ANOVA. The analysis revealed only an effect of Condition, $F(1, 77) = 38.08$, $p < .001$, $\eta_p^2 = 0.33$. Participants were more accurate at detecting repeating items in the correlated conditions ($M = 96.55$, $SE = .01$) than in the random conditions ($M = 68.20$, $SE = .04$).

To assess learning of auditory and visual sequences, we submitted the proportion of correct responses at test to a Modality (Auditory vs. Visual) × Condition (Correlated vs. Random) between-subjects ANOVA. The analysis revealed an effect of Modality, $F(1, 79) = 5.85$, $p < .05$, $\eta_p^2 = 0.07$, an effect of Condition, $F(1, 79) = 4.64$, $p < .05$, $\eta_p^2 = 0.06$, and a Modality × Condition interaction, $F(1, 79) = 3.80$, $p = .055$, $\eta_p^2 = 0.05$. As shown in Figure 3B, the interaction suggested that while there were no differences in learning the auditory sequences in the correlated and random conditions, $p = .90$, participants were more likely to learn the visual sequences in the correlated

[1] Two participants were excluded from the analysis because they did not make a single response during training.

Table 1. Accuracy detecting repeating elements during training and proportion of correct responses at test across the different experiments

Exp.	Training stimulus	Task	Auditory repetition detection (*SE*)	Auditory statistical learning (*SE*)	*d*
1	AV-correlated	Temporal	.96 (.02)	.62* (.03)	0.77
1	AV-V random	Temporal	.72 (.07)	.61* (.03)	0.96
1	Unimodal A	Temporal	.72 (.07)	.60* (.03)	0.80
2	AV-correlated	Spatial-temporal	.91 (.05)	.65* (.05)	0.63
2	AV-V random	Spatial-temporal	.81 (.06)	.66* (.05)	0.70
2	Unimodal A	Spatial-temporal	.87 (.03)	.68* (.06)	0.62
			Visual repetition detection (*SE*)	Visual statistical learning (*SE*)	*d*
1	AV-correlated	Temporal	.97 (.01)	.60* (.03)	0.49
1	AV-A random	Temporal	.59 (.06)	.49 (.02)	−0.07
1	Unimodal V	Temporal	.88 (.05)	.60* (.04)	0.79
2	AV-correlated	Spatial-temporal	.84 (.09)	.75* (.04)	1.46
2	AV-A random	Spatial-temporal	.76 (.08)	.56 (.04)	0.33
2	Unimodal V	Spatial-temporal	.92 (.04)	.63* (.05)	0.53

Notes. * Denotes that discrimination of familiar and novel sequences exceeded chance, $p < .05$. Cohen's d denotes the size of this effect.

condition than in the random condition, $t(40) = 3.28$, $p < .005$.

While Experiment 1 shows differences in repetition detection and statistical learning between the correlated and random conditions, with performance on the statistical learning task also interacting with sensory modality, it is unclear what is driving these effects. To determine if correlated cross-modal cues facilitate learning or if the randomized stream attenuated learning, we conducted two control experiments with 50 new participants to examine repetition detection and statistical learning when auditory and visual sequences were presented unimodally during training and test. Means and standard errors for the two unimodal conditions are reported in Table 1 and these data are represented by horizontal bars in Figures 3A and 3b. As can be seen in the left panel of Figure 3A, relative to the unimodal auditory baseline, correlated visual cues facilitated auditory repetition detection, independent sample t, $t(38) = 3.38$, $p < .005$, whereas, randomizing the visual sequence had no effect on auditory repetition detection, $p = .99$. To examine benefits and costs of auditory input on visual repetition detection, we compared visual repetition detection in the cross-modal conditions to the unimodal baseline (see the right panel of Figure 3A). Relative to the unimodal visual baseline, randomizing the auditory stream attenuated visual repetition detection, independent sample t, $t(48) = -3.02$, $p < .005$, and correlated auditory cues had no significant effect, $p = .14$.

