Visual Statistical Learning: Getting Some Help from the Auditory Modality

Christopher W. Robinson (robinson.777@osu.edu)
Center for Cognitive Science
The Ohio State University
208F Ohio Stadium East, 1961 Tuttle Park Place
Columbus, OH 43210, USA

Vladimir M. Sloutsky (sloutsky.1@osu.edu)
Center for Cognitive Science
The Ohio State University
208C Ohio Stadium East, 1961 Tuttle Park Place
Columbus, OH 43210, USA

Abstract

Presenting information to multiple sensory systems can (depending on conditions) either facilitate or hinder processing. The current study examined the effects of cross-modal presentation on statistical learning. The first experiment examined statistical learning within auditory and visual modalities, while using comparable experimental conditions across the two modalities. The findings suggest that the auditory modality was better suited for learning the temporal structure in the input. Experiments 2 and 3 examined whether the presence of auditory stimuli would facilitate or hinder visual statistical learning. The results suggest that auditory stimuli that share the same statistics as visual stimuli can facilitate the acquisition of visual sequences.

Keywords: Cognitive Development, Attention, Language Acquisition, Psychology, Human Experimentation.

Introduction

There are many occasions where information is presented to multiple sensory modalities. Under some conditions (e.g., when auditory-visual stimuli share the same amodal relation such as rhythm or rate), cross modal presentation is likely to facilitate processing of the amodal relation (Bahrick & Lickliter, 2000 see Lewkowicz, 2000; Lickliter & Bahrick, 2000, for reviews). At the same time, under other conditions (e.g., when auditory-visual pairings are arbitrary), cross-modal presentation is likely to hinder processing of arbitrary auditory-visual stimuli (e.g., Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; in press-a; in press-b; Sloutsky & Napolitano, 2003).

Research examining the processing of arbitrary, auditory-visual pairings has demonstrated several interesting phenomena. Of particular interest to this study is the finding that auditory input often affects visual processing, whereas, visual input is less likely to affect the processing of auditory stimuli (Sloutsky & Robinson, in press). While this finding may be fundamental for understanding the early interactions between the auditory and visual systems, Sloutsky and Robinson only examined the processing of a single visual stimulus. Thus, it is uncertain whether these effects can be generalized to the learning of other types of information.

Learning a new domain not only entails encoding and storing representations of single images and sounds, but it also involves detecting the statistical regularities between the elements within that domain. This ability appears to be in place very early in development. For example, to determine if infants could use the statistical information to segment words in a speech stream, Saffran, Aslin, and Newport (1996) presented infants with a continuous stream of syllables (e.g,. pa-bi-ku-ti-bu-do-go-la-tu). Although there were no acoustic or prosodic cues provided to infants, they were able to segment the stream into word-like units by using the statistical regularities from the input.

Statistical learning does not appear to be specific to linguistic stimuli, but rather it seems to be governed by a domain-general learning mechanism capable of operating on a variety of stimuli such as musical notes and geometric shapes (Fiser & Aslin, 2002a; 2002b; Kirkham, Slemmer, & Johnson, 2002; Saffran, Johnson, Aslin, & Newport, 1999; Turke-Browne, Junge, & Scholl, 2005).

More recently, however, there have been several studies examining how these processes interact with the modality of input. While it is often believed that the mechanism underlying statistical learning is domain-general and not intimately tied to a specific modality, there are reasons to suggest that statistical learning may be driven by several different sensory subsystems (Conway & Christiansen, 2005; 2006). These findings raise many interesting questions. For example if statistical learning of visual and auditory stimuli are guided by different subsystems, how do

these systems interact with one another? Does auditory input hinder visual statistical learning, as is found with studies examining the processing of arbitrary, auditory-visual pairings (e.g., Sloutsky & Napolitano, 2003)? Or does activation of multiple subsystems facilitate visual statistical learning as is typical found with processing of amodal relations (Bahrick & Lickliter, 2000)? Alternatively, these effects may be specific to the nature of the input. The primary goal of the current study is to begin answering some of these questions.

The current study consists of three experiments. Using comparable experimental conditions, Experiment 1 examined statistical learning within the auditory and visual modalities. Based on the recent findings of Conway and Christiansen (2005), it was hypothesized that the auditory modality would be more likely than the visual modality to abstract the temporal structure in the input. Experiments 2 and 3 examined how these processes interact with one another: Does auditory input have any effect on visual statistical learning?

Experiment 1

Method

Participants Thirty adults (9 women and 21 men, *M* = 20.3 years, *SD* = 2.4 years) participated in this experiment. Adults consisted of undergraduate students from The Ohio State University participating for course credit. The majority of adults were Caucasian. Fourteen adults were familiarized and tested on speech sounds and 16 adults were familiarized and tested on geometric shapes.

Stimuli The auditory and visual stimuli were modeled after previous research examining statistical learning of speech sounds and geometric shapes in young infants (Fiser & Aslin, 2002b; Saffran, Aslin, & Newport, 1996). The auditory stimuli consisted of 12 different syllables (see Figure 1 for examples). Each syllable was produced in isolation by a female experimenter and presented from a Dell Dimension 8200 computer with Presentation software at 65 – 70 dB.

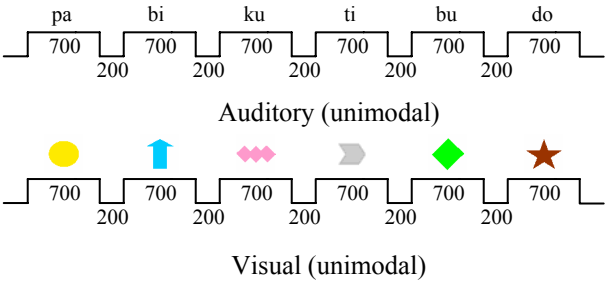


Figure 1: Overview of training phases in Experiment 1. Values denote time in milliseconds.

The visual stimuli consisted of 12 different shapes varying in color. Individual shapes were approximately 8 cm by 8 cm and were presented centrally on a computer screen. The auditory and visual stimuli were presented for 700 ms with a 200 ms inter-stimulus interval (see Figure 1 for overview of training). Although the stream of elements appeared to be random, they actually consisted of four sets of triplets (see Figure 2 for auditory and visual triplets).

Familiarization Stimuli

	Auditory Stimuli		Visual Stimuli	
Triplet	Corpus 1	Corpus 2	Corpus 1	Corpus 2
1	pabiku	dapati		
2	tibudo	labibu		
3	daropi	tupido		
4	golatu	goroku		

Testing Stimuli

Triplet	Auditory Stimuli	Visual Stimuli
1	pabiku	
2	tibudo	
3	daropi	
4	golatu	
5	dapati	
6	labibu	
7	tupido	
8	goroku	

Figure 2: Auditory and visual stimuli presented during the familiarization and testing phase. Corpus 1 and Corpus 2 varied between subjects.

Each Triplet was presented 16 times throughout familiarization, and the familiarization triplets were counter-balanced across subjects. In corpus 1, participants were familiarized to triplets ABC, DEF, GHI, and JKL. In corpus 2, participants were familiarized to triplets AEI, DGJ, BHK, and CFL. During the test phase, each participant was presented with all eight triplets (i.e., four triplets from corpus 1 and four triplets from corpus 2). Thus, the four familiar triplets for participants in corpus 1 (ABC,

DEF, GHI, and JKL) were completely novel foils for adults familiarized to corpus 2 and the familiar triplets for participants in corpus 2 (AEI, DGJ, BHK, and CFL) were completely novel foils for participants familiarized to corpus 1. The triplets were presented in a predetermined order, which was restricted by two criteria. First, triplets could not occur twice in succession (e.g., T_1, $T_{1...}$). Second, alternating triplets was not allowed (e.g., T_1, T_2, T_1, $T_{2...}$). The same predetermined order was used for the auditory and visual modalities, and the overall duration of familiarization was approximately 3 minutes in duration.

Procedure The entire experiment consisted of two phases: Familiarization phase and testing phase. During the familiarization phase, participants were presented with learning sequences. They were also given a distracter tasks, which was similar to the one used by Turke-Browne, Junge, and Scholl (2005). In the visual condition, the distracter task was to press the spacebar every time they saw two shapes in a row that were the same. In the auditory condition, the distracter task was to press the spacebar every time they heard two sounds in a row that were the same. Eight times throughout familiarization, participants were presented with a repeated element (e.g., ABCCDEF). The repeated elements were always the third element in a Triplet, and the repeated element always occurred between two triplets, as opposed to being embedded within a Triplet.