To examine benefits and costs of cross-modal input on statistical learning, we compared learning in the correlated and random conditions to the respective unimodal baselines. As can be seen in Figure 3B (the left panel), correlated and random visual sequences had no effect on auditory statistical learning, $ps > 72$. In contrast, as can be seen in the right panel of Figure 3B, whereas correlated auditory cues had no effect on visual statistical learning compared to the unimodal visual condition, $p = .95$, randomizing the auditory sequence attenuated visual statistical learning,

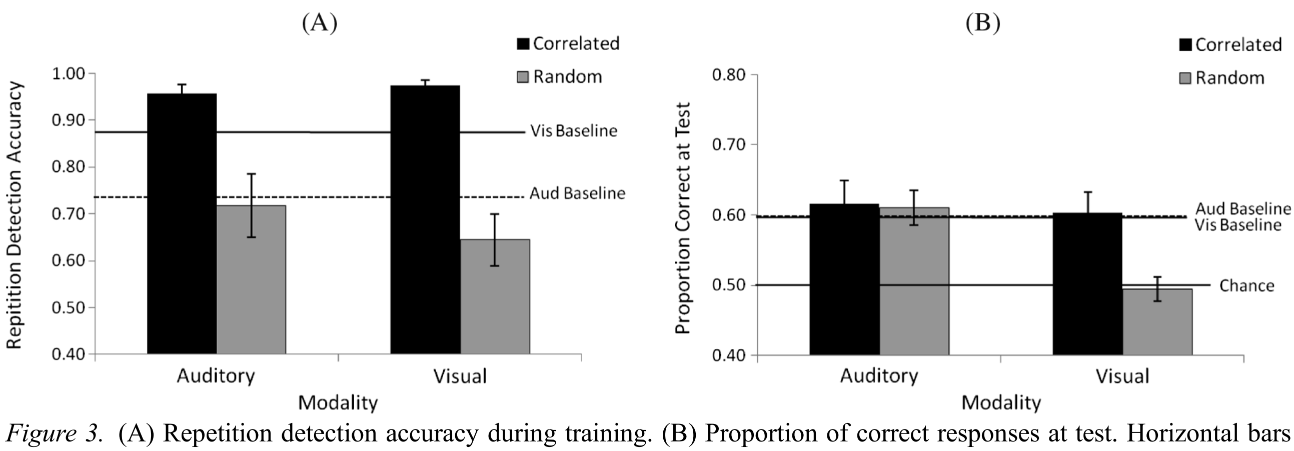

Figure 3. (A) Repetition detection accuracy during training. (B) Proportion of correct responses at test. Horizontal bars denote means from the Auditory (Aud baseline) and Visual (Vis baseline) unimodal baselines and error bars denote standard errors.

$t(50) = 2.28$, $p < .05$. Furthermore, the $\text{Visual}_{\text{structured}}/\text{Auditory}_{\text{random}}$ condition was the only condition where participants failed to learn the sequences, performance at test did not differ from the chance level, $p = .74$.

The results from Experiment 1 are consistent with auditory dominance effects found in young children. In particular, the presence of visual input had no cost on auditory processing, in fact it may have even facilitated auditory repetition detection. In contrast, auditory input attenuated visual repetition and statistical learning; however, only when the auditory input was not correlated with visual input. Therefore, a concurrent auditory task is more likely to attenuate visual statistical learning than a concurrent visual task is to attenuate auditory statistical learning.

While the current findings point to auditory dominance, it could be argued that the observed evidence of auditory dominance stems from using a purely temporal task, which could be better suited for the auditory modality than for the visual modality (cf. Conway & Christiansen, 2005; Welch & Warren, 1980). In Experiment 2, we addressed this possibility by using a visual-spatial statistical learning task, which is well suited for visual statistical learning (Kirkham, Slemmer, Richardson, & Johnson, 2007).

Experiment 2

In Experiment 2, we presented participants with a spatial-temporal, cross-modal statistical learning task where the same visual image appeared in different spatial locations. As in Experiment 1, for approximately half of the participants, auditory and visual sequences were correlated during training and we tested them on either the unimodal auditory sequences or on the unimodal visual sequences. For the remaining participants, we randomized the elements in one of the streams and tested repetition detection and statistical learning in the structured sequence.

Participants, Design and Procedure

We recruited 76 participants for the current experiment. Thirty-eight of the participants were trained on correlated AV sequences, and we tested 17 of the participants on the unimodal visual sequence and 21 of the participants on the unimodal auditory sequence. For the remaining participants, we randomized one of the sequences during training and then tested them on the structured sequence. Twenty participants were trained on an $\text{Auditory}_{\text{structured}}/\text{Visual}_{\text{random}}$ sequence and then we tested them on the auditory sequence, and 18 participants were trained on a $\text{Visual}_{\text{structured}}/\text{Auditory}_{\text{random}}$ sequence and then we tested them on the visual sequence.