After the familiarization phase, participants were presented with the testing phase. The testing phase started with a cover story, which was created for young children. Below is the cover story for the visual condition: *The pictures that you just saw were words made by an Arbo. Arbos live on a planet called Yodo, and they use shapes to talk to each other. Arbo's words are made up of 3 shapes, and different words can be made by changing the order of the shapes. In the next part you will be presented with words. Some of the words will be made by an Arbo and some will be made by a Luthop. Although Luthops use the same shapes to make their words, they make very different words. You have not seen words made by a Luthop. Your task is to determine if the words were made by an Arbo or by a Luthop. Arbo's and Luthop's languages are very similar to each other. The only way you can tell the difference between the two languages is by the order of the shapes, so you will need to pay close attention to the order of shapes.*

The cover story for the auditory condition was identical, except that all references to shapes were replaced with sounds.

At this point, participants moved to the testing phase where they were tested on the four triplets from corpus 1 and on the four triplets from corpus 2. The order of the testing trials was randomized for each subject, and each Triplet was presented twice, resulting in 16 test trials. Participants had to determine if a given sequence of three shapes or three sounds was made by an Arbo (similar to familiarization sequence) or made by a Luthop (different from familiarization sequence). No feedback was provided.

Results and Discussion

During familiarization, participants noticed 86% of the repeating elements, and no difference was found between the auditory and visual conditions, t (28) = 1.31, p = 20.

The primary analyses focused on discrimination of the familiar triplets from the novel foils. Discrimination was assessed as a difference between hits (i.e., correct acceptance of familiar triplets) and false alarms (i.e., erroneous acceptance of foils). Discrimination greater than zero reflects above-chance discrimination, whereas, discrimination equal to zero reflects at-chance discrimination. While participants discriminated the triplets from the foils in the auditory condition, (M = .20, SE = .07), one-sample t compared to zero, t (13) = 2.90, p = .012, discrimination of the visual stimuli did not exceed chance (M = .07, SE = .09), one-sample t compared to zero, t (15) = 0.77, p = .45.

While it is possible that participants simply needed more exposure to the visual sequence to successfully discriminate the familiar triplets from the foils, these findings are consistent with previous research examining statistical learning in different modalities (Conway & Christiansen, 2005).

Experiment 2

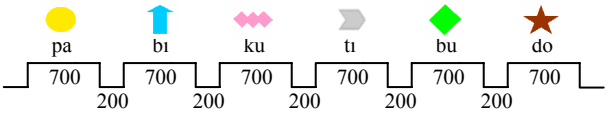
The primary goal of Experiment 2 is to examine the effect of auditory input on visual statistical learning, and the effect of visual input on auditory statistical learning. More specifically, the current study was designed to examine whether correlated auditory cues (i.e., where the same statistics are found in both the auditory and visual modalities) would affect visual statistical learning.

Method

Participants Thirty adults (11 women and 19 men, M = 19.2 years, SD = 0.7 years) participated in this experiment. Demographics and subject recruitment were identical to Experiment 1. Fourteen participants were familiarized to a correlated AV sequence and tested on auditory sequences (presented unimodally). Sixteen participants were familiarized to the same correlated AV sequence and tested on the visual sequences (presented unimodally).

Stimuli and procedure The auditory and visual stimuli were identical to Experiment 1 (see Figure 2).

In contrast to Experiment 1, the auditory and visual stimuli were correlated during the familiarization phase. For example, every time participants were presented with ABC in the visual modality, they also heard ABC in the auditory modality (see Figure 3 for overview of training).



Correlated (bimodal)

Figure 3. Overview of training phase in Experiment 2. Values denote time in milliseconds.

The familiarization sequence lasted approximately 3 minutes, and each triplet was presented 16 times. As in Experiment 1, participants were presented with a distracter task. In this task, they were directed to both modalities: They were told to press the spacebar when they saw the same two shapes in a row or when they heard the same sound twice in a row. Given that the auditory and visual stimuli were correlated in the current experiment, participants could perform quite well on the distracter task: Focusing on the sounds, on the shapes, or on both modalities could lead to accurate performance in the distracter task.