Auditory stimuli and timing of AV sequences were identical to Experiment 1, whereas, visual stimuli consisted of a red circle that appeared in 12 different locations on the computer monitor (see Figure 4). For example, for triplet 1 in Set 1, a red circle appeared in Location A for 700 ms, disappeared for 200 ms, then appeared in Location B for 700 ms, disappeared for 200 ms, and then reappeared in Location C for 700 ms. Each triplet was defined by a unique sequence of locations. Training continued until each triplet was presented 16 times. In the randomized conditions, items were not presented as triplets and we randomized the order of individual elements in the sequence. Consistent with previous visual statistical learning tasks (Fiser & Aslin, 2002a), participants in all conditions were given a two-interval forced-choice test. On each test trial, participants were presented with either two visual sequences or two auditory sequences (presented sequentially), and they had to determine which sequence was familiar.

Results and Discussion

The main results of Experiment 2 are presented in Table 1 and Figure 5. As in Experiment 1, to examine how correlated cross-modal cues and a randomized stream affected the detection of repeating items, we submitted repetition detection accuracy of the structured sequences to a Modality (Auditory vs. Visual) × Condition (Correlated vs. Random) between-subjects ANOVA.[2] The analysis revealed no significant effects or interactions, $ps > 15$.

To determine how correlated cross-modal cues and a randomized sequence affect statistical learning, we submitted the proportion of correct responses at test to a Modality (Auditory vs. Visual) × Condition (Correlated vs. Random) between-subjects ANOVA. The analysis revealed an effect of Condition, $F(1, 72) = 3.66$, $p = .06$, $\eta_p^2 = 0.05$, and a Modality × Condition interaction, $F(1, 72) = 3.98$, $p < .05$, $\eta_p^2 = 0.05$. The interaction indicated that, similar to Experiment 1, there were no differences in learning the auditory sequences in the correlated and random conditions, $p = .96$ (see the left panel of Figure 5B), whereas participants were more likely to learn the visual sequences in the correlated condition than in the random condition (see the right panel of Figure 5B), $t(33) = 3.12$, $p < .005$.

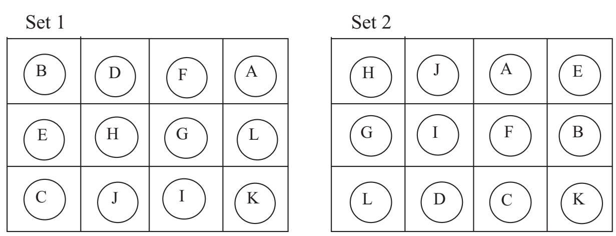

Figure 4. Visual-spatial triplets presented in Experiment 2: Triplet 1 = ABC, Triplet 2 = DEF, Triplet 3 = GHI, and Triplet 4 = JKL. Letters represent locations of elements in each triplet.

[2] Eighteen participants were excluded from the analysis because they did not make a single response during training or because they made a response to every individual item.

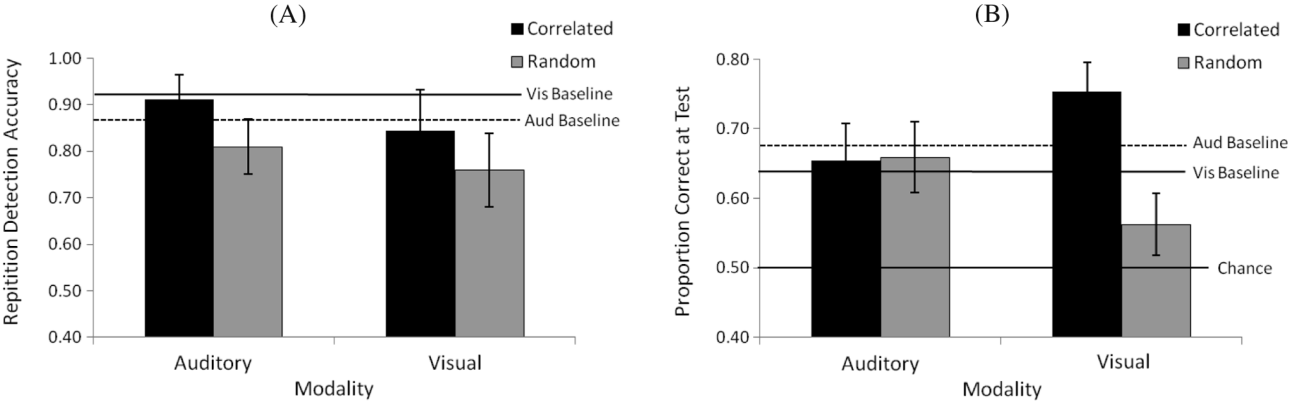

Figure 5. (A) Repetition detection accuracy during training. (B) Proportion of correct responses at test. Horizontal bars denote means from the Auditory (Aud baseline) and Visual (Vis baseline) unimodal baselines and error bars denote standard errors.

We also examined potential benefits and costs of correlated and random cross-modal cues on statistical learning by comparing performance in the cross-modal conditions to the respective unimodal baselines ($N = 46$). As can be seen in the left panel of Figure 5B, correlated and random visual sequences had no effect on auditory statistical learning, *p*s > 75. In contrast, as can be seen in the right panel of Figure 5B, relative to the unimodal visual condition, there was a marginal facilitation effect in the correlated condition, $t(40) = 1.80$, $p = .08$. While the difference between the unimodal visual condition and the random condition did not reach significance, $p = .33$, as in Experiment 1, there was no evidence that adults learned the visual sequences when the sequences were paired with a randomized auditory stream, one-sample *t*-test compared to .5, $t(17) – 1.42$, $p = .18$. However, they ably learned the same visual sequences when presented unimodally or when paired with a correlated auditory sequence, *p*s < .014.

These results corroborate and further expand the findings of Experiment 1. First, there was no evidence that the presence of visual information (correlated or random) attenuated processing in the auditory modality. And second, despite the fact that the secondary visual task was particularly conducive for visual statistical learning, especially in the correlated condition, results were similar to those of Experiment 1. Although participants ably learned the visual sequences when sequences were presented in isolation and when paired with correlated auditory sequences, this was not the case when task demands were increased. Specifically, increasing the task demands by randomizing the auditory stream attenuated this learning. Recall that in both experiments the $\text{Visual}_{\text{structured}}/\text{Auditory}_{\text{random}}$ condition was the only condition where participants failed to learn the sequences.

General Discussion

The reported research presents novel evidence enhancing our understanding of cross-modal processing. When auditory and visual sequences were presented unimodally or when they conformed to the same statistical pattern, participants ably performed the primary and the secondary tasks – they simultaneously detected repetitions and extracted statistics in auditory and visual input. In fact, correlated cross-modal cues may have even facilitated auditory repetition detection (Experiment 1) and visual statistical learning (Experiment 2). Across both reported experiments there was no evidence that increasing task demands attenuated auditory statistical learning (see Figures 3B and 5B). However, increasing the task demands attenuated visual statistical learning. In particular, in both experiments there was a significant cost on visual statistical learning when visual sequences were paired with a randomized auditory stream, compared to when the same visual sequences were paired with a correlated auditory sequence. Furthermore, out of all of the conditions tested, the $\text{Visual}_{\text{structured}}/\text{Auditory}_{\text{random}}$ condition was the only condition where participants failed to learn the sequences. These are novel findings pointing to (a) the ability to learn cross-modal statistics and (b) asymmetrical costs in statistical learning, suggesting a possibility of auditory dominance in adults.

Could it be that the current findings reflect the fact that statistical learning is better suited for the auditory modality? (e.g., Conway & Christiansen, 2005; Welch & Warren, 1980) While we cannot completely eliminate this possibility, there are three reasons why we believe this is not the case. First, Experiments 1 and 2 demonstrate that learning of unimodal auditory and unimodal visual sequences was comparable, which indicates that auditory and visual sequences were equally learnable. Second, even when the spatial statistical learning task was introduced (Experiment 2), visual statistical learning was still more likely to fail under the increased task demand condition than auditory statistical learning. And finally, the current findings have been corroborated in a recent ERP study that is not based on processing of temporal input (Robinson, Ahmar, & Sloutsky, 2010). In this study participants were presented with an "oddball" task that consisted of auditory, visual, and cross-modal stimuli. The variable of interest was "oddball" detection – the detection of infrequent stimuli among frequent standards. Findings dovetail nicely with the current study: while auditory input delayed behavioral and neurophysiological responses associated with the detection of visual oddballs (as evidenced by slower RTs and longer peak latency of P300), visual input did not delay the detection of auditory oddballs.

The current findings in conjunction with the recent ERP study examining the effect of cross-modal input on oddball detection (Robinson et al., 2010) show support for auditory dominance in adults, however, these findings differ considerably from approximately 40 years of research showing evidence of visual dominance in adults (see Sinnett et al., 2007; Spence, 2009, for reviews). While it was previously argued that auditory dominance found in children and visual dominance found in adults may reflect developmental effects (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003), finding evidence of auditory dominance in adults suggests that other factors are at play. More specifically, we believe that auditory and visual dominance research may be tapping into different processes. In particular, visual dominance has primarily been found in detection and identification tasks where participants are required to make an explicit response (although see Ngo, Sinnett, Soto-Faraco, & Spence, 2010). Furthermore, in most if not all of the studies supporting visual dominance, auditory input elicited one response and visual input elicited a different response, which may have resulted in a modality-specific response bias. Furthermore, there are reasons to believe that visual dominance effects may not reflect how multimodal stimuli are perceived but reflect how quickly participants initiate their response (see Spence, 2009 for a related discussion). The current findings are consistent with this account. When using a dependent measure that was not directly tied to a specific modality (i.e., in our experiments repetition detection was not pertinent to the task of statistical learning and responses at test were not tied to a sensory modality), there was no evidence that visual input attenuated auditory processing.