After familiarization, participants were presented with a cover story. With the following exception, the cover story was identical to Experiment 1: *The shapes and sounds that you just saw and heard were words made by an Arbo. Arbos live on a planet called Yodo, and they use shapes and sounds to talk to each other. Arbo's words are made up of 3 shapes and 3 sounds, and different words can be made by changing the order of the shapes and sounds. However, Arbos do not need to pair the shapes and sounds together to talk. They can also talk to each other by simply using shapes or sounds.*

At this point participants moved to the testing phase, which was identical to Experiment 1. More specifically, even though they were trained on a correlated AV sequence, participants were only tested on the sounds or on the shapes, as was the case in Experiment 1.

Results and Discussion

During familiarization, participants noticed 96% of the repeating elements, and no difference was found between the two conditions, t (28) = 0.91, p = 37.

As in Experiment 1, discrimination of the familiar triplets from the foils was assessed as a difference between hits (i.e., correct acceptance of familiar triplets) and false alarms (i.e., erroneous acceptance of foils). In contrast to the visual condition of Experiment 1 where participants were at-chance at discriminating the familiar triplets from the foils, discrimination of the visual sequences (M = .23, SE = .08), t (15) = 3.15, p = .007, and the auditory sequences (M = .20, SE = .06), t (13) = 2.77, p = .016, both exceeded chance. While these effects were small, the finding is remarkable given that Experiment 2 increased the amount of information that adults had to learn. This suggests that the presence of the auditory stimuli during familiarization facilitated processing of the visual sequences and this had lasting effects on the way the visual sequences were perceived.

Experiment 3

Why did the auditory input facilitate visual statistical learning? One possible explanation is that the auditory stimuli were simply more engaging and pairing the auditory stimuli with the pictures in Experiment 2 simply made the task more interesting and subsequently increased performance. Alternatively, it is possible that this effect stemmed from the correlated auditory cues helping participants detect the statistics in the visual modality. Experiment 3 distinguished between these accounts by randomizing the auditory sequence. According to the latter account, breaking the correlation between the auditory and visual stimuli should attenuate the facilitation effect. According to the former account, performance in the current experiment should be comparable to Experiment 2 because auditory input was paired with visual input in both experiments.

Method

Participants Ten adults (2 women and 8 men, M = 19.9 years, SD = 0.9 years) participated in this experiment. Demographics and subject recruitment were identical to previous experiments.

Stimuli and procedure The auditory and visual stimuli were identical to previous experiments (see Figure 2). Stimuli during the familiarization phase were presented cross-modally, whereas, visual sequences were presented in isolation at test (same as in the visual condition of Experiment 2). In contrast to Experiment 2, the auditory sequence was randomized, while the visual sequence followed the same statistics as in previous experiments.

Results and Discussion

Although stimuli were presented cross-modally during the familiarization phase in the current experiment, adults did not discriminate familiar triplets from foils, (M = -0.01, SE = .03), t (9) = -0.36, p = .73. See Figure 4 for means and standard errors across Experiments 1 - 3. This suggests that the facilitation effect in Experiment 2 resulted from

correlated auditory stimuli facilitating visual statistical learning, as opposed to the auditory stimuli simply making the task more engaging.

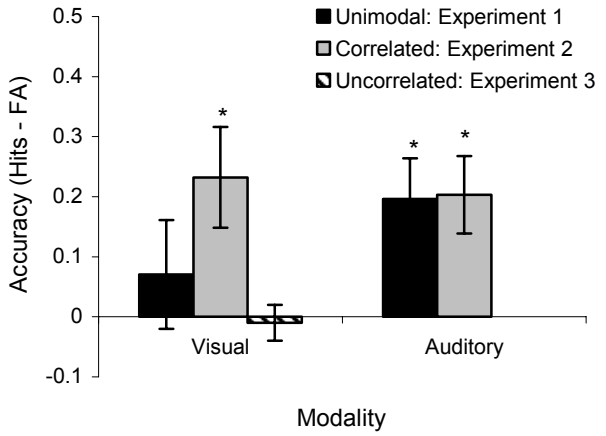


Figure 4: Discrimination accuracy broken up by modality in Experiments 1-3. Note: "*" greater than 0, $p < .05$.