At the same time, Sloutsky and colleagues have argued that auditory dominance effects stem from auditory input attenuating or delaying the processing of visual input (see Robinson et al., 2012; Robinson & Sloutsky, 2010a, for reviews). Support for this account has primarily come from research examining processing of arbitrary AV pairings in infants and young children; however, the current study provides support for this claim in adults. In particular, increasing the demands of the primary task attenuated learning in the visual modality while having no cost on learning in the auditory modality. This asymmetry occurs in implicit learning tasks such as the one reported here but also in explicit tasks where participants are asked to quickly respond to infrequent auditory and visual input. When participants made the same response to infrequent auditory and visual oddballs, auditory dominance was observed; whereas, visual dominance was observed when participants made different responses to auditory and visual input (Chandra, Robinson, & Sinnett, 2011). This recent finding provides further evidence that visual dominance effects may be occurring during the response/decision phase, whereas auditory dominance effects may stem from attenuated processing of visual input.

In summary, the current findings have important implications for answering the question of how people process multisensory information. While the current study found some evidence that correlated cross-modal cues facilitate processing, there are many situations when modalities provide different information, and in these situations modalities may compete (Rapp & Hendel, 2003). For adults, this competition is often resolved in favor of the visual modality (see Sinnett et al., 2007, for a review). However, the current study in conjunction with Robinson et al. (2010) demonstrates that this competition can be also resolved in favor of the auditory modality. While future research is needed, we believe that cross-modal competition can occur at different points in the course of processing, with auditory input attenuating visual processing and visual input dominating the response.

Acknowledgments

This research was supported by grants from the NSF (BCS-0720135), IES (R305B070407), and NIH (R01HD056105) to VMS and from NIH (1R03HD055527) to CWR. The opinions expressed do not represent views of the awarding organizations.

References

Chandra, M., Robinson, C. W., & Sinnett, S. (2011). Coexistence of multiple modal dominances. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2604–2609). Austin, TX: Cognitive Science Society.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*, 409–412.

Colonius, H., & Diederich, A. (2006). The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review, 113*, 148–154.

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*, 24–39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17*, 905–912.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*, 162–169.

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*, 15822–15826.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object recognition in humans. *Cerebral Cortex, 12*, 1031–1039.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*, 473–490.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition, 83*, B35–B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development, 78*, 1559–1571.

Miller, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247–279.

Molhom, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex, 14*, 452–465.

Ngo, M. K., Sinnett, S., Soto-Faraco, S., & Spence, C. (2010). Repetition blindness and the Colavita effect. *Neuroscience Letters, 480*, 186–190.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*, 157–171.

Rapp, B., & Hendel, S. K. (2003). Principles of cross-modal competition: Evidence from deficits of attention. *Psychological Bulletin and Review, 10*, 210–219.

Robinson, C. W., Ahmar, N., & Sloutsky, V. M. (2010). Evidence for auditory dominance in a passive oddball task. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2644–2649). Austin, TX: Cognitive Science Society.

Robinson, C. W., Best, C. A., Weng, D., & Sloutsky, V. M. (2012). The role of words in cognitive tasks: What, when, and how? *Frontiers in Psychology, 3*, 1–8.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401.

Robinson, C. W., & Sloutsky, V. M. (2010a). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 135–141.

Robinson, C. W., & Sloutsky, V. M. (2010b). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology, 107*, 351–358.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by adults and infants. *Cognition, 70*, 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception, 36*, 1445–1453.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion in Neurobiology, 11*, 505–509.

Sinnett, S., Soto-Faraco, S., & Spence, C. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychologica, 128*, 153–161.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: Revisiting the Colavita effect. *Perception & Psychophysics, 69*, 673–686.

Sloutsky, V. M., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74*, 822–833.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in visual processing: From overshadowing to attentional tuning. *Cognitive Science, 32*, 354–377.

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research, 176*, 245–258.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*, B25–B34.

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

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667.

Received May 21, 2012
Revision received July 9, 2012
Accepted July 19, 2012
Published online October 10, 2012

Chris Robinson

Department of Psychology and Center for Cognitive Science
The Ohio State University
239 Psychology Building
1835 Neil Ave.
Columbus, OH 43210
USA
E-mail robinson.777@osu.edu