General Discussion

The results point to several important findings. First, under comparable experimental conditions, adult participants were more likely to detect the temporal structure in the auditory modality than in the visual modality (Experiment 1). Second, when auditory and visual input shared the same statistics, visual statistical learning was enhanced (Experiment 2). Even though the amount of information to be learned increased in Experiment 2, this increase in processing demands increased rather than decreased the likelihood of learning the visual sequence. At the same time, the more efficient learning in Experiment 2 did not come with a cost (i.e., attenuated processing in the auditory sequence). Finally, Experiment 3 eliminated the possibility that cross-modal facilitation resulted from the auditory stimuli making the task more engaging.

These are novel findings pointing to cross-modal facilitation in statistical learning. Recall that previous research indicated that under some conditions, cross modal presentation of stimuli is likely to facilitate visual (and auditory) processing (Bahrick & Lickliter, 2000 see Lewkowicz, 2000; Lickliter & Bahrick, 2000, for reviews), whereas under other conditions, cross-modal presentation is likely to hinder visual processing (e.g., Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; in press-a; in press-b; Sloutsky & Napolitano, 2003). However, these studies were conducted primarily with infants and young children. Current research sheds light on cross-modal processing later in development, indicating that cross-modal presentation of stimuli facilitates visual statistical learning in adults.

At the same, the study also raises a number of important questions for future research. The most important issue concerns the mechanism(s) that may underlie the current effects. Experiments 2 and 3 demonstrate the importance of the auditory and visual sequences sharing the same underlying statistics. Thus, the underlying mechanism, at least in adults, appears to be sensitive to covariation across modalities. Examining whether infants and young children also benefit from correlated cues will provide some insight into the developing interactions between the auditory and visual systems.

Experiments 1 and 2 demonstrate that correlated auditory input facilitates visual statistical learning, whereas, correlated visual input has no significant effect on auditory statistical learning (see Figure 4). These findings provide preliminary evidence that facilitation effects are asymmetrical. However, this will have to be further tested in future research.

Finally, experiments, such as the ones reported here, will be fundamental for understanding how attention is allocated within and between modalities. For example, when adults are given two visual sequences and are asked to selectively attend to one of the sequences, they fail to learn the statistics in the unattended visual sequence (Turke-Browne, et al., 2005). This demonstrates the importance of selective attention within the visual modality, and possibly within any modality. However, it will also be important to understand how attention is allocated to cross-modal stimuli and to examine the role of selective attention in cross-modal statistical learning. Consistent with the current study, we have preliminary evidence for a similar asymmetry in selective attention: Selectively attending to the auditory modality has an effect on visual statistical learning, whereas, selectively attending to the visual modality has no effect on auditory statistical learning.

In summary, many studies have examined how humans and non-humans detect statistical regularities in different modalities. However, much of our experiences are multi-modal and there is relatively little known about how attention is allocated to multi-modal stimuli or how different modalities interact with one another to acquire new knowledge. The current study begins to answer some of these important questions.

Acknowledgments

This research has been supported by grants from the NSF (REC 0208103) and from the Institute of Education Sciences, U.S. Department of Education (R305H050125) to Vladimir M. Sloutsky.

References

Bahrick, L.E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190–201.

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.

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.

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.

Lewkowicz, D.J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281-308.

Lickliter, R., & Bahrick, L.E. (2000). The development of infant intersensory perception: Advantages of a comparative convergent-operations approach. *Psychological Bulletin*, *126*, 260-280.

Napolitano, A.C., & Sloutsky, V.M. (2004). Is a Picture Worth a Thousand Words? The Flexible Nature of Modality Dominance in Young Children. *Child Development*, *75*, 1850-1870.

Roberts, K., & Jacob, M. (1991). Linguistic versus attentional influences on nonlinguistic categorization in 15-month-old infants. *Cognitive Development, 6,* 355-375.

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. (in press-a). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*.

Robinson, C,W., & Sloutsky, V.M. (in press-b). Linguistic labels and categorization in infancy: Do labels facilitate or hinder?. *Infancy*.

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.

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. (in press). The role of words and sounds in visual processing: From overshadowing to attentional tuning. *Cognitive Science*.

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